PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Dynamics of working memory drift and information flow across the cortical hierarchy

Hsin-Hung Li[a,b], Wei Ji Ma[a,c], and Clayton E. Curtis[a,c,1]

Edited by Earl K. Miller, Massachusetts Institute of Technology, Cambridge, MA; received July 15, 2025; accepted December 15, 2025 by Editorial Board Member Charles D. Gilbert

Working memory is supported by widespread and distributed brain regions spanning across the cortical hierarchy. However, how working memory content evolves and is transmitted across cortical regions remains largely unknown. Here, we investigated the flow of working memory information across the cortex using time-resolved fMRI decoding. Across multiple regions in visual and parietal cortex, we found that decoded working memory content drifted over time from the memoranda toward the later errors in memory reports, consistent with the dynamics of memory drift predicted by attractor models. These behaviorally predictive neural memory errors emerged earliest in higher-order dorsal visual area (V3AB) and intraparietal sulcus (IPS0) and then later in early visual cortex (V1), suggesting a propagation of mnemonic information originated in higher-level visual area. Interareal correlation analyses revealed that during memory maintenance, information flowed predominantly in a top–down manner—from higher to lower visual areas, whereas during passive viewing, feedforward dynamics prevailed. Together, our findings demonstrate that working memory maintenance involves temporally structured drift dynamics and feedback-dominated information flow across the cortical hierarchy, providing a mechanistic link between neural population dynamics and the formation of memory errors.

working memory | fMRI | dynamics | visual cortex

Working memory is the cognitive process that holds task-relevant information online over a period of time. Working memory is constrained by limited resources, and its precision decreases over delay (1–5). Previous studies have conceptualized the maintenance of working memory content and the emergence of its error as a drift-diffusion process where the accrual of noise over time perturbs working memory (6–8). Mechanistically, this process has been modeled using neural networks with attractor dynamics, where memory content is stored in the presence of neural noise [(9–12), reviewed in ref. 13]. Findings from single-cell recordings (14–16) and EEG studies in humans (17, 18) are broadly in line with attractor dynamics for working memory. For instance, during working memory delay intervals, population vectors encoding memorized locations from neurons in macaque lateral prefrontal cortex (PFC) drift in the direction of errors in memory-guided saccades generated after the delay (14); Similarly, neural activity in mouse anterior lateral motor cortex during memory delays converge toward discrete end points that correspond to specific movement directions (15).

Although the dynamics of working memory have been rigorously characterized in theoretical models and animal neurophysiology, analogous findings from human neuroimaging remain relatively sparse. More critically, because most prior studies focused on single brain regions—particularly the prefrontal cortex—the question of how working memory representations are transmitted across the cortical hierarchy, and how errors in memory evolve across brain regions, remains unknown.

Here, we used time-resolved fMRI decoding to investigate the temporal dynamics of spatial working memory representations across the cortical hierarchy. We found that memorized locations can be decoded in all the ROIs along the dorsal visual pathway. We used a memory-guided saccade task like the one used in macaque studies because it yields a continuous metric of memory errors whose direction and amplitude can be quantified. In preview, we found that neural decoding errors drifted from the veridical targets toward errors in the direction of later memory-guided saccades, in line with drift dynamics. Furthermore, by correlating the working memory representations between cortical regions, we found that information flow across cortical regions was task dependent. Specifically, during memory delays working memory information was dominated by a feedback flow from higher-level visual cortex to early visual cortex.

Significance

Working memory stores information necessary for decision-making and planning. While it is now clear that the contents of working memory are widely distributed across the brain, how this information evolves and dynamically flows between regions remains unknown. Leveraging time-resolved functional magnetic resonance imaging (fMRI), we demonstrated that memory representations remarkably emerged first in high-level visual and parietal cortices and later propagated to early visual cortex, revealing a top–down flow of information. Critically, we also demonstrated that errors in memory result from a drift of neural representations of working memory over time, again in higher and then lower visual cortical areas. These results provide key missing evidence for theories of the neural dynamics that underlie working memory.

Author affiliations: [a]Department of Psychology, New York University, New York, NY 10003; [b]Department of Psychology, The Ohio State University, Columbus, OH 43210; and [c]Center for Neural Science, New York University, New York, NY 10003

Author contributions: H.-H.L., W.J.M., and C.E.C. designed research; H.-H.L. performed research; H.-H.L. analyzed data; W.J.M. and C.E.C. supervised the research; and H.-H.L. and C.E.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. E.K.M. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: clayton.curtis@nyu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2518110123/-/DCSupplemental.

Published January 23, 2026.

Results

In a memory-guided saccade task, participants maintained fixation at the screen center and held the location of a briefly presented target in memory over a 12-s retention interval (Fig. 1). Target dot was 12° away from the fixation, with its polar angle pseudorandomly assigned to span an imaginary circle across trials. After the delay, participants reported the target location by generating a memory-guided saccade. Each participant also completed a pRF (population receptive field) session, allowing us to define regions of interest (ROIs) by individual's visual field maps. Based on prior studies (19–26), we focused on the ROIs that exhibited the strongest decodable working memory signals. These ROIs include primary visual cortex (V1), extrastriate cortex along the dorsal visual pathway including visual cortex (V2, V3, V3AB) and intraparietal sulcus (IPS0, IPS1, IPS2, IPS3). The current study used data from experiments previously published (22, 23).

Temporal Dynamics of Working Memory Representations. To investigate the neural dynamics of working memory, we used a Bayesian method (22, 27, 28) to decode target locations (polar angles) for single trials using BOLD signals at each time point during the memory delays (TR = 750 ms). Information about target locations, quantified as the circular correlation between decoded and actual target locations, emerged rapidly in all ROIs (Fig. 2 *A* and *B*). In V3 and V3AB, this correlation became significant at about 2.25 s (the third TR) into the delay, while decoding in other ROIs emerged at about 3 s (the fourth TR). In most ROIs, target information peaked at about 4 to 5 s into the delay, consistent with stimulus encoding, and then maintained above-chance throughout the rest of the memory delay. To assess whether these decodable signals were specific to working memory maintenance, we conducted a passive viewing experiment, in which the "target" was a continuous flicker presented throughout the delay, and therefore had no memory demands. Participants were instructed to ignore the flicker while concentrating on an attentionally demanding task at fixation. In this condition, decoding performance also peaked at early time points but then returned to chance levels midway through the delay, indicating that sustained decodability relies on the engagement of the working memory (*SI Appendix*, Fig. S1).

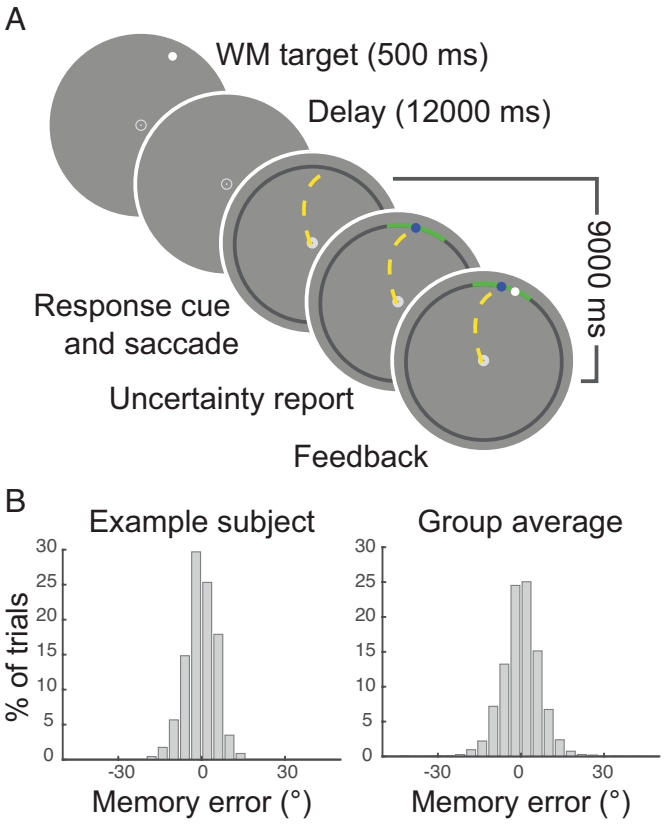

Fig. 1. Tasks and behavioral performance. (*A*) In the memory-guided saccade task, each trial starts with a target presented at 12° eccentricity and a pseudorandomly assigned polar angle. Participants have to maintain their gaze at the central fixation point throughout the delay, and use a saccadic eye movement to report the memorized location after the onset of the response cue (the onset of the black ring). After the memory report, participants report their memory uncertainty by adjusting a CI. Here, we focused on the memory reports while the results regarding the uncertainty report were described in a published study (22). (*B*) Histograms illustrate the distribution of memory errors in degree polar angle.

