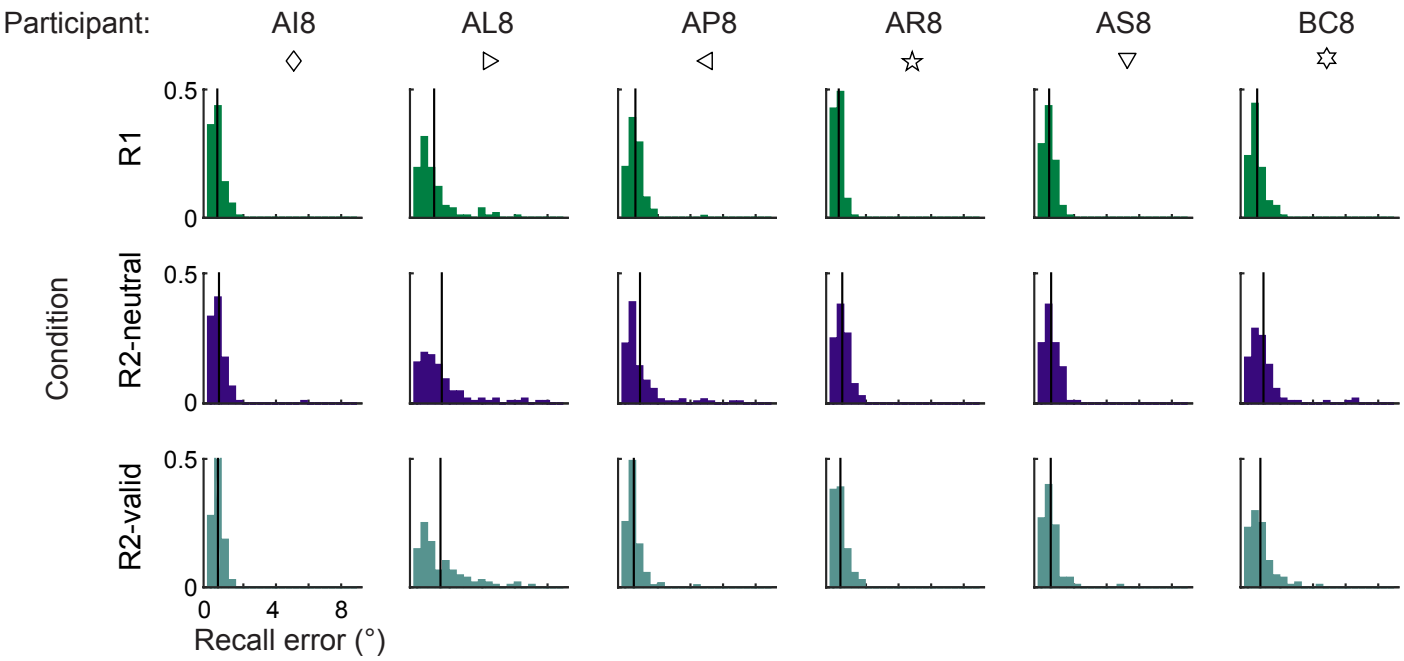


Figure S1 Recall performance recovers when one of two items is cued (related to Figure 1)
Histograms of recall error across all trials for each participant and condition for data presented in Fig. 1C. Y axis indicates "proportion of trials". Same participant identifiers used as in previous reports to facilitate comparison of data across experiments (Ester et al., 2015; Sprague and Serences, 2013; Sprague et al., 2014). Note that only 1 participant (AI) previously participated in a spatial WM experiment. All results presented in this report hold when excluding this participant.

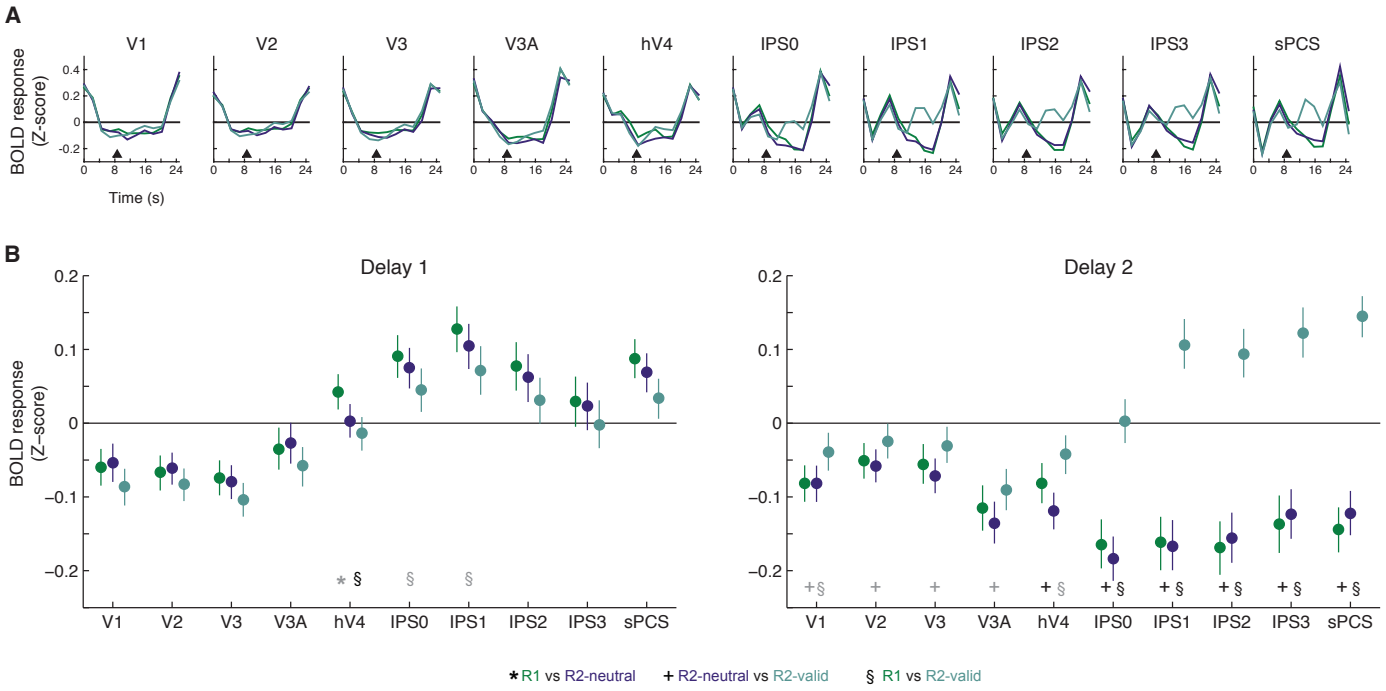


Figure S2 Univariate BOLD responses from each ROI (related to Figures 1, 3)

(A) Mean BOLD activation timecourse (event-related average, time-locked to beginning of WM delay periods) averaged across all trials, all participants, and all voxels within each ROI. Replicating previous work (Emrich et al., 2013; Harrison and Tong, 2009; Riggall and Postle, 2012; Serences et al., 2009; Sprague et al., 2014), we observe no substantial activation in occipital ROIs (V1-hV4) in the univariate BOLD signal. For subsequent analyses, we identified time points primarily corresponding to the delay period before the cue (Delay 1, 6.75-9.00 s), and the delay period after the cue (Delay 2; 15.75-18.00 s).

(B) Mean delay-period activation during Delay 1 (left) and Delay 2 (right) as a function of WM condition. During Delay 1, we found trends towards increased activation with set size (R2-neutral>R1 and/or R2-valid>R1) in IPS0 and IPS1. We also observed significantly higher activation during R2-valid trials in hV4 as compared to R1, but not R2-neutral, trials. During Delay 2, we observed significant cue-related activation (R2-valid>R1 and/or R2-valid>R2-neutral) in hV4, IPS0-IPS3, and sPCS, as well as trends towards this effect in all other regions. Significant tests reflect FDR-correction for all comparisons. Trends defined as $p < 0.05$, uncorrected for multiple comparisons. Error bars 95% confidence intervals via resampling all trials, with replacement, 1,000 times (see Supplemental Experimental Procedures: statistical procedures). All p-values for this analysis presented in Table S1.

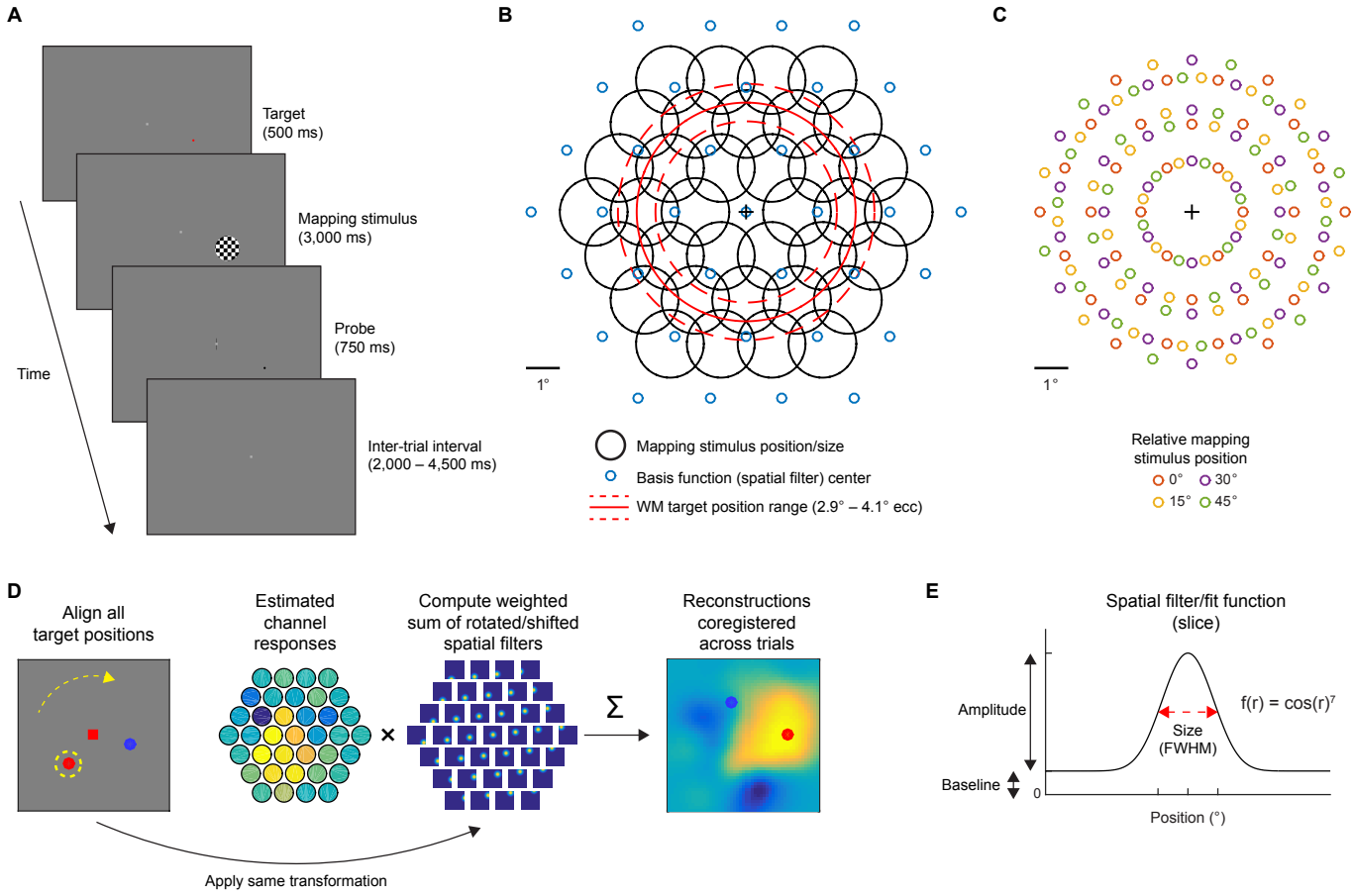


Figure S3 IEM procedures: mapping task, stimulus layout, and reconstruction coregistration (related to Figure 2 and Experimental Procedures)

