Redundant prefrontal hemispheres adapt storage strategy to working memory demands

Melanie Tschiersch[1,2], Akash Umakantha[3,4] , Ryan C. Williamson[3,4], Matthew A. Smith[3,4,5], Joao Barbosa[6,7,8]* , Albert Compte[1]*

[1]IDIBAPS, Barcelona, Spain

[2]Programa de doctorat en Biomedicina, Universitat de Barcelona (UB), Barcelona, Spain

[3]Center for the Neural Basis of Cognition, Pittsburgh PA, USA

[4]Carnegie Mellon University Neuroscience Institute, Pittsburgh PA, USA

[5]Carnegie Mellon University Department of Biomedical Engineering, Pittsburgh PA, USA

[6]Laboratoire de Neurosciences Cognitives et Computationnelles, INSERM U960, Ecole Normale Superieure - PSL Research University, 75005, Paris, France

[7]Cognitive Neuroimaging Unit, INSERM, CEA, CNRS, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France

[8]Institut de neuromodulation, GHU Paris, psychiatrie et neurosciences, centre hospitalier Sainte-Anne, pôle hospitalo-universitaire 15, Université Paris Cité, Paris, France

* Equal contribution

Keywords: attractor networks, working memory, prefrontal cortex, serial dependence, bilateral field advantage, multi-area recordings

Abstract

The prefrontal hemispheres must coordinate dynamically to maintain a unified representation of visual space. Recently, two opposing theories using distinct storage strategies have been proposed: A high-capacity *specialized* architecture, where each hemisphere governs contralateral behavior, and a fail-safe *redundant* one, where both hemispheres jointly guide behavior across the visual space. To determine the neural architecture underlying interhemispheric coordination, we analyzed simultaneous bilateral prefrontal cortex recordings from three macaque monkeys performing a visuo-spatial working memory task. Both hemispheres equally predicted behavioral imprecisions across the visual field and decoding errors were weakly correlated between hemispheres, suggesting a redundant, weakly coupled organization. Attractor network simulations showed that redundancy improved precision when task demands were below memory capacity, while weak interhemispheric coupling increased capacity in more demanding tasks by allowing hemispheric specialization. This interhemispheric architecture reconciles previous findings thought to support distinct models into a unified architecture, revealing a versatile interhemispheric architecture that adapts to varying cognitive demands.

Introduction

The world is spatially continuous, yet the inputs to the primary visual cortex (V1) are divided at the vertical meridian, with each hemisphere of the brain processing only the contralateral visual field (1). These divided hemifield representations are stitched together by callosal projections between the V1 hemispheres, which selectively connect neurons with receptive fields along the vertical midline (2–4). While this division is well understood in early visual areas, it remains a critical open question how interhemispheric interactions in higher-order regions such as the prefrontal cortex (PFC) support working memory across the full visual space (5).

While V1 neural receptive fields are strictly contralateral, prefrontal neurons can be tuned to either hemifield (6–8), though more neurons exhibit a contralateral preference (9,10). To explain how this organization can lead to integrated working memory across space, recent studies propose two alternative theories. In a specialized architecture, each hemisphere is predominantly responsible for contralateral information (Fig. 1a) (11). In contrast, in a redundant architecture both hemispheres represent the entire spatial field and equally guide behavior across both hemifields (Fig. 1b) (5). Each architecture helps explain previously conflicting empirical findings and also offers intuitive functional advantages.

A useful analogy is that of data storage in two hard drives. The strategy of maintaining unique files in each drive (specialization) maximizes storage capacity, while duplicating files across drives (redundancy) ensures robustness against failure. For everyday data, capacity may take precedence, but for critical files, backup is essential. Similarly, the optimal neural architecture for spatial working memory may depend on task demands.

Evidence for the specialized architecture arises from behavioral and neural findings in multi-item working memory studies in human and non-human primates. Behaviorally, this is reflected in the *bilateral field advantage*: the ability to memorize more items (higher capacity) when items are located in different visual hemifields compared to when items are confined to one hemifield (12–19). For example, humans can attentionally track up to eight items presented bilaterally, but only four when presented within a single hemifield (12), suggesting independent storage slots for each hemisphere.

This hemispheric specialization is further supported by electrophysiological evidence, where the prefrontal neural code for a memorized item is only diminished when additional items are presented in the same, but not the opposite hemifield (14,18). Similarly, distractors affect memorized items more strongly when they are presented in the same rather than the opposite hemifield, in both behavior and electroencephalography recordings (20). Additionally, working memory representations transfer between prefrontal hemispheres following gaze shifts across the vertical meridian (21). Taken together, these studies suggest a model of hemispheric specialization with largely independent working memory resources in each hemisphere, which selectively communicate (11).

Conversely, optogenetic studies in mice suggest a redundant architecture, since short-term memory can recover from unilateral but not bilateral prefrontal perturbations (5,22,23). Additionally, frontal neural receptive fields encompass the full visual space, albeit with a contralateral preference, in mice (22,24,25) and monkeys (6–9,26). These findings suggest that each hemisphere is sufficient to store memories and guide behavior across the full visual space (Fig. 1b), challenging the specialized perspective. It is unclear if this conflicting evidence corresponds to differences in task design, brain regions, or animal species, or if it can be integrated into a unified framework.

A fundamental conceptual tool to interpret neural recordings in spatial working memory are bump attractor networks (27–29). These models can replicate working memory features such as delay-dependent precision (28,30), history effects (31–34), location-dependent distraction (20,28), feature-binding through synchronization (35), and limited multi-item working memory capacity and interference based on within-network attractor competition (36–40). This model thus constitutes a prime tool to address the computational basis of spatial working memory, which we leverage here to clarify the architecture underlying interhemispheric working memory.

In this study, we mainly focus on two of these key experimental observations that characterize working memory. First, we analyze within-trial memory diffusion, known to explain response inaccuracies of working memory precision through random diffusion in persistent neural activity (28,30). Second, we study across-trial history drift, explaining the attraction of current memory

reports to the previous one (serial dependence) (41–48) based on synaptic plasticity traces (31–34,49). We selected these effects as they operate on different timescales and involve distinct mechanisms, providing complementary insights into the dynamics of interhemispheric memory processing.

We investigated the role of prefrontal hemispheres in these phenomena by analyzing simultaneous bilateral recordings from monkey prefrontal cortex during a visuo-spatial working memory task. Similar to previous work (6–9), our neural analyses showed that working memory across visual space was stably and simultaneously represented by both prefrontal hemispheres in a lateralized manner. Importantly, activity from both hemispheres equally guided behavior, consistent with redundant memories. Neural decoding errors were weakly, but significantly correlated across hemispheres, suggesting weak interhemispheric connections that served to coordinate the working memory representations. We modeled this architecture through two redundant bump attractor networks of spiking neurons, building on recent multi-network frameworks (20,35,50–52) and thus replicated findings of memory diffusion and history drift. Importantly, simulations revealed that this architecture balances the complementary advantages of storage redundancy and specialization across different task complexities, offering a unified network architecture that reconciles previously conflicting findings.

Results

Quantifying task-related behavior: Working memory precision and serial dependence.

We investigated the prefrontal hemispheric organization of spatial working memory by analyzing behavior and bilateral population recordings from area 8Ar of three macaque monkeys performing a visuo-spatial oculomotor delayed response task (53). While the monkeys were fixating, a stimulus appeared in one of 4, 8, or 16 possible locations (depending on the session), which had to be remembered for a variable delay period (1.5-3 s) and reported through a saccade to the remembered location (Fig. 1c, top).

We analyzed two main behavioral effects: working memory precision, defined as the variability in memory recall of stimulus locations, and serial dependence, the systematic attraction of current reports towards previously remembered items. These behavioral effects have been mechanistically linked to the diffusion of persistent activity (28,30) and drift based on synaptic plasticity (31,32) through model-driven analyses of electrophysiological recordings (Fig. 1c, bottom) and can offer complimentary insights on working memory processing.

Monkeys exhibited good accuracy overall, but absolute precision errors significantly increased with longer delays (Fig. 1d,e), as previously reported (54–56), suggesting memory diffusion.

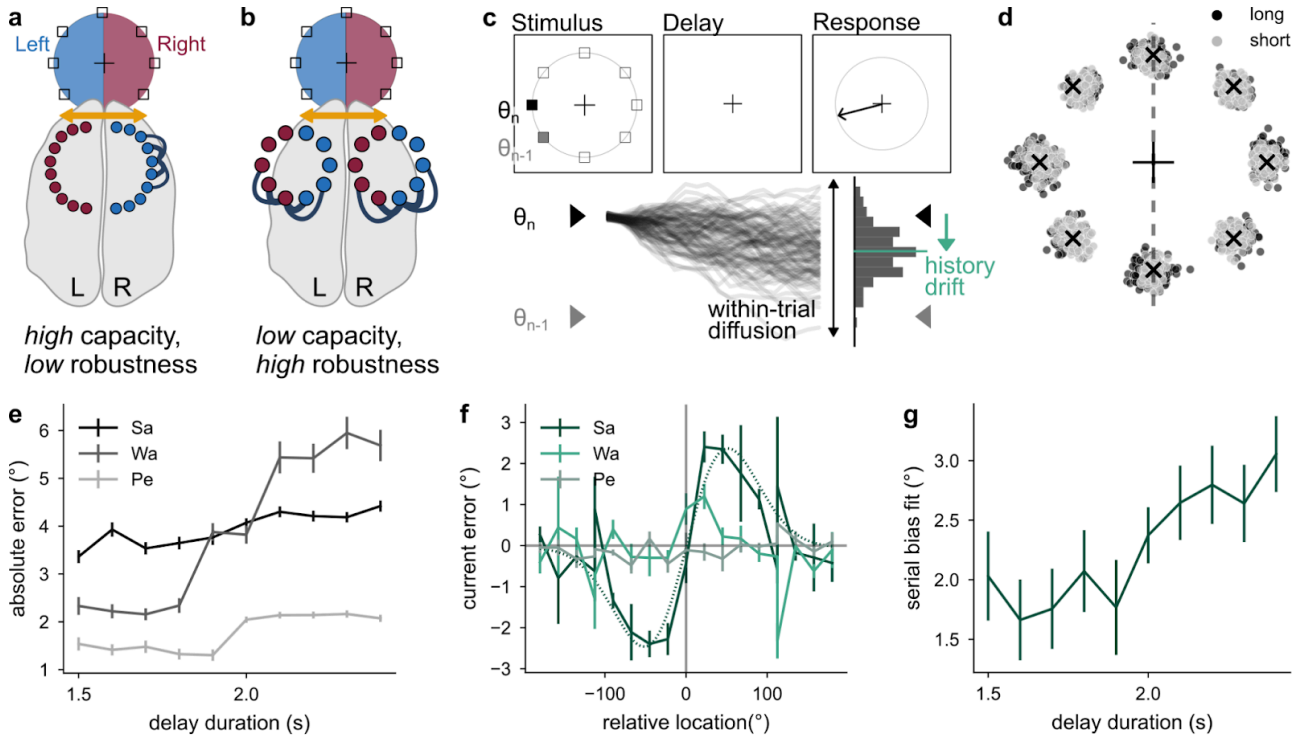


Figure 1: Competing interhemispheric models and task behavior. *a) In a specialized model, each hemisphere primarily represents contralateral memoranda. This architecture supports high capacity via distribution of resources, but low-robustness to memory disruption. Blue and red circles represent left and right-preferring neurons; yellow arrow denotes interhemispheric communication. b) In a redundant model, both hemispheres represent both hemifields. This supports robustness via shared coding spaces at the cost of reduced capacity. c) Top: Behavioral task schematic indicating a previous (θ_{n-1}) and current (θ_n) stimulus location. Stimulus ring and unfilled squares were not visible to the animal. Bottom: Errors may arise through random within-trial diffusion based on noisy memories and based on across-trial history drift (towards θ_{n-1}) based on synaptic plasticity. d) Example session (monkey Sa): Responses (dots) clustered around targets (crosses) with growing response errors with longer delay lengths. e) Absolute errors grew with delay length (OLS model: Sa: β = 1.02 °/s, $p < 10^{-14}$. Pe: β = 0.99 °/s, $p < 10^{-16}$. Wa: β = 4.5 °/s, $p < 10^{-48}$, across*

*monkeys: $p < 10^{-14}$). f) Serial dependence showed idiosyncratic differences between animals: strong for Sa ($p < 10^{-70}$), weak for Wa ($p = 0.05$), absent for Pe ($p = 0.76$). Model fits shown as dashed lines (OLS model (see Methods): Sa: $\sigma = 0.9$ rad: $\beta = 2.92°$; Wa: $\sigma = 0.45$ rad: $\beta = 0.73°$; Pe: $\sigma = 2.15$ rad: $\beta = 0.04°$). g) Serial dependence grew with delay for monkey Sa, but not the other monkeys (OLS model: "DoG1*delay"-interaction (see Methods): Sa: $\beta = 1.03$ °/s, $p = 0.001$; Pe: $\beta = 0.21$ °/s, $p = 0.2$; Wa: $\beta = 0.05$ °/s, $p = 0.97$). Error bars are s.e.m..*

We also observed serial dependence (SD, Fig. 1f), replicating findings in both humans (31,33,41–49,57–64) and monkeys (31,54,65). We measured the effect size by fitting the response error using a first-order derivative of Gaussian to fit the serial dependence curve (DoG1 curve; hyperparameters in Extended Data Fig. 6a-c; see Methods). We found varying degrees of serial dependence across monkeys (Fig. 1f). The monkey with the strongest serial dependence showed a clear build-up of serial dependence with longer delays (Fig. 1g), as previously reported (33,41,44,65).

