Neural population dynamics of human working memory

Hsin-Hung Li[1,2] and Clayton E. Curtis[1,2]

1. Department of Psychology, New York University, New York, NY 10003
2. Center for Neural Science, New York University, New York, NY 10003

Correspondence: clayton.curtis@nyu.edu

Keywords: working memory, fMRI, dimension reduction, PCA, population receptive field

Conflicts of interest: The authors declare no competing financial interests.

Number of pages: 24
Number of figures: 7
Number of words for abstract: 246
Number of words for introduction: 550
Number of words for discussion: 1348
Conflict of interest statement: The authors declare no conflict of interest

Acknowledgements: We thank New York University's Center for Brain Imaging for technical support. This research was supported by the National Eye Institute (R01 EY-016407 and R01 EY-027925 to C.E. Curtis). Hsin-Hung Li was supported by the Swartz Foundation Postdoctoral Fellowship.

Abstract

Working memory (WM) depends on the activity of neural populations to maintain internal representations of information that are no longer available in the environment. The neural population dynamics that support WM in the human cortex remain largely unexplored. Utilizing fMRI and time-resolved multivariate analyses, we found evidence of both stable and dynamic WM representations during a classic memory-guided saccade task. Using dimensionality reduction techniques, we identified robust neural subspaces in early visual, parietal and even frontal cortex that were stable throughout the delay period with memorized locations organized within a two-dimensional representation of visual space. The dynamic subspace, which was largely restricted to early visual cortex, had both an early and late delay component during which the WM representation changed. Leveraging models of the receptive fields of voxels comprising each of the visual field maps, we visualized how the dynamic population code changed over time. Early in the trial, the dynamic code in V1 was dominated by a narrowly tuned activation among voxels with receptive fields centered on the location of the peripheral memory target. Late in the trial, the dynamic code was driven by diffuse activation among voxels with receptive fields along a line between the fovea and the peripheral location of the target. Together, the timing, shape, and spatial distribution of the WM dynamics in early visual cortex suggest that the early component encodes feedforward visual inputs, while the late component encodes abstracted task-relevant mnemonic representations resulting from feedback signals from higher-level brain areas.

Significance statement

Recent empirical evidence of neural dynamics during working memory (WM) challenges established canonical theories of how WM representations are stored in the brain. Utilizing fMRI and time-resolved multivariate analyses, we identified both stable WM codes in visual and association cortex, as well as changing, dynamic codes in early visual cortex. Each of these population codes were composed of low-dimensional topological representations of visual space. Characteristics of the WM dynamics in early visual cortex suggest that feedback signals from higher-brain areas recode sensory inputs into a format matching the demands of the memory-guided task.

Working memory (WM) allows us to store information over a brief period of time, increasing the duration of neural representations available to support computations and decisions. Early studies on the neural mechanisms underlying WM mostly focused on the prefrontal cortex (PFC) of non-human primates (Fuster and Alexander, 1971; Kubota and Niki, 1971; reviewed in Curtis and Sprague, 2021). Neurophysiological studies demonstrated that neurons in PFC exhibit elevated and sustained activity during memory delays suggesting a mechanism by which stimulus information is maintained (Funahashi et al., 1989, 1993a, 1993b; Miller et al., 1996; Romo et al., 1999; Constantinidis et al., 2001a; Wasmuht et al., 2018; Panichello and Buschman, 2021).

Implicit in the studies of persistent activity is the idea that the population neural response that encodes a specific feature value is stable through memory delay. However, later empirical studies reported that during memory delays, some neurons in PFC exhibit temporal dynamics more complicated than those predicted by stable persistent activity (Baeg et al., 2003; Brody et al., 2003; Shafi et al., 2007; Barak et al., 2010; Harvey et al., 2012). Recent multivariate analyses revealed that both stable and dynamic codes of WM coexist in the macaque PFC at the population level (Murray et al., 2017; Spaak et al., 2017).

In the past decade, a number of human studies have demonstrated that one can decode the content of WM from fMRI voxel activity patterns across widespread brain regions including visual, parietal, and prefrontal cortex, as well as subcortical regions (Harrison and Tong, 2009; Serences et al., 2009; Jerde et al., 2012; Riggall and Postle, 2012; Emrich et al., 2013; Ester et al., 2013, 2015; Lee et al., 2013; Xing et al., 2013; Sprague et al., 2014, 2016; Sreenivasan et al., 2014; Bettencourt and Xu, 2016; Christophel et al., 2017, 2018; Lorenc et al., 2018; Rahmati et al., 2018, 2020; Rademaker et al., 2019; Brissenden et al., 2021; Hallenbeck et al., 2021; Li et al., 2021; Kwak and Curtis, 2022). Intriguingly, in these studies, decoding accuracy (or decoding precision) was typically the highest in visual cortex. In contrast, the patterns of activity in PFC cannot decode WM content (Christophel et al., 2012; Riggall and Postle, 2012) or when it does it is only slightly above chance (Ester et al., 2015; Li et al., 2021). Whether and how the findings regarding the WM dynamics in macaque PFC can be extrapolated to other brain regions and to the human brain remains a critical but open question.

Here, we used time-resolved multivariate analysis to investigate the dynamics of spatial WM in the human brainTo preview, we demonstrated the coexistence of stable and dynamic population codes for spatial WM. Importantly, the stability of WM varied across brain regions with early visual cortex exhibiting the strongest dynamics. Leveraging the receptive field properties of visual field maps (Dumoulin and Wandell, 2008; Mackey et al., 2017), we visualized how the neural representations of memorized items evolved over time. In V1, where we observed prominent dynamics, activation was limited to the location of the peripheral target early in the delay and over time spread toward the fovea along the polar angle of the target. We propose that the dynamics of WM are mainly driven by a recoding of sensory inputs into a format most proximal to the demands of the task.

Results

To facilitate direct comparisons with the existing monkey neurophysiological studies (e.g., Funahashi et al., 1989, 1993a; Constantinidis et al., 2001a; Wimmer et al., 2014; Murray et al., 2017; Spaak et al., 2017), we used a memory-guided saccade task to study spatial WM. In each trial, a brief (500 ms) target dot was presented in the periphery, followed by a 12 s delay period. The polar angle of the target was chosen pseudo-randomly from 1 of 32 positions spanning the full circle. After the delay, participants reported the remembered location with a memory-guided saccade (Figure 1A). Participants were able to make precise memory reports close to the target locations (Fig 1B). In addition to the memory-guided saccade experiment, participants underwent a population receptive field (pRF) mapping session (Mackey et al., 2017), which allowed us to define four retinotopic visual (V1, V2, V3, and V3AB) and four parietal (IPS0, IPS1, IPS2, and IPS3) areas as the regions of interest (ROIs).

Coexistence of stable and dynamic WM codes. We first characterized the dynamics of WM neural representations using a temporal generalization analysis (Stokes et al., 2013; King et al., 2016; Spaak et al., 2017), in which we tested whether the decoder trained by the voxel activity pattern at one time point can be generalized to decode the location of the WM target using the voxel activity at other time points. This analysis generated a two-dimensional (trained time by test time) cross-decoding matrix for each ROI (Figure 1C). For a stable WM code, the location of the WM target should be decodable even when the training data and the testing data came from different time points. We found a stable WM code in all the ROIs (pink dashed lines in Figure 1C), quantified by the above-chance decoding performance among the off-diagonal elements of the cross-decoding matrices. In addition, across all the ROIs, WM content was present shortly (about 1 to 2 sec) after the delay onset (or target offset), and remained decodable throughout the delay period.

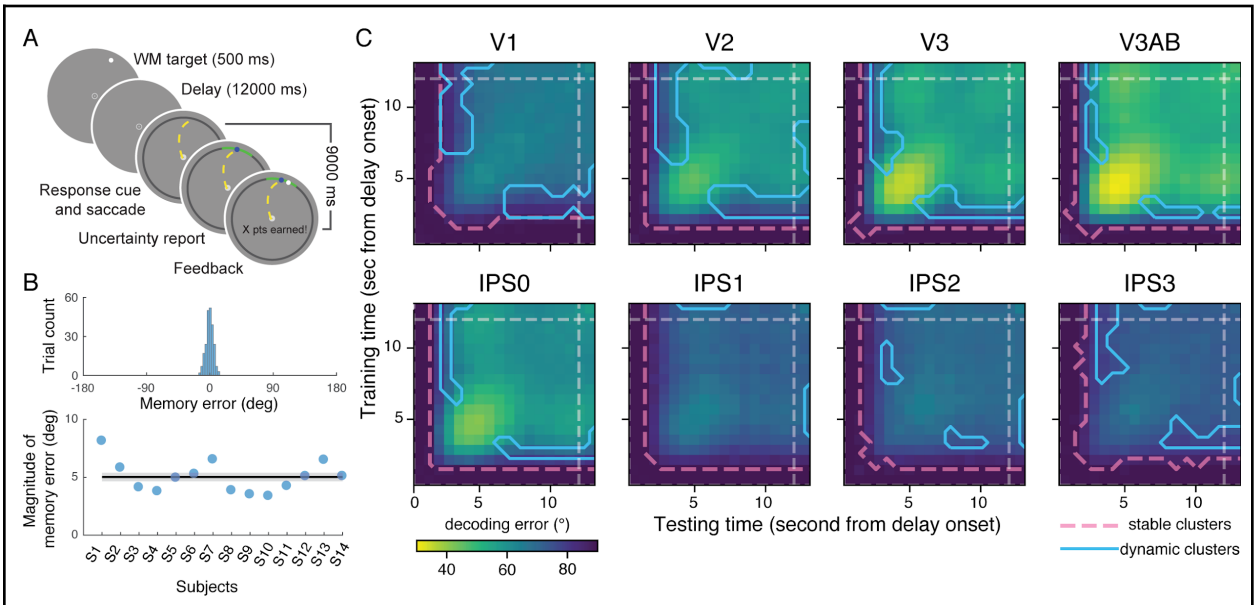


Figure 1. Behaviors and temporal generalization analysis. (A) Procedures. Participants maintained fixation while remembering the location of a target dot presented at a pseudorandom polar angle and at