We next focused on neural responses that predicted behavioral memory reports beyond the target's location. We asked whether decoding errors were predictive of behavioral memory errors by correlating the two. Our goal is to identify the neural signals that best correlate with overall behavioral memory errors; therefore, we computed the difference between the reported locations and target locations. This error measure is contributed by both variability and bias (*Discussion* and *SI Appendix*, Fig. S2 and *Text*). Thereby, the "drift" in the present study represents the overall memory error, similar to the "memory drift" in some prior literature, which refers to any shifts of bump activity within a trial (8, 14, 17). This is different from some previous studies that distinguish stimulus-specific drift (bias) from memory errors that are caused by random fluctuations (6, 29, 30).

We found that the time courses of these error correlations (Fig. 2*D*) diverged markedly from those based on the target locations alone (Fig. 2*B*). In extrastriate visual cortex V2, V3, V3AB and intraparietal sulcus area IPS0, these error correlations gradually ramped up over the delay. To further visualize this drift, we binned trials according to the direction of behavioral memory errors (clockwise CW vs. counterclockwise CCW relative to the target), which confirmed that decoding errors drifted in the direction of upcoming behavioral errors in memory (Fig. 2*F*). We also observed significant error correlations in early visual cortex V1, but the latency of these correlations were significantly delayed relative to those in extrastriate visual cortex and IPS0 (cluster-based permutation test $P < 0.05$, Fig. 2 *D* and *E*). In V1, neural decoding errors were not predictive of behavioral errors in memory until the TR right before the response cue (Fig. 2 *D* and *E*). Overall, these error correlations are consistent with the predictions by drift-diffusion processes and neural networks with attractor-like dynamics (11, 14). We observed such dynamics in multiple visual maps across the cortical hierarchy, but with systematic differences in their timing. In the time courses of both Fig. 2 *B* and *D*, working memory-related signals in high-level visual area and IPS0 preceded those in early visual cortex, such as V1. This temporal order suggests that the neural representations that are predictive of memory error are first formed in higher-order regions and subsequently propagated backward to early visual cortex. To quantify this information flow during the memory delay, we focused on two areas that showed the earliest and latest dynamics—V3AB and V1, respectively.

While sustained univariate activity has been observed in the prefrontal cortex during working memory tasks in human fMRI studies, the amount of decodable stimulus information in this region is typically low. We report the decoding results from the superior precentral sulcus (sPCS), which likely contains the human homologue of the macaque frontal eye field (FEF), in *SI Appendix*, Fig. S3. This region is among the few prefrontal areas in which both our group and others have found decodable working memory content (22, 26, 31–34). We observed above-chance decoding during a time

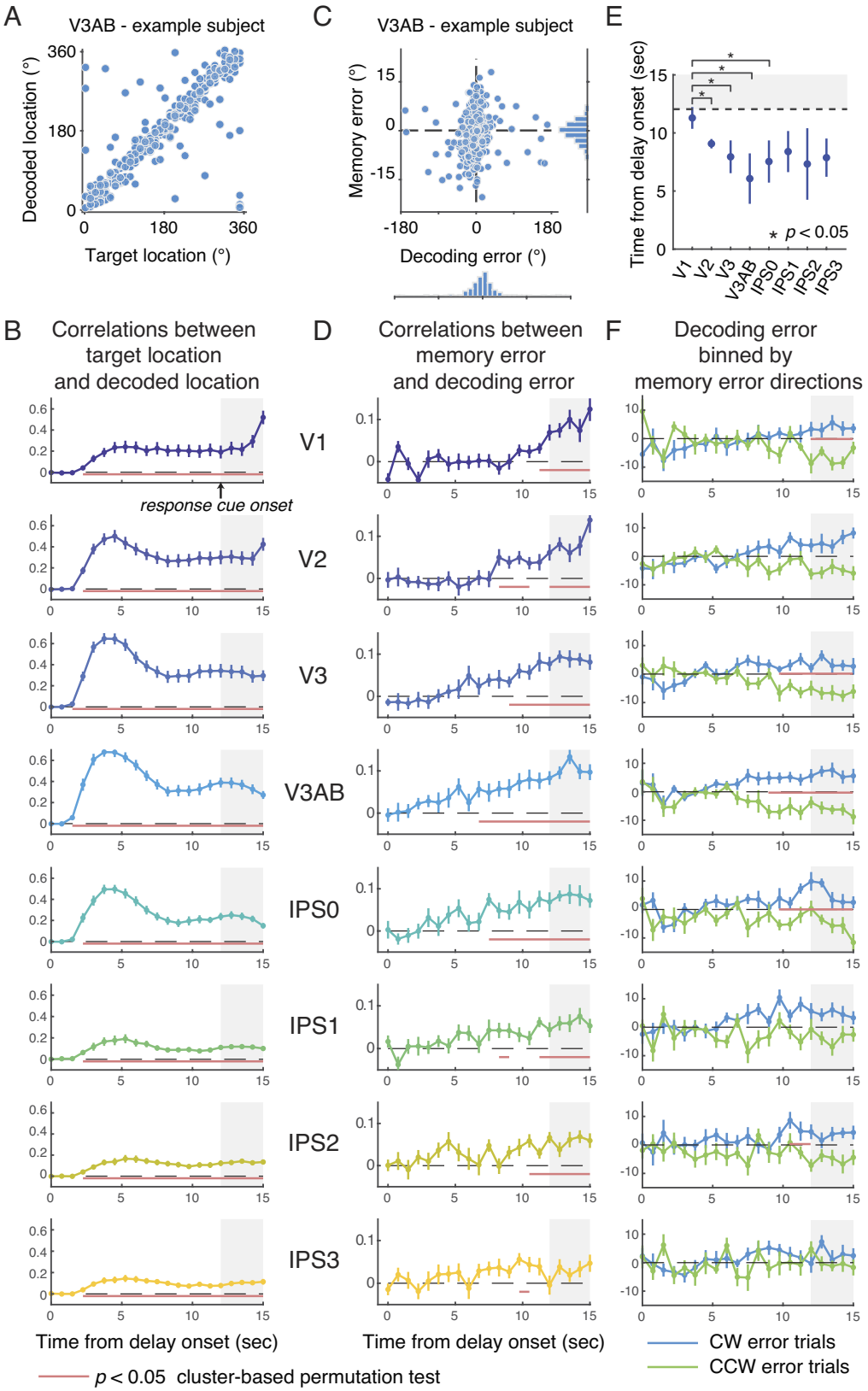

Fig. 2. The temporal dynamics of working memory signals. (*A*) Decoded location and target location from V3AB of an example participant. (*B*) The time courses of the circular correlations between decoded locations and target locations. The red horizontal lines indicate the clusters where correlations are significantly above zero based on cluster-based permutation tests. The gray shaded area marks the time from the onset of the response cue. The data points represent mean ± 1 SEM. (*C*) Decoding error and behavioral memory error of an example participant. (*D*) The time courses of the circular correlations between decoding errors and behavioral memory errors. (*E*) The onset times of significant correlations between decoding errors and behavioral memory errors. The error bars represent ±1 bootstrapped SD. (*F*) The time courses of decoding errors for trials in which the participants made clockwise (blue) and counterclockwise (green) errors.

window in the delay period and again during the response period, although the overall decoding performance was substantially lower than that in visual and parietal cortices (*SI Appendix*, Fig. S3*A*). The decoding error in sPCS did not predict behavioral memory error (*SI Appendix*, Fig. S3*B*). The absence of error–error correlation likely reflects the lower signal-to-noise ratio for time-resolved decoding in this region, as in our previous analyses using time-averaged activity (rather than single time points), the error–error correlation remained significant across multiple experiments (22). We discuss in the Discussion section how measurement resolution and task-related factors may contribute to reduced working memory decoding performance in the prefrontal cortex.