Each hemisphere stably represented the full visual space with a contralateral preference.

While the monkeys performed this task, neural activity was recorded from bilateral 96-channel Utah arrays in area 8Ar of the prefrontal cortex (see Methods). We recorded an average of 139.2 ± 3.8 multi-units in each of the 41 sessions across both hemispheres. We found tuned persistent activity in trial-averaged data collected during the delay period of the task (example multi-unit in Fig. 2a), as previously reported (9,30,31,66). We leveraged the large population recordings to obtain a single-trial view of the population activity by showing the evolving firing rates of multi-units sorted by their memory fields (cross-validated, see Methods). In an example session with 158 multi-units, we observed clear, tuned persistent activity in some individual trials from stimulus onset until the response (example trial in Fig. 2b, different stimulus locations Extended Data Fig. 1, all trials from this session Extended Data Fig. 2). As expected, we saw clear tuning during the delay in the trial-averaged activity (example session in Fig. 2c).

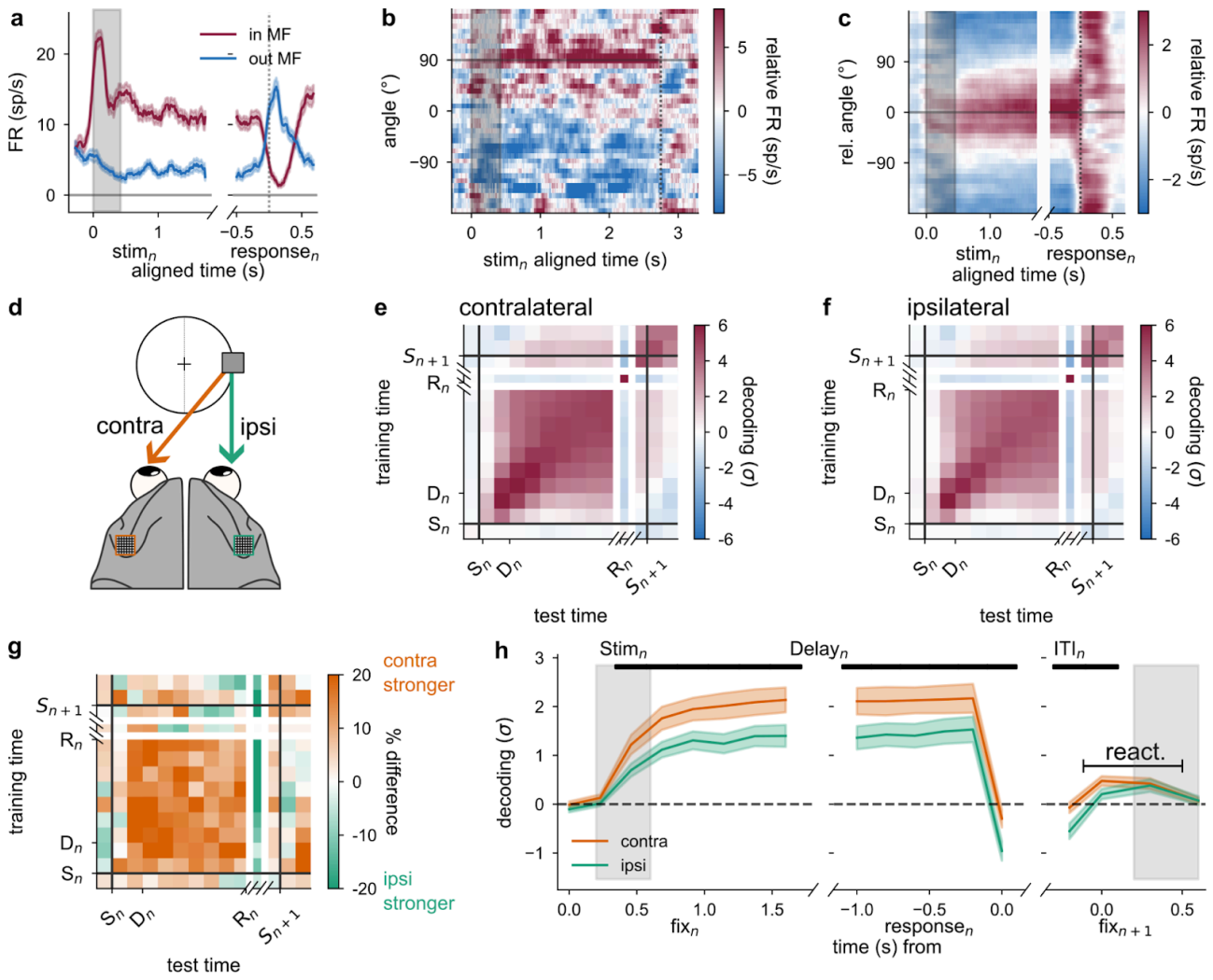


Figure 2: Hemispheres represent the full visual field with a contralateral preference. *a) Example multi-unit: Firing rate averages of cues in or out of the memory field (calculated on mean delay activity) showed clear memory tuning. 55.71% of 5708 multi-units were selective during the delay (Sa: 86.58%, Pe: 23.63%, Wa: 68.95%, bootstrapped test, see Methods). b) Single-trial tuned persistent population activity from stimulus start (gray shade; stimulus at 90°) to response (vertical dotted line), with multi-units sorted by preferred location (estimated on left-out trials, Methods). c) Session average of panel b, after recentering around the stimulus location. d) Schematic of the bilateral arrays in area 8Ar in a right stimulus trial, rendering the left (right) hemisphere contralateral (ipsilateral). e) Cross-temporal decoder accuracy in contralateral trials (average of left/right hemispheres and sessions), expressed as z-scored distance from shuffle baseline (σ). Task events: stimulus (S_n), delay (D_n), response (R_n), next stimulus (S_{n+1}). White lines indicate discontinuity because of variable delays; black lines indicate stimulus presentation. f) same as panel e for ipsilateral trials. g) Percent lateralization in decoding ((contra-ipsi)/contra*100), showing stronger contralateral coding (14.94% stronger in the delay). h) Delay decoder (trained on average delay activity) tested on contralateral and ipsilateral trials separately revealed emerging reactivations at the next trial's fixation (fix_{n+1}) in both hemispheres (t-test: ipsilateral: t = 2.54σ, p = 0.016, contralateral: t = 4.86σ, $p < 10^{-4}$, paired t-test: difference contra-ipsi: t = 1.8σ, p = 0.08). Error bars are s.e.m..*

We investigated the extent of hemispheric lateralization across sessions, by linearly decoding the stimulus location from the population activity in each hemisphere (see Methods; firing rates in Extended Data Fig. 3). Decoders were trained on 80% of the trials at each time step and the mean-squared error between predictions and stimulus was computed on contralateral and ipsilateral left-out trials separately (schematic Fig. 2d). We compared the mean-squared error with a stimulus-shuffled decoder (*decoding*: z-score to shuffle; see Methods) and found a stable delay code in both contralateral (Fig. 2e) and ipsilateral (Fig. 2f) trials. Contralateral outperformed ipsilateral decoding by 15% (averaged delay difference, Fig. 2g).

In the following, we used a decoder trained on the average neural activity during the delay period (Fig. 2h, *Delay Decoder*, Methods) to focus on memory encoding rather than other task parameters, such as motor responses (67). This Delay Decoder revealed a reactivation of memory-related activity during fixation of the upcoming trial (line at fix_{n+1}, Fig. 2h), previously shown to correlate with serial dependence strength (31,49,68–70). Notably, reactivations were present in both hemispheres (Fig. 2h), indicating little lateralization in their occurrence. Based on previous proposals in a single hemisphere (31,49), this suggests that each hemisphere contains enough persistent activity to produce reliable synaptic traces which can be reactivated at the start of the next trial.

Both hemispheres equally reflected trial-by-trial behavioral variability with weakly correlated memory representations.

To address how the prefrontal cortex related to behavioral response, we correlated the neural decoder errors with the corresponding behavioral response errors (30,54). A positive correlation during the delay would indicate that the multi-units can predict behavioral imprecisions before they occur. Furthermore, a steady increase throughout the delay suggests slowly diffusing memories as proposed by attractor dynamics (30,54). Indeed, we found steadily increasing correlations when pooling multi-units from both hemispheres (Fig. 3a).

To understand how hemispheric memory representations are combined to provide unified behavior, we separately trained decoders on each hemisphere and performed the same analysis. The left

and right hemispheres showed no difference in neural-behavioral correlations (Fig. 3b, Extended Data Fig. 4b). Surprisingly, neither did ipsilateral vs. contralateral trials (Fig. 3c, Extended Data Fig. 4a), even though the memory code was more precise in contralateral trials (Fig. 2g). This analysis showed that both hemispheres equally predicted small behavioral imprecisions in the full visual field and were equally weighted in guiding behavior. Therefore, our analyses did not support the specialized model (Fig. 1a), but rather pointed towards a redundant architecture (Fig. 1b), with both hemispheres encoding stimuli and guiding behavior in the full visual field.

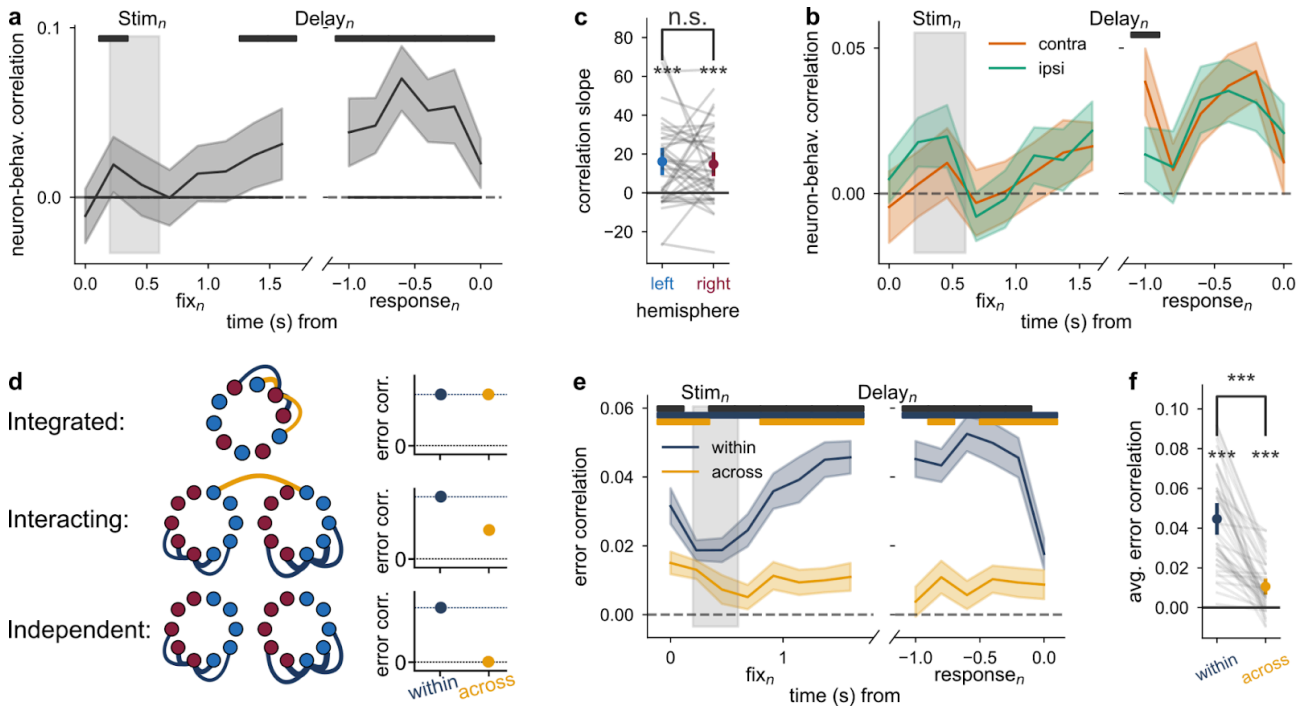


Figure 3: Memories were stored redundantly in both hemispheres with interacting representations. *a) Memory diffusion was quantified by correlating decoder and behavioral response errors. Neural prediction errors became increasingly aligned with behavioral errors throughout the delay (t-test of slope: t = 4.95 1/s, $p < 10^{-4}$). Error bars are 95% C.I. b) Same as panel a, but showing no significant difference in slope for left and right hemisphere predictions separately (paired t-test: t = 0.375 1/s, p = 0.71). c) Same as panel a-b, but for ipsilateral vs contralateral trials showing no significant difference between conditions (paired t-test of slopes: t = -0.38 1/s, p = 0.7). d) Schematic of different network architectures predicting different across-hemisphere error correlations. e) Decoder prediction error correlations within and across hemispheres. f) Delay-averaged error correlations from panel e. Error correlations within and across hemispheres differed significantly from zero (t-test: within: t = 11.19, $p< 10^{-13}$; across: t = 5.19, $p < 10^{-5}$), but correlations across hemispheres were weaker than within the same hemisphere (paired t-test: t = 9.07, $p < 10^{-10}$). Error bars are +/- 1 s.e.m. unless otherwise specified. Top bars denote significant differences.*

We further aimed to specify the interhemispheric organization underlying the redundant working memory network. A redundant architecture can range from being so tightly coupled that the hemispheres are essentially integrated (Fig. 3d, top) to two uncoupled hemispheres that each represent the full visual space (Fig. 3d, bottom). These alternatives have different predictions for memory diffusion in each hemisphere. If both hemispheres are integrated, memories would be perfectly mirrored with identical encoding errors (Fig. 3d, top). On the other hand, unconnected hemispheres would diffuse independently leading to uncorrelated memory errors (Fig. 3d, bottom, right).