(A) Participants performed 4 runs of a spatial mapping task during each fMRI scanning session. On each trial, we presented a single WM target stimulus (red dot) for 500 ms, followed immediately by a flickering checkerboard (1.083° radius; 6 Hz full-field flicker) overlapping the WM target location. After 3,000 ms, a probe stimulus (black dot) appeared slightly offset to either the left or right, or above or below, the remembered position (distance varied across runs to equate difficulty) for 750 ms. Simultaneously, a horizontal or vertical bar appeared at fixation, indicating the participant must make a 2AFC "left/right" or "above/below" judgment in response to the question "was the probe dot [left/above] or [right/below] [of] the remembered position?" before the end of the inter-trial interval (2-4.5 s). All stimulus features are drawn to scale. Participants performed on average 87.62% correct (target/probe separation distance adjusted across runs to maintain sufficient task difficulty).
(B) The position of the mapping stimulus varied on each trial along a hexagonal grid (black circles), both inside and outside the range of eccentricities used for the main WM task (red ring). This enabled us to reconstruct images of the contents of spatial WM across the entire visual field subtended by the projector screen inside the scanner (Fig. 2), despite only remembering items from a small range of positions in the WM task (Fig. 1). Blue dots indicate the center of spatial filters used for image reconstruction (Fig. 2).
(C) On each run of the spatial mapping task, we rotationally offset the position of the mapping stimuli by a fixed angular amount. Across sessions, we adjusted the "baseline" angle by 5° (session 1 arrangement shown here).
(D) On each trial of the primary WM task (Fig. 1), the WM targets appeared pseudo-randomly within the red dashed ring in (B). To align data across trials in "information space", we rotated basis functions so as to zero-out the polar angle component of the WM target coordinate (1-d reconstructions & representational fidelity analyses; Figs. 5-6, S5-S7). Then, for analyses in which we precisely aligned target positions (Figs. 7-8, S8), we also shifted them horizontally to precisely align the target position to the coordinate x = 3.5°, y = 0° (see red dot, Fig 7A, C). For example, if a target appeared at 42° polar angle (up and to the right) and 3.7° eccentricity, we first rotated each basis function by 42° polar angle clockwise, then shifted all basis functions horizontally 0.2° to the left, before computing reconstructions. This means that we used a slightly different set of basis functions for computing each trial's reconstructions (same set of basis functions used for each time point of each trial), eliminating any potential idiosyncrasies caused by the exact set of filter centers we used.
(E) Once we averaged coregistered reconstructions from all trials (on each resampling iteration, see Experimental Procedures: Statistical procedures), we fit a surface function (slice shown), which was shaped identically to each spatial filter, to the mean reconstruction. We allowed the function to vary in its size, baseline, and amplitude. Its position was constrained to be nearby the maximum pixel of the average reconstruction (see Supplemental Experimental Procedures).

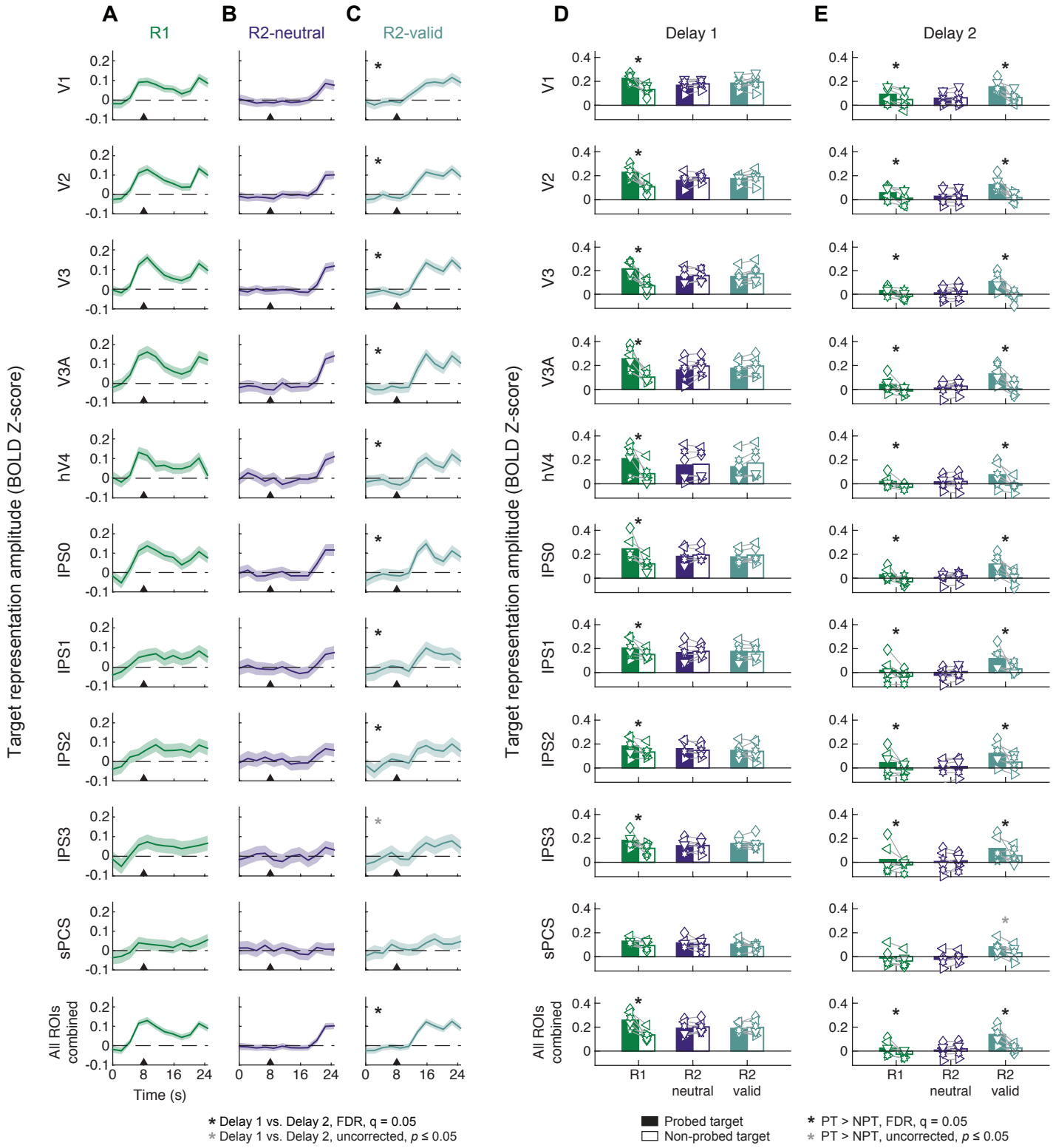
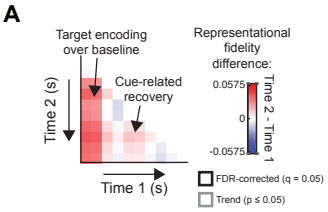
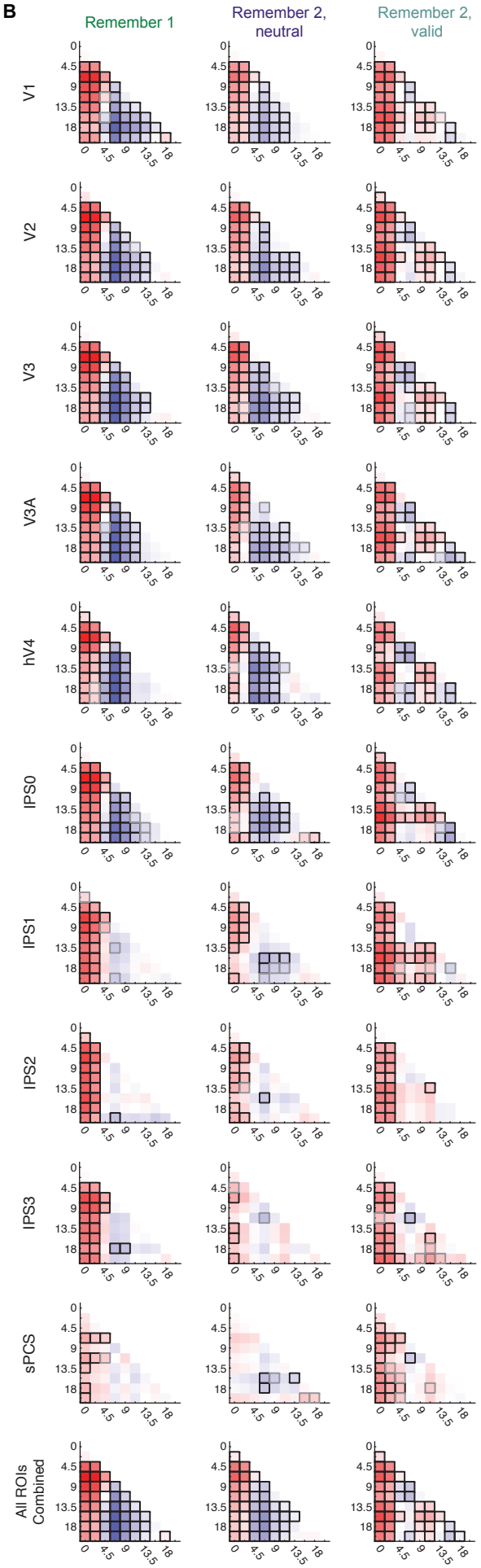


Figure S4 Informative cue shifts target representations from R2- to R1-like state (related to Figures 5-6)
As an alternative visualization of the time course of WM target representations to those shown in Figures 3 and 5, we extracted the activation from each reconstruction within a 0.5° radius circular aperture centered at the exact target positions for each trial. We call this signal the "reconstruction activation", as it reflects BOLD activation patterns transformed into visual field coordinates and extracted at the relevant visual field position. Then, we computed the difference between the activation at the probed target location and the non-probed target location (on R1 trials, the probed target was always the target in WM, on R2-neutral trials, the probed target was the one queried at the end of the trial; on R2-valid trials, the probed target was the validly-cued target in WM).
(A) On R1 trials, the remembered target representation shows elevated activation relative to the non-remembered target representation throughout the entire 16 s delay interval, despite the weakening target representations as visualized in re-

constructions in Figs 3-4.
(B) On R2-neutral trials, both target representations are equal throughout the delay period, with the queried target representation becoming stronger once the response period begins (16.0 s).
(C) On R2-valid trials, we see a transition from R2-neutral-like target representations (both are equal, and so the difference is near zero) during the first delay period to R1-like target representations (the remaining target representations recover) during the second delay period. Black triangle at 8.0 s indicates beginning of second delay interval. Units are BOLD Z-score. Dashed lines mark 95% CI via resampling, see Experimental Procedures: Statistical procedures.
(D-E) We also computed mean delay-period reconstruction activation separately for probed (filled bars) and non-probed (open bars) target positions for each participant individually (each symbol reflects a single participant, as in Fig. 1C; Figure S1) within Delay 1 (D) and Delay 2 (E).
Asterisks in panels A-C indicate a significant change between Delay 1 and Delay 2 (two-tailed); asterisks in panels D-E indicate that the probed target representation activation is greater than the non-probed target representation activation (one-tailed). All statistics computed using a resampling procedure (see Experimental Procedures: Statistical procedures). Black asterisks indicate a significant difference after FDR-correction for multiple comparisons ($q = 0.05$); gray asterisks indicate a non-significant trend defined using an uncorrected threshold of $\alpha = 0.05$. All p-values from this analysis available in Table S6.