Information Flow during Working Memory. To investigate the information flow between cortical regions, we next computed trial-wise interareal correlations of decoded target locations between these two regions (Fig. 3*A*). In Fig. 3*B*, using V1 as the seed region, we correlated its decoded target locations at each time point with those from V3AB at later time points (a feedforward flow). Conversely, in Fig. 3*C*, using V3AB as the seed, we correlated its decoded values with those from V1 at later time points (a feedback flow). Throughout the delay period, interareal correlations were consistently stronger when V3AB served as the seed, indicating a predominant flow of information from V3AB to V1 (cluster-based permutation test $P < 0.05$, Fig. 3*D*).

We further visualized the interareal correlations in a complementary format by using V3AB as the reference (seed) region. Specifically, we correlated the decoded locations from V3AB at each time point (y-axis in Fig. 4 *A–D*) with those from V1 across a range of temporal lags (x-axis in Fig. 4 *A–D*), capturing activity in V1 that either preceded or followed that of V3AB (see also the caption of Fig. 4*A*). When correlations were summed across the delay period, we again found that correlations were stronger when V1 lagged behind V3AB, confirming a predominant feedback flow from V3AB to V1 during working memory (Fig. 4*E*). To assess whether this directionality was task-dependent, we applied the same analyses during the passive viewing experiment. In contrast to the working memory task, information flow during passive viewing was stronger in the feedforward direction—from V1 to V3AB—than in the feedback direction (Fig. 4 *C* and *G*; also see Fig. 3 *E–G*), highlighting that feedback dynamics were specific to working memory. We repeated this analysis using trial-wise decoding errors instead of decoded locations (Fig. 4 *B* and *D*). While these correlations were overall weaker, they remained positive—indicating that a portion of the neural variability (or noise) was shared between V3AB and V1 and was detectable in BOLD signals. Crucially, during working memory but not in the passive viewing experiment, the feedback flow from V3AB to V1 remained stronger than the feedforward flow, suggesting that shared noise or variability also traveled predominantly from higher-level visual cortex to early visual areas during memory maintenance (Fig. 4 *F* and *H*).

Drift of Bump Activity underlying Memory Error. In network models with drift dynamics, memory maintenance is often described as a localized "bump" of activity centered on the memorized target (11, 14, 35). To visualize such bump activity, we projected neural responses in V3AB into the coordinates of a two-dimensional visual field, where the center corresponds to the fixation point (*Methods*). We visualized the neural activity for three bins of trials sorted by the direction of behavioral memory errors (Fig. 5*A*). Reconstructed neural responses showed a bump at the target's polar angle (Fig. 5*B*). Comparisons between CW and CCW trials revealed significant differences at locations flanking the target, consistent with two bumps drifting in opposite directions toward the direction of errors (CW or CCW) in the later memory-guided saccades (*Rightmost* panel, Fig. 5*B*).

Previous studies have suggested that memory errors may arise from both a drift of the center of the bump of activity and a decrease in the gain of the bump (8, 36). To investigate these two possibilities, we collapsed the two-dimensional reconstructions to one-dimensional response functions over polar angle, visualizing unimodal bumps of activity centered around the polar angle of the targets (Fig. 5*C*). By fitting von-Mises functions to the polar angle response functions, we found that the memory errors were mainly driven by a drift in the centers of the bumps ($p < 0.01$, the main effect of bins in permutation ANOVA; Fig. 5*D*).

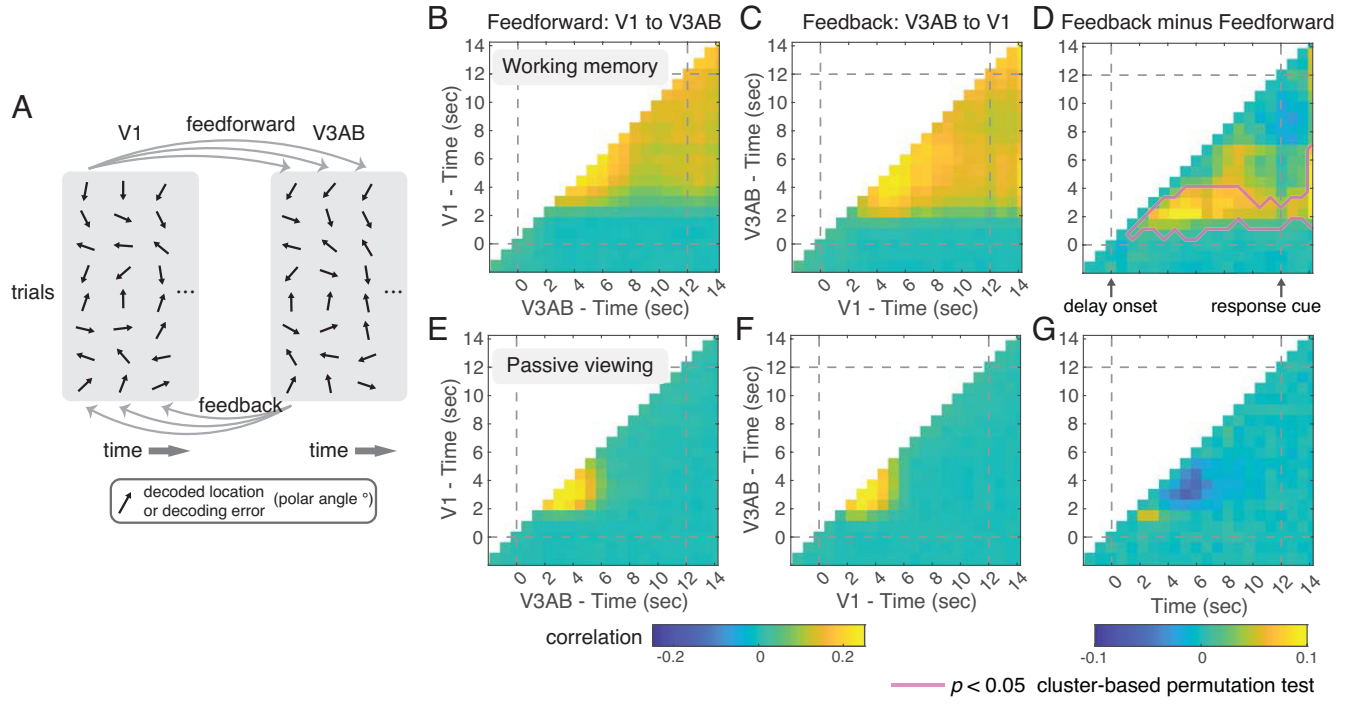

Fig. 3. Flow of memory information quantified by interareal correlations. (*A*) Information flows between V1 and V3AB calculated as the correlations of working memory signals (decoded locations or decoding errors) across the two regions and across different time points. Feedforward flow is represented by the correlations between V1's working memory signals at a given time point and V3AB's working memory signals at later time points. Feedback flow is represented by the correlations between V3AB's working memory signals at a given time point and V1's working memory signals at later time points. (*B*) Interareal correlations based on decoded locations in the feedforward direction. (*C*) Interareal correlations based on decoded locations in the feedback direction. (*D*) The interareal correlations in the feedback direction minus that in the feedforward direction. The pink outline indicates a cluster where the difference is significant based on a cluster-based permutation test. (*E–G*) Similar to (*B–D*) but for the passive viewing experiment.