12° eccentricity from fixation. At the end of each trial, participants made saccadic eye movements to report their memory, and adjusted the length of an arc for uncertainty report. (B) Top panel: the distribution of (signed) memory error of an example subject. Bottom panel: the mean magnitude (absolute value) of memory error of each participant. The horizontal line represents the group mean with ±1 s.e.m. (C) The temporal generalization analysis. Decoders were trained to decode the target location from the fMRI BOLD response. The temporal generalizability of the decoders trained with voxel activity patterns of each time point were tested with the data of all the time points. Pink dashed lines: the stable clusters, the cluster that exhibited above-chance decoding performance. Blue solid lines: the dynamic clusters, the off-diagonal elements that exhibited lower decoding performance than (both of) their corresponding diagonal elements.

The presence of a stable WM representation does not exclude the coexistence of dynamic population codes. To test if neural dynamics existed during WM, we asked if the off-diagonal elements showed worse decoding performance than their corresponding diagonal elements. The reduction of decoding performance on the off-diagonal elements would indicate a change of WM representations over time. In most ROIs (V1, V2, V3, V3AB, IPS0, IPS2 and IPS3), we observed clusters of off-diagonal elements exhibiting reduced decoding performance, indicating a dynamic WM neural code (light blue lines in Figure 1C) coexisted with a stable neural code. Such dynamics were most obvious in the early visual cortex from V1 to V3, whereas only a small number of off-diagonal elements showed significant reductions in the higher-level visual and parietal cortex. When present, these significant clusters showed up as elongated horizontal stripes at the lower right of the cross-decoding matrices (and a mirrored version as a vertical stripe at the upper left, Figure 1C). These patterns suggest a dynamic change from a shorter early time window, which presumably reflects the encoding of WM target, to a late time window that reflects the maintenance of target information over a longer time scale.

Stable subspaces. The existence of a stable code indicates a coding subspace within which the representation of a target location is stable over time and remains discriminable from other target locations. We characterized such a stable subspace by applying principal component analysis (PCA) on time-averaged BOLD responses. This analysis identified the dimensions that maximize the variance across target locations and discarded the dynamical aspect of voxel activity patterns by averaging across time (Murray et al., 2017). Given that our target stimuli only varied in a two dimensional space, we focused on a subspace defined by the top two principal components that maximized the variance explained. After defining the subspace with time-averaged data, we projected the data of each single time point during the delay into this stable subspace. We found that in this subspace, target locations were topologically organized, preserving their spatial relationships in the visual field, and the neural activity remained stable across time in this subspace (Figure 2).

Dynamic subspaces. The results of the temporal generalization analysis (Figure 1C) indicated that there existed an early and a late time epoch between which the voxel activity patterns changed across time. We identified an early subspace and a late subspace by applying PCA to the voxel activity of an early and a late time window. The two subspaces can be thought of as two hyperplanes on which the memorized locations are spatially organized (Figure 3A). Next, we characterized how the early and the late subspaces oriented in the high dimensional space

by computing their principal angles (see Methods). The principal angles fell within a range between 35 to 55 degrees. Thus, the two subspaces were not orthogonal (a principal angle at 90°). Nevertheless, the principal angles between the early and the late subspaces (solid blue lines in Figure 3B) were larger than zero, and were larger than the angles between two subspaces that were estimated by the data sampled from the same time window (which mainly reflected the reliability of subspace estimation; dashed lines in Figure 3B). The only exception was V3AB whose principal angle computed across time points was not distinguishable from the angle computed within the late time window. These results demonstrated that the neural subspaces used to encode the memorized locations changed substantially, though not completely, across the early and late time windows.

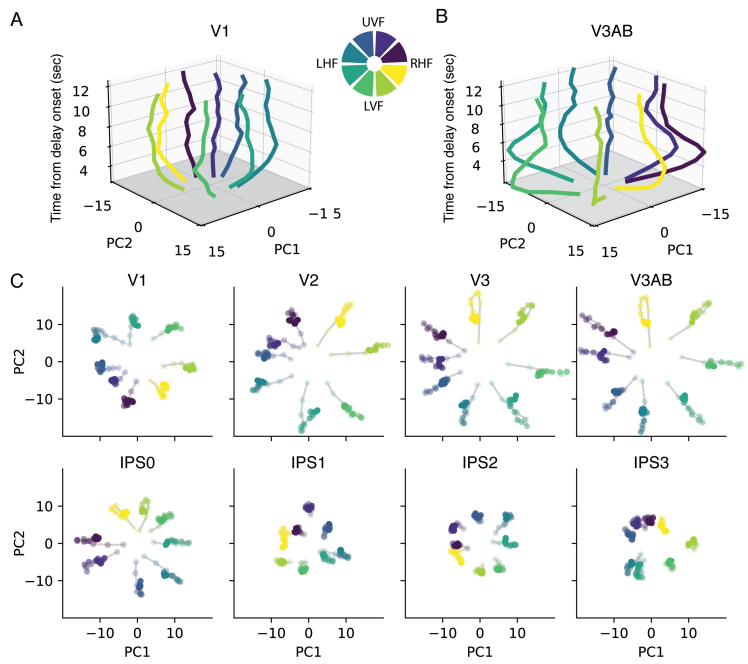


Figure 2. Stable subspaces. The stable coding subspaces were identified by applying PCA to time-averaged voxel activity patterns. (A) V1's response at each time point during the delay period was projected into the stable subspace, the top two principal components (x-axis and y-axis) obtained from PCA. The z-axis represents time from delay onset. Each curve represents one bin of target location. The legend indicates the location, in the visual field, represented by each of the eight colors. (B) Same as (A) but for V3AB. (C) Similar to (A) and (B) but without visualizing the time axis (z-axis); thereby is equivalent to the bird-eye's view of (A) and (B). Each dot represents data from one time point during the delay, with more saturated colors representing later time points.

Given that we selected the early and late time windows based on the dynamic clusters obtained in the temporal generalization analysis, we expected that the principal angle between the early and the late subspaces to be larger than zero. This principal angle summarizes the stability of WM dynamics of each ROI, and allows a comparison between brain regions. We found that the principal angles varied across ROIs ($F(7, 91) = 4.90$, $p < .00$), and the main effect of ROI was stronger than that predicted by the angles within time windows (Figure 3C). That is, the dynamics of WM representations were not the same across brain regions. They underwent greater changes in early visual cortex, and were more stable in higher-level visual cortex and parietal cortex. This was confirmed by pairwise comparisons showing that the early visual cortex (V1, V2 and V3) had larger principal angles than higher-level visual cortex V3AB and parietal region IPS0 (Figure 3D).