We used the decoder prediction errors as a proxy for the memory imprecisions of each hemisphere. Specifically, we repeatedly trained four decoders on random halves of the multi-units recorded from each hemisphere (sub-hemispheres: $left_A$, $left_B$, $right_A$, $right_B$) and obtained their prediction errors. We correlated the prediction errors either within (*within*: $left_A$ vs. $left_B$, and $right_A$ vs. $right_B$) or across hemispheres (*across*: $left_A$ vs. $right_A$, and $left_B$ vs. $right_B$). We found that errors were more strongly correlated within than across hemispheres (Fig. 3e,f), suggesting separate functional networks for each hemisphere, which separately accumulate memory errors.

However, interhemispheric decoder error correlations were significantly above zero (Fig. 3e,f, across), suggesting either connections across hemispheres or a common input to both hemispheres. To address a plausible source of common input, we considered uninstructed eye movements while monkeys fixated (e.g., microsaccades). Such movements could co-vary memory imprecisions and also reflect brain-wide arousal signals (53,71,72), accounting for two types of common input. Indeed, eye gaze locations correlated with the memorized stimulus (Extended Data Fig. 5a-b) (71,73–75). However, gaze errors did not correlate with neural prediction errors (Extended Data Fig. 5b), making it unlikely that the eye movements provided common input related to memory errors. Indeed, even when removing uninstructed eye movements from the decoder analysis (see Methods), the same pattern of across-hemisphere error correlations remained (t-test across: $t = 4.2$, $p = 0.00015$, Extended Data Fig. 5c, d). Our analyses therefore suggest that the

observed interhemispheric error correlations arise from interacting hemispheres (Fig. 3d, middle) and not from a gaze confound.

Tuned across-area connections led to shared memory diffusion in a working memory model.

To understand how redundant, interacting hemispheres can support the observed interhemispheric working memory effects, we implemented the proposed architecture (Fig. 3d, middle) in a computational network model (Fig. 4a). We used the bump-attractor framework (28), which can model memory diffusion consistent with the observed neural-behavior correlations (Fig. 3a) (30).

Specifically, we modeled each hemisphere as a bump-attractor network of spiking excitatory and inhibitory leaky-integrate-and-fire neurons covering the full visual space. Brief external stimuli were delivered to each ring (henceforth called area) at the beginning of each trial (Fig. 4b, S_n, S_{n+1}), which were maintained in each area through recurrent excitation after the stimulus offset, leading to tuned persistent activity (Fig. 4b, D_n, D_{n+1}).

We studied the effects of different interhemispheric interactions by manipulating the strength and structure of excitatory connections between the two areas (Fig. 4a), expressed as a percentage of the within-area connectivity (see Methods). In particular, we compared unconnected areas (*0.0%*), randomly connected areas (*random*, 1.0% of within connectivity), and varying levels of tuned across-area connections, where neurons with similar receptive fields were more strongly connected across areas (*0.2% - 1.0%*).

Irrespective of connectivity type, memory precision was higher when responses were based on the averaged population activity of both areas (system), compared to on a single area (area, Fig. 4c), highlighting that hemispheric redundancy improved memory performance. Precision within each area improved with tuned across-area connectivity (Fig. 4c, grey) due to enhanced population tuning and therefore stability of each area's memory bump (Fig. 4d). The system-wide memory precision on the other hand remained stable across connectivity types (Fig. 4c, black),

To relate the data to the model, we asked which model can replicate the correlated memory errors observed in the neural analyses (Fig. 3e). Neither the uncoupled model (Fig. 4e, *0.0%*) nor the model with random 1.0% connectivity (Fig. 4e, *random*) produced across-area error correlations.

Only increasing degrees of tuned connectivity increased error correlations (Fig. 4e), suggesting that the prefrontal hemispheres are connected via tuned interhemispheric connections, in line with noise correlation analyses in these data (Fig S2 in (53)).

Together, our simulations show that redundancy enhances memory precision and suggest that interhemispheric connections in the monkey PFC are tuned, supporting local memory stability.

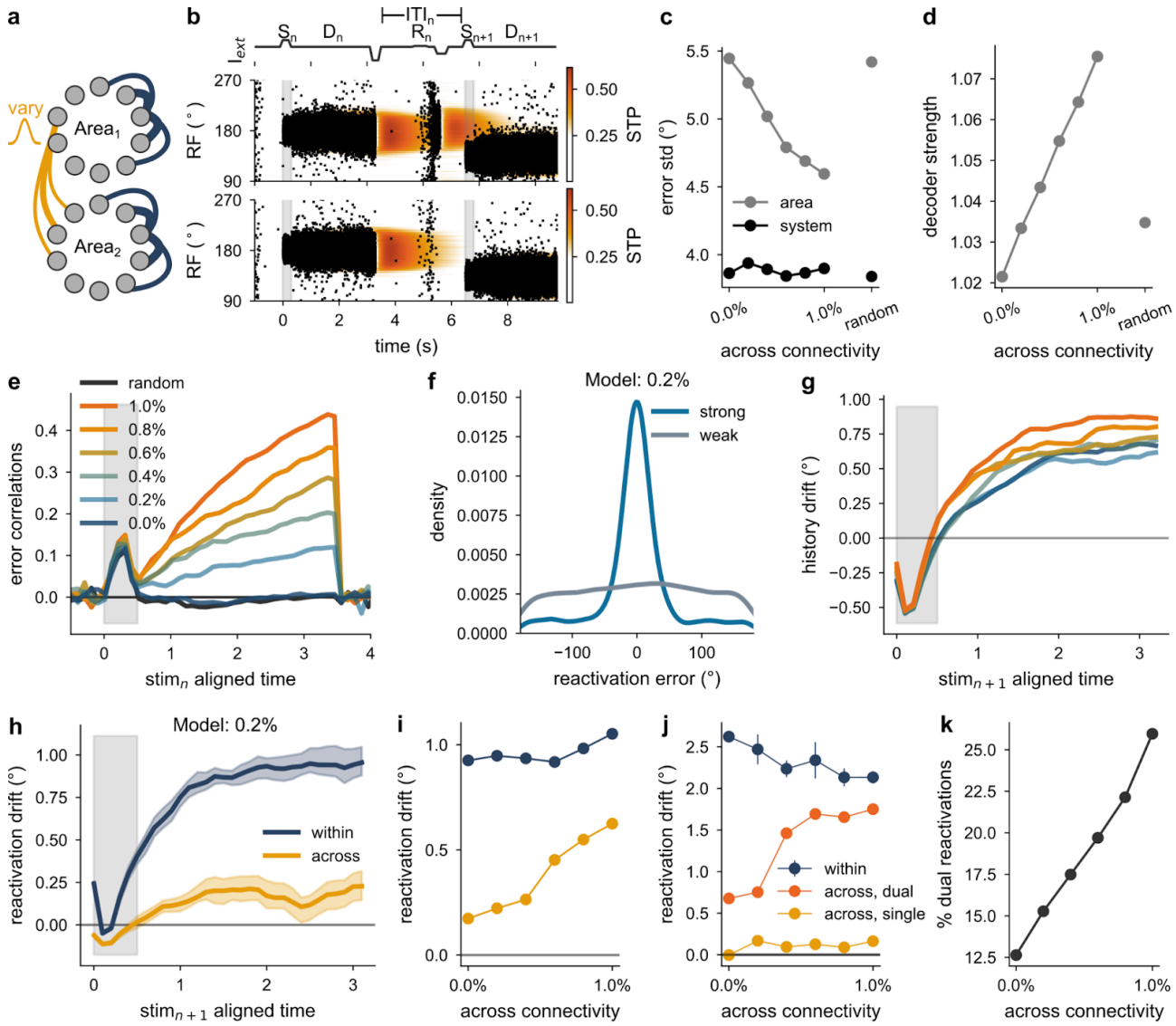


Figure 4: Tuned across-area connectivity increased coordination of within-trial memory diffusion and across-trial memory drift. *a) Model schematic: Two bump-attractor networks with short-term plasticity (STP) and varying across-area connectivity. b) Example simulation of two consecutive trials. Only Area$_1$ but not Area$_2$ showed a reactivation. Top: External stimulus strength. 'S' , 'D', 'R' and 'ITI' mark stimulus onset, delay, reactivation, and inter-trial interval respectively. Middle and Bottom: Spiking activity (dots) and STP dynamics (heatmap) for Area$_1$, Area$_2$, respectively. c) Increasing tuned interarea connectivity reduced response errors of each area, while system-wide*

response errors remained consistently low. d) Tuned connectivity enhanced memory stability (decoding strength). e) Only models with tuned across-area connections exhibited error correlations. f) Reactivations occurred on 38% of trials, with stronger reactivations linked to tighter replication of the previous stimulus location. g) Neural delay activity increasingly drifted towards the previous stimulus (history drift) over time and with stronger connectivity. h) Drift of the neural delay activity was stronger towards reactivations within the same hemisphere than across hemispheres (exemplary for the 0.2% model). i-j) Stronger across-area connectivity increased shared reactivation drift (taken at the end of the delay), particularly in trials with dual reactivations. k) The percentage of dual reactivations increased with connectivity strength.

Increasing across-area connectivity in a working memory model increased shared history drift.

We further used this model to investigate memory dynamics at longer timescales, namely across-trial memory drift underlying serial dependence. We could analyze this in the model, since within-area excitatory connections were facilitating (76), increasing synaptic efficacy after pre-synaptic spiking (short-term plasticity, see Methods) and leading to serial dependence in this task (31–34,49).

Guided by previous work with a single area (31–33,49), we simulated sequential trials separated by an “activity-silent” inter-trial interval (Fig. 4b, ITI_n). Towards the end of the inter-trial interval, a nonspecific excitatory current (a “reactivation” current (R_n), Fig. 4b) was applied to both networks, generating a bump reactivation in 38% of trials. Due to synaptic plasticity, reactivations were typically centered at the location of the previous stimulus (Fig. 4f) and further replenished the synaptic plasticity (heatmap between R_n and S_{n+1}, Fig. 4b, top), leading to stronger attraction to the previous memory in the upcoming trial (31). Furthermore, we modeled adaptation, a repulsive sensory history effect originating in lower-level visual areas (48,77–79), by shifting stimulus inputs away from the previous trial’s location (33,34) (see Methods). Overall, we found a delay-dependent increase in serial dependence from an initial repulsion to a later attraction as observed in human experimental data (33,41,44) for all connectivity strengths (Fig. 4g).

To clarify the impact of across-area connectivity strengths on history drift, we analyzed how the mnemonic activity (Fig. 4b, D_{n+1}) in each area drifted towards the previous reactivations (Fig. 4b, R) of the same area (*within*) or across areas. In the weakest-tuned model, the attraction was much stronger within than across areas (Fig. 4h), but increasing the interarea connectivity increased

shared history drift (Fig. 4i). This effect was driven by trials in which reactivations occurred concurrently in both hemispheres (Fig. 4j, *dual*), with the history drift remaining private when reactivations only occurred in a single area (Fig. 4j, *single*). Therefore, shared history drift was driven by an increased amount of *dual* reactivation trials (Fig. 4k).

In sum, our simulations showed that reactivations and with them across-trial history drift were affected by increasing across-area connectivity strength. Notably, history effects could be transferred across areas only if across-area connectivity strength was high and simultaneous reactivations occurred in both areas (*dual* trials). Based on these insights from the model we next evaluated inter-hemispheric history effects in the neural data.

Each prefrontal hemisphere has private history effects, suggesting weak interhemispheric connectivity.

We first confirmed that neural activity during memory maintenance in monkey PFC was biased towards previous stimulus locations. This across-trial history drift is a neural correlate of serial dependence previously found using non-invasive recordings in humans (42,48,64,79) and is parsimoniously explained by bump-attractor models (Fig 4g) (31).

Neural decoder errors were biased towards previous stimuli (Fig. 5a), mirroring behavioral serial dependence (Fig. 1f). Interestingly, the neural code was repelled from the previous stimulus during stimulus presentation but became attracted at the end of the delay period (Fig. 5a, example monkey Sa). To obtain a time-resolved estimation of the neural serial dependence, we fit a first-order derivative-of-Gaussian curve to the predicted errors at each time (dashed lines in Fig. 5a; hyperparameters in Extended Data Fig. 6d-f; see Methods) and observed a gradual change from repulsion to attraction throughout the delay (Fig. 5b). Notably, the neural history drift scaled with the behavioral serial dependence strength across monkeys (Extended Data Fig. 7a), though more subjects are needed to establish a correlation at the population level. From now on, the analyses focus on the monkeys with positive behavioral and neural serial dependence (n=2, Sa, Wa).

We tested whether this neural history drift was private to each hemisphere. For this, we computed the attraction of the neural code towards the previous reactivation location decoded from either the same or the opposite hemisphere, as in the model (Fig. 4h). First, we verified that reactivations could be extracted using a Delay decoder (see fix_{n+1} in Fig. 2h) and were tuned to the prior stimulus location in strongly reactivated trials (Fig. 5c). Then, we assessed how the neural code drifted towards the prior reactivations (DoG1 fit, hyperparameters in Extended Data Fig. 6g-i; see Methods). We exclusively found memory drift to reactivations occurring in the same hemisphere (*within*), but not the opposite hemisphere (*across*, Fig. 5d). Based on our model simulations, this finding of hemisphere-specific history drift suggests weak interactions between the hemispheres (Fig. 4h,i).

As in the model (Fig. 4j), we also investigated how different reactivation types affected the reactivation bias. We split the data into *null*, *single* and *dual* reactivation trials based on a 30% reactivation strength threshold for each hemisphere (Extended Data Fig. 7b). We observed that in *null* trials no significant reactivation bias emerged, whereas both single and dual reactivation trials showed within- but not across-hemispheric attraction (Extended Data Fig. 7c). These findings pointed towards a model with very weak interhemispheric interactions (Fig. 4j, e.g. 0.2%).