Figure S5 Cue-related representation recovery does not depend on time points chosen (related to Figures 5-6)





In all figures reporting mean delay-period activation, representational fidelity, or reconstructions, we used the 3rd and 4th imaging volumes following target presentation (6.75-9.00 s) for Delay 1, and the 7th and 8th volumes for Delay 2 (15.75 and 18.00 s, approximately 8 s following Delay 1, per timing of task events). We used these time points for Delay 1 following observations from our previous report (Sprague, Ester & Serences, 2014) that sensory-related representations in image reconstructions are almost entirely absent 6.75 s following target presentation on trials in which WM maintenance was not required. Delay 2 was accordingly chosen to be ~8 s after Delay 1 concluded. However, inspection of representational fidelity time courses in Fig. 5B suggests that these time points may not reveal the maximum cue-related restoration in representational fidelity. For some regions, the Delay 1 time points average a high-fidelity and lower-fidelity representation, and the Delay 2 time points average a lower-fidelity and higher-fidelity representation. This may conservatively bias our results, and we may miss significant recovery of WM representations in some areas. We addressed this by comparing the difference in representational fidelity between each pair of time points across trials (resampled) against 0, two-tailed. We found at least one pair of significant time points suggestive of post-cue representational recovery in every ROI except sPCS, demonstrating that our general observation of cue-related representational recovery is not contingent on the time points chosen; and, if anything, our choice was conservative. Allowing for different time courses for different ROIs revealed that this phenomenon is widespread throughout the cortex.

(A) Example plot of one region and condition (V1, R2-valid). Each cell plots the signed difference between the representational fidelity for the probed WM target at each paired set of time points (T2 [row] – T1 [column]); for brevity we only plot below-diagonal cells (above-diagonal cells are the negative of their below-diagonal counterparts). Starting at the (blank) diagonal cell, moving down a column indicates the difference between that time point and each successive time point. Left columns are mostly positive, indicating that target representations emerge early in the trial and remain positive. In this R2-valid example, the positive block below the diagonal for later time points corresponds to a relative increase in fidelity between an earlier time point and later time points, consistent with a recovery in the cued WM representation.

(B) Paired time point comparisons plotted for all conditions (columns) and ROIs (rows). Black squares highlight significant cells (two-tailed, FDR-corrected across all ROIs, conditions, and time point pairs, q = 0.05, All ROIs combined corrected separately, see Experimental Procedures); gray squares highlight trends (defined as α = 0.05, uncorrected). All p-values for this analysis available in Table S8, available online as an Excel file.

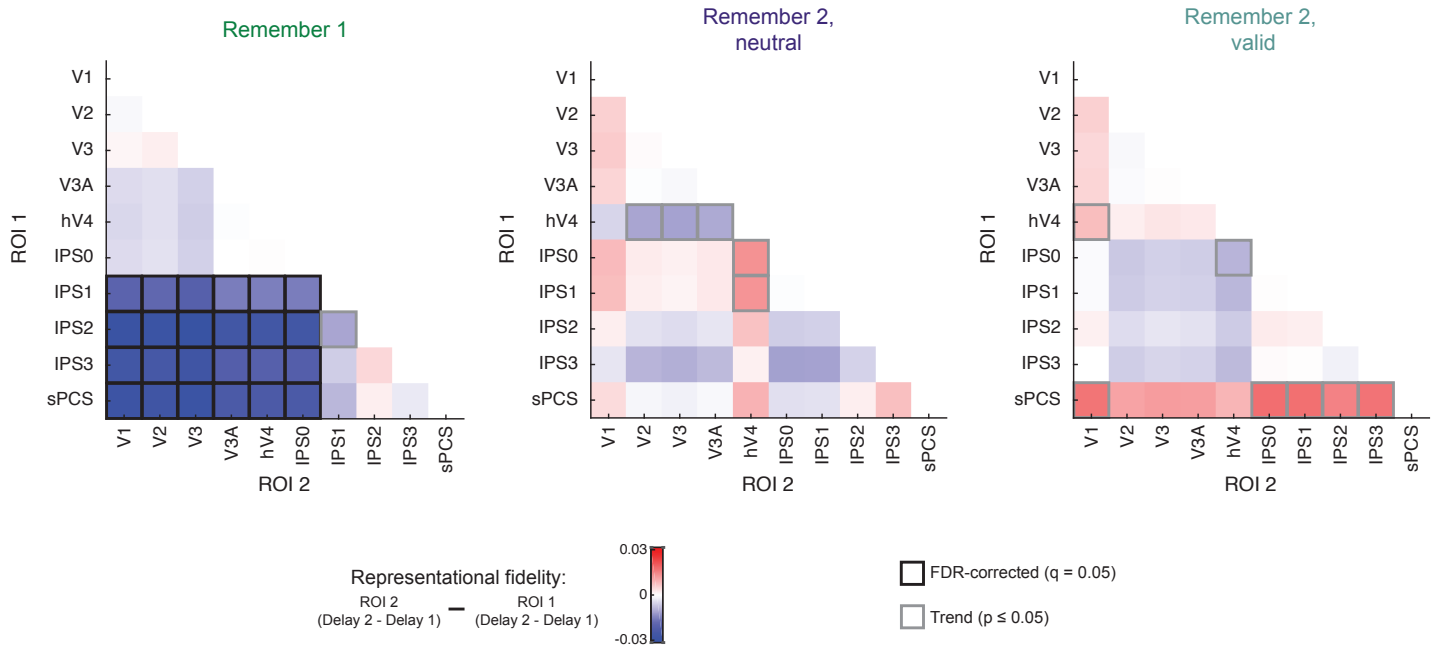


Figure S6 Comparison of delay period-related changes between ROIs (related to Figure 6)

For each ROI pair, we computed the difference in the change in WM representational fidelity (Figs. 5-6) from Delay 1 to Delay 2 (computed as Delay 2 – Delay 1: positive differences reflect increased fidelity; negative differences reflect decreased fidelity) between each non-matching ROI pair. We computed this difference score on each of 1,000 resampling iterations, drawing from all trials concatenated across participants with replacement (Experimental Procedures), and compared the resampled distribution against 0 (two-tailed). Significant differences (FDR-corrected within each condition, q = 0.05) are highlighted with black boxes; trends (defined as p ≤ 0.05, uncorrected) are highlighted with gray boxes. In R1 trials, IPS1, anterior parietal (IPS2-IPS3) and frontal (sPCS) representations remained more stable throughout the entire delay period than did visual (V1-hV4) and posterior parietal (IPS0) representations (see Figs. 3-7 for visualizations of representations in each condition). All p-values for this analysis are reported in Table S7, available online as an Excel file.

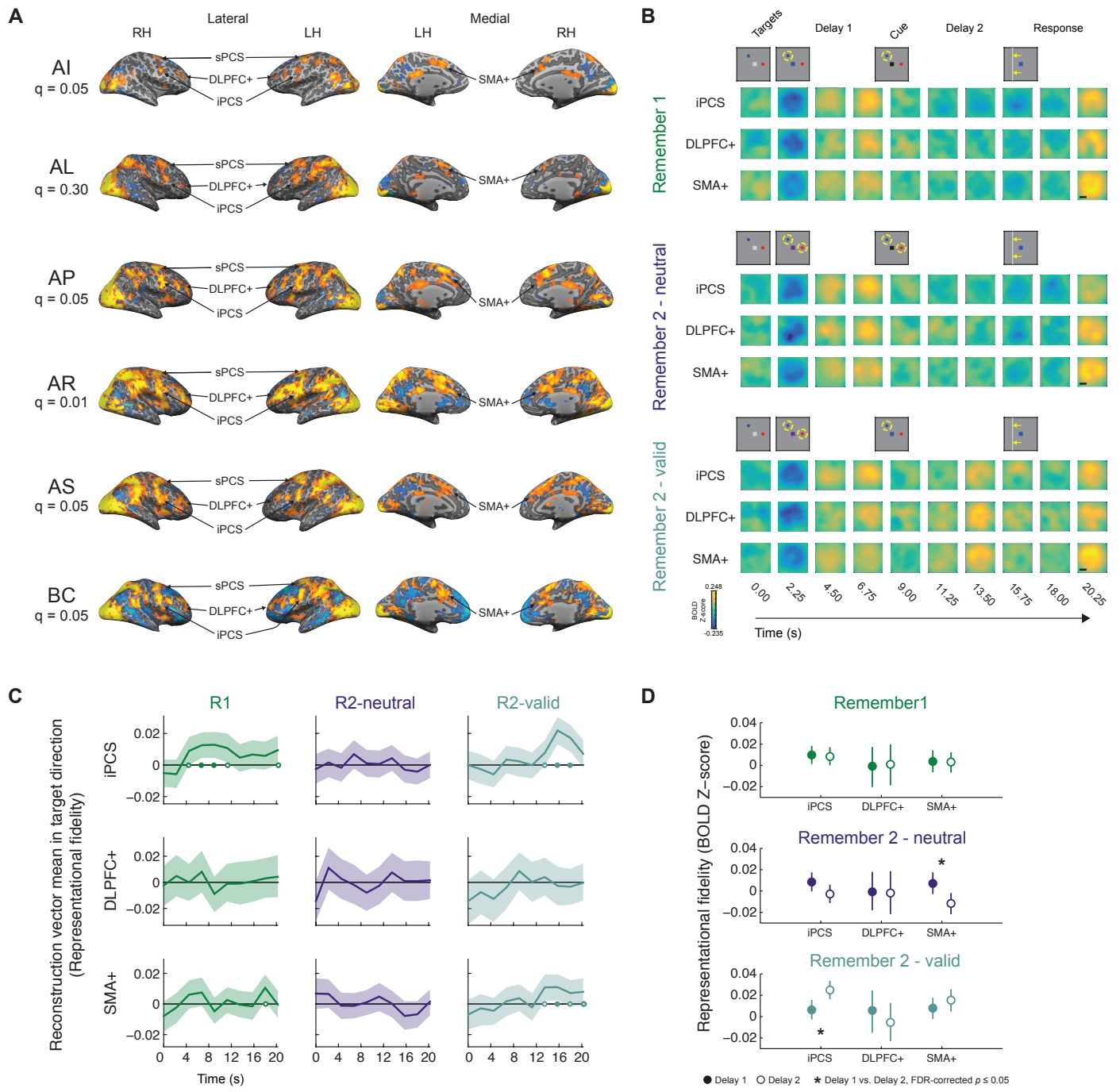


Figure S7 Exploratory analysis of additional prefrontal cortex regions of interest (related to Figs. 3-6)

Previous human neuroimaging work has identified WM representations of a single remembered feature (e.g., orientation or stimulus category) in regions of visual (Albers et al., 2013; Bettencourt and Xu, 2015; Christophel et al., 2012; Ester et al., 2009; Harrison and Tong, 2009; Pratte and Tong, 2014; Riggall and Postle, 2012; Saber et al., 2015; Serences et al., 2009), parietal (Bettencourt and Xu, 2015; Christophel et al., 2012; Ester et al., 2015; Riggall and Postle, 2012; Saber et al., 2015), and frontal cortex (Ester et al., 2015; Lee et al., 2013). Interestingly, prefrontal cortex (PFC) representations seem to depend on the type of information maintained (Lee et al., 2013) and/or the analysis method used (Ester et al., 2015). As an exploratory analysis, we sought to evaluate to what extent spatial WM representations are carried by a subset of PFC regions identified using our spatial localizer task (see Supplemental Experimental Procedures). We identified several additional regions of interest in each participant's prefrontal cortex (PFC) using their visual localizer data (used to restrict voxels analyzed for all analyses).