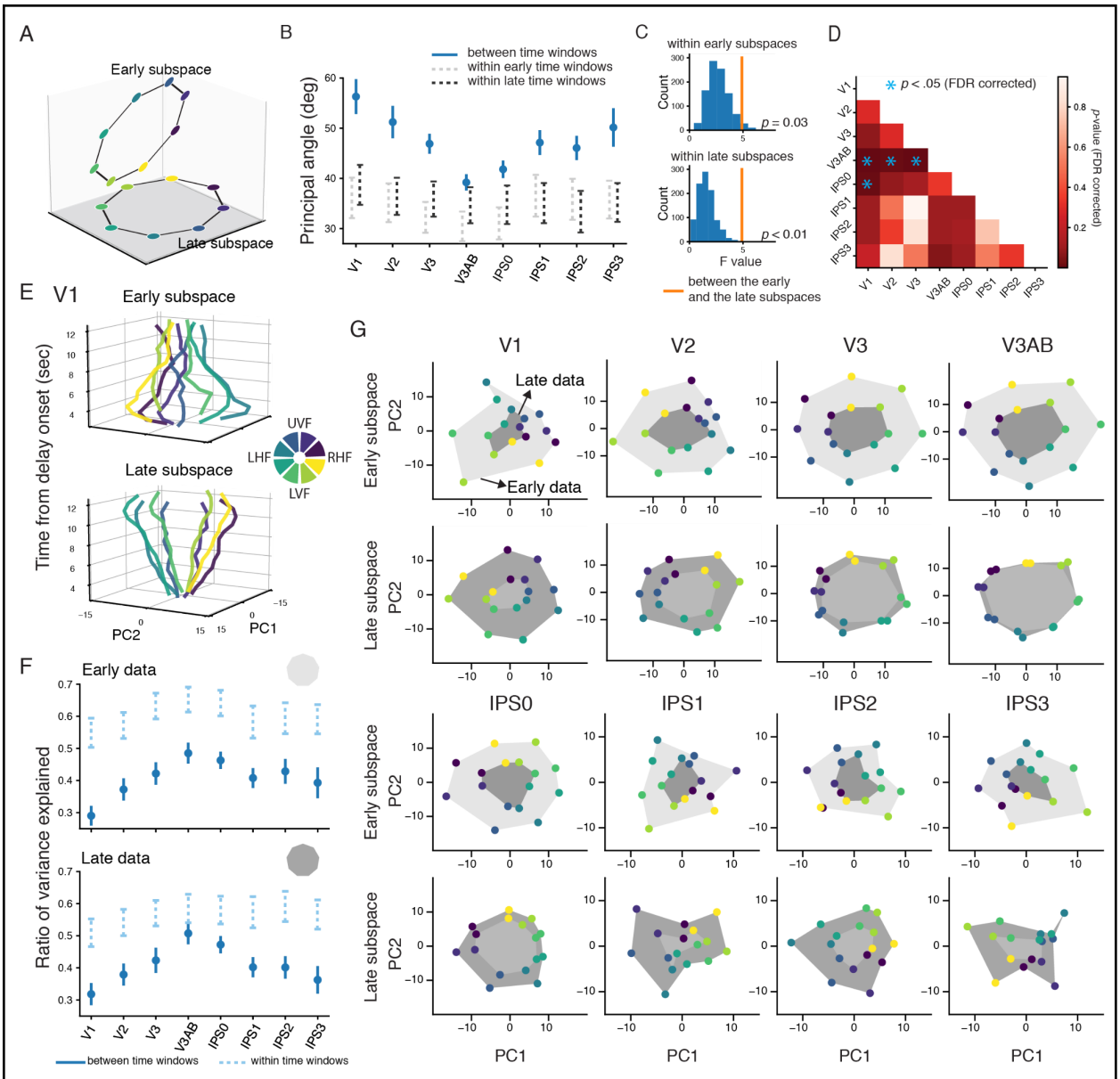


Figure 3. Dynamic subspaces. Early and late neural coding subspaces were identified by applying PCA to voxel activity patterns in two different time windows. (A) Cartoon illustration of an early and a late subspace that encodes the WM target locations in the high dimensional space. (B) Principal angles were computed to characterize the angles between the early and late subspaces. The dark blue data points represent the principal angle of each ROI (group mean with ±1 s.e.m). The dashed lines represent 95% confidence intervals of the distribution of the principal angles computed between the two subspaces estimated by resampling the data from either the same early (light gray dashed lines) or the late (dark gray dashed lines) time windows. (C) The main effect of the principal angle between the late and the early subspaces is larger than that predicted by the principal angle within the same time window. In a bootstrapping procedure, in each iteration, we resampled the data, estimated two early (or two late) subspaces and computed the principal angle between these two subspaces. We then computed the main effect of ROI on the principal angle by ANOVA, leading to a distribution of the F values (blue histograms). P values were computed by comparing the distributions of the F values with the empirical F value computed using the principal angle between the early and the late subspaces (orange vertical lines). (D) Pairwise comparisons of the principal angles between ROIs. (E) V1's

response of each time point during the delay was projected into the early (top) and the late (bottom) dynamic subspace, where the top two principal components (x-axis and y-axis) were derived from PCA. The z-axis represents the time from delay onset. Each curve represents one bin of the target locations. (F) Ratio of variance explained (RVE) of each ROI. The dark blue data points represent the RVE (group mean with ±1 s.e.m) of each ROI. The light blue dashed lines represent the 95% confidence interval of the distribution of the RVE computed by resampling the data from the same time window. (G) Projections of the early (data points connected with the light gray areas) and the late (data points connected with the dark gray areas) voxel activity patterns in the early (top panel of each ROI) and the late (bottom panel of each ROI) dynamic subspaces. See Methods for the details of the principal angle and the RVE.

We further visualized the dynamics of neural activity by projecting the voxel activity patterns of the early or late time window into the early or late subspaces (Figure 3G; see Figure 3E for a version where the voxel activity of each single time window were projected into the two subspaces). The spread of target locations was larger when projected to the subspace of their own time window, compared to the subspace of the other time window (Figure 3G). To quantify this reduction of the variance explained, we computed the ratio between the variances explained by each of the two subspaces (RVE, dark blue data points in Figure 3F; see Methods). We also computed the RVE within time windows by resampling the trials and computing the subspace twice using the data from the same time window (dashed light blue lines in Figure 3F represent the 95% confidence intervals of the baseline values). The RVEs computed across time windows were smaller than the RVEs within the time windows indicating a significant reduction of variance explained when the data of one time window were projected to the subspace of the other time window. Similar to the principal angle, the (across-time) RVE varied across ROIs for both the data of the early (main effect for ROI, $F(7, 91) = 3.47$, $p < .005$) and the late time window (main effect for ROI, $F(7, 91) = 4.56$, $p < .001$).

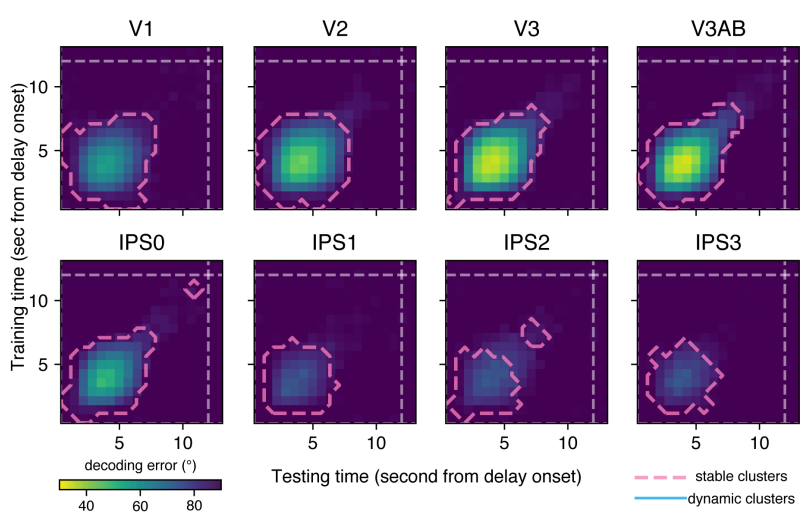


Figure 4. Temporal generalization for the passive viewing experiment. Plotted in the same format as that in Figure 1C. Pink dashed lines: stable clusters that exhibited above-chance decoding performance. Blue solid lines: dynamic clusters, the off-diagonal elements that exhibited lower decoding performance than both of their corresponding diagonal elements.

Passive viewing. A subset of the subjects (n=6) participated in a passive viewing experiment in which they were asked to perform a task at central fixation while a high-contrast flickering peripheral 'WM target', treated as an irrelevant stimulus, was present continuously throughout

the delay. In this case, the location of the peripheral target was only decodable for a much shorter period of time early in the delay without sustaining through the delay (Figure 4), and only a stable cluster was identified**.** Thus, the decodable neural activity we analyzed in the late time window cannot be considered a mere decay of sensory-driven responses.

WM dynamics. Now that we demonstrated that the neural population code for WM changes over time, we next turned to addressing how it changes. To visualize how neural populations represent memorized locations, for each ROI and time window, we utilized the voxels' pRF center and pRF size, and computed an activation map by projecting population responses into the coordinates of two-dimensional visual field space (Methods).

For both the early and late time windows, we found peak responses in the maps at the target's polar angle (Figure 5A,B,F,G, in which the activation maps for different target locations were circular shifted to be aligned at the right of horizontal meridian), consistent with previous fMRI studies that visualized neural responses in the visual field space (Kok and de Lange, 2014; Rahmati et al., 2018; Favila et al., 2020; Kwak and Curtis, 2022; Zhou et al., 2022). In V1, where we find the strongest dynamics, the spatial pattern of the population responses clearly differed in the early and late time windows. In the early time window, the activation peaked at the target's location, both in polar angle and eccentricity (Figure 5A). However, in the late time window, the activation at the target's peripheral location declined while activity spread in eccentricity toward the fovea along the target's polar angle (Figure 5B). A horizontal slice across the target location showed that instead of being localized at central fixation, this increment at perifoveal locations during the late time window extended toward the target, exhibiting a line-like pattern connecting fixation and the target location (dark line in Figure 5C; also see Figure 5D and 5E). In V3AB, the ROI with the greatest stability, we found that the peak in activation remained at the target's peripheral location over the course of the trial (Figure 5F-5J; also see Extended Data Figure 5-1 for the activation maps of the other ROIs).

Our results based on the reconstructed activation maps suggested that the dynamic neural code observed in early visual cortex during WM resulted from dynamics that differed in the neural populations selective for near and far eccentricity. To further investigate this idea, we trained two decoders to predict the target location (polar angle) during memory delay, a near-eccentricity decoder whose input data were the responses of the voxels whose pRF centers were within 6° from the fixation (orange lines in Figure 5K), and a far-eccentricity decoder that utilized the response of the voxels whose pRF centers were larger than 6° from the fixation (blue lines in Figure 5K). Early visual cortex (V1, V2 and V3) contained information about the WM target (indexed by the decoding error) that was more precise at the far than the near eccentricity in the early time window, but this difference disappeared in the late time window. These patterns were confirmed by significant two-way interactions between the time windows and the eccentricity in two-way ANOVAs (V1: $F(1, 13) = 22.94$, $p < .001$; V2: $F(1, 13) = 33.22$, $p < .001$; V3: $F(1, 13) = 37.23$, $p < .001$; FDR corrected for the number of ROIs). A two-way interaction was also significant in higher-level visual cortex V3AB, where the near-eccentricity decoder outperformed the far-eccentricity decoder ($F(1, 13) = 6.85$, $p < .05$). Even so, V3AB and IPS differed from

early visual cortex, as the far-eccentricity decoder always outperformed the near-eccentricity decoder regardless of time windows.