The model further predicted that private reactivation drift depends on most trials exhibiting reactivations in a *single* hemisphere (Fig. 4i-k). We tested this prediction using a Gaussian Mixture Model (GMM; see Methods) assuming four possible reactivation patterns: *null,* $single_{left}$, $single_{right}$ and *dual* reactivations. Across 250 runs with different initial parameters, the GMM produced eight distinct class means, all located away from the diagonal in the decoder-strength space (Fig. 5e). This suggests two key points: (1) the underlying distribution is not well explained by four classes, as different initial parameters caused class means to shift; and (2) *dual* reactivations were rare, as they would appear on the diagonal in this graph.

We further found that a three-class GMM consistently identified three off-diagonal class means across all 250 re-runs, accounting for *null,* $single_{left}$ and $single_{right}$ reactivation types (Extended Data Fig. 7 d-f). Taken together, these findings suggest that three (and not four) classes underlay the distribution, with no evidence for a *dual* reactivation class. In contrast, during active

memory-maintenance in the delay period, the four-class GMM analysis reliably estimated all four classes, including a *dual* class (Fig. 5f).

We conclude that the prefrontal hemispheres interacted through tuned but weak interhemispheric coupling, leading to correlated memory diffusion but private history drift and reactivations. This is particularly intriguing since it allowed the networks to interact when both hemispheres stored similar items (delay code strength correlations, Fig. 5f) but perform separately when inputs only triggered events in one hemisphere (reactivations remaining private, Fig. 5d,e).

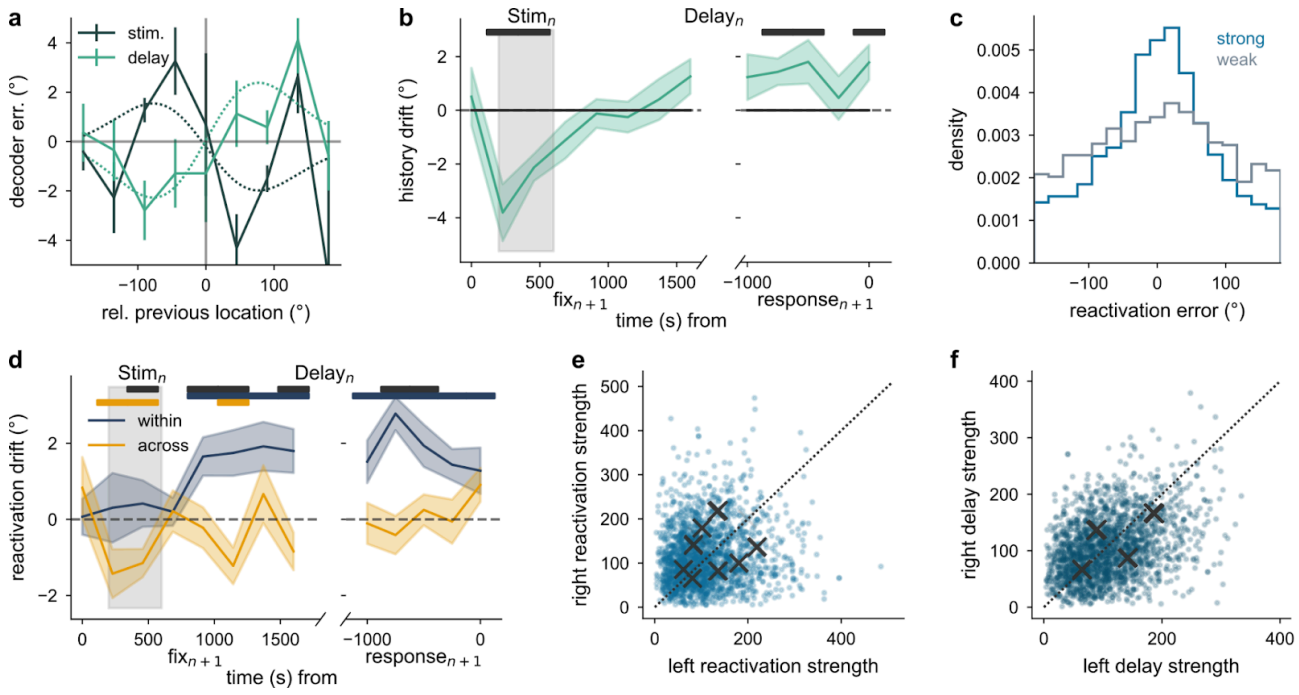


Figure 5: Private across-trial history drift in monkey PFC indicates weak but specific interhemispheric coupling. *a) Example monkey (Sa): Decoder error as a function of the relative distance of previous and current stimulus locations, showing neural repulsion during stimulus presentation but attraction at the end of the delay. Dashed lines indicate DoG1 curve fit (see Methods). b) Time-resolved neural history drift (fit from panel a) showed a continuous increase from repulsion to attraction in monkeys with significant behavioral serial dependence (Sa, Wa). c) Example monkey (Sa): Precision of reactivation location relative to the previous stimulus location was higher in stronger reactivated trials (cut-off at 30% of each session's decoder strength; see Methods). d) Attraction towards the previous reactivation location was private within each hemisphere (within, t-test: $t = 6.86°$, $p < 10^{-6}$) while remaining at zero across hemispheres (across, t-test: $t = 0.21°$, $p = 0.83$) with a significant difference between conditions (within-across, rel. t-test: $t = 5.54°$, $p < 10^{-4}$). Average of the two monkeys with significant behavioral serial dependence. e) Distribution of decoded reactivation strengths in the left and right hemispheres. Gaussian Mixture Model (250 repeats) estimated class means (crosses)*

off-diagonally, indicating few dual reactivations. f) Same as panel e for delay prediction strengths estimated a class on the diagonal indicating covarying representations. Error bars represent +/- 1 s.e.m.; black top bar indicates significance from 0.

Reconciling competing models: Weakly connected hemispheres enable both redundancy and hemispheric specialization.

Our findings suggest weakly connected, redundant hemispheres consistent with shared memory diffusion and private history drift. These results seemingly contradict prior evidence supporting specialized hemispheres. Specifically, previous behavioral and electrophysiological studies have shown reduced memory interference when multiple items are presented across rather than within a hemifield, a phenomenon known as the bilateral field advantage (12–19).

However, even in our proposed redundant architecture, we observed that reactivations remained private to each hemisphere, indicating that the prefrontal hemispheres were capable of maintaining separate memory representations if they were independently activated. We hypothesized that a lateralization of inputs to the PFC from upstream areas (22,80) could preferentially activate each hemisphere for contralateral stimuli and thus promote hemispheric specialization during multi-item tasks (19,81).

To test this hypothesis, we simulated a simple multi-item working memory task in the previously proposed weakly-connected model (Fig. 4, *0.2%*). We simultaneously presented two stimuli to the network, which were recalled at the end of the delay. The only modification to the previous architecture was the addition of the input lateralization. Each area was assigned a hemisphere ($Area_1$ = left, $Area_2$ = right) and received a weaker input for ipsilateral than contralateral stimuli.

We tested two equidistant stimulus conditions. Stimuli were presented either bilaterally (one per hemifield; Fig. 6a) or unilaterally (both in the same hemifield; Fig. 6b). In bilateral trials, each area received one stronger contralateral and one weaker ipsilateral input (Fig. 6a, solid vs dashed arrows). Due to the single-item capacity of each area, this resulted in the memory maintenance of only the stronger, contralateral item in each area and therefore the successful maintenance of both items across the entire network in a specialized manner (Fig. 6c, example trial). In contrast, in

unilateral trials, both stimuli were contralateral (or ipsilateral) to the same hemisphere, such that each hemisphere received two equally strong inputs (Fig. 6b). This created an equal competition between both items, causing the forgetting of a random item in each area. Crucially, both areas could forget the same item (example trial, Fig. 6d) and therefore store one item redundantly, which reduced the system-wide working memory capacity in unilateral compared to bilateral trials, replicating the well-known bilateral field advantage (Fig. 6e, 1,500 simulations).

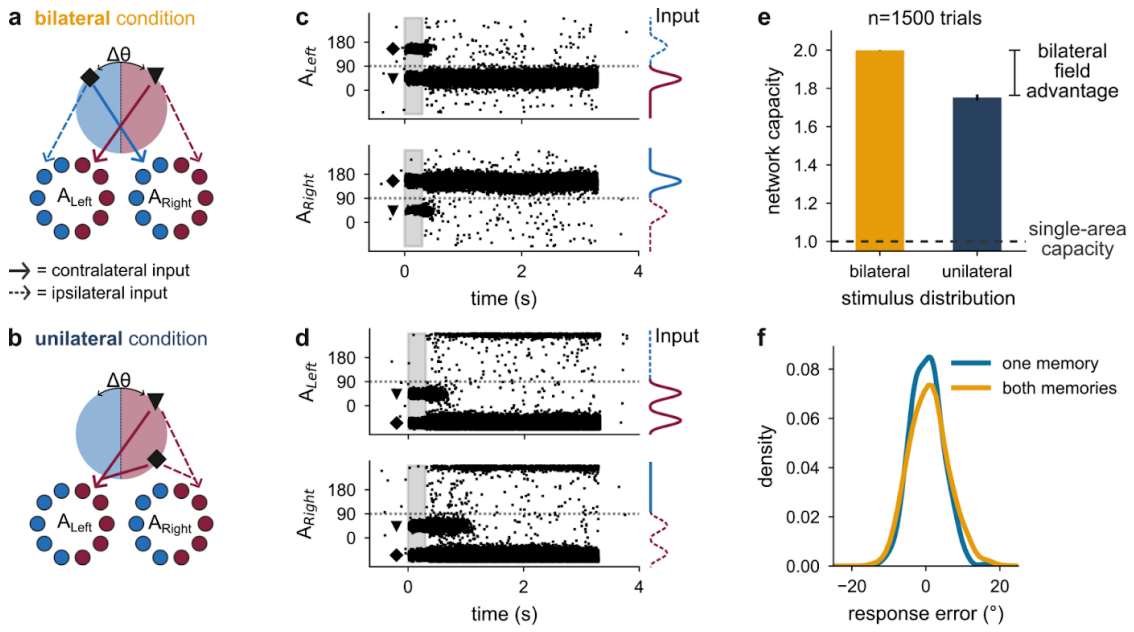


Figure 6: Stimulus lateralization in a redundant network model enabled hemispheric specialization in a multi-item task. *a) The interhemispheric network model from Fig. 4 (0.2% connected) was simulated for two simultaneously presented items, one in each hemifield. Inputs were weaker to the ipsilateral hemisphere (dashed arrows; 20% reduced vs. contralateral, solid arrows). Shapes are illustrative only. b) Same model with two items presented unilaterally. Both areas received two equally strong inputs, since each hemisphere was contralateral (ipsilateral) to both items. c) Example simulation from panel a. Each area retained its contralateral item, leading to both items being retained (A_{Left} retains triangle, A_{Right} retains diamond). Triangle and diamond mark item locations; dashed horizontal lines indicate hemifield boundary (at 90°); colored Gaussians represent input strength. d) Example simulation from panel b. Both areas retained the same item (diamond), causing the forgetting of the other item (triangle) due to equal stimulus inputs in each area. e) Working memory capacity (trial-averaged number of retained items) for bilateral and unilateral stimuli. Capacity was higher in bilateral vs. unilateral trials, with both conditions outperforming the capacity of a single network (bilateral = 2 items, unilateral = 1.75 items, single=1 item). f) Response precision was reduced when both items were remembered*

(bilateral condition, σ = 7.02°) compared to when one item was remembered (unilateral subset when one item was remembered, σ = 4.32°).

Conversely, the memory precision of the network was higher when both areas redundantly stored the same item (Fig. 6f, blue) compared to when each area specialized in storing a unique item (Fig. 6f, yellow). This aligned with earlier findings (Fig. 4c), suggesting that redundant memory representations improve working memory precision.

Taken together, redundant, weakly connected hemispheres with lateralized inputs flexibly optimize working memory performance based on task demands. In simple below-capacity tasks, the redundancy allows for increased precision, while in more demanding above-capacity tasks independent stimulus activation allows for increased capacity through hemispheric specialization.

Discussion

We explored the prefrontal architecture supporting interhemispheric spatial working memory. While mnemonic spatial representations exhibited a contralateral bias in monkey PFC (Fig. 2g,h), each hemisphere predicted trial-to-trial behavioral errors equally well (Fig. 3b,c) and had weakly correlated encoding errors (Fig. 3e,f). These findings align with a network model where each hemisphere operates as a separate attractor network, with tuned but weak interhemispheric connectivity (Fig. 4a-e). This architecture allows for shared memory diffusion (Fig. 4e) as well as private history effects such as reactivations and memory drift (Fig. 4f-k), consistent with the experimental data (Fig. 5). Finally, we demonstrate that this architecture can represent working memory items redundantly or independently, depending on working memory demands (Fig. 6).

Importantly, this study addresses a pending debate about interhemispheric working memory organization. On the one hand, multi-item working memory studies suggest largely independent resources in each prefrontal hemisphere, supporting a lateralized organization (Fig. 1a; reviewed in (11)). On the other hand, optogenetic studies in mice indicate a redundant organization, since the disruption of *both* hemispheres is necessary for memories to be irrevocably disrupted (5,22,23). Our model reconciles these views by proposing a redundant, weakly connected network with lateralized inputs. When memory demands are low, both hemispheres are redundantly activated,

improving memory precision and robustness. When demands are high, competition for lateralized resources promotes hemisphere-specific encoding and increases working memory capacity. Importantly, this flexibility emerges from input lateralization alone, without requiring task-dependent changes in connectivity, as proposed in previous studies (18,22,25,82). Taken together, this network flexibly adapts to varying cognitive demands with a fixed, parsimonious architecture.