(A) Activation maps plotted separately for each participant on their individual inflated white/gray matter boundary surface, thresholded as indicated (chosen for each participant to maximize visibility and/or distinctness of activation clusters). Labels with arrows indicate clusters used to identify each PFC ROI (sPCS: superior precentral sulcus; iPCS: inferior

precentral sulcus; DLPFC+: dorsolateral prefrontal cortex and surrounding activation; SMA+ supplementary motor area and surrounding activation, likely including pre-SMA and human supplementary eye fields).

(B) Reconstructions from PFC activation patterns computed through time, as in Fig. 3. Cartoons above each panel indicate coregistered positions of targets, and yellow dashed circles indicate remembered positions at each point during the trial. Only iPCS appears to have an approximate representation of WM targets, and most prominently after the valid cue during R2-valid trials.

(C) Timecourse of representational fidelity (as in Fig. 5B). Filled symbols on horizontal axis indicate significant representations (one-tailed, FDR corrected across all conditions, PFC ROIs, and time points, $q = 0.05$); open symbols indicate trends ($p \leq 0.05$, uncorrected). Shaded regions denote 95% confidence intervals, computed via resampling all trials with replacement (1,000 iterations).

(D) Comparison of representational fidelity between each delay period (as in Fig. 6C). Asterisks indicate significant differences between Delay 1 and Delay 2 (two-tailed, FDR corrected across all conditions and PFC ROIs, resampling test with 1,000 iterations). All p-values for (D) available in Table S3. Because this is an exploratory analysis, correction for multiple comparison for these p-values was conducted independently from correction for p-values used for a priori retinotopic ROIs V1-IPS3 and localizer-defined sPCS, which we have analyzed in a previous report (Sprague et al., 2014).

The absence of WM representations in DLPFC+ is not necessarily surprising in the present study, as its role in the maintenance of spatial positions in humans has recently come into question (Mackey et al., 2016). Additionally, we presented WM targets at ~3.5° eccentricity in this experiment. Insofar as PFC neurons have larger receptive field sizes (often > 20° diameter, Mohler et al., 1973; Zirnsak et al., 2014), a larger stimulus display may result in more robust identification of spatial WM representations.

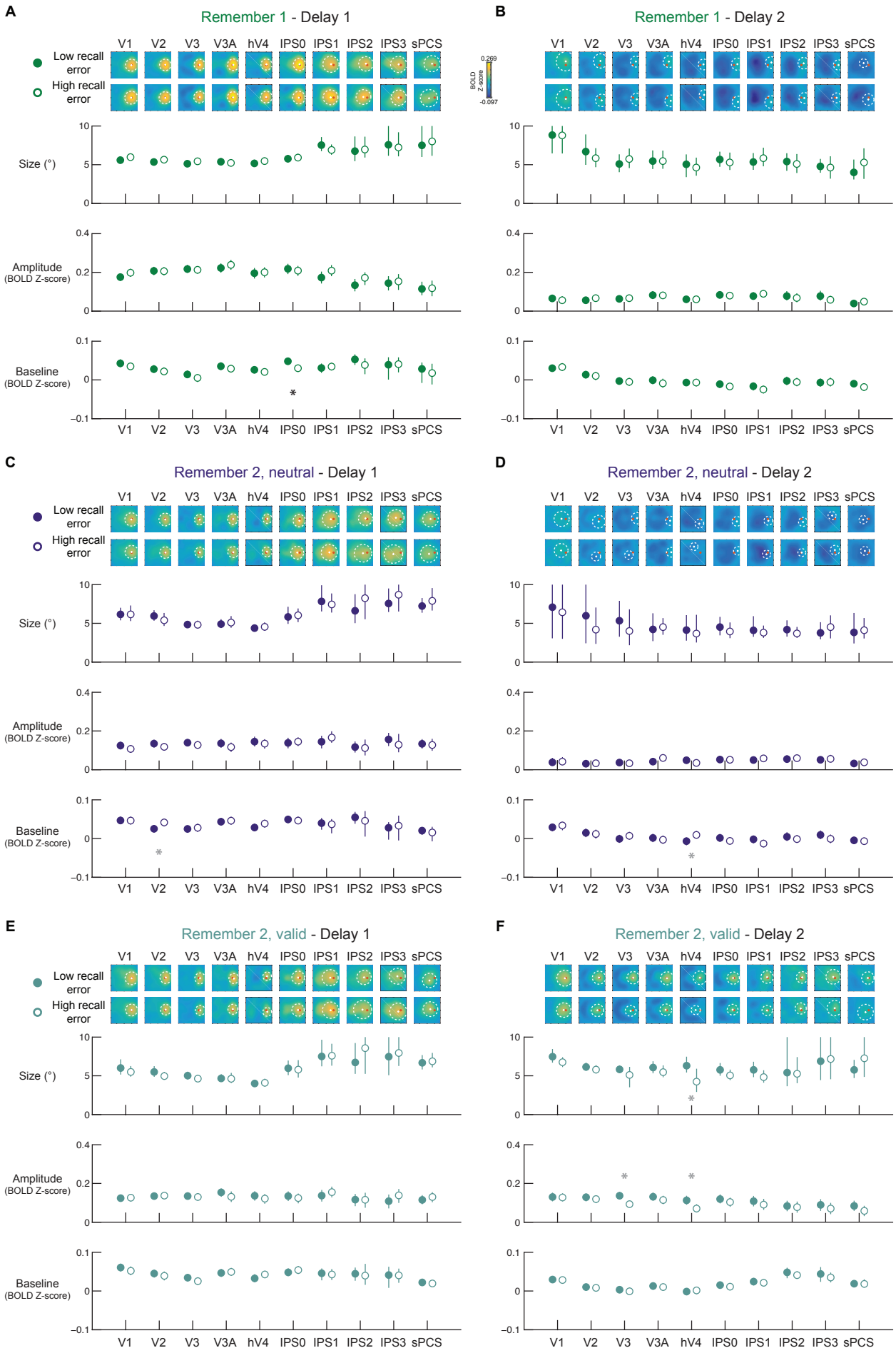

A
Remember 1 - Delay 1
B
Remember 1 - Delay 2
C
Remember 2, neutral - Delay 1
D
Remember 2, neutral - Delay 2
E
Remember 2, valid - Delay 1
F
Remember 2, valid - Delay 2
V1 V2 V3 V3A hV4 IPS0 IPS1 IPS2 IPS3 sPCS
Low recall error
High recall error
BOLD Z-score
0.269
-0.097
Size (°)
Amplitude (BOLD Z-score)
Baseline (BOLD Z-score)

Figure S8 Comparison of quantified WM target representations across performance bins for each individual ROI (related to Figure 8)

Data plotted as in Figure 8, with trials sorted based on behavioral recall performance. All resampling and fitting procedures are identical to those used for Figure 8. All error bars 95% confidence intervals over resampled fitting iterations. Black asterisks indicate significant difference between low- and high-recall error trials for that WM condition, delay period, and fit parameter, FDR-corrected for multiple comparisons within each parameter ($q = 0.05$). Gray asterisks are trends, thresholded at $\alpha = 0.05$, uncorrected for multiple comparisons. All p-values available in Table S5.

(A) Remember 1 trials, Delay 1. IPS0 baseline was significantly higher on low recall error trials ($p < 0.001$).

(B) Remember 1 trials, Delay 2.

(C) Remember 2-neutral trials, Delay 1. V2 baseline trended to be lower for low recall error trials ($p = 0.02$).

(D) Remember 2-neutral trials, Delay 2. In hV4, baseline trended to be smaller for low recall error trials ($p = 0.01$).

(E) Remember 2-valid trials, Delay 1.

(F) Remember 2-valid trials, Delay 2. In V3 and hV4, amplitude trended towards being larger on low recall error trials ($p = 0.002$ and $p = 0.004$, respectively), and size trended towards being larger on low recall error trials in hV4 ($p = 0.048$).

Figure S2	**Delay 1**			**Delay 2**		
Comparison:	R1 vs R2-neutral	R2-neutral vs. R2-valid	R1 vs R2-valid	R1 vs R2-neutral	R2-neutral vs. R2-valid	R1 vs R2-valid
V1	0.754	0.068	0.134	0.954	*0.018*	*0.018*
V2	0.736	0.182	0.324	0.642	*0.036*	0.13
V3	0.766	0.118	0.056	0.396	*0.014*	0.184
V3A	0.676	0.104	0.24	0.326	*0.024*	0.284
hV4	*0.02*	0.35	**0.002**	*0.034*	**0**	*0.04*
IPS0	0.468	0.152	*0.046*	0.396	**0**	**0**
IPS1	0.326	0.16	*0.016*	0.854	**0**	**0**
IPS2	0.496	0.192	0.062	0.676	**0**	**0**
IPS3	0.808	0.29	0.214	0.652	**0**	**0**
sPCS	0.33	0.072	**0**	0.346	**0**	**0**

Table S1 (related to Figure S2)

P-values for comparisons of mean delay-period activation over all voxels within each ROI between WM conditions (two-tailed). All *p*-values reflect pair-wise comparisons between conditions (R1 vs R2-neutral, R2-neutral vs. R2-valid, and R1 vs. R2-valid). P-value of 0 indicates $p < 0.001$, the minimum p-value achievable per our resampling procedure with 1,000 iterations. Bold numbers indicate significant differences after FDR correction for all comparisons ($q = 0.05$, all comparisons and all individual ROIs). Italicized numbers indicate trends, defined using $\alpha = 0.05$, uncorrected. Significant comparisons and trends are shown graphically in Figure S2. FDR threshold for V1-sPCS is $p \leq 0.002$

Representational fidelity (Fig. 6A-B and Fig. S7)	Delay 1 (one-tailed)			Delay 2 (one-tailed)		
	R1	R2-neutral	R2-valid	R1	R2-neutral	R2-valid
V1	**0**	**0**	**0**	**0**	**0**	**0**
V2	**0**	**0**	**0**	**0**	**0**	**0**
V3	**0**	**0**	**0**	**0**	**0**	**0**
V3A	**0**	**0**	**0**	**0**	0.109	**0**
hV4	**0**	**0**	**0**	**0**	**0**	**0**
IPS0	**0**	**0**	**0**	**0**	**0.034**	**0**
IPS1	**0**	**0**	**0**	**0**	0.462	**0**
IPS2	**0**	**0**	**0**	**0**	0.35	**0**
IPS3	**0**	**0**	**0.001**	**0**	**0.013**	**0**
sPCS	**0**	**0**	**0**	**0**	0.784	**0**
All ROIs combined	**0**	**0**	**0**	**0**	**0**	**0**
iPCS	*0.017*	*0.028*	0.092	*0.027*	0.755	**0**
DLPFC+	0.52	0.541	0.272	0.469	0.572	0.721
SMA+	0.237	0.085	0.07	0.261	0.991	**0.002**

Table S2 (related to Figure 6 and Figure S7)

P-values for comparisons between representational fidelity computed separately within each WM delay (one-tailed, against the null hypothesis that representational fidelity ≤ 0). P-value of 0 indicates $p < 0.001$, the minimum p-value achievable per our resampling procedure with 1,000 iterations. Bold numbers indicate significant differences after FDR correction for all comparisons ($q = 0.05$, all conditions and all individual *a priori* ROIs [V1-sPCS], and separately for "All ROIs combined" and the exploratory PFC ROIs [Figure S7; iPCS-SMA+] across all conditions, see Experimental Procedures). Italicized numbers indicate trends, defined using $\alpha = 0.05$, uncorrected. Significant comparisons and trends are shown graphically in Figure 6A-B. FDR threshold for V1-sPCS is $p \leq 0.034$, for All ROIs combined is $p < 0.001$, and for PFC ROIs is $p \leq 0.002$.