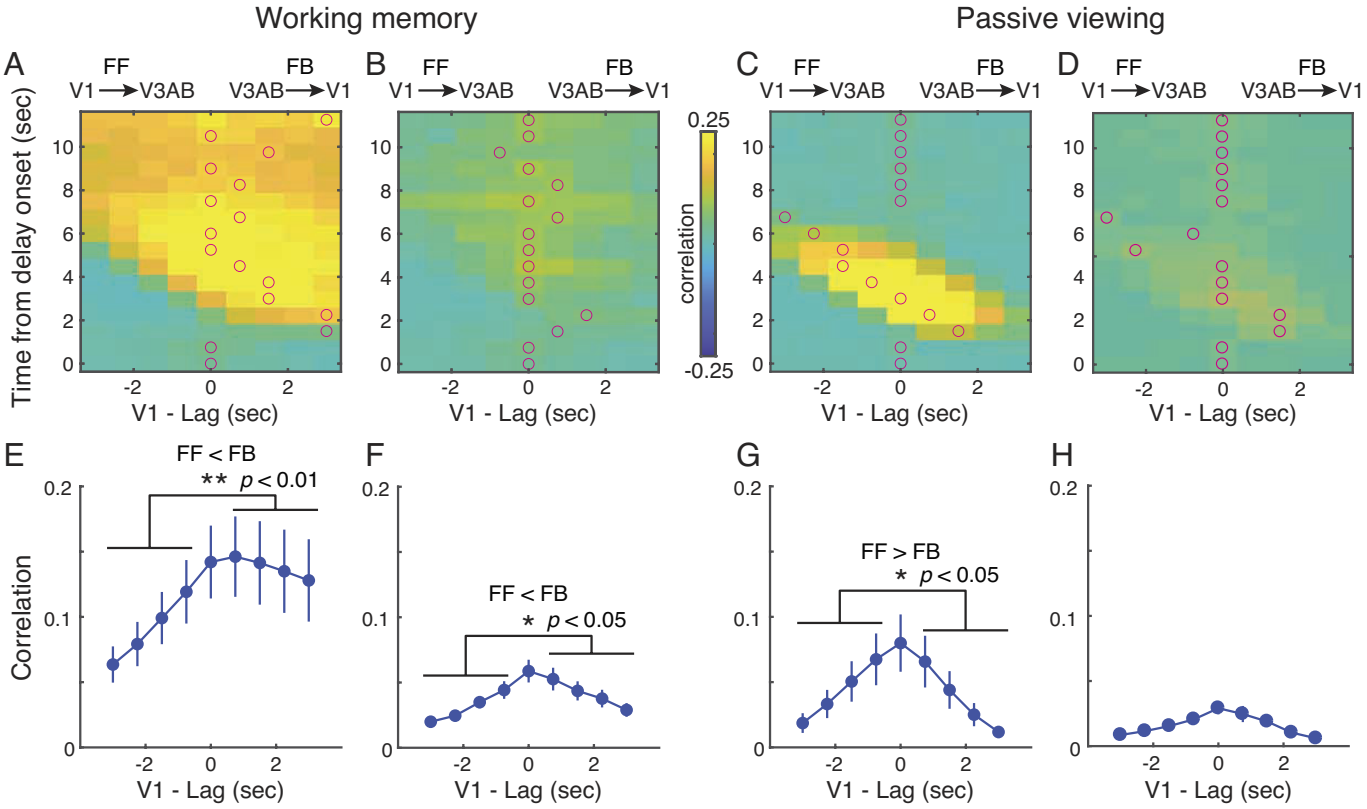

Fig. 4. Time lags in interareal correlations. (*A*) Interareal correlations based on decoded locations. Similar to Fig. 3 *B* and *C*, but the results are realigned so V3AB is the reference time axis (the *y*-axis), and the decoded locations from V3AB at each time point are correlated with the decoded location from V1 with different time lags (*x*-axis). Thus, the left half of the image represents feedforward (FF) flow (V1 ahead of V3AB) and the right half of the image represents feedback (FB) flow (V1 lags behind V3AB). In each row, red circles denote peaks in the interareal correlations for each reference time point (row). (*B*) Similar to (*A*), but here the interareal correlations are computed using decoding errors instead of decoded locations. (*C*) Similar to (*A*) but for the passive viewing condition. (*D*) Similar to (*B*) but for the passive viewing condition. (*E*–*H*) The averaged interareal correlation for each time lag, computed by averaging over each column in the images in (*A*–*D*), respectively. Paired *t* tests were used to compare the overall magnitude of the FF flow (average of the four data points with negative lags) and FB flow (average of the four data points with positive lags). The data points represent mean ± 1 SEM.

Discussion

We investigated how neural representations of working memory content evolve over time across the cortical hierarchy in humans. We observed drift dynamics, with decoding errors systematically drifting toward the direction of subsequent errors in memory, across multiple visual maps in visual and parietal cortex. This error-predictive signal emerged first in V3AB and IPS0 and then later in early visual cortex. Interareal correlations of working memory content during working memory maintenance further confirmed that information flowed from high-level visual cortex to early visual cortex. Together, these results provide evidence that working memory maintenance involves drift dynamics, not just within, but critically across the cortical hierarchy, and is supported by a feedback cascade from parietal-occipital regions to early visual cortex.

Early studies on the neural mechanisms that support working memory mostly restricted their focus to the lateral PFC [(37, 38), reviewed in refs. 39 and 40]. However, accumulating evidence from recent studies has led to the emerging view that representations of working memory content are distributed across multiple brain regions (20, 22, 23, 31, 41–45). Whether or how neural representations of working memory may differ and serve different roles across brain regions remains an open question (43, 46–48), but one that deserves attention. Our results in early visual cortex are generally consistent with a recent study that performed time-resolved decoding on a single ROI combining V1 to V3, where the decoded orientation drifted toward the direction of stimulus-specific memory error during the delay (29). Here, we found that among visual field maps along the dorsal visual pathway, all of which exhibited above-chance performance in decoding the working memory target, and more importantly, they showed different temporal dynamics when we analyzed the signals predictive of behavioral memory errors.

The finding that V3AB and IPS0—two regions situated at the boundary between the occipital (visual) and parietal cortices—exhibited the earliest error correlations (Fig. 2 *D* and *E*) suggests that they may serve as a driving source that broadcasts the representations of working memory content to early visual cortex after the disappearance of the target stimulus. Interestingly, in our previous study, we observed that the neural code of working memory was the most stable in V3AB and IPS0, whereas the neural code of memory targets in V1 underwent significant changes between the stimulus-evoked period to memory maintenance (23); another recent study reported that the format of working memory in early visual cortex was "aligned" to that of the IPS during memory delay (47). Together, these findings provide converging evidence that working memory representations in early visual cortex are shaped by those originating from higher-level dorsal visual and posterior parietal regions.