This architecture seems to strike a functional trade-off between redundancy and specialization reminiscent of strategies used when storing data in multiple hard drives. When storage space allows, maintaining back-ups reduces the probability of information loss. As storage space becomes limited, prioritizing unique content over duplicates becomes more efficient. The prefrontal hemispheres seem to implement this principle by combining a redundant structure with hemisphere-specific inputs, prioritizing either robustness or capacity depending on task demands.

While our analyses support weak interhemispheric connections, other studies suggest stronger coupling. Specifically, neural selectivity in the mouse ALM can reemerge from the opposite hemisphere after unilateral optogenetic disruption (22,23,83). In contrast, in our weakly connected model neural activity does not recover if an area is silenced (simulations not shown). This discrepancy may reflect task-dependent increases in background input that facilitate faster recovery during active engagement (83) or differences in targeted brain regions. Notably, the optogenetic studies were conducted in the mouse premotor cortex (22,23,25,83), an area known to have stronger callosal connections than prefrontal cortex (84,85). These distinctions highlight the need to investigate how memory lateralization varies across the cortical hierarchy, adding a lateral dimension underexplored in studies that emphasize the role of distributed brain processing in working memory maintenance (39,51,52,86,87), serial dependence (88), and multi-item working memory (7,14,20).

Finally, our findings can have implications for several neuropsychiatric disorders, such as schizophrenia and autism, which show a reduction in callosal connectivity (92,93). The observed deficits in working memory performance in these disorders (32,94,95) should be researched in light of the changes in interhemispheric connectivity.

In conclusion, our study bridges paradoxical findings from single- and multi-item working memory studies by providing a single functional architecture underlying prefrontal hemispheric organization. The proposed redundant, weakly connected architecture accounts for memory effects across task complexities and timescales, encompassing memory precision, history effects, and the bilateral field advantage in multi-item working memory.

Acknowledgements

We are grateful to Megan McDonnell for helpful advice and discussion and assistance in data processing. Grant PID2021-125453OB-I00 funded by MICIU/AEI /10.13039/501100011033 and by FEDER, UE. Grant AC20/00071 funded by Instituto de Salud Carlos III (AC). MT is supported by MICIU/AEI (FPI program). Data collection for this work was supported by NSF NCS BCS 1734916/1954107 (to MAS and Byron Yu), NIH R01 MH118929 (MAS and BY), NIH R01 EB026953 (MAS, BY and Brent Doiron), NIH R01 EY022928 (MAS), and the Simons Foundation (BY).

References

1. Gattass R, Sousa AP, Rosa MG. Visual topography of V1 in the Cebus monkey. J Comp Neurol. 1987 May 22;259(4):529–48.

2. Hubel DH, Wiesel TN. Cortical and callosal connections concerned with the vertical meridian of visual fields in the cat. J Neurophysiol. 1967 Nov;30(6):1561–73.

3. Peiker C, Wunderle T, Eriksson D, Schmidt A, Schmidt KE. An Updated Midline Rule: Visual Callosal Connections Anticipate Shape and Motion in Ongoing Activity across the Hemispheres. J Neurosci. 2013 Nov 13;33(46):18036–46.

4. Schmidt KE, Lomber SG, Innocenti GM. Specificity of Neuronal Responses in Primary Visual Cortex Is Modulated by Interhemispheric CorticoCortical Input. Cereb Cortex. 2010 Dec 1;20(12):2776–86.

5. Svoboda K, Li N. Neural mechanisms of movement planning: motor cortex and beyond. Curr Opin Neurobiol. 2018 Apr 1;49:33–41.

6. Chen J, Zhang C, Hu P, Min B, Wang L. Flexible control of sequence working memory in the macaque frontal cortex. Neuron. 2024 Oct 23;112(20):3502-3514.e6.

7. Panichello MF, Buschman TJ. Shared mechanisms underlie the control of working memory and attention. Nature. 2021 Mar 31;592(7855):601.

8. Xie Y, Hu P, Li J, Chen J, Song W, Wang XJ, et al. Geometry of sequence working memory in macaque prefrontal cortex. Science. 2022 Feb 11;375(6581):632–9.

9. Funahashi S, Bruce CJ, Goldman-Rakic PS. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol. 1989 Feb;61(2):331–49.

10. Rainer G, Asaad WF, Miller EK. Memory fields of neurons in the primate prefrontal cortex. Proc Natl Acad Sci U S A. 1998 Dec 8;95(25):15008–13.

11. Brincat SL, Miller EK. Cognitive independence and interactions between cerebral hemispheres. Neuropsychologia. 2025 Jun;212:109153.

12. Alvarez GA, Cavanagh P. Independent resources for attentional tracking in the left and right visual hemifields. Psychol Sci. 2005 Aug;16(8):637–43.

13. Awh E, Pashler H. Evidence for Split Attentional Foci. J Exp Psychol Hum Percept Perform. 2000 Apr 1;26:834–46.

14. Buschman TJ, Siegel M, Roy JE, Miller EK. Neural substrates of cognitive capacity limitations. Proc Natl Acad Sci. 2011 Jul 5;108(27):11252–5.

15. Delvenne JF. The capacity of visual short-term memory within and between hemifields. Cognition. 2005 Jul 1;96(3):B79–88.

16. Delvenne JF, Castronovo J, Demeyere N, Humphreys GW. Bilateral Field Advantage in Visual Enumeration. PLoS ONE. 2011 Mar 23;6(3):e17743.

17. Dimond S, Beaumont G. Use of two cerebral hemispheres to increase brain capacity. Nature. 1971 Jul 23;232(5308):270–1.

18. Matsushima A, Tanaka M. Different Neuronal Computations of Spatial Working Memory for Multiple Locations within versus across Visual Hemifields. J Neurosci. 2014 Apr

16;34(16):5621.

19. Umemoto A, Drew T, Ester EF, Awh E. A bilateral advantage for storage in visual working memory. Cognition. 2010;117(1):69–79.

20. Raya DV, Gupta S, Sridharan D. Distractors induce space-specific neural biases in visual working memory [Internet]. bioRxiv; 2025 [cited 2025 Jun 26]. p. 2023.12.24.573161. Available from: https://www.biorxiv.org/content/10.1101/2023.12.24.573161v2

21. Brincat SL, Donoghue JA, Mahnke MK, Kornblith S, Lundqvist M, Miller EK. Interhemispheric transfer of working memories. Neuron. 2021 Mar 17;109(6):1055-1066.e4.

22. Chen G, Kang B, Lindsey J, Druckmann S, Li N. Modularity and robustness of frontal cortical networks. Cell. 2021 Jul 8;184(14):3717-3730.e24.

23. Li N, Daie K, Svoboda K, Druckmann S. Robust neuronal dynamics in premotor cortex during motor planning. Nature. 2016 Apr;532(7600):459–64.

24. Li N, Chen TW, Guo ZV, Gerfen CR, Svoboda K. A motor cortex circuit for motor planning and movement. Nature. 2015 Mar;519(7541):51–6.

25. Yin X, Wang Y, Li J, Guo ZV. Lateralization of short-term memory in the frontal cortex. Cell Rep. 2022 Aug 16;40(7):111190.

26. Viswanathan P, Nieder A. Comparison of visual receptive fields in the dorsolateral prefrontal cortex and ventral intraparietal area in macaques. Eur J Neurosci. 2017;46(11):2702–12.

27. Amari S ichi. Dynamics of pattern formation in lateral-inhibition type neural fields. Biol Cybern. 1977 Jun 1;27(2):77–87.

28. Compte A, Brunel N, Goldman-Rakic PS, Wang XJ. Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model. Cereb Cortex. 2000 Sep 1;10(9):910–23.

29. Wilson HR, Cowan JD. Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons. Biophys J. 1972 Jan;12(1):1–24.

30. Wimmer K, Nykamp DQ, Constantinidis C, Compte A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. Nat Neurosci. 2014 Mar;17(3):431–9.

31. Barbosa J, Stein H, Martinez RL, Galan-Gadea A, Li S, Dalmau J, et al. Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. Nat Neurosci. 2020 Aug;23(8):1016–24.

32. Kilpatrick ZP. Synaptic mechanisms of interference in working memory. Sci Rep. 2018 May 18;8(1):7879.

33. Stein H, Barbosa J, Rosa-Justicia M, Prades L, Morató A, Galan-Gadea A, et al. Reduced serial dependence suggests deficits in synaptic potentiation in anti-NMDAR encephalitis and schizophrenia. Nat Commun. 2020 Aug 25;11(1):4250.

34. Bliss DP, D'Esposito M. Synaptic augmentation in a cortical circuit model reproduces serial dependence in visual working memory. PLoS ONE. 2017 Dec 15;12(12):e0188927.

35. Barbosa J, Babushkin V, Temudo A, Sreenivasan KK, Compte A. Across-Area Synchronization Supports Feature Integration in a Biophysical Network Model of Working Memory. Front Neural Circuits [Internet]. 2021 [cited 2022 May 10];15. Available from:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8489684/

36. Almeida R, Barbosa J, Compte A. Neural circuit basis of visuo-spatial working memory precision: a computational and behavioral study. J Neurophysiol. 2015 Sep;114(3):1806–18.

37. Bouchacourt F, Buschman TJ. A Flexible Model of Working Memory. Neuron. 2019 Jul 3;103(1):147-160.e8.

38. Buschman TJ. Balancing Flexibility and Interference in Working Memory. Annu Rev Vis Sci. 2021 Sep 15;7:367–88.

39. Edin F, Klingberg T, Johansson P, McNab F, Tegnér J, Compte A. Mechanism for top-down control of working memory capacity. Proc Natl Acad Sci. 2009 Apr 21;106(16):6802–7.

40. Wei Z, Wang XJ, Wang DH. From Distributed Resources to Limited Slots in Multiple-Item Working Memory: A Spiking Network Model with Normalization. J Neurosci. 2012 Aug 15;32(33):11228–40.

41. Bliss DP, Sun JJ, D'Esposito M. Serial dependence is absent at the time of perception but increases in visual working memory. Sci Rep. 2017 Nov 7;7(1):14739.

42. Fischer C, Kaiser J, Bledowski C. A direct neural signature of serial dependence in working memory. eLife [Internet]. 2024 Jul 7 [cited 2024 Oct 15];13. Available from: https://elifesciences.org/reviewed-preprints/99478

43. Fischer J, Whitney D. Serial dependence in visual perception. Nat Neurosci. 2014 May;17(5):738–43.

44. Fritsche M, Mostert P, Lange FP de. Opposite Effects of Recent History on Perception and Decision. Curr Biol. 2017 Feb 20;27(4):590–5.

45. Fritsche M, de Lange FP. The role of feature-based attention in visual serial dependence. J Vis. 2019 Nov 27;19(13):21.

46. Kiyonaga A, Scimeca JM, Bliss DP, Whitney D. Serial Dependence across Perception, Attention, and Memory. Trends Cogn Sci. 2017 Jul;21(7):493–7.

47. Liberman A, Fischer J, Whitney D. Serial Dependence in the Perception of Faces. Curr Biol CB. 2014 Nov 3;24(21):2569–74.

48. Sheehan TC, Serences JT. Attractive serial dependence overcomes repulsive neuronal adaptation. PLOS Biol. 2022 Jun 9;20(9):e3001711.

49. Stein H, Barbosa J, Soldevilla DL, Rosa-Justicia M, Morató A, Galan-Gadea A, et al. Neural signatures of reduced serial dependence in anti-NMDAR encephalitis and schizophrenia [Internet]. OSF; 2024 [cited 2024 Oct 15]. Available from: https://osf.io/kz73s

50. Ardid S, Wang XJ, Compte A. An Integrated Microcircuit Model of Attentional Processing in the Neocortex. J Neurosci. 2007 Aug 8;27(32):8486–95.

51. Froudist-Walsh S, Bliss DP, Ding X, Jankovic-Rapan L, Niu M, Knoblauch K, et al. A dopamine gradient controls access to distributed working memory in monkey cortex [Internet]. bioRxiv; 2020 [cited 2024 Oct 18]. p. 2020.09.07.286500. Available from: https://www.biorxiv.org/content/10.1101/2020.09.07.286500v1

52. Mejías JF, Wang XJ. Mechanisms of distributed working memory in a large-scale network of macaque neocortex. eLife. 2022 Feb 24;11:e72136.

53. McDonnell ME, Umakantha A, Williamson RC, Smith MA, Yu BM. Interactions across hemispheres in prefrontal cortex reflect global cognitive processing [Internet]. bioRxiv; 2025 [cited 2025 Jun 13]. p. 2025.06.12.659406. Available from: https://www.biorxiv.org/content/10.1101/2025.06.12.659406v1

54. Papadimitriou C, White RL, Snyder LH. Ghosts in the Machine II: Neural Correlates of Memory Interference from the Previous Trial. Cereb Cortex N Y N 1991. 2017 Apr 1;27(4):2513–27.

55. Ploner CJ, Gaymard B, Rivaud S, Agid Y, Pierrot-Deseilligny C. Temporal limits of spatial working memory in humans. Eur J Neurosci. 1998;10(2):794–7.

56. White JM, Sparks DL, Stanford TR. Saccades to remembered target locations: an analysis of systematic and variable errors. Vision Res. 1994 Jan 1;34(1):79–92.