Representational fidelity (Fig. 6C; Fig. S7D)	Delay 1 vs. Delay 2		
	R1	R2-neutral	R2-valid
V1	**0**	**0**	**0**
V2	**0**	**0**	0.072
V3	**0**	**0**	*0.046*
V3A	**0**	**0**	0.072
hV4	**0**	*0.044*	0.282
IPS0	**0**	**0**	**0**
IPS1	*0.028*	**0**	**0.018**
IPS2	0.848	**0**	**0.054**
IPS3	0.362	0.122	0.07
sPCS	0.538	**0**	0.39
All ROIs combined	**0**	**0**	**0.004**
iPCS	0.84	0.062	**0.006**
DLPFC+	0.918	0.928	0.404
SMA+	0.934	**0.008**	0.286

Table S3 (related to Figure 6 and Figure S7)

P-values for comparisons of representational fidelity between Delay 1 and Delay 2 (two-tailed). P-value of 0 indicates $p < 0.001$, the minimum p-value achievable per our resampling procedure with 1,000 iterations. Bold numbers indicate significant differences after FDR correction for all comparisons ($q = 0.05$, all conditions and all individual *a priori* ROIs [V1-sPCS], separately for "All ROIs combined", and separately for exploratory PFC ROIs [iPCS-SMA+] across all conditions, see Experimental Procedures). Italicized numbers indicate trends, defined using $\alpha = 0.05$, uncorrected. Significant comparisons and trends are shown in Figure 6C and Figure 7D. FDR threshold for V1-sPCS is $p \leq 0.028$, for All ROIs combined is $p \leq 0.002$, and for PFC ROIs is $p \leq 0.008$.

Fig. 7	Param:	**Size**			**Amplitude**			**Baseline**		
Delay	ROI	R1 vs R2-neutral	R2-neutral vs R2-valid	R1 vs R2-valid	R1 vs R2-neutral	R2-neutral vs R2-valid	R1 vs R2-valid	R1 vs R2-neutral	R2-neutral vs R2-valid	R1 vs R2-valid
1	V1	0.27	0.278	0.78	**0**	0.328	**0**	0.144	*0.022*	**0.002**
1	V2	0.506	0.092	0.182	**0**	0.24	**0**	0.138	0.07	**0**
1	V3	*0.032*	0.878	*0.02*	**0**	0.94	**0**	**0.002**	0.422	**0**
1	V3A	0.3	0.31	*0.04*	**0**	0.214	**0**	**0.018**	0.512	**0**
1	hV4	**0**	0.594	**0**	**0**	0.106	**0**	**0.016**	0.446	**0.002**
1	IPS0	0.858	0.852	0.702	**0**	0.368	**0**	0.052	0.664	**0.01**
1	IPS1	0.304	0.852	0.464	**0.016**	0.5	**0.004**	0.408	0.346	0.082
1	IPS2	0.706	0.702	0.3	**0.012**	0.556	*0.042*	0.396	0.198	0.594
1	IPS3	0.15	0.576	0.354	0.938	0.394	0.242	0.21	0.43	0.592
1	sPCS	0.788	0.488	0.478	0.422	0.872	0.474	0.418	0.5	0.354
1	All ROIs	0.822	0.256	0.108	**0**	0.818	**0**	**0**	0.494	**0**
2	V1	0.572	0.964	*0.022*	0.072	**0**	**0**	0.89	0.762	0.386
2	V2	0.626	0.842	0.152	**0**	**0**	**0**	0.804	0.368	0.378
2	V3	0.498	0.728	0.41	**0**	**0**	**0**	0.104	0.462	0.354
2	V3A	*0.016*	0.096	0.292	**0.004**	**0**	**0**	0.35	*0.028*	**0**
2	hV4	0.282	0.066	0.244	**0.002**	**0**	**0**	*0.048*	0.48	0.212
2	IPS0	0.06	0.1	0.492	**0**	**0**	**0.002**	**0.012**	**0**	**0**
2	IPS1	**0.002**	*0.03*	0.112	**0**	**0**	0.118	**0.004**	**0**	**0**
2	IPS2	*0.046*	0.116	0.92	0.074	*0.042*	0.626	0.224	**0**	**0**
2	IPS3	0.398	*0.024*	0.072	0.084	**0.016**	0.366	*0.044*	**0**	**0**
2	sPCS	0.584	*0.016*	0.114	0.122	**0**	**0.016**	0.07	**0**	**0**
2	All ROIs	0.17	0.23	0.578	**0**	**0**	**0**	**0.018**	**0**	**0**
FDR thresh:	V1-sPCS	0.002			V1-sPCS	0.016		V1-sPCS	0.018	
	All ROIs	< 0.001			All ROIs	< 0.001		All ROIs	0.018	

Table S4 (related to Figure 7)

P-values for comparisons of parameters (size, amplitude, baseline) for best-fit surfaces to coregistered reconstructions between conditions for each delay period individually (two-tailed). All p-values reflect pair-wise comparisons between conditions (R1 vs R2-neutral, R2-neutral vs. R2-valid, and R1 vs. R2-valid). P-value of 0 indicates $p < 0.001$, the minimum p-value achievable per our resampling procedure with 1,000 iterations. Bold numbers indicate significant differences after FDR correction for all comparisons ($q = 0.05$, all conditions and all individual ROIs, and separately for "All ROIs combined" across all conditions, see Experimental Procedures). Italicized numbers indicate trends, defined using $\alpha = 0.05$, uncorrected. Significant comparisons and trends are shown graphically in Figure 7.

Fig. 8 and Fig. S8	Parameter:	**Size**			**Amplitude**			**Baseline**		
Delay	ROI	R1	R2-neutral	R2-valid	R1	R2-neutral	R2-valid	R1	R2-neutral	R2-valid
1	V1	0.2	0.996	0.374	0.124	0.254	0.874	0.232	0.952	0.232
1	V2	0.178	0.304	0.206	0.954	0.232	0.836	0.434	*0.02*	0.48
1	V3	0.144	0.964	0.25	0.734	0.348	0.748	0.196	0.638	0.202
1	V3A	0.554	0.756	0.854	0.312	0.258	0.19	0.364	0.722	0.704
1	hV4	0.344	0.616	0.862	0.742	0.532	0.414	0.396	0.14	0.132
1	IPS0	0.602	0.692	0.856	0.568	0.734	0.55	**0**	0.598	0.43
1	IPS1	0.294	0.706	0.908	0.06	0.284	0.392	0.59	0.818	0.784
1	IPS2	0.818	0.372	0.278	0.094	0.796	0.982	0.202	0.584	0.696
1	IPS3	0.708	0.416	0.722	0.676	0.344	0.246	0.938	0.72	0.906
1	sPCS	0.698	0.442	0.808	0.936	0.746	0.448	0.578	0.744	0.67
1	All ROIs	0.236	0.71	0.286	0.972	*0.036*	0.524	0.056	0.396	0.824
2	V1	0.96	0.886	0.146	0.472	0.796	0.884	0.624	0.492	0.872
2	V2	0.462	0.644	0.364	0.268	0.762	0.46	0.598	0.72	0.756
2	V3	0.496	0.44	0.164	0.698	0.636	0.002	0.722	0.268	0.526
2	V3A	0.968	0.78	0.268	0.938	0.102	0.266	0.294	0.556	0.688
2	hV4	0.608	0.646	*0.048*	0.976	0.164	*0.004*	0.988	*0.01*	0.634
2	IPS0	0.568	0.454	0.174	0.744	0.97	0.28	0.348	0.226	0.436
2	IPS1	0.576	0.756	0.144	0.428	0.474	0.326	0.29	0.116	0.622
2	IPS2	0.674	0.42	0.96	0.526	0.63	0.76	0.724	0.416	0.356
2	IPS3	0.878	0.412	0.9	0.202	0.698	0.394	0.906	0.178	0.362
2	sPCS	0.364	0.696	0.252	0.35	0.514	0.206	0.178	0.776	0.962
2	All ROIs	0.714	0.432	*0.03*	0.422	0.96	**0**	0.374	0.382	0.88
FDR thresh:	V1-sPCS	<0.001			V1-sPCS	< 0.001		V1-sPCS	<0.001	
	All ROIs	.0083 (Bonferroni)			All ROIs	0.0083		All ROIs	0.0083	

Table S5 (related to Figure 8 and Figure S8)

P-values for comparisons of parameters (size, amplitude, baseline) for best-fit surfaces to coregistered reconstructions between low recall error and high recall error trials (two-tailed, always equal number of trials in each bin per participant and session). P-value of 0 indicates $p < 0.001$, the minimum p-value achievable per our resampling procedure with 1,000 iterations. Bold numbers indicate significant differences after FDR correction for all comparisons within each parameter ($q = 0.05$, all conditions and all individual ROIs, and separately for "All ROIs combined" across all conditions, see Experimental Procedures; FDR thresholds indicated at bottom of table). Italicized numbers indicate trends, defined using $\alpha = 0.05$, uncorrected. Significant comparisons and trends are shown graphically in Figure 8 and Figure S8. For the All ROIs combined comparisons, use of a threshold derived with Bonferroni's method produces identical significant comparisons.

Fig. S4A-C	V1	V2	V3	V3A	hV4	IPS0	IPS1	IPS2	IPS3	sPCS	All ROIs combined
Remember 1	0.522	0.904	0.556	0.226	0.246	0.262	0.31	0.126	0.35	0.756	0.932
Remember 2 - neutral	0.764	0.3	0.556	0.788	0.57	0.944	0.382	0.2	0.682	0.234	0.862
Remember 2 - valid	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	*0.028*	0.38	**0**

Table S6A (related to Figure S4A-C)

P-values for comparisons of target activation differences (probed target – non-probed target) between Delay 1 and Delay 2 (two-tailed). P-value of 0 indicates $p < 0.001$, the minimum p-value achievable per our resampling procedure with 1,000 iterations. Bold numbers indicate significant differences after FDR correction for all comparisons ($q = 0.05$, all conditions and all individual ROIs, and separately for "All ROIs combined" across all conditions, see Experimental Procedures). Italicized numbers indicate trends, defined using $\alpha = 0.05$, uncorrected. Significant comparisons and trends are shown in Figure S4A-C. FDR thresholds for V1-sPCS and for All ROIs combined are $p < 0.001$. Identical comparisons remain significant when correcting with Bonferroni's method.