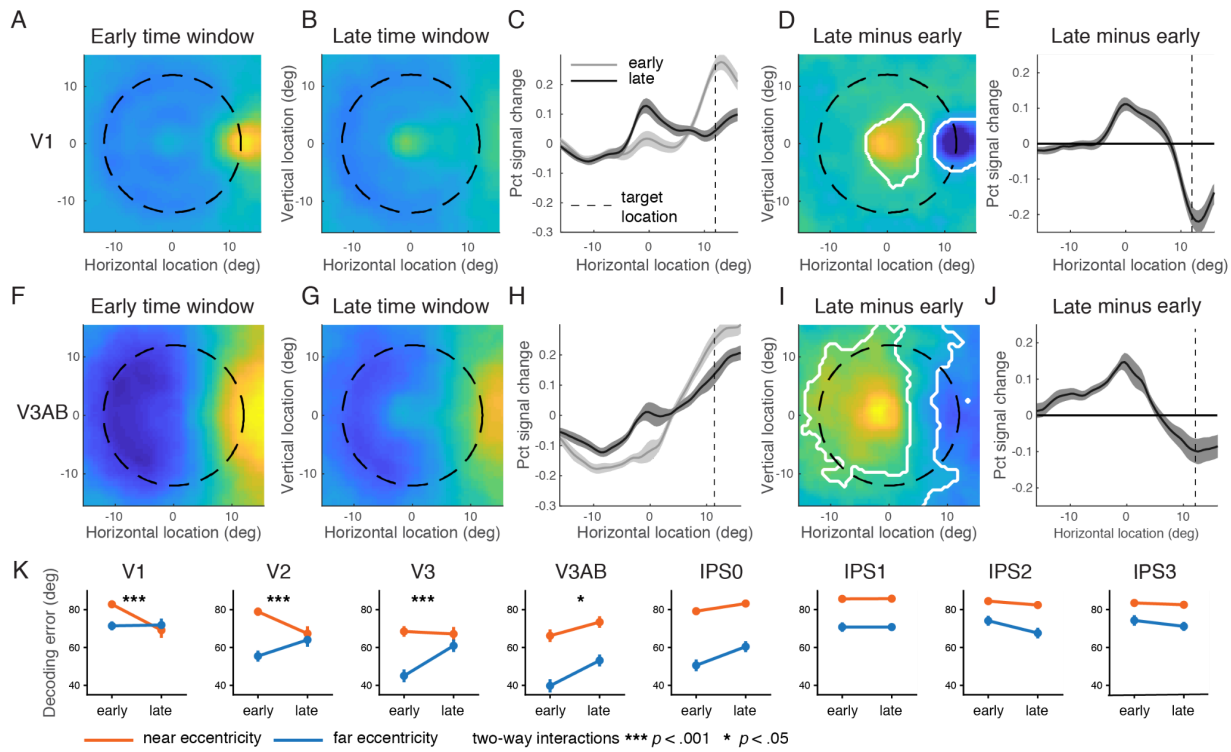


Figure 5. Visualizing WM dynamics. (A) Activation maps visualize the projection of voxel activity patterns onto two dimensional visual field space. The activation map for V1 in the early time window. (B) The activation map for V1 in the late time window. (C) Horizontal slices of the early (light gray solid line) and the late (dark gray solid line) activation maps at the target location, corresponding to the row at 0° vertical location in (A) and (B). The vertical dashed line indicates the horizontal position of the target. (D) The difference of activation computed as the activation in the late time window minus the activation in the early time window. The white contours outline significant increment or decrement of activation based on cluster-based permutation test ($p < 0.05$). (E) The horizontal slice from (D) at the target location (vertical location at 0°). (F-J) Same as (A-E) but for V3AB. (K) Decoding error of the decoders using either the voxels selective for near eccentricity (red lines) or far eccentricity (blue lines). Data plotted in C, E, H, J, and K represent group mean with ±1 s.e.m.

We computed polar angle response functions by collapsing the activation maps across eccentricity (see Methods). For both the early and late time windows, all ROIs had response functions centered at target polar angle (Figure 6), indicating that there was no systematic shift in voxels' preferred polar angle across time. By fitting the response functions with von Mises functions, we observed significant gain reductions in V1, V2, V3, V3AB, IPS0 (Figure 6B) in the late time window compared to the early time window. In addition, the width of the response functions widened in the late time window in V1 and V2, but not in the other ROIs (Figure 6C). We further computed polar angle response functions separately for the near and the far eccentricity. We found that in V1 and V2, the gain of the response functions were reduced in the far but not near eccentricity (Extended Data Figure 6-1). Thus, information about the target's polar angle remained robust across time in near-eccentricity neural populations in early visual

cortex (Figure 4K). Overall, by projecting population activity into visual field space, we found that the dynamic code of WM observed in early visual cortex was driven by a spread of the target polar angle information across eccentricity, and a widening of polar angle response functions during the delay. Together, these dynamics in early visual cortex suggest that neural activity early in the trial encodes feedforward visual inputs, while activity later during the delay may only preserve task-relevant features (i.e., target polar angle and not eccentricity) that result from feedback signals from higher brain areas (see Discussion).

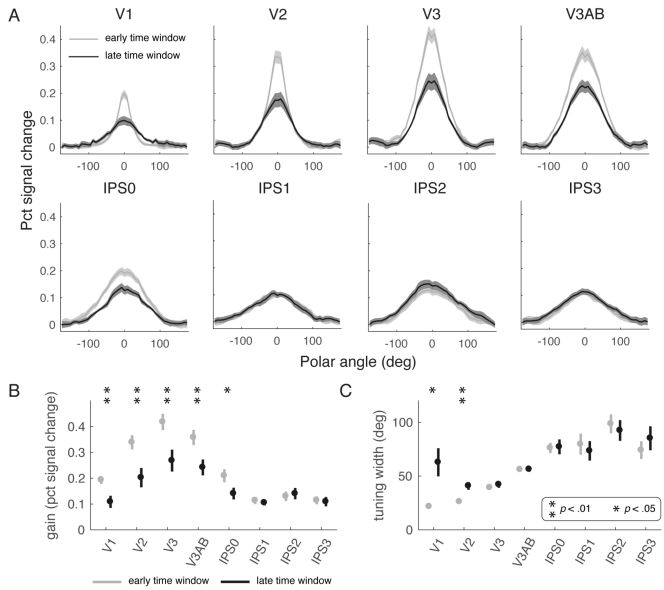


Figure 6. Polar angle response functions. (A) Polar angle response functions computed by averaging the activation maps over eccentricity. (B) Gain of the response functions obtained by fitting von Mises distributions to the response functions. (C) The tuning width of the response functions. All the data plotted represent group mean with ±1 s.e.m.

Neural code during WM is stable in PFC. Previous studies of the neural population codes during WM in nonhuman primates have largely focused on the neural activity in PFC. Here, we extended our analyses above to two frontal regions, iPCS and sPCS, the regions where we previously observed topographic organization in the pRF mapping session (consistent with Mackey et al., 2017). We found decodable WM content in both frontal regions but their decoding errors were larger than all the other ROIs (Figure 7A). In the temporal generalization analysis, we found a stable cluster throughout the memory delay in each of the ROIs, without significant dynamical clusters (Figure 7B and 7C). Projecting the voxel activity pattern of each time point into the stable subspace extracted by PCA, we found that the locations of the targets remained largely stable and separable within the stable subspace (Figure 7D and 7E). However, their spatial topology (in the visual field space) was not well-maintained like those observed in the visual and the parietal cortices.

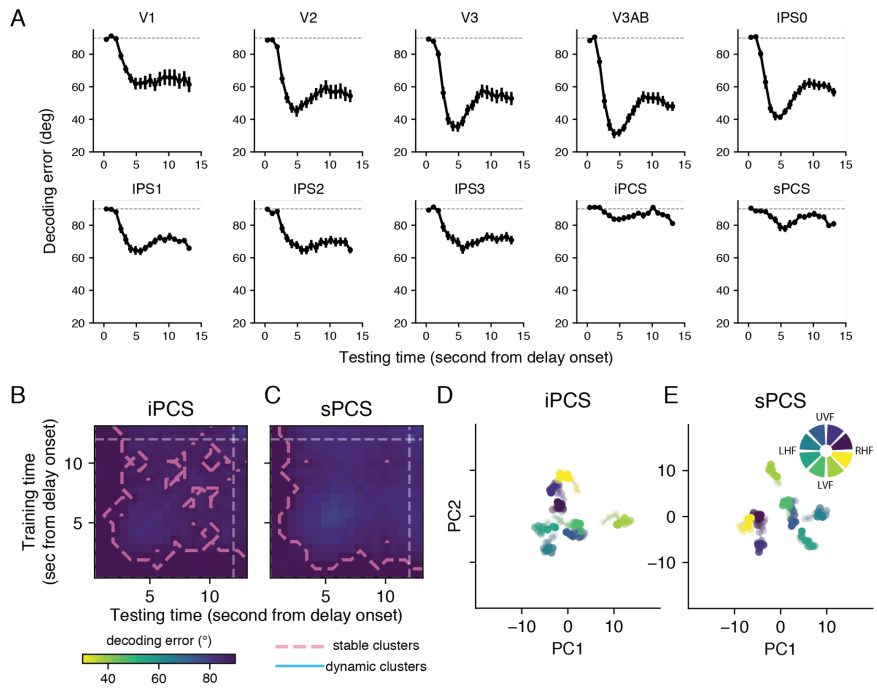


Figure 7. WM decoding in PFC. (A) Decoding error of each ROI including two ROIs in the prefrontal cortex. These data are equivalent to the diagonal elements of the decoding matrices from the temporal generalization analysis (Figure 1C and 8B-C). Data points represent group mean with ±1 s.e.m. (B-C) Temporal generalization analysis for iPCS and sPCS. Plotted in the same format as Figure 1C. (D-E) Stable subspaces. Plotted in the same format as Figure 2C. The stable coding subspaces were identified by applying PCA to time-averaged voxel activity patterns. Voxel activity pattern of each time point during the delay period was projected in the stable subspace, the top two principal components (x-axis and y-axis) obtained from PCA.