This idea that information flows from V3AB to V1 during working memory was corroborated by interareal correlations based on decoded location and errors. Previous neurophysiological studies have used regression (49) or canonical correlation (50) to investigate the information flow between cortical regions. Our approach differed in that we first converted multivariate neural responses to the task-relevant dimension—target's polar angle—and used it

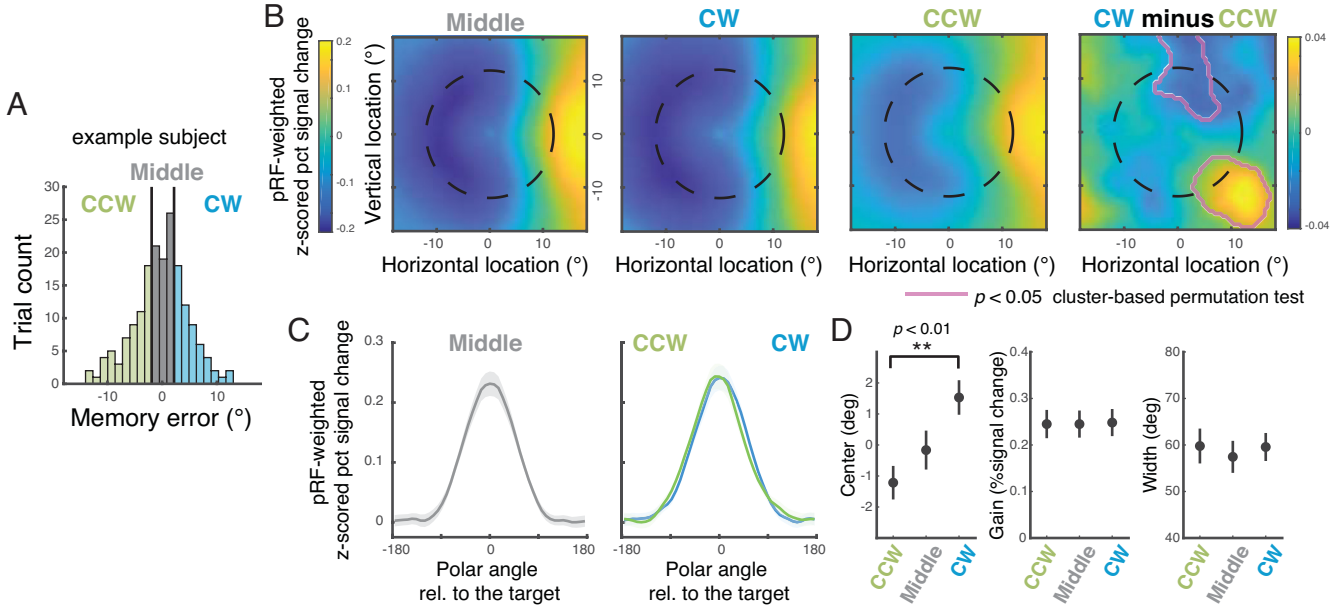

Fig. 5. Visualizations of population activity in V3AB. (*A*) Each participant's trials were placed into three bins with an equal number of trials based on the direction and magnitude of memory errors. (*B*) Activation maps visualize the projection of voxel activity patterns onto two-dimensional visual field space. The first three images: Each image represents the activation map reconstructed using the activity of V3AB averaged over the late delay duration (7 to 12 s from delay onset). The right-most image, the difference between the activation map of the CW trials and the CCW trials. (*C*) Polar angle response functions for trials with small errors ("Middle" bin, gray curve), CW errors (blue curve), and CCW errors (green curve). (*D*) Best-fit parameters for polar angle response functions. The data points represent mean ± 1 SEM.

to quantify the information flows between pairs of brain regions (Fig. 3*A*). Note that even though the interareal correlations computed based on decoded locations were associated with the stimulus locations by definition, they were task-dependent. Indeed, the interareal correlations were much stronger during the working memory tasks compared to a passive viewing condition. Moreover, only during working memory did we observe a dominance of feedback flow (Fig. 3), indicating that this feedback flow is specific to working memory function. The interareal correlations computed based on decoding errors were smaller in magnitude compared to those computed using decoded locations, but were still significantly above zero (Fig. 4*B*). These results indicate that while a major proportion of the neural (or measurement) noise was independent between ROIs, some variability may be shared across brain regions. Our analysis showed that such variability shared between brain regions also exhibited stronger feedback flow than feedforward flow during working memory delay. Overall, our results infer a directional or even causal influence of working memory representations in V3AB and IPS0 on those in V1. Future studies using neurostimulation approaches could directly test this directional relationship.

The present findings echo some previous neuroimaging studies that also reported decoding error or multivariate neural activity to correlate with behavioral errors (17, 22, 51). In the literature of perceptual decision-making, this type of choice-related signal, conditioned upon the stimulus, is often considered a strong indicator of a tight link between the measured neural responses and behaviors (52–55). Our results here are generally consistent with neurophysiology studies reporting that choice-related or choice probability is weaker in early sensory cortex, and stronger in higher-level regions (52, 54–59). Different from our results, a previous fMRI study reported that the behaviorally consistent bias in neural decoding was higher in the early visual cortex (EVC) than in IPS regions (51). However, the EVC in this study included combined striate (V1) and multiple extrastriate visual areas (V2 to V4). It is unknown whether the behaviorally consistent bias they found in EVC was mainly driven by extrastriate cortex like V3. Similarly, their IPS area contained a combined ROI including IPS0 to IPS4. Thus, it remains unclear if our results are inconsistent or not. We found that the error-related signals were early and prominent in IPS0, but were much weaker from IPS1 and beyond. Aggregating all IPS subdivisions into a single ROI may obscure the specific contributions of individual subregions.

Other approaches, such as Granger causality, have been used to infer directionality of information flow in fMRI data. In general, all such methods, including ours, rely on lagged correlation or autoregressive modeling to test whether activity in one region at an earlier time predicts activity in another region at a later time. The key difference lies in what constitutes the data being analyzed. As an exploratory test, we applied a Granger causality test to the univariate time series and found no significant directionality between V1 and V3AB during working memory (*SI Appendix*, Fig. S4 and *Text*). This outcome is not unexpected given that univariate fMRI signals may contain many components unrelated to working memory content and are more susceptible to the characteristics of hemodynamic responses. We acknowledge that there are various implementations and extensions of Granger causality (60), but our decoding-based approach provides a simple and interpretable framework: By projecting neural activity onto the decoded working memory representations, we directly test the flow of mnemonic information across cortical areas.

When we visualized the neural response of V3AB in the coordinates of the visual field, we observed a bump of activity that drifted in the direction of errors in memory (Fig. 5). The drift we observed is predicted by neural networks with drift dynamics [(9–12), reviewed in ref. 13] and resembled results from electrophysiological recordings from macaque PFC while performing a similar memory-guided saccade task (14). Previous studies have postulated that errors in memory may be caused by drift of the bump of activity or a reduction of the gain (8, 36). In our study, focusing on a set size of one, we found that memory errors were primarily attributable to the drift of memory representations.

These results are in line with the attractor neural networks where the decay of memory is mainly attributed to the drift of bump activity at the stored feature value (11, 14, 61). However, this does not preclude the possibility that changes in gain contribute to memory variability under different conditions. For instance, in a recent study, we investigated spatial working memory with a set size of two and incorporated spatial attentional cues to manipulate behavioral relevance of the items. We observed that task relevance affected the precision of working memory representations through the change of neural gains (32). These findings suggest that multiple mechanisms may influence working memory quality. Specifically, both drift and changes in the gain (amplitude) of neural activity can contribute to memory variability, with their relative impact depending on task demands and context.

In the present study, the memory errors we analyzed reflected contributions from both variability and bias. By decomposing memory errors into these two components, we observed an "oblique effect" where variability was lower in cardinal than in diagonal polar angle, and repulsion biases around some cardinal axes (*SI Appendix*, Fig. S2 and *Text*). These patterns are consistent with previous studies on memory-guided saccade (62), and partially align with those reported in the orientation domain (63–67). Unlike Panichello et al. (6) and Yu et al. (30), our goal was not to dissociate bias and variability but to identify neural signals that best predict overall memory errors. Therefore, the *drift* in our study simply refers to dynamics of memory error—the deviation of memory representations from the target location over time, similar to the notions in refs. 8, 14, and 17. Yu et al. (30) examined univariate fMRI activity that covaried with fitted drift rate and diffusion rate across participants and memory loads. Such univariate analyses likely capture control-related signals that modulate working memory strength. In contrast, our study focuses on the dynamics of neural representations of memory content, which is known to require multivariate approaches to detect with fMRI. In addition, Yu et al. (30) investigated changes of drift or diffusion rates driven by increasing memory load, whereas our experiment did not manipulate load.