Condition:	Remember 1		Remember 2 - neutral		Remember 2- valid	
Delay:	Delay 1	Delay 2	Delay 1	Delay 2	Delay 1	Delay 2
V1	**0**	**0**	0.919	0.874	0.826	**0**
V2	**0**	**0**	0.959	0.61	0.796	**0**
V3	**0**	**0**	0.626	0.887	0.912	**0**
V3A	**0**	**0**	0.974	0.961	0.981	**0**
hV4	**0**	**0**	0.588	0.868	0.929	**0**
IPS0	**0**	**0**	0.937	0.934	0.828	**0**
IPS1	*0.006*	**0**	0.727	0.968	0.518	**0**
IPS2	**0.012**	**0**	0.173	0.772	0.504	**0**
IPS3	**0.013**	**0**	0.191	0.393	0.304	**0.001**
sPCS	0.127	0.068	0.153	0.733	0.217	*0.029*
All ROIs combined	**0**	**0**	0.925	0.963	0.85	**0**

Table S6B (related to Figure S4D)

P-values for comparisons between probed target (PT) activation and non-probed target (NPT) activation computed separately within each WM delay (one-tailed, against the null hypothesis that PT ≤ NPT). P-value of 0 indicates $p < 0.001$, the minimum p-value achievable per our resampling procedure with 1,000 iterations. Bold numbers indicate significant differences after FDR correction for all comparisons ($q = 0.05$, all conditions and all individual ROIs, and separately for "All ROIs combined" across all conditions, see Experimental Procedures). Italicized numbers indicate trends, defined using $\alpha = 0.05$, uncorrected. Significant comparisons and trends are shown in Figure S4D. FDR threshold for V1-sPCS is $p \leq 0.013$ and for All ROIs combined is $p < 0.001$.

Table S7 (related to Figure S5)

(Available in Sprague Ester Serences_TableS7.xlsx)

P-values for comparisons of restoration in representational fidelity (Delay 2 – Delay 1) between each pair of ROIs. All tests are two-tailed against the null hypothesis of no difference in restoration effect between each pair of ROIs. Italicized numbers indicate trends, defined using $\alpha = 0.05$, uncorrected. Bold indicates significant differences, FDR-corrected within each condition ($q = 0.05$; FDR thresholds for R1: $p \leq 0.014$, R2-neutral: $p < 0.001$, R2-valid: $p < 0.001$).

Table S8 (related to Figure S6)

(Available in Sprague Ester Serences_TableS8.xlsx)

P-values for comparisons of representational fidelity between each pair of time points (Time 2 – Time 1) for each condition and ROI. All tests are two-tailed against the null hypothesis of no difference in representational fidelity between each pair of time points. FDR threshold for V1-sPCS: $p \leq 0.028$, All ROIs combined: $p \leq 0.028$ (indicated with black squares in Figure S6).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Participants

We recruited *n* = 6 participants (5 female; aged 22-29 yrs) naïve to the purpose of the experiment from the UC San Diego community. We used a small sample size, but acquired substantial data from each participant to maximize sensitivity to subtle WM representations, similar to our previous report (Sprague et al., 2014). Participant identifiers are identical to those used in previous reports to facilitate comparison of data across experiments (Ester et al., 2015; Sprague and Serences, 2013; Sprague et al., 2014). Participants AI and AL participated in the experiments reported in Sprague & Serences (2013). Participant AI participated in the experiments reported in Sprague et al (2014). Participants AI, AL and AP participated in Ester et al (2015). Participants gave written informed consent as approved by the UCSD Institutional Review Board and received monetary compensation for their time ($20/hr for fMRI sessions, $10/hr for behavioral sessions).

Spatial WM retro-cueing task

All participants underwent 3 fMRI scanning sessions and 1 retinotopic mapping scanning session, each lasting 2 hrs. Participants also completed 2-4 behavioral sessions, each lasting 1-1.5 hrs. The size of the stimulus display was fixed across all behavioral and scanning sessions. However, the size of the screen, which constantly contained a gray background, differed (inside scanner: 18.18° × 13.64°, aspect ratio 4:3; outside scanner: 44.71° × 25.15°, aspect ratio 16:9).

We adapted a spatial WM task reported previously (Sprague et al., 2014). On each trial, we presented 2 target stimuli (a red and a blue dot, 0.15° diameter) for 500 ms at pseudorandom locations 3.5° from fixation on average. Following target presentation, the fixation point (square, 0.2°/side) immediately changed color to either red, blue, or purple. A red or blue fixation cue (1/3 of trials) indicated the target to be maintained in WM over the delay interval (Remember 1). A purple fixation cue (2/3 of trials) indicated both targets should be maintained in WM (Remember 2). After an 8,000 ms delay interval (Delay 1), the fixation cue changed color once again. On Remember 1 trials, the cue always changed to black, indicating participants should maintain the encoded target in WM over the subsequent second delay interval. On ½ of Remember 2 trials (1/3 of trials overall), the fixation cue turned black, providing a neutral cue as to which target was relevant for behavior (Remember 2-neutral condition). On the remaining ½ of Remember 2 trials (1/3 of trials overall), the fixation cue changed from purple to either red or blue, cueing the participants with 100% validity to remember only one of the targets (Remember 2-valid condition). Following this cue change, participants continued to maintain 1 or 2 items in WM over an additional 8,000 ms delay interval (Delay 2).

At the end of each trial after both delay intervals, participants recalled the exact horizontal or vertical coordinate of the item cued by the color of the fixation point. The response coordinate was randomly chosen on every trial so that participants could not implement a uni-dimensional encoding scheme (i.e., encode only *x* or *y* coordinate). Participants responded by adjusting the position of a gray horizontal bar up or down (for *y* coordinate trials) or a vertical bar left or right (for *x* coordinate trials) using either a computer keyboard or an MR-compatible button box (bar thickness: 0.02°). We took the adjusted bar coordinate at the end of a 3,000 ms response window as the participant's response.

Target locations were drawn from an isoeccentric ring 3.5° from fixation at 60° polar angle intervals along the ring, where the starting angle was jittered by up to ±15° on each trial. The position of the second target relative to the first target was always offset from the first target by 60°, 120°, or 180° in either direction (clockwise or counterclockwise, see colored discs in Fig. 1B). This resulted in a minimum target separation distance of 2.3° and a maximum separation distance of 8.2°. By using random target positions on each trial, we ensured that participants maintained precise spatial locations in WM rather than using alternative coding strategies, like verbally labeling the location(s). Additionally, constraining relative target positions within one of several discs allowed for comparison of data from trials with similar target arrangements (Figs. 3-4).

We counterbalanced trials for target position (1 of 6 discs), relative target position (1 of 3 relative angular separation distances, Fig. 1B), and memory condition (Remember 1, Remember 2-neutral, or Remember 2-valid), resulting in 54 trials per full counterbalanced repetition. Each full set of trials (or "super-run") was broken up into 3 runs, each with 18 trials, each 19.5 s long. Trials were separated by a random inter-trial interval chosen from a uniform distribution from 3 to 6 s.

Spatial mapping task

Inside the scanner, participants completed 4 runs per session of a spatial mapping task used to estimate voxel-level encoding models for reconstruction analyses described below. On each trial, participants remembered the exact position of

a single target stimulus over a 3,000 ms delay interval during which a flickering checkerboard disc (6 Hz, full-field flicker, 1.083° radius, 1.474 cycles/°; Figure S3A) appeared nearby the memorized location. Following checkerboard presentation, participants indicated whether a probe stimulus (black dot) was either to the left or right or above or below the remembered stimulus position, as cued by an oriented bar at fixation (horizontal bar: respond left vs. right; vertical bar: respond above vs. below; probe and response bar presented for 750 ms). Participants could respond until the beginning of the next trial (after 2,000 – 4,500 ms inter-trial interval, uniform distribution). We maintained performance at ~75% correct by adjusting the target-probe separation distance between runs, but due to a programming error, accuracy was computed incorrectly during scanning ("null" trials were counted as incorrect responses, so accuracy on task trials was actually ~89%, not ~75% as computed within the stimulus presentation script, Figure S3 caption). To ensure participants did not just encode one target coordinate dimension (*x* or *y*), we jittered the irrelevant coordinate on each trial by a small amount, preventing a scenario in which the presentation of the probe stimulus added certainty to the position maintained in WM. Each run included 6 null trials (no target/mapping stimulus/probe presented) during which participants passively fixated until the subsequent trial began.

During each run of this spatial mapping task the checkerboard stimuli were presented at each of 36 positions arrayed along a hexagonal grid (see Figure S3B-C) and the target position was randomly chosen from a uniform disc centered at the checkerboard position with radius 0.542°. On each run, we rotated the angular orientation of the entire hexagonal grid by 15° polar angle (Figure S3C). Across sessions, we rotated the "baseline" angular orientation of the grid by 5° polar angle. This resulted in $4 \times 3 \times 36 = 432$ unique stimulus positions across all scanning sessions. We used different grid orientations (and thus stimulus positions) on each scanning run to maximize the number of unique stimulus positions so that we could estimate as robust a spatial encoding model as possible (see below), as well as to ensure our model was not identifying peculiarities specific to a given set of mapping stimulus positions.

Localizer task

To focus our neuroimaging analyses to voxels responsive during spatial WM maintenance over the area subtended by our display setup, we scanned each participant on 6-8 runs (AI: 6, AL: 7, AS: 8, AR: 7, AP: 7, BC: 8) of a visual spatial WM localizer task similar to one we have described before (Sprague et al., 2014). On each trial we presented a flickering radial checkerboard annulus in one visual hemifield extending from 0.8° to 6.0° from fixation (1.25 cycles/° from fixation, 12° per polar angle cycle, 6 Hz contrast-reversing) for 10 s. During the stimulus interval, we presented 2 spatial WM trials in which participants remembered the precise position of 1 red dot over a 3 s delay interval. At the end of each delay interval, participants responded whether a green probe stimulus was to the left or to the right, or above or below, the remembered target position as indicated by a horizontal or vertical bar at fixation, respectively. WM targets could only appear within the stimulated hemifield. We maintained performance at ~75% by adjusting the task difficulty (target/probe separation distance) across trials. Stimulus epochs were separated by 3 – 5 s ITIs (uniform distribution). Each run contained 4 null trials that were the same duration as normal trials but did not contain checkerboard stimuli.

Behavioral analysis

For the main WM task, we defined behavioral recall error as the absolute distance along the relevant coordinate dimension (*x* or *y*) between the position of the response bar at the conclusion of the response window and the actual coordinate of the recalled target. We averaged all recall errors across all trials from scanning sessions within each participant.

In fMRI analyses in which we split trials based on behavioral performance, we computed the median recall error within each WM condition (R1, R2-neutral, R2-valid) within each scanning session. Trials with recall error greater than or equal to the median value were labeled "high recall error" and trials with recall error less than the median value were labeled "low recall error" (Figure 8 and Figure S8, Table S5).

fMRI acquisition

We scanned all participants on a 3 T research-dedicated GE MR750 scanner located at the UCSD Keck Center for Functional Magnetic Resonance Imaging with a 32 channel send/receive head coil (Nova Medical, Wilmington, MA) using identical sequences to those we have reported previously (Sprague and Serences, 2013; Sprague et al., 2014). We acquired functional data using a gradient echo planar imaging (EPI) pulse sequence (19.2 × 19.2 cm field of view, 96 × 96 matrix size, 31 3-mm-thick slices with 0-mm gap, obliquely-oriented through occipital, parietal & dorsal frontal cortex, TR = 2,250 ms, TE = 30 ms, flip angle = 90°, voxel size 2×2×3 mm, xyz).