DISCUSSION

In the human brain, we found evidence of both stable and dynamic WM representations (Figure 1) during a simple memory-guided saccade task. Using PCA, we identified robust subspaces in early visual, parietal and even frontal cortex that were stable throughout the delay period with components that topologically corresponded to two-dimensional visual space (Figure 2). The dynamic subspace, which was largely restricted to early visual cortex, had an early and a late delay component between which the WM representation changed over time (Figure 3). Leveraging models of the receptive fields of voxels comprising each of the visual field maps, we measured how the population code changed over time. The early code was dominated by a narrowly tuned activation among voxels with receptive fields centered on the location of the memory target. The late code was driven by a more diffuse activation among voxels with receptive fields along a line between foviation and the peripheral location of the target (Figure 5). Together, the timing, spatial distribution, and shape of the dynamic code suggests that the early component encodes the visual inputs, while the late component maintains the task-relevant target information.

Neural dynamics of WM representations. In general, our findings provide critical insights into the largely unknown dynamics of WM representations in the human brain, building on previous work that utilized fMRI decoding to study the neural mechanisms of WM (e.g., Harrison and Tong, 2009; Serences et al., 2009; Christophel et al., 2017; see review in Curtis and Sprague, 2021). Using time-resolved multivariate analyses, we demonstrated that neural populations in human visual and association cortex represent WM content by a largely stable code throughout memory delay. This stable code coexists with a dynamic code that is mainly driven by a transition from stimulus encoding to WM maintenance. Our results show striking similarities, but also differences, to those observed in neurophysiological recordings in monkey PFC. Applying the same temporal generalization analysis to data from Watanabe and Funahashi (Watanabe and Funahashi, 2007), Spaak et al. (2017) reported the coexistence of stable and dynamic coding in monkey's PFC for spatial WM. Moreover, in the cross-decoding matrices, the dynamic coding they observed exhibited as elongated clusters like ours, indicating that the dynamic coding was mainly driven by the difference between a brief early encoding phase and a longer late maintenance phase (Figure 2 and 6 in Spaak et al., 2017). In a reanalysis of Constantinidis et al., (Constantinidis et al., 2001b) and Romo et al., (Romo et al., 1999), Murray et al. (2017) reported that a stable subspace (extracted by applying PCA to time-averaged firing rate of the neurons in monkey PFC) explained a comparable amount of stimulus variance relative to a dynamic subspace (extracted by applying PCA to firing rate each single time points) during the delay. Only around the time of stimulus presentation did the stable subspace show a disadvantage compared to the dynamic subspace. On the one hand, these neurophysiological results from monkey PFC align perfectly with our findings using fMRI from the human brain. Both provide converging evidence for stable neural codes that coexist with dynamic codes that are predominantly driven by the transformation of sensory-driven responses into WM representations. On the other hand, we find evidence for these coexisting stable and dynamic states in early visual cortex, not PFC. While we cannot know whether these differences are due to differences in species or differences in measurements of neural activity, they parallel a vast literature in humans demonstrating the importance of early visual cortex for WM (Curtis and Sprague, 2021).

Building on previous studies that investigated the dynamics of WM code in a single brain region (Wolff et al., 2015; Murray et al., 2017; Spaak et al., 2017), we compared the stability of WM representations across multiple ROIs and found that the dynamics of WM, driven by a transition from sensory-driven response to WM maintenance, vary across brain regions, with early visual cortex exhibiting the strongest dynamics (Figure 3B-3D). When quantifying the polar angle response function evoked by the WM target, the widths of the functions in V1 and V2 were broader in the late time window compared to early in the delay (Figure 6), and in contrast no such changes were observed in higher-level cortical regions. These results suggest that WM-related activation in early visual cortex may rely on feedback signals originating from the higher-level visual cortex, parietal, and/or frontal cortex. During target encoding, polar angle response function in V1 and V2 was narrow and pointed, owing to the small receptive field sizes in these areas. Feedback signals originating from higher-level cortical areas, with larger receptive fields, might broaden the widths of the response functions in early visual cortex during WM. Overall, a shift from bottom-up sensory signals to top-down WM-related feedback led to

the dynamics we observed in early visual cortex. These results are consistent with laminar recording in macaque V1 (van Kerkoerle et al., 2017) and fMRI measurements of human V1 (Saber et al., 2015; Rahmati et al., 2018; Kwak and Curtis, 2022) showing that the persistent activity during WM in V1 has top-down origins. Beyond WM, two recent studies also employed top-down feedback signals to explain the decline of the precision of spatial representations in early visual cortex during mental imagery (Breedlove et al., 2020) and episodic memory (Favila et al., 2019), relative to stimulus encoding.

Format of WM representations. Researchers have suspected that mnemonic codes are somehow different from perceptual codes given that classifiers trained on fMRI patterns during visual stimulation often showed lower decoding performance during WM maintenance (Harrison and Tong, 2009; Rademaker et al., 2019; Kwak and Curtis, 2022). Moreover, WM representations remain decodable in the presence of visual distractors (Lorenc et al., 2018; Rademaker et al., 2019; Hallenbeck et al., 2021) suggesting that because these representations do not cause catastrophic interference they perhaps have unique representational formats (Libby and Buschman, 2021).

At the population level in visual cortex, early in the delay we observed strong activity in voxels with receptive fields overlapping the visual target. Later in the delay, during the epoch dominated by a stable WM code, we found increased activity across a range of eccentricities that were more foveal than the target location (Figure 5). Increased activation towards the fovea might reflect the requirement to fixate and monitor the central visual stimulus for the signal to initiate the delayed memory-guided saccade. However, we could decode target angle even when we restricted our analysis to voxels with receptive fields just a few degrees from the fovea (Figure 5K). Thus, even near foveal voxel activation contained information about the target's polar angle. These results—emergence of target's polar angle information at foveal locations—led to early and late neural coding subspaces that were partially overlapping but not orthogonal, and thereby explain why stable and dynamic code coexist in the temporal generalization analysis.

Perhaps the spread across eccentricity of the target's polar angle is similar to the spread of memorized orientation information across a retinotopic map (Ester et al., 2009) Alternatively, the activation we observed from central fixation to the eccentric target in the late time window (Figure 5B-5D) might represent the planned trajectory of the memory-guided saccade. In this case, the sensory properties of the visual target were reformatted into a mnemonic code more proximal to the behavioral demands of the task. Recently, we demonstrated that voxel activity patterns in visual cortex are recoded into a line-like spatial coding scheme when subjects are asked to remember and estimate the orientation of a grating or direction of moving dots (Kwak and Curtis, 2022). As a summary statistic of each stimulus type, a line is both efficient and useful when behaviorally estimating memorized angles. These findings together with our current findings indicate that WM representations are surprisingly flexible. Within early visual cortex, sensory representations appear to transform into representations that depend on the type of memory-guided behavior. By implementing different task demands under the same sensory

inputs, future studies can investigate how the dynamics of WM neural code are affected by the level of abstraction or reformation of sensory information required in memory-guided behavior.

METHODS

Procedures. The details of the main experiment, the passive viewing experiment and the retinotopic mapping sessions have been previously reported in (Li et al., 2021). In the main experiment, participants performed a memory-guided saccade task in the fMRI scanner. Each trial started with the onset of the WM target, a light gray dot (0.65° diameter) with a duration of 500 ms. The target was at 12° eccentricity and the polar angle of the target was pseudo-randomly sampled from 1 of 32 locations evenly tiling a full circle. The target was followed by a 12-second delay period, during which the participants were required to maintain their gaze at the fixation point while remembering the location of the target dot. After the delay, the fixation point changed from a light gray circle into a gray filled dot, serving as the response cue. In addition, an black ring whose radius matches the eccentricity of the target was presented. Upon the onset of the response cue, participants reported the location of the target by making a saccadic eye movement onto the black ring. The reported location was first read out by the eye tracker, and a dot was presented at the reported location. Participants were allowed to further use a manual dial to adjust the reported location, and they pressed a button to finalize the memory report. Upon the button press, an arc centered at the reported location appeared on the ring. The participants were asked to use the manual dial to adjust the length of the arc in a post-estimation wager, in which they should reflect the uncertainty of their WM by the length of the arc, the longer the arc the more uncertain. Participants finalized the uncertainty report by a button press, after which a white dot was presented at the true target location as the feedback. Participants could earn points if the true target location fell within the arc (see details in Li et al., 2021). The results of uncertainty reports are detailed in an earlier study (Li et al., 2021).