Our decoding results in sPCS are broadly consistent with prior human neuroimaging findings—decoding performance is above chance but low in magnitude (22, 31–33, 39). These results contrast with neurophysiological studies in nonhuman primates that emphasize the PFC as a key region for working memory. Which regions within PFC contribute to working memory and the extent of PFC's involvement in working memory are under active debate, especially when comparing human neuroimaging with nonhuman primate electrophysiology. Limited fMRI spatial resolution may obscure fine-scale feature selectivity or cortical organization within PFC, reducing decoding sensitivity. In a recent 7T MRI study from our group (34), using higher spatial resolution (0.9 mm^3), with an almost identical task, decoding from FEF improved compared to the present dataset (circular correlation between the target and the decoded location r = 0.33 vs. 0.10 in ref. 22) but remained lower than in visual and parietal cortex (r = 0.4 to 0.75 in Li et al. (22)). It is possible that resolution alone does not fully explain the lower PFC decoding. Recent work shows that decoding performance increases in rule-based categorization tasks relative to fine-grained estimation like here (33) and may also increase with long-term training (68). Together, both measurement resolution and task demands influence the strength of working memory representations in PFC.

Methods

Subjects. We analyzed the datasets previously reported in refs. 22 and (23). There were 16 participants in total: 14 were the same participants reported in Experiment 2 in ref. 22, and two additional, nonoverlapping participants were from the control experiment reported in ref. 23. All participants had normal or corrected-to-normal vision. Written informed consent was obtained prior to participation, following protocols approved by the University Committee on Activities Involving Human Subjects at New York University. Participants were compensated at a rate of $30 per hour.

Procedures. Participants completed a memory-guided saccade paradigm while undergoing fMRI scanning. Each trial began with the presentation of a working memory (working memory) target–a light gray dot (0.65° in diameter)–for 500 ms. Targets have a fixed eccentricity at 12°, and their polar angle was pseudorandomly drawn from 32 evenly spaced positions covering a full circle. After the target offset, there was a 12-s delay. During the delay, participants were instructed to maintain fixation at the central fixation point while retaining the spatial location of the target in memory. At the end of the delay, the fixation point changed its appearance (from an empty circle to a filled gray dot) and a black circular ring appeared, centered at fixation and matching the eccentricity of the target. These changes served as the go cue, instructing the participants to indicate the remembered location by executing a saccade to the memorized location. The initial saccade landing location was recorded by the eye tracking, and a dot was shown at that point. Participants could then refine this memory report using a manual dial and confirm their final memory estimate with a button press. Upon confirmation, a visual arc appeared on the ring, centered on the reported location. Participants adjusted the length of this arc using the manual dial to indicate their subjective uncertainty–the longer the arc, the greater the expressed uncertainty. This confidence report was finalized with another button press. Feedback was then provided by displaying a white dot at the actual target location. Participants earned points if the true location fell within the arc boundaries (for task details, see ref. 22). Analyses of the uncertainty reports were presented in a prior publication (22). Two participants from the control experiment reported in ref. 23 completed the same procedure, except that the go cue consisted only of the black ring indicating the target's eccentricity, without a change in the fixation point.

Apparatus and Eye Tracking. Visual stimuli were projected using a VPixx ProPix LCD projector positioned behind the MRI bore. Participants viewed the display via an angled mirror with a field of view of 52° horizontally and 31° vertically. A gray circular aperture (30° in diameter) remained visible on the screen throughout the experiment. Eye movements were monitored using an EyeLink 1000 Plus infrared video-based eye tracker (SR Research) inside the scanner bore, with a sampling rate of 500 Hz. During and between scanning runs, gaze data were monitored in real time. The eye tracker is calibrated at the start of each fMRI session and at the beginning of a run when needed.

MRI Data Acquisition and Preprocessing. MRI data were collected on a Siemens Prisma 3T scanner equipped with a 64-channel head/neck coil. Functional images were acquired with 44 slices and a voxel size of 2.5 mm (4× simultaneous-multi-slice acceleration; FoV was 200 × 200 mm; no in-plane acceleration; TE/TR: 30/750 ms, flip angle: 50°, Bandwidth: 2,290 Hz/pixel; 0.56 ms echo spacing; P→A phase encoding). To correct for susceptibility-induced distortions, spin-echo images were intermittently collected during the session in both forward and reverse phase-encoding directions (TE = 45.6 ms, TR = 3,537 ms; three volumes per direction) using the same slice prescription but without multiband acceleration. Functional data for retinotopic mapping were acquired in a separate session using a higher-resolution multiband protocol with 56 slices (4× multiband acceleration), a voxel size of 2 mm isotropic, and a FoV of 208 × 208 mm. Imaging parameters for this session were TR = 1,200 ms, TE = 36 ms, flip angle = 66°, bandwidth = 2,604 Hz/pixel, echo spacing = 0.51 ms, and phase encoding in the P→A direction. In each participant's retinotopic mapping session, we also collected 2 or 3 T1-weighted whole-brain anatomical scans (MPRAGE sequence; 0.8 mm^3).

Quantification and Statistical Analysis.

Behavioral data analysis. As participants were allowed to manually refine their saccadic response using a dial, the final adjusted position of the dot was taken as the participant's memory report. Eye position data were analyzed offline. First, the raw gaze coordinates were smoothed using a Gaussian kernel, then transformed into eye velocity by computing velocity based on a five-point moving window of adjacent samples. Saccades were identified when eye velocity exceeded the median velocity by more than five SD and lasted at least 8 ms. Trials were excluded

if the primary saccade could not be reliably detected, if gaze deviated from fixation by more than 2.5°, or if the memory error exceeded three SD.

Generative model and Bayesian decoding. We applied a generative model-based approach (22, 27) to decode the spatial location (polar angle) of the working memory (working memory) targets from the fMRI BOLD signal. The time series data were first converted to percent signal change and then normalized (z-scored) within each run. The decoding procedures followed those described in ref. 22, with one key difference: While the previous study decoded the average BOLD response over a late delay period (i.e., normalized percent signal change averaged from 5.25 to 12.00 s after delay onset), here we performed decoding at each individual time point (TR = 750 ms). Below, we summarize the generative model and decoding procedure.

In the generative model, the multivoxel BOLD response associated with a given stimulus location (polar angle) was assumed to follow a multivariate normal distribution. The expected response (mean) of each voxel for a particular stimulus was defined by its tuning curve, representing the voxel's response profile across polar angles. To estimate each voxel's tuning, we used a weighted linear combination of eight basis functions that uniformly spanned the polar angle space. These basis functions were raised cosine functions.

$$f(\mathrm{s})_k = \left\lfloor \cos\left(\mathrm{s} - \phi_k\right) \right\rfloor,$$

where $\lfloor\ \rfloor$ represents half-wave rectification, and ϕ_k is the center of the kth channel. The response of ith voxel b_i given a stimulus s is then modeled as

$$b_i(\mathrm{s}) = \sum_{k=1}^{8} W_{ik}\left(f_k(\mathrm{s}) + \eta_k\right) + v_i,$$

where $\mathbf{W}$ is a weighting matrix that determines the weights of each basis function for each voxel. Thus, for each training dataset, we assumed that the voxel activity pattern followed a multivariate normal distribution $p\left(\mathbf{b}|s\right) \sim \mathcal{N}\left(\mathbf{W}f(\mathrm{s}), \Omega\right)$, in which $\Omega = \lambda\Omega_0 + (1-\lambda)\Omega_{\text{sample}}$.