57. Barbosa J, Compte A. Build-up of serial dependence in color working memory. Sci Rep. 2020 Jul 2;10(1):10959.

58. Cicchini GM, Mikellidou K, Burr DC. The functional role of serial dependence. Proc R Soc B Biol Sci. 2018 Oct 31;285(1890):20181722.

59. de Azevedo Neto RM, Bartels A. Disrupting Short-Term Memory Maintenance in Premotor Cortex Affects Serial Dependence in Visuomotor Integration. J Neurosci. 2021 Nov 10;41(45):9392–402.

60. Fornaciai M, Park J. Attractive Serial Dependence in the Absence of an Explicit Task. Psychol Sci. 2018 Mar;29(3):437–46.

61. Fornaciai M, Park J. Serial dependence generalizes across different stimulus formats, but not different sensory modalities. Vision Res. 2019 Jul 1;160:108–15.

62. Manassi M, Whitney D. Continuity fields enhance visual perception through positive serial dependence. Nat Rev Psychol. 2024 May;3(5):352–66.

63. Markov YA, Tiurina NA, Pascucci D. Serial dependence: A matter of memory load. Heliyon. 2024 Jul 15;10(13):e33977.

64. St John-Saaltink E, Kok P, Lau HC, de Lange FP. Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. J Neurosci Off J Soc Neurosci. 2016 Jun 8;36(23):6186–92.

65. Papadimitriou C, Ferdoash A, Snyder LH. Ghosts in the machine: memory interference from the previous trial. J Neurophysiol. 2015 Jan 15;113(2):567–77.

66. Constantinidis C, Funahashi S, Lee D, Murray JD, Qi XL, Wang M, et al. Persistent Spiking Activity Underlies Working Memory. J Neurosci. 2018 Aug 8;38(32):7020–8.

67. Calangiu I, Kollmorgen S, Reppas J, Mante V. Prospective and retrospective representations of saccadic movements in primate prefrontal cortex. Cell Rep. 2025 Feb 25;44(2):115289.

68. Luo M, Zhang H, Luo H. Reactivated past decisions repel early sensory processing and attract late decision-making [Internet]. bioRxiv; 2024 [cited 2024 Oct 15]. p. 2024.02.26.582221. Available from: https://www.biorxiv.org/content/10.1101/2024.02.26.582221v1

69. Ranieri G, Benedetto A, Ho HT, Burr DC, Morrone MC. Evidence of Serial Dependence from Decoding of Visual Evoked Potentials. J Neurosci. 2022 Oct 12;JN-RM-1879-21.

70. Zhang H, Luo H. Feature-specific reactivations of past information shift current neural encoding thereby mediating serial bias behaviors. PLOS Biol. 2023 Mar 24;21(3):e3002056.

71. Gu Q, Zhang Q, Han Y, Li P, Gao Z, Shen M. Microsaccades reflect attention shifts: a mini review of 20 years of microsaccade research. Front Psychol [Internet]. 2024 Feb 19 [cited 2024 Mar 25];15. Available from: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2024.1364939/full

72. Johnston R, Snyder AC, Khanna SB, Issar D, Smith MA. The eyes reflect an internal cognitive state hidden in the population activity of cortical neurons. Cereb Cortex. 2022 Aug 1;32(15):3331–46.

73. Linde-Domingo J, Spitzer B. Geometry of visuospatial working memory information in miniature gaze patterns. Nat Hum Behav. 2024 Feb;8(2):336–48.

74. Mostert P, Albers AM, Brinkman L, Todorova L, Kok P, Lange FP de. Eye Movement-Related Confounds in Neural Decoding of Visual Working Memory Representations. eNeuro [Internet]. 2018 Jul 1 [cited 2024 Nov 21];5(4). Available from: https://www.eneuro.org/content/5/4/ENEURO.0401-17.2018

75. Willeke KF, Tian X, Buonocore A, Bellet J, Ramirez-Cardenas A, Hafed ZM. Memory-guided microsaccades. Nat Commun. 2019 Aug 16;10(1):3710.

76. Mongillo G, Barak O, Tsodyks M. Synaptic theory of working memory. Science. 2008 Mar 14;319(5869):1543–6.

77. Dragoi V, Sharma J, Sur M. Adaptation-induced plasticity of orientation tuning in adult visual cortex. Neuron. 2000 Oct;28(1):287–98.

78. Gibson JJ, Radner M. Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. J Exp Psychol. 1937;20(5):453–67.

79. Hajonides JE, van Ede F, Stokes MG, Nobre AC, Myers NE. Multiple and Dissociable Effects of Sensory History on Working-Memory Performance [Internet]. bioRxiv; 2021 [cited 2022 May 10]. p. 2021.10.31.466639. Available from: https://www.biorxiv.org/content/10.1101/2021.10.31.466639v1

80. Pasternak T, Lui LL, Spinelli PM. Unilateral Prefrontal Lesions Impair Memory-Guided Comparisons of Contralateral Visual Motion. J Neurosci. 2015 May 6;35(18):7095–105.

81. Zhang Y, Ye C, Roberson D, Zhao G, Xue C, Liu Q. The bilateral field advantage effect in memory precision. Q J Exp Psychol 2006. 2018 Mar;71(3):749–58.

82. Pinotsis DA, Buschman TJ, Miller EK. Working Memory Load Modulates Neuronal Coupling. Cereb Cortex. 2019 Apr 1;29(4):1670–81.

83. Inagaki HK, Fontolan L, Romani S, Svoboda K. Discrete attractor dynamics underlies persistent activity in the frontal cortex. Nature. 2019 Feb;566(7743):212–7.

84. Caminiti R, Ghaziri H, Galuske R, Hof PR, Innocenti GM. Evolution amplified processing with temporally dispersed slow neuronal connectivity in primates. Proc Natl Acad Sci. 2009 Nov 17;106(46):19551–6.

85. Schwartz ML. Organization and Development of Interhemispheric Connections of the Prefrontal Cortex in Rhesus Monkey. In: Reeves AG, Roberts DW, editors. Epilepsy and the Corpus Callosum 2 [Internet]. Boston, MA: Springer US; 1995 [cited 2024 Nov 11]. p. 29–47. Available from: https://doi.org/10.1007/978-1-4899-1427-9_4

86. Christophel TB, Klink PC, Spitzer B, Roelfsema PR, Haynes JD. The Distributed Nature of Working Memory. Trends Cogn Sci. 2017 Feb;21(2):111–24.

87. Dotson NM, Hoffman SJ, Goodell B, Gray CM. Feature-Based Visual Short-Term Memory Is Widely Distributed and Hierarchically Organized. Neuron. 2018 Jul;99(1):215-226.e4.

88. Tschiersch M. Causal evidence for the higher-order origin of serial dependence suggests a multi-area account. J Neurophysiol. 2022 Aug 1;128(2):336–8.

89. Hikosaka O, Wurtz RH. Visual and oculomotor functions of monkey substantia nigra pars reticulata. IV. Relation of substantia nigra to superior colliculus. J Neurophysiol. 1983 May;49(5):1285–301.

90. Vallat R. Pingouin: statistics in Python. J Open Source Softw. 2018 Nov 19;3(31):1026.

91. Talluri BC, Kang I, Lazere A, Quinn KR, Kaliss N, Yates JL, et al. Activity in primate visual cortex is minimally driven by spontaneous movements. Nat Neurosci. 2023 Nov;26(11):1953–9.

92. Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK. Single-trial neural dynamics are dominated by richly varied movements. Nat Neurosci. 2019 Oct;22(10):1677–86.

93. Liu B, Nobre AC, van Ede F. Microsaccades transiently lateralise EEG alpha activity. Prog Neurobiol. 2023 May 1;224:102433.

94. Semedo JD, Jasper AI, Zandvakili A, Krishna A, Aschner A, Machens CK, et al. Feedforward and feedback interactions between visual cortical areas use different population activity patterns. Nat Commun. 2022 Mar 1;13(1):1099.

95. Gerstner W, Kistler WM, Naud R, Paninski L. Neuronal Dynamics - From single neurons to networks and models of cognition and beyond [Internet]. 2014 [cited 2024 Jan 22]. Available from: https://neuronaldynamics.epfl.ch/

96. Stimberg M, Brette R, Goodman DF. Brian 2, an intuitive and efficient neural simulator. Skinner FK, Calabrese RL, Skinner FK, Zeldenrust F, Gerkin RC, editors. eLife. 2019 Aug 20;8:e47314.

Methods

Behavioral task and recordings

Three male adult rhesus monkeys (Macaca mulatta) performed an oculo-motor delayed response task (89), where they had to fixate on the center of the screen while a peripheral stimulus was shown for 420 ms in one of 4, 8 or 16 different locations (7 (3,453 trials), 12 (5,666 trials), 22 (9,065 trials) sessions, respectively). After a delay period the monkeys made a saccade to the remembered stimulus location to report their response and gain a water reward for correct trials. These data were collected as part of a separate study (53). The details of the experiments (task, training, surgeries, recordings) are briefly summarized here. All experimental procedures were approved by the Institutional Animal Care and Use Committees of the University of Pittsburgh and Carnegie Mellon University.

The delay lengths were randomly varied between 1.5s and 3s, where each session uniformly covered a 0.5s interval of delay lengths (except session Sa1, which covered 1.5s) and each monkey had a different range of delay lengths (Sa: 1.5s - 3s, Pe: 1.5s - 2.6s, Wa: 1.5s - 2.4s). Between trials, a mostly self-timed inter-trial interval of 0.9s - 2.2s separated individual trials. Fixation was maintained within an invisible window around the fixation point (Monkey Sa: 2.4 or 2.9, Monkey Pe: 2.4, Monkey Wa: 2.8 degrees radius) throughout the entire trial until response. However, we here denote the fixation time as the time from fixation onset until stimulus presentation. Eye Tracking was performed with an EyeLink 1000 eye tracker system with a sampling rate of 1 kHz.

All three monkeys were implanted with a 96-channel multi-electrode recording device (“Utah” array, Blackrock Neurotech) in each hemisphere of the prefrontal cortex anterior to the arcuate sulcus and dorsal to the medial sulcus (area 8Ar), which were used to simultaneously record activity during the task. The recordings led to a total number of 5708 recorded multi-units across both hemispheres for all animals and 41 sessions (Sa: 137.9+/-16.4 units/session, Pe: 157.9+/-15.4 units/session, Wa: 108.4+/-10.8 units/session).

Behavioral data analysis

Precision analysis

Each trial's saccadic endpoint response was parameterized in polar coordinates taking the fixation point as origin and only the angular dimension θ_r was considered for further analysis. In each trial, the angular error (θ_e) was computed as the circular distance between the stimulus angle (θ_s) and the response (θ_r). We subtracted the mean of θ_e in each target location for each subject, session and delay to account for systematic biases (except for the analyses shown in Fig. 1d,e). For each monkey, precision was estimated by fitting an ordinary least squares (OLS) model on the absolute response error with the delay (abs(θ_e) ~ delay) across sessions using the *statsmodels.formula.api* Python package.

Serial dependence plots

To visualize serial dependence (Fig. 1f), we plotted the current trial's error (θ_e^{curr}) against the circular distance between the previous and the current stimulus location ($\Delta_{prev,curr} = circdist(\theta_s^{prev}, \theta_s^{curr})$). To account for possible systematic errors, we plotted a flipped form of the serial dependence by multiplying θ_e^{curr} with the $sign(\Delta_{prev,curr})$ and plotting it over $abs(\Delta_{prev,curr})$ (57).

Serial dependence strength fit

To measure the strength of serial dependence β, both in behavior and neural analyses, we fitted behavioral or decoder errors to $\Delta_{prev,curr}$ using an OLS model: $\theta_e \sim \beta\, F(\Delta_{prev,curr})$. The typical shape of serial dependence curve $F(x)$ was parameterized using a negative first-order normalized derivative of Gaussian (DoG1) (33):

$$f'(\sigma, x) = -\frac{x}{\sigma^3\sqrt{2\pi}} \cdot exp(-\frac{x^2}{2\sigma^2})$$

$$f'_{normalized}(\sigma, x) = \frac{f'(\sigma, x)}{max|f'(\sigma, x)|}$$

For each monkey, we determined $F(x) = f'_{normalized}(\sigma_{opt}, x)$, where σ_{opt} was obtained for each monkey by minimizing the model's BIC within a range of σ values ($\sigma = [0.3, 3.0, 0.05]$ rad) for all trials from that monkey (Extended Data Fig. 6, a-c for behavior, d-f for neural history drift, g-i for neural reactivation drift).

Neural data analysis

Memory fields

To compute the preferred location of each multi-unit during memory maintenance (memory field (MF)), we obtained its mean firing rate in each trial r_i (spike count between stimulus offset and go-cue divided by delay duration) for each stimulus location ($\theta_{s,i}$) and computed the population vector (30):

$$MF = \left(\sum_{i=0}^{trials} r_i exp\left(j \cdot \theta_{s,i}\right)\right)\left(\sum_{i=0}^{trials} r_i\right)^{-1}$$

The preferred orientation of each multi-unit θ_{pref} is given by the angle of the complex-valued MF and the tuning strength by the absolute value of the MF.

To determine if a multi-unit's MF tuning was significantly above chance, we computed the MF for 1000 random shuffles of stimulus locations. A multi-unit's MF was classified as significant when its tuning strength exceeded 97.5% of the tuning strengths of the shuffles.

Visualizing PFC bumps

To visualize the representation of the stimulus in PFC activity (Fig. 2b,c), both in individual trials and as an average over trials, we displayed the firing rate of each multi-unit over time, where the multi-units were sorted by their θ_{pref}. To avoid the circularity of generating a bump during the delay solely by sorting by θ_{pref}, we estimated θ_{pref} in trials not used for display using a 5-fold cross validation for trial average, and a leave-one-out cross validation for single trial visualization. To correct for heterogeneous baseline firing rates of each multi-unit, we subtracted the average firing rate (across time and trials) from each multi-unit's activity.