A subset of participants (n=6) joined an additional passive viewing experiment. The timing of this experiment was similar to the main experiment. Instead of a brief and dim WM target stimulus, we presented a salient high-contrast flickering checkerboard (0.875 deg radius; 1 cycle/deg spatial frequency; 8 Hz flicker) at the same locations as the main experiment. The checkerboard was presented for 12.5 seconds (throughout the WM target period and the delay in Figure 1A), during which the fixation point, a '+' symbol, changed its weight-height ratio, and the participants were asked to attend to the fixation and detect the changes by button presses.

Each participant was scanned for a 1.5-2 hour retinotopic mapping session. The procedures of the retinotopic mapping session followed those used in Mackey et al (2017). In short, during the retinotopic mapping session, participants maintained their gaze at the fixation point while a bar sweeping across the screen 12 times per run in various directions. Participants were required to attentively track the bar and performed a motion discrimination task based on the random dot kinematograms presented within the bar apertures (see details in Mackey et al., 2017). We fit a pRF model with compressive spatial summation to the BOLD time series of the retinotopic mapping session (Dumoulin and Wandell, 2008; Kay et al., 2013). We projected the voxels'

preferred phase angle and eccentricity on the cortical surface and defined ROIs by visual inspection (primarily based on the reversal of voxels' preferred phase angle). We define bilateral dorsal visual ROIs V1, V2, V3, V3AB, IPS0, IPS1, IPS2, IPS3 and two frontal ROIs, iPCS and sPCS.

Setup and eye tracking. Visual stimuli were presented by an LCD (VPixx ProPix) projector located behind the scanner bore. Participants viewed the stimuli through an angled mirror with a field of view of 52° by 31°. We presented a gray circular aperture (30° diameter) on the screen throughout the experiments. Eye position was recorded with a sampling rate at 500 Hz using an EyeLink 1000 Plus infrared video-based eye tracker (SR Research) mounted inside the scanner bore. We monitored gaze data and adjusted pupil/corneal reflection detection parameters as necessary during and/or between each run.

Behavioral data analysis. As participants were allowed to manually adjust the reported location after the saccades, the final dot location after the manual adjustment was used as the participants' memory report. Eye position was analyzed offline. The raw eye position data were first smoothed with a Gaussian kernel, and was converted into eye velocity using the eye positions of the five neighboring time points. Saccades were detected when the eye velocities exceeded the median velocity by 5 SDs with a minimum duration of 8 ms. Trials with ill-defined primary saccade, or with the magnitude of memory error or uncertainty report (arc length) larger than mean 3 standard deviations were excluded from analyses.

Temporal generalization. We decoded the location (polar angle) of the target from the BOLD response. We focused on the BOLD response measured from 0 to 13.5 second from the delay onset (18 TRs in total). Here, decoding is a regression problem where we aimed to predict the target location y from single-trial BOLD response X (an $n_{trial} \times n_{voxel}$ matrix) by estimating weights w. As polar angle is a circular variable, we trained two regressions to predict $y_{sin} = sin(y)$ and $y_{cos} = cos(y)$, and the predicted target location was computed as $\hat{y} = atan2(\hat{y}_{sin}, \hat{y}_{cos})$. We used support vector regression (with linear kernel) in the scikit-learn Python library to estimate the weights. In a 10 fold cross-validation procedure, all the trials from a subject were separated as the training set (9/10 of the trials) and the testing set (1/10 of the trials). Regression weights were estimated using the training set's BOLD response of a particular time point. The weights were then applied to predict the target location using the testing trials' BOLD response. Note that the testing was not only applied to the time point same as the time point of the training data, but was applied to all the time points of the data of the testing trials to investigate the generalizability of the decoders. The performance of the decoder was quantified as (the absolute value of) decoding errors averaged across all the trials for each participant. In FIgure 1C, we report the mean decoding error averaged over participants.

Statistical tests for the stable and dynamic code. The stable code was represented by above-chance decoding even when the training and the testing data were from different time points. Thereby we defined the stable code as the elements in the cross-decoding matrix where the decoding error was smaller than 90° (the mean decoding error under the null hypothesis that the decoder generates random predictions uniformly distributed across the entire polar angle

space). We first applied t-test for each element in the cross-decoding matrix, and the t-scores were subject to cluster-based permutation test for identifying clusters exhibiting above-chance decoding performance. The permutation test was done by randomly permuting the decoders' predicted target location and computing the t-scores summed over the element in the most-significant cluster. This procedure was repeated 1000 times resulting in a null-hypothesis distribution of the summed t-scores, which was used to decide whether a cluster was statistically significant in the cross-decoding matrix (see details in Maris and Oostenveld, 2007). The dynamic code was defined as the off-diagonal elements whose decoding performance was significantly poorer than their corresponding on-diagonal elements. Note that for an off-diagonal element to be included in a cluster, its decoding performance had to be lower than both of its corresponding on-diagonal elements. The cluster-based permutation test was done by randomly permuting the locations of the on- and off-diagonal elements. Overall, the definitions and the statistical tests for the stable and dynamic code were similar to those used in a previous monkey neurophysiological study (Spaak et al., 2017).

Stable and dynamic subspaces. We used PCA (principal component analysis) to define low-dimensional subspaces that encode target locations. We defined a data matrix $\mathbf{X}$ with a size of $n_{stimulus}$ by n_{voxel}, which represented the voxel activity patterns averaged across all the trials from the same stimulus location. For all the PCA conducted in this study, for the purpose of higher signal-to-noise ratio, we binned four neighboring target locations together resulting in 8 stimulus locations for the data matrix ($n_{stimulus} = 8$). n_{voxel} was the number of voxels of each ROI. $\mathbf{X}$ had column-wise zero empirical mean, as the mean of each column was removed. When visualizing the subspaces (Figure 2, 3E and 3G), for each ROI, we concatenated the voxels across all participants, and applied the PCA on the participant-aggregated ROI. When computing the indices that quantified the stability of the subspaces—the principal angle and the ratio of variance explained—PCA was applied to individuals' ROIs and the indices were computed for each participant.

For the stable subspaces, we disregarded the time-varying information by averaging the data across all the time points that fell within the stable cluster during the delay (pink dashed lines in Figure 1C). Overall, all the time points within the delay were included except the first 2 or 3 TRs depending on the ROI. To obtain the principal components (PCs), we applied eigendecomposition on the covariance matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X} = \mathbf{W}\Lambda\mathbf{W}^{\mathrm{T}}$, where each column of $\mathbf{W}$ was a unit-length vector, with a size of n_{voxel} by 1, representing the weights of each PC and Λ was a diagonal matrix containing the corresponding eigenvalues. Throughout this study, we used the first two PCs with the largest eigenvalues to define the subspaces; thereby we focused on the weight matrix $\mathbf{W}$ with a size of n_{voxel} by 2. To visualize the dynamics of population neural responses in the stable subspace, we projected the data of each time point into the stable subspace by computing $\mathbf{T} = \mathbf{X}\mathbf{W}$, where $\mathbf{X}$ was the data matrix of a single time point and $\mathbf{T}$, with a size of 8 by 2, was the projection of the voxel activity pattern of each stimulus location in the subspace defined by the top two PCs, PC1 and PC2. $\mathbf{T}$ is often referred to as PC scores in the context of PCA.

For the dynamic subspaces, we defined an early time window and a late time window. The early time window was set at 2.25 to 4.5 second from delay onset. This time window (with a duration of 3 TRs) was defined based on the dynamic clusters in the cross-decoding matrix. The significant dynamic

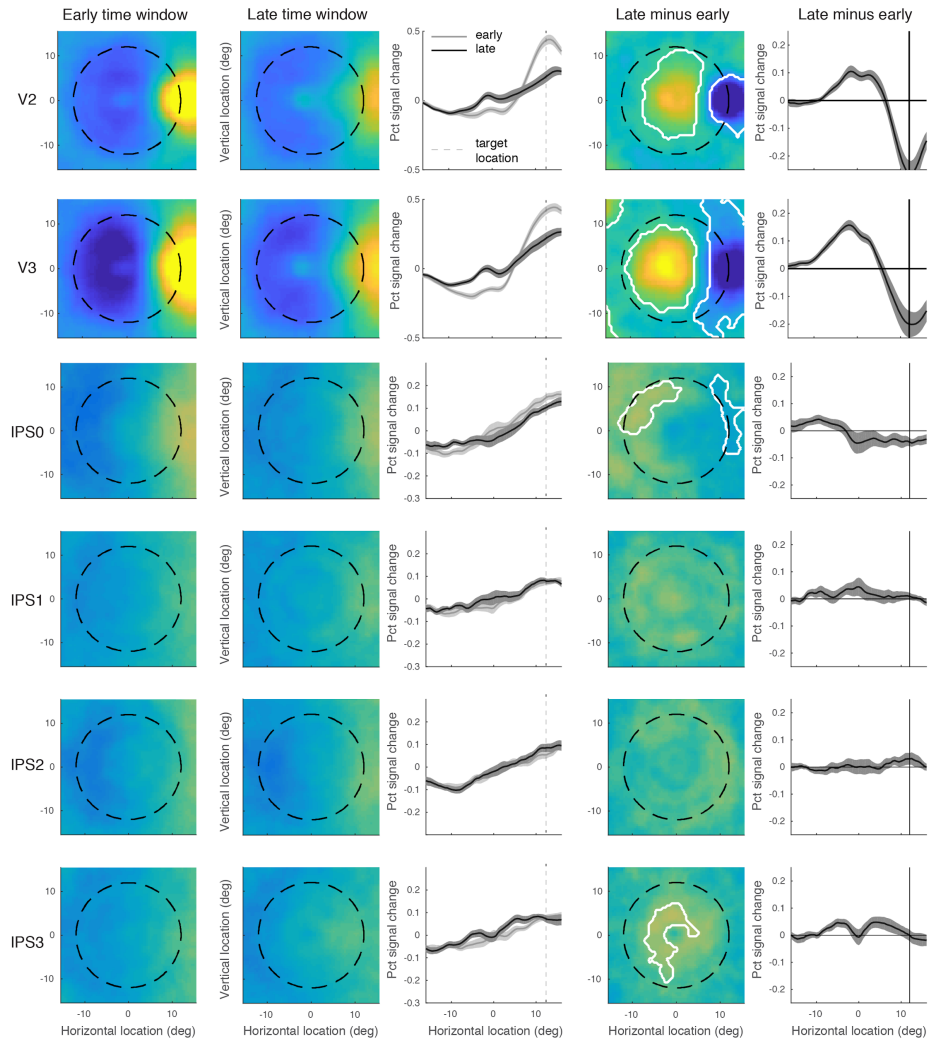
clusters indicated that the decoders trained at these early time points exhibited lower performance when decoding the data at the later time points (blue solid lines in Figure 1C). The late time window, with a duration of 3 TRs, was defined as the last 2.25 second of the delay period. The early and the late subspaces were obtained by applying PCA to the data matrices, $\mathbf{X_e}$ and $\mathbf{X_l}$, which were the BOLD response averaged over either the early or the late time windows.