When the number of variables (voxels) is larger than the number of observations (trials), the sample covariance is not invertible. To ensure a stable estimation of the covariance matrix, the covariance matrix was modeled as the sample covariance matrix $\mathbf{\Omega}_{\text{sample}}$ "shrunk" (69) to a target covariance matrix–the theoretically important covariance matrix $\mathbf{\Omega}_0$. The free parameter λ determined the degree of shrinkage. Here, the target covariance matrix $\mathbf{\Omega}_0$ is assumed to have a simple structure computed as a weighted sum over a diagonal matrix, a rank-1 covariance matrix, and the covariance depending on the tuning functions $\mathbf{W}$ of the voxels (see details in refs. 22 and 27).

For each participant and each region of interest (ROI), we performed decoding of the target's polar angle using a leave-one-run-out cross-validation approach. In the training phase, the generative model was fit to all trials except those from a single held-out run. The model estimated free parameters based on this training set, and then for each trial in the held-out run (the test set), decoding was performed using Bayes' rule:

$$p\left(\mathrm{s}|\mathbf{b};\theta_j\right) = \frac{p\left(\mathbf{b}|\mathrm{s};\theta_j\right)p(\mathrm{s})}{\int p\left(\mathbf{b}|\mathrm{s};\theta_j\right)p(\mathrm{s})ds}.$$

We assumed the prior p(s) to be uniform, and we approximated the continuous posterior probability function by sampling 1,000 steps evenly spanning the location space. Finally, we computed the mean of the decoded posterior probability function to represent the decoded location (polar angle) of each single trial in the test set.

Statistical Tests. To provide robust statistical tests across multiple contiguous time points or spatial locations, we applied cluster-based permutation tests whenever feasible. In Fig. 2*B*, to identify time points where the decoded location correlated with the target location, we computed the circular correlation between the decoded and target locations for each participant. At each time point, we performed a one-sample *t* test to evaluate whether the correlation differed from zero. Neighboring time points with uncorrected significance ($P < 0.05$) were grouped into clusters. For each cluster, we summed the t-values of all time points within the cluster. The significance of each cluster was then determined by comparing its summed t-value to a null distribution generated via a permutation procedure. Specifically, we randomly permuted the decoded target locations, repeated the same analysis (*t* tests at each time point and clustering of significant points), and computed the maximum cluster-level t-sum for each permutation. This procedure was repeated 1,000 times to generate a null distribution of maximum clusters' t-sums. Clusters, originally identified in the real data, with t-sums exceeding the 95th percentile of the null distribution were deemed significant and are marked with red horizontal lines in Fig. 2*B*. The same analysis was used for assessing correlations between decoding error and behavioral memory error (error–error correlation; Fig. 2*D*). To estimate the CI for the onset latency of error–error correlations and compare the latency between ROIs, we performed a bootstrap procedure: Participants were resampled with replacement, and for each resampled dataset, the cluster-based test was applied to identify the earliest time point that was deemed significant. This process was repeated 1,000 times, yielding a distribution of onset latencies for each ROI (Fig. 2*E*). In other analyses where we compared two different conditions (Figs. 2*F* and 3*D*, and the *Rightmost* panel of Fig. 5*B*), we used the same cluster-based method, with the null distribution generated by randomly shuffling condition labels across participants. Full details of the cluster-based permutation procedure can be found in ref. 70.

Data, Materials, and Software Availability. Anonymized preprocessed fMRI data and behavioral data have been deposited in Open Science Framework (https://osf.io/jgmqu/) (71).

ACKNOWLEDGMENTS. NIH R01 EY-027925 to W.J.M. and C.E.C.; R01 EY-016407 and R01 EY-033925 to C.E.C; Swartz Foundation Postdoctoral Fellowship to H.-H. L.

1. H. Shin, Q. Zou, W. J. Ma, The effects of delay duration on visual working memory for orientation. *J. Vis.* **17**, 10 (2017).
2. Y. Pertzov, S. Manohar, M. Husain, Rapid forgetting results from competition over time between items in visual working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 528–536 (2017).
3. W. A. Phillips, On the distinction between sensory storage and short-term visual memory. *Percept. Psychophys.* **16**, 283–290 (1974).
4. D. McKeown, T. Mercer, Short-term forgetting without interference. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 1057–1068 (2012).
5. S. Magnussen, E. Idås, S. H. Myhre, Representation of orientation and spatial frequency in perception and memory: A choice reaction-time analysis. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 707–718 (1998).
6. M. F. Panichello, B. DePasquale, J. W. Pillow, T. J. Buschman, Error-correcting dynamics in visual working memory. *Nat. Commun.* **10**, 3366 (2019).
7. A. Fennell, R. Ratcliff, A spatially continuous diffusion model of visual working memory. *Cogn. Psychol.* **145**, 101595 (2023).
8. S. Schneegans, P. M. Bays, Drift in neural population activity causes working memory to deteriorate over time. *J. Neurosci.* **38**, 4859–4869 (2018).
9. D. J. Amit, N. Brunel, Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex* **7**, 237–252 (1997).
10. X. J. Wang, Synaptic basis of cortical persistent activity: The importance of NMDA receptors to working memory. *J. Neurosci.* **19**, 9587–9603 (1999).
11. A. Compte, N. Brunel, P. S. Goldman-Rakic, X. J. Wang, Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10**, 910–923 (2000).
12. F. Bouchacourt, T. J. Buschman, A flexible model of working memory. *Neuron* **103**, 147–160.e8 (2019).
13. X.-J. Wang, 50 years of mnemonic persistent activity: Quo vadis? *Trends Neurosci.* **44**, 888–902 (2021).
14. K. Wimmer, D. Q. Nykamp, C. Constantinidis, A. Compte, Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* **17**, 431–439 (2014).
15. H. K. Inagaki, L. Fontolan, S. Romani, K. Svoboda, Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* **566**, 212–217 (2019).
16. A. Finkelstein *et al.*, Publisher correction: Attractor dynamics gate cortical information flow during decision-making. *Nat. Neurosci.* **24**, 897 (2021).
17. M. J. Wolff, J. Jochim, E. G. Akyürek, T. J. Buschman, M. G. Stokes, Drifting codes within a stable coding scheme for working memory. *PLoS Biol.* **18**, e3000625 (2020).
18. G.-Y. Bae, Neural evidence for categorical biases in location and orientation representations in a working memory task. *Neuroimage* **240**, 118366 (2021).
19. S. A. Harrison, F. Tong, Decoding reveals the contents of visual working memory in early visual areas. *Nature* **458**, 632–635 (2009).
20. R. L. Rademaker, C. Chunharas, J. T. Serences, Coexisting representations of sensory and mnemonic information in human visual cortex. *Nat. Neurosci.* **22**, 1336–1344 (2019).
21. T. C. Sprague, E. F. Ester, J. T. Serences, Restoring latent visual working memory representations in human cortex. *Neuron* **91**, 694–707 (2016).
22. H.-H. Li, T. C. Sprague, A. H. Yoo, W. J. Ma, C. E. Curtis, Joint representation of working memory and uncertainty in human cortex. *Neuron* **109**, 3699–3712.e6 (2021).
23. H.-H. Li, C. E. Curtis, Neural population dynamics of human working memory. *Curr. Biol.* **33**, 3775–3784.e4 (2023).
24. O. Gosseries *et al.*, Parietal-occipital interactions underlying control- and representation-related processes in working memory for nonspatial visual features. *J. Neurosci.* **38**, 4357–4366 (2018).
25. Q. Yu, W. M. Shim, Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory. *Neuroimage* **157**, 97–107 (2017).
26. G. E. Hallenbeck, T. C. Sprague, M. Rahmati, K. K. Sreenivasan, C. E. Curtis, Working memory representations in visual cortex mediate distraction effects. *Nat. Commun.* **12**, 4714 (2021).