Time-resolved population representations were obtained by averaging both along the time and the neuron axes. Along the time axes, we smoothed firing rates with a sliding average window (step size 10ms; window size 100ms for trial averages, and 200ms for single trials).

Along the neural axis, we also performed a moving average smoothing (bin width W = $\frac{2\pi}{N}$, with N the number of different target stimuli used in the session; step size $\frac{W}{5}$).

Finally, prior to averaging representations across trials we re-centered the multi-unit's relative to the stimulus angle in each trial, so that multi-units preferring each trial's stimulus location were placed at the axis origin.

Population decoding of stimulus location

We used linear regression to decode stimulus location from population activity in the trial. For each unit and trial, spikes were binned into independent 200ms time bins aligned to fixation or response onset. Due to variable delay lengths, the fixation- and response-aligned delay activity share a partial overlap since a section of the shortest delay duration (for each session) from stimulus start or delay-end is taken into account.

All decoders were trained using a 5-fold stimulus-stratified cross validation approach. In the training set (80% of the trials), at each time t the complex weights β_t of the neural activity regressors X_t were estimated to reproduce the complex stimulus location y_s by least-squares estimation:

$$\beta_t = (X_t^{\mathrm{T}} \cdot X_t)^{-1} \cdot X_t^{\mathrm{T}} \cdot y_s$$

where both β_t and y_s are complex vectors, describing the coordinates $(a + ib)$ of the trained receptive field and stimulus locations, respectively.

The computed weights were then used to predict the stimulus location $(\widehat{y_s})$ based on the neural activity in the previously left out trials (test set, 20% of trials):

$$\hat{y}_s = X_t \cdot \beta_t$$

As a metric for the decoding strength of the trained decoder, we then evaluated the mean squared error (MSE) of the angular predictions ($\hat{\theta}_s = arg(\hat{y}_s)$) against the angular stimulus locations ($\theta_s = arg(y_s)$):

$$\bar{\theta}_e = circdist(\hat{\theta}_s, \theta_s)\ , MSE = mean(\bar{\theta}_e{}^2)\ .$$

To account for the different number of stimuli in each session and any potential unbalanced trial distribution, we also tested the previously trained model repeatedly (50 times) against randomly shuffled θ_s. To compute a decoding score, we z-scored the MSE against the shuffled MSEs ($MSE_{shuffle}$) and provided the output as a measure of standard deviation σ. The decoding score is thus obtained as:

$$decoding = \frac{mean(MSE_{shuffle}) - MSE}{std(MSE_{shuffle})}$$

where we changed the sign of the score to facilitate interpretability, so that high values correspond to high accuracy.

Cross-temporal decoders

We trained a linear regression to fit the stimulus location as a linear combination of the neural activity in each time point using a stratified 5-fold cross validation. For each trained time point, we tested the model on all time points of the left-out trials. The cross-temporal decoder matrices shown in Fig. 2e-g only shows fixation-aligned activity for simplicity.

Single-trial decoders

While the cross-temporal decoders were used to gain a general impression of the data, the single-trial decoders were meant to give a precise prediction for individual trials. We focused on two different decoders: A Same-time-Decoder for estimates based on each time step (train, test in same time point; diagonal of cross-temporal decoder) and a Delay-Decoder which only contains the memory code (train on average delay activity, test in each time point). The Delay-Decoder

allowed for the detection of memory reactivations at the next trial's fixation onset. As shown before (31,49), the Same-time-Decoder included response-related activity. For single-trial decoders, we performed a 20-times repeated 5-fold cross validation. We performed the repeats since single trials were noisy and averaging across repeats led to more stable predictions.

Decoder-behavior error correlations

We computed trial-by-trial correlations between neural decoder prediction errors $\bar{\theta}_e$ and behavioral response errors θ_e to view how prefrontal population codes reflected the small clockwise and counter-clockwise response errors of the monkeys. We either trained one decoder on all multi-units across hemispheres (combined analysis, Fig. 3a) or one per hemisphere (independent analysis, left vs. right hemisphere, Fig. 3b,c). The obtained prediction errors (combined, left, right) were each circularly correlated (*circ_corrcc* from *pingouin* package (version 0.4.0) (90); uniform correction) with the behavioral response errors in each session in each time step and averaged across sessions (error bars taken across sessions). In the independent analysis, we performed the correlations either on only ipsilateral vs. contralateral trials relative to the respective hemisphere, or for the left and right hemisphere separately on all trials, to compare which hemisphere predicts behavior.

Decoder-decoder error correlations

We assessed task-related coupling between simultaneously recorded neural populations of the two hemispheres by estimating the trial-by-trial error correlations between predictions of the respective decoders (Fig. 3e,f). We compared within hemisphere to across hemisphere error correlations. To do so, we repeatedly (5 times) trained four independent decoders on a random half of units from each hemisphere ($left_{A,}$ $left_B$, $right_{A.}$ $right_B$) respectively. The obtained decoder errors ($\bar{\theta}_e$) were then circularly correlated between the different decoders, either within a hemisphere ($left_A$ vs. $left_B$, $right_A$ vs. $right_B$) or across hemispheres ($left_{A/B}$ vs. $right_{A/B}$, respectively) at each time point for each repeat and session. The correlations were averaged across the five repeats and afterwards across sessions (error bars are across sessions).

Uninstructed eye-movement controls

We used a Same-Time Decoder to decode the stimulus location from uninstructed eye movements (Extended Data Fig. 5a). We computed the decoder as mentioned previously, but we replaced the multi-unit activity by the complex gaze location (see section *Population decoding of stimulus location*). Since the acquisition rate of the eye tracker and electrophysiology was the same, we binned the eye locations in the same manner as the neural data.

We also correlated the gaze location with the stimulus location (Extended Data Fig. 5b, *angle*) and the gaze error with the performed response error (Extended Data Fig. 5b, *error*), to understand if gaze location could correlate performance errors across hemispheres. Since gaze and later response errors were not correlated we excluded this confound.

Additionally, we tested if removing the information of the complex gaze location from the neural decoder had an impact on error correlations (Extended Data Fig. 5c,d). To do so, we added the gaze location as an additional regressor to the neural activity in each of the four sub-hemispheric decoders (complex gaze location is treated like an additional multi-unit). However, we computed the error predictions by testing on a partial model, where gaze information was discarded. This gave us neural predictions that did not overlap with stimulus encoding through uninstructed eye movements. The gaze-excluded predictions were correlated within and across sub-hemispheres ($left_{A,}$ $left_{B}$, $right_{A.}$ $right_{B}$) as described above (Extended Data Fig. 5c, d).

Taken together, these analyses control for uninstructed eye movements as a potential source of common input, but other uninstructed movements also elicit shared neural variability across brain regions although less in monkeys (91) than in mice (92). However, the shared neural information underlying correlations in memory imprecisions is highly specific and can only be conceivably induced by movements that cause a consistent rotation of mnemonic representations across hemispheres. We therefore argue that eye gaze is the main confound to address, as it has been shown to change spatial encoding during the delay period (71,75,93). Since removing eye-confounds did not change the correlations, we argue for a task-related communication

929 subspace (94) underlying information transfer between the prefrontal hemispheres, likely realized
930 through tuned interhemispheric connections anatomically supported by the corpus callosum.

931

932 **Neural history drift**

933 To estimate the attraction (or repulsion) of the neural population code in the current trial's delay
934 period towards (or away from) the previous trial stimulus location, we plotted the single-trial
935 Sametime Decoder prediction errors ($\overline{\theta}_e$) against the distance of the previous and current stimulus
936 ($\Delta_{prev,curr}$) (Fig. 5a). We quantified this history drift by fitting a negative first-order normalized
937 derivative of Gaussian (DoG1) to this curve, as for the behavior (see section *Serial dependence*
938 *strength*). The hyperparameter σ_{opt} was chosen based on the model BIC (see section *Serial*
939 *dependence strength;* Extended Data Fig. 6, d-f, Sa: 1.35 rad, Wa: 1.75 rad, Pe: 3.0 rad). We
940 applied this fitting procedure to $\overline{\theta}_e$ in each time step (200 ms) through the stimulus and delay
941 periods (Fig. 5b).

942 To account for auto-correlation between sequential stimuli (consecutive trials were less likely to
943 occur in the same location) we computed the decoder errors as the circular distance between $\hat{\theta}_s$
944 and a shuffle control ($\hat{\theta}_{s,shuffle}$). We obtained the shuffled predictions $\hat{\theta}_{s,shuffle}$ by shuffling the
945 neural activity of trials with the same current stimulus location randomly (50 repeats) and multiplied
946 them with the decoder weights. This corrected for uneven sequential stimuli distribution, while
947 retaining the drift towards a specific previous stimulus location (as only trials with the same current
948 stimulus are shuffled). This shuffle was only performed for completeness to extrapolate the drift to
949 the previous trial. It did not affect the drift of the current trials shown in Fig. 5b.

950

951 **Reactivations**

952 Fig. 2h shows that reactivations occur immediately before stimulus presentation. We therefore
953 estimated the location and strength of reactivations in individual trials using a Single-trial
954 Delay-Decoder on neural activity averaged over the two time bins in the time window starting from

fixation start until the current stimulus start. The angle of this decoder's prediction provides the reactivation location, while its absolute value reflects the reactivation strength.

Reactivation drift

The reactivation drift was computed in a similar manner to the Neural history drift but instead of considering the relative location of the previous trial's stimulus, we considered the relative distance of the current stimulus to the previous reactivation location ($\Delta_{react,\, curr}$). Therefore both components of the reactivation drift, the relative reactivation locations and the delay prediction errors, depended on the neural predictions from either hemisphere. We could therefore compute the drift of the population activity during the delay towards the reactivation either within the same hemisphere or across hemisphere (Fig. 5d).

Gaussian Mixture Model

To analyze if reactivations co-occur in the two hemispheres we applied the Gaussian mixture model (GMM) from *scikit-learn* to the scatter plot of trial-by-trial reactivation decoder strengths from the right and left PFC hemispheres (Fig. 5e). We decided on modeling four classes as the hypothesis was for the model to find a class for each: *null,* $single_{left}$, $single_{right}$ and *dual* reactivations. We determined the covariance constraint by comparing the BIC of the different models (best fit: diagonal covariance) and repeated the GMM fit 250 times with different initial parameters. The fact that *dual* reactivations (a class along the diagonal indicating strong decoder strength for both hemispheres) did not appear in any of our fits of reactivation strengths (Fig. 5e) but were instead systematically observed if instead of reactivations we considered the strength of delay-period codes (Fig. 5f) indicated that reactivations occurred privately in each hemisphere.

<u>Bump-attractor model</u>

We adapted the previously proposed bump attractor model implementation from (31) based on (95) in Brian2 (96). We simulated 1024 excitatory and 256 inhibitory leaky-integrate-and-fire spiking neurons for each area representing the left and right hemispheres. Within each area, all neurons were densely connected (all-to-all connectivity) with AMPAR, NMDAR and $GABA_AR$ synaptic

dynamics (decay time constants $\tau_{AMPA} = 2ms$, $\tau_{NMDA} = 100ms$, $\tau_{GABA} = 10ms$). The strength of excitation between excitatory neurons in an area depended on the distance between their respective receptive fields through a Gaussian connectivity profile (std 14.4 deg). All other within-area connections were untuned.

Most within-area parameters were kept as in the original implementation (31), but we changed $J_{EE}^{+} = 1.73$, $\eta_{poisson} = 0.925Hz$.

Short-term plasticity

The model also incorporated short-term synaptic plasticity in the excitatory-to-excitatory connections within each area (31), following the dynamics of facilitation (u) and depression (x) proposed by (76):

$$\frac{dx}{dt} = \frac{(1-x)}{\tau_x} - ux\delta(t - t_{sp})$$

$$\frac{du}{dt} = \frac{(U-u)}{\tau_u} + U(1 - u)\delta(t - t_{sp})$$

where t_{sp} marks the spike times and $\delta(t)$ is the Dirac-Delta function. The constants were chosen as $\tau_x = 0.25s$, $\tau_u = 1.5s$ and $U = 0.15$, and the variables were initialized as $u_0 = 0.08$, $x_0 = 0.9$.

Synaptic weights between excitatory neurons were multiplied by $u \cdot x$. The addition of short-term plasticity to the bump-attractor model allows for activity-silent storage of previous spiking activity and therefore allows for serial dependence through the inter-trial interval as previously shown (31).

We also implemented a perceptual repulsion to the previous stimulus by shifting the stimulus input a maximum of 0.75° away from the previous stimulus through a DoG1 ($\sigma = 0.5$ rad) of the distance between the previous and current item $\Delta_{prev,curr}$, similar to (33,34).

Across-area connectivity

We connected the two area networks for the left and the right hemispheres via excitatory connections either in a tuned (scaled version of the within-area excitatory-to-excitatory

connectivity) or random manner. For simplicity, across-area connections were only mediated by AMPAR excitatory-to-excitatory (EE, tuned) and excitatory-to-inhibitory (EI, untuned) synaptic dynamics as in (35). To stay within a similar parameter regime when varying across-area connectivity, we compensated the across-area connectivity by concomitant decreases in within-area EE and EI AMPAR connectivity. We varied the across-area connectivity between 0.2% and 1% of the within-area baseline model strengths (in 0.2% steps).