To anatomically coregister images across sessions, and within each session, we also acquired a high resolution anatomical scan during each scanning session (FSPGR T1-weighted sequence, TR/TE = 11/3.3 ms, TI = 1,100 ms, 172 slices, flip angle = 18°, 1 mm^3 resolution). For all sessions but one, anatomical scans were acquired with ASSET acceleration. For the remaining session, we used an 8 channel send/receive head coil and no ASSET acceleration to acquire anatomical images with minimal signal inhomogeneity near the coil surface, which enabled improved segmentation of the gray-white matter boundary. We transformed these anatomical images to Talairach space and then reconstructed the gray/white matter surface boundary in BrainVoyager 2.6.1 (BrainInnovations) which we used for identifying ROIs.

fMRI preprocessing

We preprocessed fMRI data similarly to our previous report (Sprague et al., 2014). We coregistered functional images to a common anatomical scan across sessions (used to identify gray/white matter surface boundary as described above) by first aligning all functional images within a session to that session's anatomical scan, then aligning that session's scan to the common anatomical scan. We performed all preprocessing using FSL (Oxford, UK) and BrainVoyager 2.6.1 (BrainInnovations). Preprocessing included unwarping the EPI images using routines provided by FSL, then slice-time correction, three-dimensional motion correction (six-parameter affine transform), temporal high-pass filtering (to remove first-, second- and third-order drift), transformation to Talairach space (resampling to 2×2×2 mm resolution) in BrainVoyager, and finally normalization of signal amplitudes by converting to Z-scores separately for each run using custom MATLAB scripts. We did not perform any spatial smoothing beyond the smoothing introduced by resampling during the co-registration of the functional images, motion correction and transformation to Talairach space. All subsequent analyses were computed using custom code written in MATLAB (release 2014b, The Mathworks, Inc).

One participant (AS) changed positions inside the scanner substantially during one session. As a result, the field inhomogeneities estimated with the field map scan used for unwarping were only accurate for half of the runs during this session and could not be used to unwarp the other half of scans. To mitigate this problem with the raw data, we did not perform unwarping on any session for this participant in order to maintain consistency in the analysis procedure across sessions for this participant. This did not appear to affect any aspect of their results.

Identifying regions of interest (ROIs)

Based on our previous work, we identified 10 *a priori* ROIs using independent scanning runs from those used for all analyses reported in the text. For retinotopic ROIs (V1-V3, hV4, V3A, IPS0-IPS3), we utilized a combination of retinotopic mapping techniques. Each participant completed several scans of meridian mapping in which we alternately presented flickering checkerboard "bowties" along the horizontal and vertical meridians. Additionally, each participant completed several runs of an attention-demanding polar angle mapping task in which they detected brief contrast changes of a slowly-rotating checkerboard wedge (described in detail in Sprague and Serences, 2013). We used a combination of maps of visual field meridians and polar angle preference for each voxel to identify retinotopic ROIs (Engel et al., 1994; Swisher et al., 2007). Polar angle maps computed using the attention-demanding mapping task for most participants are available in previous publications (AI: Sprague & Serences, 2013; AL and AP: Ester et al., 2015). We combined left- and right-hemispheres for all ROIs, as well as dorsal and ventral aspects of V2 and V3 for all analyses by concatenating voxels.

We defined superior precentral sulcus (sPCS) by plotting voxels active during either the left or right conditions of the localizer task described above (FDR corrected, $q = 0.05$) on the reconstructed gray/white matter boundary of each participant's brain and manually identifying clusters appearing near the superior portion of the precentral sulcus, following previous reports (Srimal and Curtis, 2008). Additionally, for an exploratory *post-hoc* analysis of prefrontal cortex ROIs, we used activation maps from this localizer to identify significant voxels nearby the inferior aspect of the precentral sulcus (iPCS), dorsolateral prefrontal cortex (DLPFC+), and a medial region comprising the supplementary and pre-supplementary motor areas (SMA+). Because these ROIs were not always observable at the rigorous FDR-corrected threshold used to identify sPCS (an *a priori* chosen ROI), in some participants we adjusted the statistical threshold to maximize visibility and/or discriminability of the activation patches (see Fig. S7A).

The "All ROIs combined" region reported throughout the text consists of all voxels from all 10 individual *a priori* ROIs (V1, V2, V3, V3A, hV4, IPS0, IPS1, IPS2, IPS3, sPCS) concatenated together, and so all multivariate analyses involving this ROI reflect the net information content of the entire set of regions studied (reported also in Sprague et al., 2014).

fMRI analysis: univariate

For all ROI analyses, we used data from the localizer scans to identify voxels significantly active during checkerboard stimulus presentation and WM maintenance (FDR corrected, $q = 0.05$) for inclusion in further analyses. All analyses include only those voxels.

We computed BOLD time series by extracting signal at each time point averaged over all voxels within an ROI on each trial from 0 to 24.75 s (0 to 11 TRs) after the beginning of the first delay (rounded to the nearest TR), then averaging time series over all trials. We extracted mean activation levels for each delay period by averaging the TRs 6.75-9.00 s after probe onset for Delay 1 and 15.75-18.00 s after probe onset for Delay 2.

fMRI analysis: inverted encoding model

To reconstruct images of spatial WM contents, we implemented an inverted encoding model (IEM) for spatial position. This analysis involves first estimating an encoding model (sensitivity profile over the relevant feature dimension(s) as parameterized by a small number of modeled information channels) for each voxel in a region using a "training set" of data reserved for this purpose. Then, the encoding models across all voxels within a region are inverted to estimate a mapping used to transform novel activation patterns from a "test set" into activation patterns in a modeled set of information channels.

We built an encoding model for spatial position based on a linear combination of spatial filters (Sprague and Serences, 2013; Sprague et al., 2015, 2014). Each voxel's response was modeled as a weighted sum of 37 identical spatial filters arrayed in a hexagonal grid (Fig. 2A). Centers were spaced by 2.293° and each filter was a Gaussian-like function with full-width half-maximum of 2.523°:

Equation 1: $$f(r) = \left(0.5 + 0.5\cos\frac{2\pi r}{s}\right)^7 \text{ for } r < s\text{; 0 otherwise}$$

Where r is the distance from the filter center and s is a "size constant" reflecting the distance from the center of each spatial filter at which the filter returns to 0. Values greater than this are set to 0, resulting in a single smooth round filter at each position along the hexagonal grid ($s = 6.349°$; see Fig. 2A, Figure S3E for illustration of filter layout and shape; see also Sprague and Serences, 2013; Sprague et al., 2014).

This hexagonal grid of filters forms the set of information channels for our analysis. Each mapping task stimulus is converted from a contrast mask (1's for each pixel subtended by the stimulus, 0's elsewhere) to a set of filter activation levels by taking the dot product of the vectorized stimulus mask and the sensitivity profile of each filter. This results in each mapping stimulus being described by 37 filter activation levels rather than 1,024 × 768 = 786,432 pixel values. Once all filter activation levels are estimated, we normalize so that the maximum filter activation is 1.

We model the response in each voxel as a weighted sum of filter responses (which can loosely be considered as hypothetical discrete neural populations, each with spatial RFs centered at the corresponding filter position).

Equation 2: $$B_1 = C_1 W$$

Where B_1 (n trials × m voxels) is the observed BOLD activation level of each voxel during the spatial mapping task (averaged over 6.75 – 9.00 s after mapping stimulus onset; Figure S3A), C_1 (n trials × k channels) is the modeled response of each spatial filter, or information channel, on each trial of the mapping task (normalized from 0 to 1), and W is a weight matrix (k channels × m voxels) quantifying the contribution of each information channel to each voxel. Because we have more stimulus positions than modeled information channels, we can solve for W using ordinary least-squares linear regression:

Equation 3: $$\widehat{W} = (C_1^T C_1)^{-1} C_1^T B_1$$

This step is univariate and can be computed for each voxel in a region independently. Next, we used all estimated voxel encoding models within a ROI ($\widehat{W}$) and a novel pattern of activation from the WM task (each TR from each trial, in turn) to compute an estimate of the activation of each channel ($\hat{C}_2$, n trials × k channels) which gave rise to that observed activation pattern across all voxels within that ROI (B_2, n trials × m voxels):

Equation 4: $$\hat{C}_2 = B_2 \widehat{W}^T \left(\widehat{W}\widehat{W}^T\right)^{-1}$$

The Moore-Penrose pseudoinverse of the estimated weight matrix from the training set ($\widehat{W}$) is the *inverted* part of the IEM: all encoding models across all voxels are used, and this step is multivariate. This analysis is only feasible when more voxels are measured than information channels are modeled. The Moore-Penrose pseudoinverse acts as a linear mapping from data measured in voxel space (B_2) into channel space ($\hat{C}_2$), and accordingly stretches, scales and skews voxel activation patterns during this transformation, but importantly does not result in any nonlinear transformations. This analysis can be considered a directed form of dimensionality reduction in which activation patterns are transformed from an idiosyncratic activation pattern across voxels (unique to each individual participant and ROI, and thus difficult to directly compare) to a common information space, common across ROIs and participants, which allows for direct manipulation, quantification, and comparison of activation patterns in an intuitive and stimulus-referred coordinate space.

Once channel activation patterns are computed (Equation 4), we compute spatial reconstructions by weighting each filter's spatial profile by the corresponding channel's reconstructed activation level and summing all weighted filters together. This step aids in visualization, quantification, and coregistration of trials across WM target positions, but does not confer additional information.

We analyzed all data within each session: we used the 4 mapping task runs for a given session to estimate the encoding model for each voxel, then inverted that encoding model to reconstruct WM representations during all main WM task runs within that same session. Then, we averaged reconstructions over sessions within each participant.

Because WM target positions were unique on each and every trial, direct comparison of WM reconstructions on each trial is not possible without coregistration of reconstructions so that WM targets appeared at a common position across trials. To accomplish this, we adjusted the center position of the spatial filters on each trial such that we could rotate (and sometimes translate) the resulting reconstruction. For Figures 3-4, we rotated each trial such that one target (the target not queried at the end of each trial) was on average centered at $x = 3.5°$ and $y = 0°$ and the other target was in the upper visual hemifield (which required flipping ½ of reconstructions across the horizontal meridian). For Figures 7 and 8 and Figure S8, we coregistered each trial so that the queried target position was always centered at exactly $x = 3.5°$ and $y = 0°$ by first rotating the reconstruction so that the target was aligned along the positive x Cartesian axis, then horizontally translating it so that its *x* coordinate was exactly 3.5° (Figure S3D).

Because we carefully designed our task such that we presented an equal number of trials for each target separation condition (+60°, +120°, +180°, -60°, -120°, and -180° polar angle) in order to minimize the potential for participants to discover geometric regularities in the target arrangements, there was an overabundance of trials at ±180° polar angle separation distance, which led to a non-uniform distribution of positions for the non-coregistered target (that is, there were double the number of trials with non-coregistered targets at 180° polar angle from the coregistered target as there were for +60°, -60°, +120° and -120°). As a result, we excluded the second half of 180° separation condition trials from each super-run from all reconstruction-based analyses. When the other half of these trials is included, there is often a noticeable "bump" along the negative *x* axis corresponding to the greater number of trials in which a non-coregistered target appeared near that position, which renders quantification of target representations via curvefitting methods (see below) suboptimal.