27. R. S. van Bergen, J. F. M. Jehee, TAFKAP: An improved method for probabilistic decoding of cortical activity. bioRxiv [Preprint] (2021). https://www.biorxiv.org/content/10.1101/2021.03.04.433946v2 (Accessed 15 July 2025).
28. R. S. van Bergen, W. J. Ma, M. S. Pratte, J. F. M. Jehee, Sensory uncertainty decoded from visual cortex predicts behavior. *Nat. Neurosci.* **18**, 1728–1730 (2015).
29. H. Gu *et al.*, Attractor dynamics of working memory explain a concurrent evolution of stimulus-specific and decision-consistent biases in visual estimation. *Neuron* **113**, 3476–3490.e9 (2025), 10.1016/j.neuron.2025.07.003.
30. Q. Yu, M. F. Panichello, Y. Cai, B. R. Postle, T. J. Buschman, Delay-period activity in frontal, parietal, and occipital cortex tracks noise and biases in visual working memory. *PLoS Biol.* **18**, e3000854 (2020).
31. E. F. Ester, T. C. Sprague, J. T. Serences, Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron* **87**, 893–905 (2015).
32. H.-H. Li, T. C. Sprague, A. H. Yoo, W. J. Ma, C. E. Curtis, Neural mechanisms of resource allocation in working memory. *Sci. Adv.* **11**, eadr8015 (2025).
33. Z. Shao, M. Zhang, Q. Yu, Stimulus representation in human frontal cortex supports flexible control in working memory. *Elife* **13**, RP100287 (2025).
34. Z. Lu, L. T. Dowdle, K. N. Kay, C. E. Curtis, Mnemonic maps of visual space in human prefrontal cortex. bioRxiv [Preprint] (2025). https://www.biorxiv.org/content/10.1101/2025.10.17.683147v1 (Accessed 18 October 2025).
35. J. M. Esnaola-Acebes, A. Roxin, K. Wimmer, Flexible integration of continuous sensory evidence in perceptual estimation tasks. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2214441119 (2022).
36. D. Standage, M. Paré, Slot-like capacity and resource-like coding in a neural model of multiple-item working memory. *J. Neurophysiol.* **120**, 1945–1961 (2018).
37. S. Funahashi, C. J. Bruce, P. S. Goldman-Rakic, Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61**, 331–349 (1989).
38. S. Funahashi, C. J. Bruce, P. S. Goldman-Rakic, Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. *J. Neurophysiol.* **63**, 814–831 (1990).
39. C. E. Curtis, T. C. Sprague, Persistent activity during working memory from front to back. *Front. Neural Circuits* **15**, 696060 (2021).
40. S. Funahashi, Working Memory in the Prefrontal Cortex. *Brain Sci.* **7**, 49 (2017).
41. A. C. Riggall, B. R. Postle, The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *J. Neurosci.* **32**, 12990–12998 (2012).
42. T. A. Jerde, E. P. Merriam, A. C. Riggall, J. H. Hedges, C. E. Curtis, Prioritized maps of space in human frontoparietal cortex. *J. Neurosci.* **32**, 17382–17390 (2012).
43. M. F. Panichello, T. J. Buschman, Shared mechanisms underlie the control of working memory and attention. *Nature* **592**, 601–605 (2021).
44. J. Huang *et al.*, Neuronal representation of visual working memory content in the primate primary visual cortex. *Sci. Adv.* **10**, eadk3953 (2024).
45. D. Mendoza-Halliday, S. Torres, J. C. Martinez-Trujillo, Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat. Neurosci.* **17**, 1255–1262 (2014).
46. S.-H. Lee, D. J. Kravitz, C. I. Baker, Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat. Neurosci.* **16**, 997–999 (2013).
47. Y. Xu, Parietal-driven visual working memory representation in occipito-temporal cortex. *Curr. Biol.* **33**, 4516–4523.e5 (2023).
48. T. B. Christophel, P. C. Klink, B. Spitzer, P. R. Roelfsema, J.-D. Haynes, The distributed nature of working memory. *Trends Cogn. Sci.* **21**, 111–124 (2017).
49. J. D. Semedo, A. Zandvakili, C. K. Machens, B. M. Yu, A. Kohn, Cortical areas interact through a communication subspace. *Neuron* **102**, 249–259.e4 (2019).
50. J. D. Semedo *et al.*, Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nat. Commun.* **13**, 1099 (2022).
51. P. Iamshchinina, T. B. Christophel, S. Gayet, R. L. Rademaker, Essential considerations for exploring visual working memory storage in the human brain. *Vis. Cogn.* **29**, 425–436 (2021).
52. H. Nienborg, M. R. Cohen, B. G. Cumming, Decision-related activity in sensory neurons: Correlations among neurons and with behavior. *Annu. Rev. Neurosci.* **35**, 463–483 (2012).
53. M. N. Shadlen, K. H. Britten, W. T. Newsome, J. A. Movshon, A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.* **16**, 1486–1510 (1996).
54. H. Nienborg, B. G. Cumming, Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature* **459**, 89–92 (2009).
55. X.-J. Yu, J. D. Dickman, G. C. DeAngelis, D. E. Angelaki, Neuronal thresholds and choice-related activity of otolith afferent fibers during heading perception. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6467–6472 (2015).
56. K. H. Britten, W. T. Newsome, M. N. Shadlen, S. Celebrini, J. A. Movshon, A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.* **13**, 87–100 (1996).
57. J. V. Dodd, K. Krug, B. G. Cumming, A. J. Parker, Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *J. Neurosci.* **21**, 4809–4821 (2001).
58. L. Camarillo, R. Luna, V. Nácher, R. Romo, Coding perceptual discrimination in the somatosensory thalamus. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21093–21098 (2012).
59. R. L. T. Goris, C. M. Ziemba, G. M. Stine, E. P. Simoncelli, J. A. Movshon, Dissociation of choice formation and choice-correlated activity in macaque visual cortex. *J. Neurosci.* **37**, 5195–5203 (2017).
60. A. K. Seth, A. B. Barrett, L. Barnett, Granger causality analysis in neuroscience and neuroimaging. *J. Neurosci.* **35**, 3293–3297 (2015).
61. Y. Burak, I. R. Fiete, Fundamental limits on persistent activity in networks of noisy neurons. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17645–17650 (2012).
62. M. R. Burke, J. B. Clarke, J. Hedley, Effect of retinal and/or extra-retinal information on age in memory-guided saccades. *Exp. Brain Res.* **205**, 87–94 (2010).
63. S. Appelle, Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychol. Bull.* **78**, 266–278 (1972).
64. V. de Gardelle, S. Kouider, J. Sackur, An oblique illusion modulated by visibility: Non-monotonic sensory integration in orientation processing. *J. Vis.* **10**, 6 (2010).
65. X.-X. Wei, A. A. Stocker, Lawful relation between perceptual bias and discriminability. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10244–10249 (2017).
66. R. Taylor, P. M. Bays, Efficient coding in visual working memory accounts for stimulus-specific variations in recall. *J. Neurosci.* **38**, 7132–7142 (2018).
67. A. R. Girshick, M. S. Landy, E. P. Simoncelli, Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* **14**, 926–932 (2011).
68. J. A. Miller, A. Tambini, A. Kiyonaga, M. D'Esposito, Long-term learning transforms prefrontal cortex representations during working memory. *Neuron* **110**, 3805–3819.e6 (2022).
69. O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* **88**, 365–411 (2004).
70. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
71. H.-H. Li, W. J. Ma, C. E. Curtis, Data for Working memory drift. Open Science Framework. https://osf.io/jgmqu. Deposited 8 August 2025.