To quantify how the early and the late subspaces oriented in the high-dimensional neural space, we computed the principal angle, which measured the alignment between two subspaces (Björck and Golub, 1973; Gallego et al., 2018; Xie et al., 2022). The principal angle was computed using the method proposed by Björck and Golub (1973): We applied singular decomposition to the inner-product matrix $\mathbf{W_e}^{\mathrm{T}}\mathbf{W_l} = \mathbf{P_e}\mathbf{C}\mathbf{P_l}^{\mathrm{T}}$, where $\mathbf{W_e}$ and $\mathbf{W_l}$, both had a size of n_{voxel} by 2, were the weighting matrices of the early and the late subspaces obtained by PCA. The matrix $\mathbf{C}$ was a diagonal matrix whose diagonal elements were the ranked (from small to large) cosines of the principal angles θ_1 and θ_2: $\mathbf{C} = \mathrm{diag}(\cos(\theta_1), \cos(\theta_1))$. The first principal angle was reported in Figure 3B. We also computed the principal angle for two subspaces from the same time window (dashed lines in Figure 3B). This was achieved by a bootstrapping procedure with 1000 iterations. In each iteration, we resampled the trials twice, computed two subspaces using the data of the same time window and calculated the principal angle between them, resulting in a bootstrapped distribution of the principal angle within the early (or the late) time window. In addition to the principal angle, we further computed the ratio of variance explained (RVE), which quantified how much the variance explained decreased when the data of a time window was projected to the subspace of a different time window. For example, for the data of the early time window $\mathbf{X_e}$, the RVE was computed as $\mathrm{Var}(\mathbf{X_e}\mathbf{W_l})/\mathrm{Var}(\mathbf{X_e}\mathbf{W_e})$ (dark blue data points in the top panel of Figure 3F).

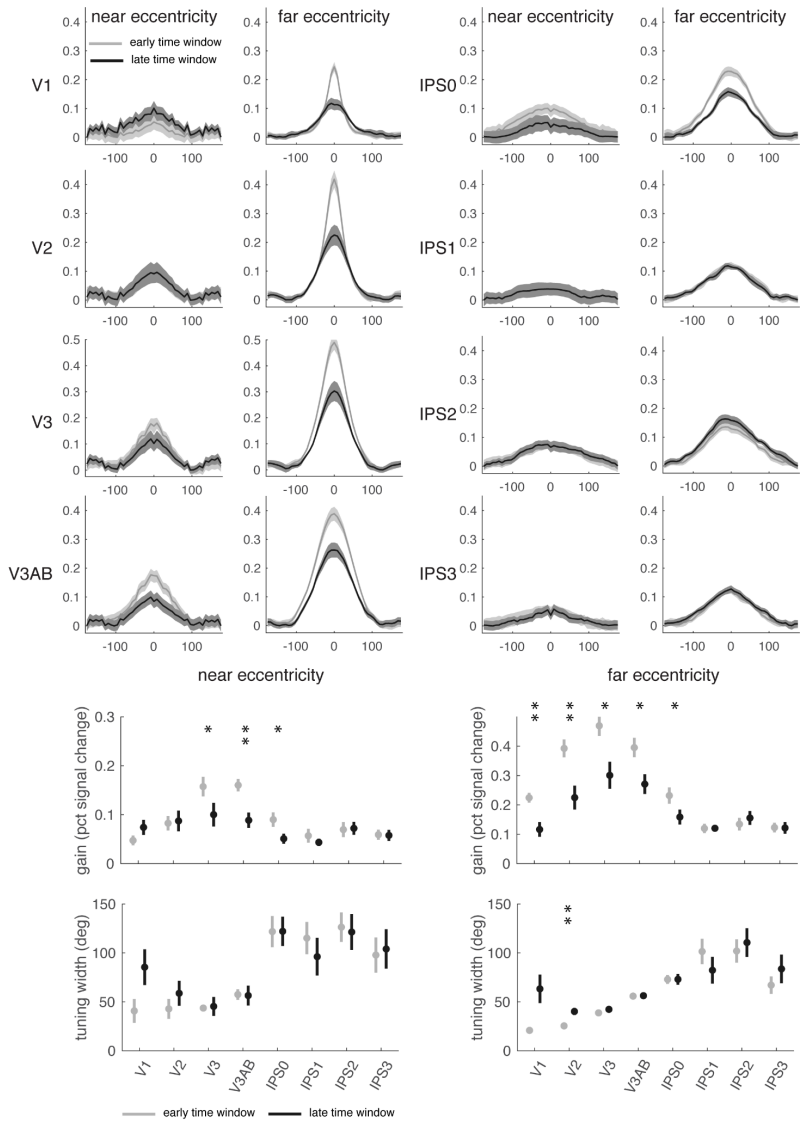
Visualizations of WM representations. We used voxels' pRF parameters to visualize how neural populations represent remembered locations. To compute the activation maps (Figure 5A,B,F and G), a grid was positioned in the visual field, centered at the fixation. The grid points evenly sampled the visual space with a step of 0.5° and the entire grid covered ±18° in both horizontal and vertical directions from the fixation point. We computed the neural activity a_i for the grid point i, whose coordinate in the visual field was (x_i, y_i), as the weighted sum of voxel responses $a_i = \frac{\sum_{v=1}^{n_{voxel}} w_{vi} r_v}{\sum_{v=1}^{n_{voxel}} w_{vi}}$. Here, r_v was the response of voxel v. All the voxels within an ROI were included except the voxels whose responses in the retinotopic mapping session can not be well-fitted by the pRF model (thresholding at variance explained by the pRF model = 10%). w_{vi} was the weight of voxel v at grid point i, which was determined by the density function of a bivariate circular gaussian distribution $N(x_i, y_i; \mathbf{u}, \sigma^{\mathbf{2}}\mathbf{I})$, in which $\mathbf{u}$ was the voxel's pRF center and σ was the voxel's pRF size. To compute an activation map for a time window and an ROI, we did the following steps: We computed an activation map for each of the 32 target locations, rotated the images to align the target at 0° polar angle, averaged the 32 images for each participant, and lastly we subtracted the grand mean from the activation map.

We computed each ROI's polar angle tuning function from the activation maps. The cartesian coordinates (x_i, y_i) of each grid point were first converted to phase angle θ_i and eccentricity r_i. We then binned the grid points based on their phase angle ranging from -180° to +180° with a

step of 8°. For the analysis reported in Figure 6, we included the grid points with eccentricity smaller than 15°. The polar angle tuning function was computed as the activity averaged over the grid points within each bin. In Extended Data Figure 6-1, we computed polar angle tuning functions separately for the near and the far eccentricity, using the grid points with eccentricity smaller than 6° or with eccentricity ranging from 6° to 15° respectively. To quantify the change of polar angle tuning functions between the early and the late time windows, the tuning functions were fitted by von Mises distributions $\frac{e^{\kappa\cos(x-\mu)}}{2\pi I_0(\kappa)}$, where $I_0(\kappa)$ was the modified Bassel function of order 0. We reported the gain of the tuning function defined as the difference between the maximum and the minimum value of the best-fit tuning function, and the tuning width represented by the fitted κ value, converted to have a unit in polar angle degree (Figure 6B).



Extended Data Figure 5-1. Visualizing WM dynamics. Plotted in the same format as Figure 5A-5J, but for the other ROIs not included in Figure 5. See the captions of Figure 5A-5J.



Extended Data Figure 6-1. Polar angle tuning functions for the near and far eccentricity. Similar to Figure 7, but here for each ROI and each time window, two polar angle tuning functions were computed separately for the near and far eccentricity, The two bottom rows are the gain and width of the tuning functions. All the data plotted represent group mean with ±1 s.e.m.