Quantifying WM representations: fidelity

We took three approaches to WM representation quantification. First, we defined a "representational fidelity" metric that quantifies the extent to which a target representation reliably appeared within a reconstruction. To accomplish this, we first reduced the reconstruction from a 2-d image to a 1-d line plot by averaging over each of 220 evenly-spaced polar angle arms subtending 2.9-4.1° eccentricity (subset illustrated in Fig. 2C). The resulting 1-dimensional reconstruction reflects the average profile along an annulus around fixation. A target representation in these reconstructions would be a "bump" near 0° after the reconstructions have been rotated to a common center (where 0° corresponds to the actual target polar angle). To reduce these 1-d reconstructions to a single number which could be used to quantify the presence of target representations (F), we computed a vector mean of the 1-d reconstruction ($r(\theta)$, where θ is the polar angle of each point and $r(\theta)$ is the reconstruction activation) when plotted as a polar plot, as projected along the *x* axis (because the reconstructions were rotated such that the target was presented at 0°; Fig. 2C):

Equation 5: $$F = mean(r(\theta)\cos\theta)$$

If *F* is reliably greater than zero, over a resampling procedure (see Statistical Procedures), this quantitatively demonstrates that the net activation over the entire reconstruction carries information above chance about the target position. This

measure is independent of baseline activation level in the reconstruction, as the mean of $r(\theta)$ is removed by averaging over the full circle. We computed timecourses of representational fidelity (Fig. 5B), as well as representational fidelity for each delay period (Fig. 6). To determine whether the cue on Remember 2-valid trials restores representations, we compared *F* between Delay 2 and Delay 1 for each ROI, and between each pair of time points individually (Fig. S5). To evaluate whether ROIs differed in the extent to which their WM representations changed over the long delay interval, we compared the difference in fidelity between Delay 2 and Delay 1 between each pair of ROIs (Fig. S6).

Quantifying WM representations: fit surfaces

Additionally, we sought to evaluate the size, amplitude, and baseline of the WM target representation(s) from each WM condition and WM delay interval to establish how the information content of the population code changed across conditions. We followed procedures developed previously (Sprague et al., 2014) whereby we resampled all trials with replacement concatenated across all sessions from all participants from a condition 1,000 times and computed a single mean coregistered reconstruction (Figs. 7-8, Figure S8) on each resampling iteration. Then, we fit the mean reconstruction with a round Gaussian-like function parameterized by its center position, size, amplitude, and baseline:

Equation 6: $$f(r) = b + a\left(0.5 + 0.5\cos\frac{2\pi r}{s}\right)^7 \text{ for } r < s\text{; 0 otherwise}$$

Where *r* is the distance from the center of the surface, *s* is the size constant (as in Eq. 1), and *a* and *b* are the amplitude and baseline, respectively. Because there are many free parameters and some reconstructions are noisy, we adopted several heuristics to constrain our optimization problem. First, we found the maximum point on the entire reconstruction and used this as the center position (Sprague et al., 2014). Then, we performed a search through different sizes of fit surface function (FWHM: 0.099° to 9.934° in 0.099° steps). At each search iteration, we used ordinary least squares linear regression to find the amplitude and baseline which minimized residual errors between the reconstruction and the fit function. Finally, we used the best-fit amplitude, baseline, and size parameters from this search procedure and the global maximum position on the reconstruction as seed values for a constrained nonlinear optimization fitting algorithm (Matlab's fmincon function) subject to several constraints: position could not deviate more than one reconstruction "pixel size" (0.235° × 0.235°) from the global maximum position; size could not surpass the range used in the grid search procedure (0.099° to 9.934°), and amplitude/baseline could each not go below -5 or above 10 (BOLD Z-score units). This entire curvefitting procedure was repeated on each resampling iteration, for each condition described in the text (R1, R2-neutral, R2-valid broken down by Delay 1 and Delay 2 for Fig. 7, each of those broken down by High and Low recall error for Fig. 8 and Figure S8), resulting in 1,000 resampled estimates of each fit parameter on each condition for each ROI. Average resampled reconstructions over all resampling iterations are shown in Figure 7A,C, Figure 8 and Figure S8.

Quantifying WM representations: target activation

As a third means of quantifying the integrity of WM representations, we evaluated the relative strengths of each target representation at each time point of the trial by extracting the average reconstruction activation within a 0.5° radius circle centered at each target position. Then, we took the difference between the reconstructed target representation activation of the target probed at the end of each trial and that of the target which was not probed at the end of each trial (on R1 trials, the probed target was always the remembered target; on R2-neutral trials, the probed target was the target queried at the end of the trial; on R2-valid trials, the probed target was always the remaining target following the valid retro-cue; Figure S4). This allowed us to directly compare the strength of the representation through time for each target in a manner which did not require fitting a surface with many free parameters.

Statistical procedures

All statistical statements reported in the text are based on resampling procedures in which a variable of interest is computed over 1,000 iterations. In each iteration, all single-trial variables from a given condition are resampled with replacement and averaged, resulting in 1,000 resampled averages for a given condition. We then subjected these distributions of resampled averages to pairwise comparisons by computing the distribution of differences between one resampled distribution (e.g., R1) and another resampled distribution (e.g., R2), yielding a new distribution of 1,000 difference values. We tested whether these difference distributions significantly differed from 0 in either direction by performing two one-tailed tests (*p* = proportion of values greater than or less than 0; null hypothesis that difference between conditions = 0) and doubling the smaller *p* value. For the supplemental analysis in which we compared the change in fidelity between the two delay periods between each pair of ROIs, we compared the distribution of differences of delay period differences against 0, two-tailed ((ROI1: Delay

2 – ROI1: Delay 1) – (ROI2: Delay 2 – ROI2: Delay 1)). For tests in which we compared whether representations were present in 1-d reconstructions using the representational fidelity measure, we performed one-tailed tests (null hypothesis that $F \leq 0$).

Because we performed 1,000 iterations of these analyses, we cannot identify p values less than 0.001, so all comparisons in which resampled difference distributions were all greater than or less than 0 are reported as $p < 0.001$. Because we performed many pairwise comparisons, we corrected all repeated tests within an analysis using the false discovery rate (Benjamini and Yekutieli, 2001) and a threshold of $q = 0.05$ (except for tests of behavioral performance, which we corrected using Bonferroni's method due to the small number of comparisons performed). All p-values for all tests are reported in Supplementary Tables. All error bars/intervals reflect 95% confidence intervals as estimated using this resampling procedure. Because PFC ROIs were examined in an exploratory manner, we corrected all tests for multiple comparisons independently for the *a priori* ROIs (V1-sPCS), and PFC ROIs (iPCS, SMA+, DLPFC+). The All ROIs Combined ROI (consisting of concatenated voxels across the *a priori* ROIs) was not independent of the constituent ROIs, which required us to independently correct for multiple comparisons within that ROI alone.

Code and data availability

In an effort to improve reproducibility, all data and code required to perform the analyses described here and to generate figures appearing in the text and supplement, as well as task scripts, are freely available in the Open Science Framework (http://osf.io/s5r6g). Additionally, tutorial and stimulus presentation scripts for implementing inverted encoding model (IEM)-based image reconstruction analyses are freely available at bit.ly/IEM_tutorial. Any questions regarding code or data can be addressed to author TCS.

Supplemental References

Albers, A.M., Kok, P., Toni, I., Dijkerman, H.C., de Lange, F.P., 2013. Shared representations for working memory and mental imagery in early visual cortex. Curr. Biol. 23, 1427–31. doi:10.1016/j.cub.2013.05.065

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29, 1165–1188.

Bettencourt, K.C., Xu, Y., 2015. Decoding the content of visual short-term memory under distraction in occipital and parietal areas. Nat. Neurosci. doi:10.1038/nn.4174

Christophel, T.B., Hebart, M.N., Haynes, J.-D.J.-D., 2012. Decoding the Contents of Visual Short-Term Memory from Human Visual and Parietal Cortex. J. Neurosci. 32, 12983–12989. doi:10.1523/JNEUROSCI.0184-12.2012

Emrich, S.M., Riggall, A.C., Larocque, J.J., Postle, B.R., 2013. Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. J. Neurosci. 33, 6516–23. doi:10.1523/JNEUROSCI.5732-12.2013

Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.-J., Shadlen, M.N., 1994. fMRI of human visual cortex. Nature 369, 525.

Ester, E.F., Serences, J.T., Awh, E., 2009. Spatially Global Representations in Human Primary Visual Cortex during Working Memory Maintenance. J. Neurosci. 29 , 15258–15265. doi:10.1523/JNEUROSCI.4388-09.2009

Ester, E.F., Sprague, T.C., Serences, J.T., 2015. Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. Neuron 87, 893–905. doi:10.1016/j.neuron.2015.07.013

Harrison, S.A., Tong, F., 2009. Decoding reveals the contents of visual working memory in early visual areas. Nature 458, 632–635. doi:10.1038/nature07832

Lee, S.-H., Kravitz, D.J., Baker, C.I., 2013. Goal-dependent dissociation of visual and prefrontal cortices during working memory. Nat. Neurosci. 16, 997–9. doi:10.1038/nn.3452

Mackey, W.E., Devinsky, O., Doyle, W.K., Meager, M.R., Curtis, C.E., 2016. Human Dorsolateral Prefrontal Cortex Is Not Necessary for Spatial Working Memory. J. Neurosci. 36, 2847–2856. doi:10.1523/JNEUROSCI.3618-15.2016

Mohler, C.W., Goldberg, M.E., Wurtz, R.H., 1973. Visual receptive fields of frontal eye field neurons. Brain Res. 61, 385–389.

Pratte, M.S., Tong, F., 2014. Spatial specificity of working memory representations in the early visual cortex. J. Vis. 14, 22. doi:10.1167/14.3.22

Riggall, A.C., Postle, B.R., 2012. The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. J. Neurosci. 32, 12990–8. doi:10.1523/JNEUROSCI.1892-12.2012

Saber, G.T., Pestilli, F., Curtis, C.E., 2015. Saccade planning evokes topographically specific activity in the dorsal and ventral streams. J. Neurosci. 35, 245–52. doi:10.1523/JNEUROSCI.1687-14.2015

Serences, J.T., Ester, E.F., Vogel, E.K., Awh, E., 2009. Stimulus-Specific Delay Activity in Human Primary Visual Cortex. Psychol. Sci. 20 , 207–214. doi:10.1111/j.1467-9280.2009.02276.x

Sprague, T.C., Ester, E.F., Serences, J.T., 2014. Reconstructions of Information in Visual Spatial Working Memory Degrade with Memory Load. Curr. Biol. doi:10.1016/j.cub.2014.07.066

Sprague, T.C., Saproo, S., Serences, J.T., 2015. Visual attention mitigates information loss in small- and large-scale neural codes. Trends Cogn. Sci. 19, 215–26. doi:10.1016/j.tics.2015.02.005

Sprague, T.C., Serences, J.T., 2013. Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nat. Neurosci. 16, 1879–87. doi:10.1038/nn.3574

Srimal, R., Curtis, C.E., 2008. Persistent neural activity during the maintenance of spatial position in working memory. Neuroimage 39, 455–468.

Swisher, J.D., Halko, M.A., Merabet, L.B., McMains, S.A., Somers, D.C., 2007. Visual topography of human intraparietal sulcus. J. Neurosci. 27, 5326–5337.

Zirnsak, M., Steinmetz, N.A., Noudoost, B., Xu, K.Z., Moore, T., 2014. Visual space is compressed in prefrontal cortex before eye movements. Nature 507, 504–507. doi:10.1038/nature13149