References

Baeg EH, Kim YB, Huh K, Mook-Jung I, Kim HT, Jung MW (2003) Dynamics of population code for working memory in the prefrontal cortex. Neuron 40:177–188.

Barak O, Tsodyks M, Romo R (2010) Neuronal population coding of parametric working memory. J Neurosci 30:9424–9430.

Bettencourt KC, Xu Y (2016) Decoding the content of visual short-term memory under distraction in occipital and parietal areas. Nat Neurosci 19:150–157.

Björck Åke, Golub GH (1973) Numerical methods for computing angles between linear subspaces. Math Comput 27:579–594.

Breedlove JL, St-Yves G, Olman CA, Naselaris T (2020) Generative Feedback Explains Distinct Brain Activity Codes for Seen and Mental Images. Curr Biol 30:2211–2224.e6.

Brissenden JA, Tobyne SM, Halko MA, Somers DC (2021) Stimulus-Specific Visual Working Memory Representations in Human Cerebellar Lobule VIIb/VIIIa. J Neurosci 41:1033–1045.

Brody CD, Hernández A, Zainos A, Romo R (2003) Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. Cereb Cortex 13:1196–1207.

Christophel TB, Hebart MN, Haynes J-D (2012) Decoding the contents of visual short-term memory from human visual and parietal cortex. J Neurosci 32:12983–12989.

Christophel TB, Iamshchinina P, Yan C, Allefeld C, Haynes J-D (2018) Cortical specialization for attended versus unattended working memory. Nat Neurosci 21:494–496.

Christophel TB, Klink PC, Spitzer B, Roelfsema PR, Haynes J-D (2017) The Distributed Nature of Working Memory. Trends Cogn Sci 21:111–124.

Constantinidis C, Franowicz MN, Goldman-Rakic PS (2001a) The sensory nature of mnemonic representation in the primate prefrontal cortex. Nat Neurosci 4:311–316.

Constantinidis C, Franowicz MN, Goldman-Rakic PS (2001b) Coding specificity in cortical microcircuits: a multiple-electrode analysis of primate prefrontal cortex. J Neurosci 21:3646–3655.

Curtis CE, Sprague TC (2021) Persistent Activity During Working Memory From Front to Back. Front Neural Circuits 15:696060.

Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. Neuroimage 39:647–660.

Emrich SM, Riggall AC, Larocque JJ, Postle BR (2013) Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. J Neurosci 33:6516–6523.

Ester EF, Anderson DE, Serences JT, Awh E (2013) A neural measure of precision in visual

working memory. J Cogn Neurosci 25:754–761.

Ester EF, Serences JT, Awh E (2009) Spatially global representations in human primary visual cortex during working memory maintenance. J Neurosci 29:15258–15265.

Ester EF, Sprague TC, Serences JT (2015) Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. Neuron 87:893–905.

Favila SE, Kuhl BA, Winawer J (2019) Spatial perception and memory have distinct activation profiles in human visual cortex. bioRxiv:811331.

Favila SE, Kuhl BA, Winawer J (2020) Perception and memory have distinct spatial tuning properties in human visual cortex. Cold Spring Harbor Laboratory:811331 Available at: https://www.biorxiv.org/content/10.1101/811331v2.abstract [Accessed February 12, 2021].

Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331–349.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1993a) Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas." J Neurosci 13:1479–1497.

Funahashi S, Chafee MV, Goldman-Rakic PS (1993b) Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. Nature 365:753–756.

Fuster JM, Alexander GE (1971) Neuron Activity Related to Short-Term Memory. Science 173:652–654 Available at: http://dx.doi.org/10.1126/science.173.3997.652.

Gallego JA, Perich MG, Naufel SN, Ethier C, Solla SA, Miller LE (2018) Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. Nat Commun 9:4233.

Hallenbeck GE, Sprague TC, Rahmati M, Sreenivasan KK, Curtis CE (2021) Working memory representations in visual cortex mediate distraction effects. Nat Commun 12:4717.

Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. Nature 458:632–635.

Harvey CD, Coen P, Tank DW (2012) Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature 484:62–68.

Jerde TA, Merriam EP, Riggall AC, Hedges JH, Curtis CE (2012) Prioritized maps of space in human frontoparietal cortex. J Neurosci 32:17382–17390.

Kay KN, Winawer J, Mezer A, Wandell BA (2013) Compressive spatial summation in human visual cortex. J Neurophysiol 110:481–494.

King J-R, Pescetelli N, Dehaene S (2016) Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. Neuron 92:1122–1134.

Kok P, de Lange FP (2014) Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. Curr Biol 24:1531–1535.

Kubota K, Niki H (1971) Prefrontal cortical unit activity and delayed alternation performance in monkeys. J Neurophysiol 34:337–347.

Kwak Y, Curtis CE (2022) Unveiling the abstract format of mnemonic representations. Neuron Available at: http://dx.doi.org/10.1016/j.neuron.2022.03.016.

Lee S-H, Kravitz DJ, Baker CI (2013) Goal-dependent dissociation of visual and prefrontal cortices during working memory. Nat Neurosci 16:997–999.

Libby A, Buschman TJ (2021) Rotational dynamics reduce interference between sensory and memory representations. Nat Neurosci 24:715–726.

Li H-H, Sprague TC, Yoo AH, Ma WJ, Curtis CE (2021) Joint representation of working memory and uncertainty in human cortex. Neuron 109:3699–3712.e6.

Lorenc ES, Sreenivasan KK, Nee DE, Vandenbroucke ARE, D'Esposito M (2018) Flexible Coding of Visual Working Memory Representations during Distraction. J Neurosci 38:5267–5276.

Mackey WE, Winawer J, Curtis CE (2017) Visual field map clusters in human frontoparietal cortex. Elife 6 Available at: http://dx.doi.org/10.7554/eLife.22974.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.

Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J Neurosci 16:5154–5167.

Murray JD, Bernacchia A, Roy NA, Constantinidis C, Romo R, Wang X-J (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. Proc Natl Acad Sci U S A 114:394–399.

Panichello MF, Buschman TJ (2021) Shared mechanisms underlie the control of working memory and attention. Nature 592:601–605.

Rademaker RL, Chunharas C, Serences JT (2019) Coexisting representations of sensory and mnemonic information in human visual cortex. Nat Neurosci 22:1336–1344.

Rahmati M, DeSimone K, Curtis CE, Sreenivasan KK (2020) Spatially Specific Working Memory Activity in the Human Superior Colliculus. J Neurosci 40:9487–9495.

Rahmati M, Saber GT, Curtis CE (2018) Population Dynamics of Early Visual Cortex during Working Memory. J Cogn Neurosci 30:219–233.

Riggall AC, Postle BR (2012) The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. J Neurosci 32:12990–12998.

Romo R, Brody CD, Hernández A, Lemus L (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. Nature 399:470–473.

Saber GT, Pestilli F, Curtis CE (2015) Saccade planning evokes topographically specific activity in the dorsal and ventral streams. J Neurosci 35:245–252.

Serences JT, Ester EF, Vogel EK, Awh E (2009) Stimulus-specific delay activity in human primary visual cortex. Psychol Sci 20:207–214.

Shafi M, Zhou Y, Quintana J, Chow C, Fuster J, Bodner M (2007) Variability in neuronal activity in primate cortex during working memory tasks. Neuroscience 146:1082–1108.

Spaak E, Watanabe K, Funahashi S, Stokes MG (2017) Stable and Dynamic Coding for Working Memory in Primate Prefrontal Cortex. The Journal of Neuroscience 37:6503–6516 Available at: http://dx.doi.org/10.1523/jneurosci.3364-16.2017.

Sprague TC, Ester EF, Serences JT (2014) Reconstructions of information in visual spatial working memory degrade with memory load. Curr Biol 24:2174–2180.

Sprague TC, Ester EF, Serences JT (2016) Restoring Latent Visual Working Memory Representations in Human Cortex. Neuron 91:694–707.

Sreenivasan KK, Vytlacil J, D'Esposito M (2014) Distributed and dynamic storage of working memory stimulus information in extrastriate cortex. J Cogn Neurosci 26:1141–1153.

Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic coding for cognitive control in prefrontal cortex. Neuron 78:364–375.

van Kerkoerle T, Self MW, Roelfsema PR (2017) Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. Nat Commun 8:13804.

Wasmuht DF, Spaak E, Buschman TJ, Miller EK, Stokes MG (2018) Intrinsic neuronal dynamics predict distinct functional roles during working memory. Nat Commun 9:3499.

Watanabe K, Funahashi S (2007) Prefrontal delay-period activity reflects the decision process of a saccade direction during a free-choice ODR task. Cereb Cortex 17 Suppl 1:i88–i100.

Wimmer K, Nykamp DQ, Constantinidis C, Compte A (2014) Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. Nat Neurosci 17:431–439.

Wolff MJ, Ding J, Myers NE, Stokes MG (2015) Revealing hidden states in visual working memory using electroencephalography. Front Syst Neurosci 9:123.

Xie Y, Hu P, Li J, Chen J, Song W, Wang X-J, Yang T, Dehaene S, Tang S, Min B, Wang L (2022) Geometry of sequence working memory in macaque prefrontal cortex. Science 375:632–639.

Xing Y, Ledgeway T, McGraw PV, Schluppeck D (2013) Decoding working memory of stimulus contrast in early visual cortex. J Neurosci 33:10301–10311.

Zhou Y, Curtis CE, Sreenivasan KK, Fougnie D (2022) Common Neural Mechanisms Control Attention and Working Memory. J Neurosci 42:7110–7120.