Stimulation paradigm

An example simulation can be seen in Figure 4b. A positive current ($0.3\,nA$) of a Gaussian shape ($\sigma = \pi/2$ rad) was selectively applied to the neurons centered around the chosen stimulus location for $500\,ms$. Following the stimulus extinction, circuit reverberation supported tuned persistent activity that lasted through a $3s$ delay before a negative, non-specific input to all neurons ($-\,0.6\,nA$, $500\,ms$) disrupted the delay activity and reset a low-rate baseline state. The network was left unperturbed and mostly silent for an inter-trial interval of $1500ms$, after which we applied a reactivation signal through a weak, positive, non-specific input ($0.036\,nA$, $500\,ms$). This input reactivated the previously silent code in 38% of trials and was turned off after a period of $300ms$ with a negative stimulus ($-\,0.3\,nA$, $500\,ms$). After another $1200\,ms$, the second stimulus input was applied at a random location with the same characteristics as the first stimulus input. However, its stimulus location was shifted slightly away from the previous stimulus to simulate sensory adaptation from upstream areas (see section *Short-term plasticity*). This stimulus also elicited persistent activity, which was maintained during another $3s$ delay, after which the simulation was terminated.

Model behavior

The behavioral readout (θ_{bump}) from each area of the model was determined by computing the circular mean of the excitatory neurons' preferred direction (θ_n) weighted by their firing rates (r_n):

$$\theta_{bump} = angle\left(\sum_{n=0}^{N} r_n e^{\theta_n i}\right)$$

with N being the number of excitatory neurons. The "behavioral response" of the model simulation was determined using the firing rate average in the last $250\ ms$ of each delay. We also computed neural decoder readouts using the same approach but considering r_n in $200ms$ time windows (in $100\ ms$ steps) throughout the trial. The reactivation location and strength was computed using the same decoding approach with r_n in the $250ms$ following the reactivation stimulus end.

Multi-item simulations

For the simulations in Figure 6, the same model as in the single-item working memory simulations (Fig. 4) was used with an across-area connectivity of 0.2% (which fits the data in Fig. 5) and simulated for n=1500 trials. The only change to the network was the introduction of a slight stimulus input lateralization (80): We assigned each area to a hemisphere ($Area_1$ = left, $Area_2$ = right) and varied the stimulus strength according to the target's hemifield (90-270° = left; 0-90° or 270-360° = right). When stimuli were presented contralaterally to a hemisphere, the stimulus strength was set to $0.3\ nA$, as before, but if the stimuli were presented ipsilaterally the stimulus strength was reduced to $0.24\ nA$. The results remained qualitatively true for varying the ipsilateral stimulus strength in a reasonable range.

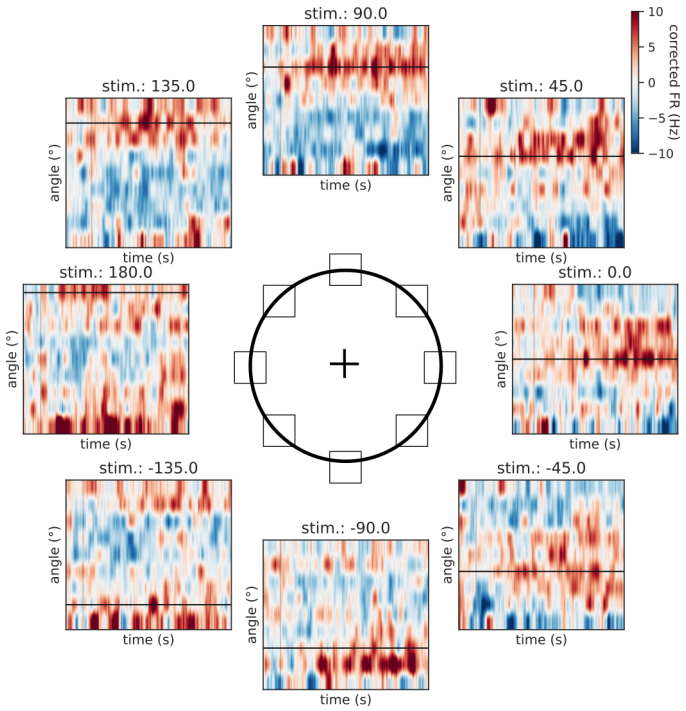
Both networks were simultaneously stimulated with two stimuli, where one stimulus was always induced at 45° (right hemifield), while the other stimulus was randomly presented in 110° distance to the first stimulus, placing it at 155° (left hemifield, "across" hemifield to the first) or 295° (right hemifield, "within" hemifield to the first). The time course of the stimulation protocol remained the same as in the single-item task.

For this proof-of-concept we reused the model from the previous analysis, which has a default capacity of one item per ring. This led to a system-wide capacity of two-items since we have two attractor networks (Area1, Area2). However, fine-tuning the model with stronger and more tuned excitation (40) or inhibition (39) can increase the working memory capacity of each bump-attractor model, which would allow modeling of higher capacities.

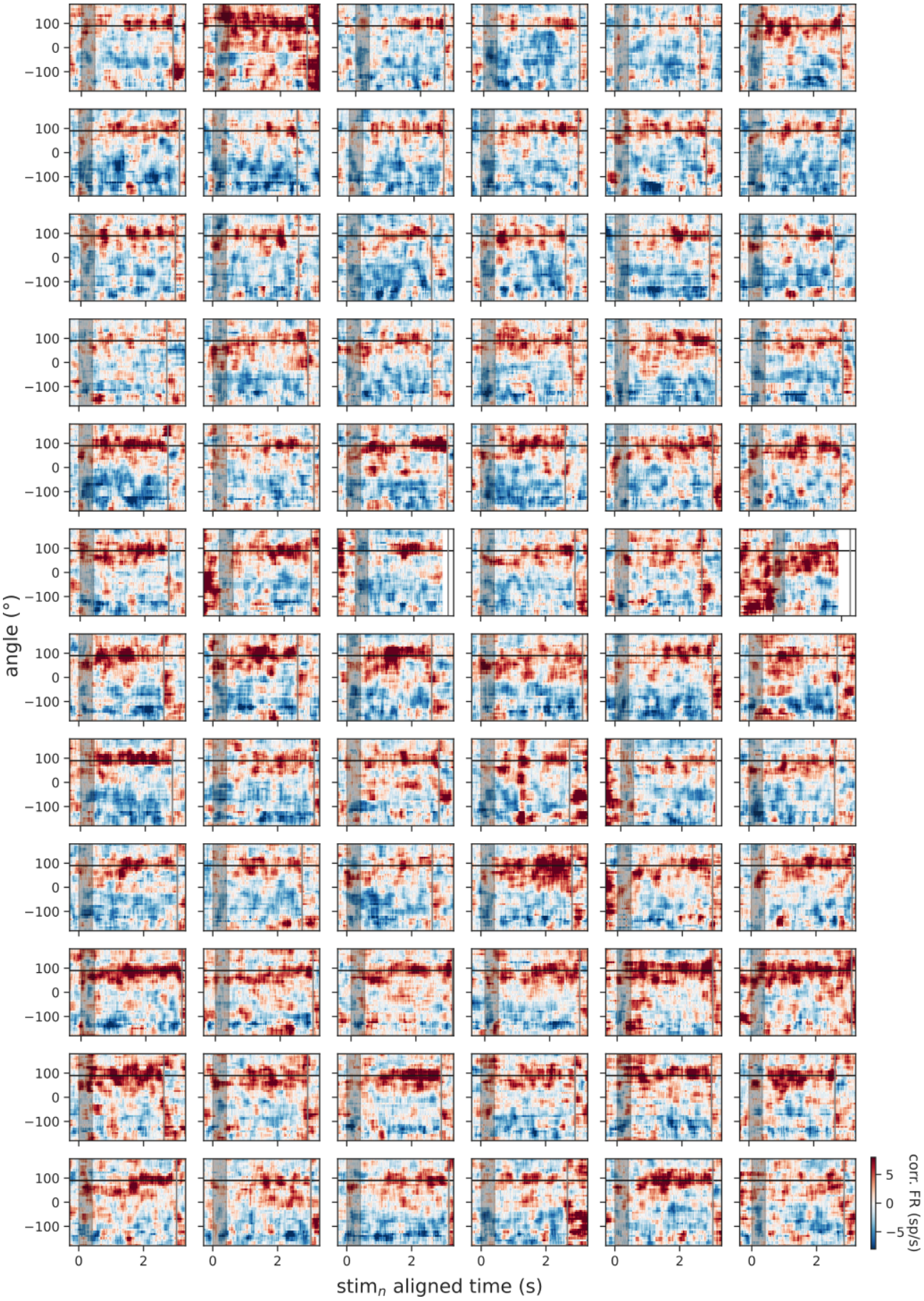
To read out two “behavioral responses” from each model simulation, we divided the network into three equally sized bins (0-120°, 120-240°, 240-360°) and decoded from the neural activity in the two bins where stimuli had been presented. This either led to a precise readout if the bump for the corresponding stimulus had been maintained or to a random location readout in case the memory was not retained. This allowed us to determine the network capacity, meaning how many items were retained in each trial (Fig. 6e).

The memory precision (Fig. 6f) was computed by determining which stimulus was retained in each area and computing the response error to the remembered stimulus. We compared simulations were one item was remembered (the same item in both areas) to simulations were both items were remembered (one item per area).

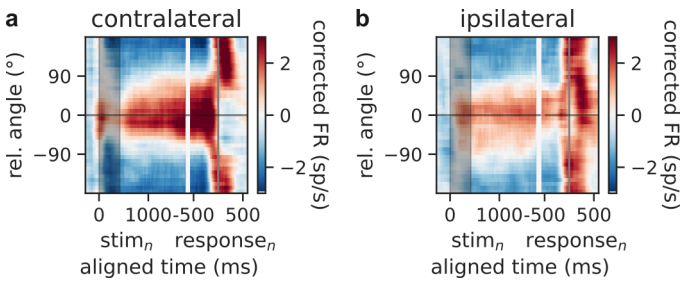
Extended Data



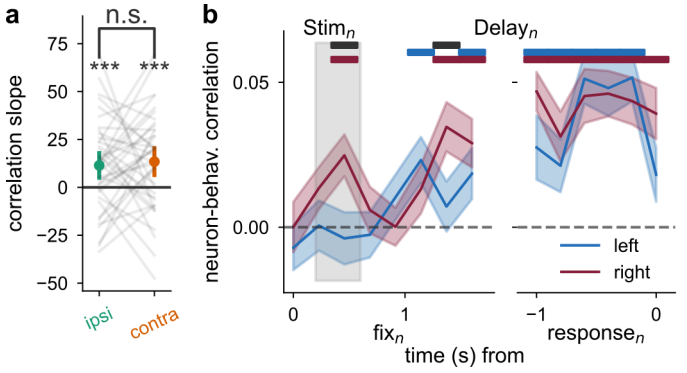
Extended Data Fig. 1: Neural firing in example single trials with varying stimulus location. *Multi-units (y-axis) are grouped by their memory fields into 8 bins across circular space and the average of the multi-unit’s mean-subtracted firing rates within each bin (win.: 200ms, step: 10ms) is shown over time. Colors are mapped for all targets in the same limits , see colorbar top right corner.*



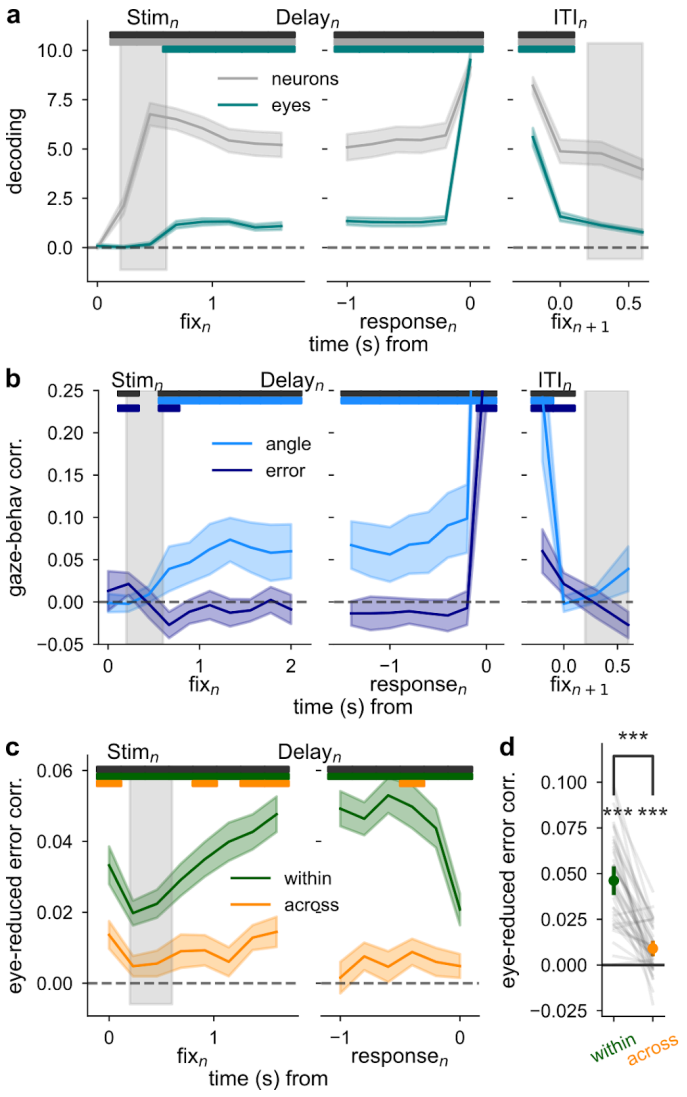
Extended Data Fig. 2: Neural firing in all single trials with stimulus location at 90° in an example session (monkey: Sa, session: 0).



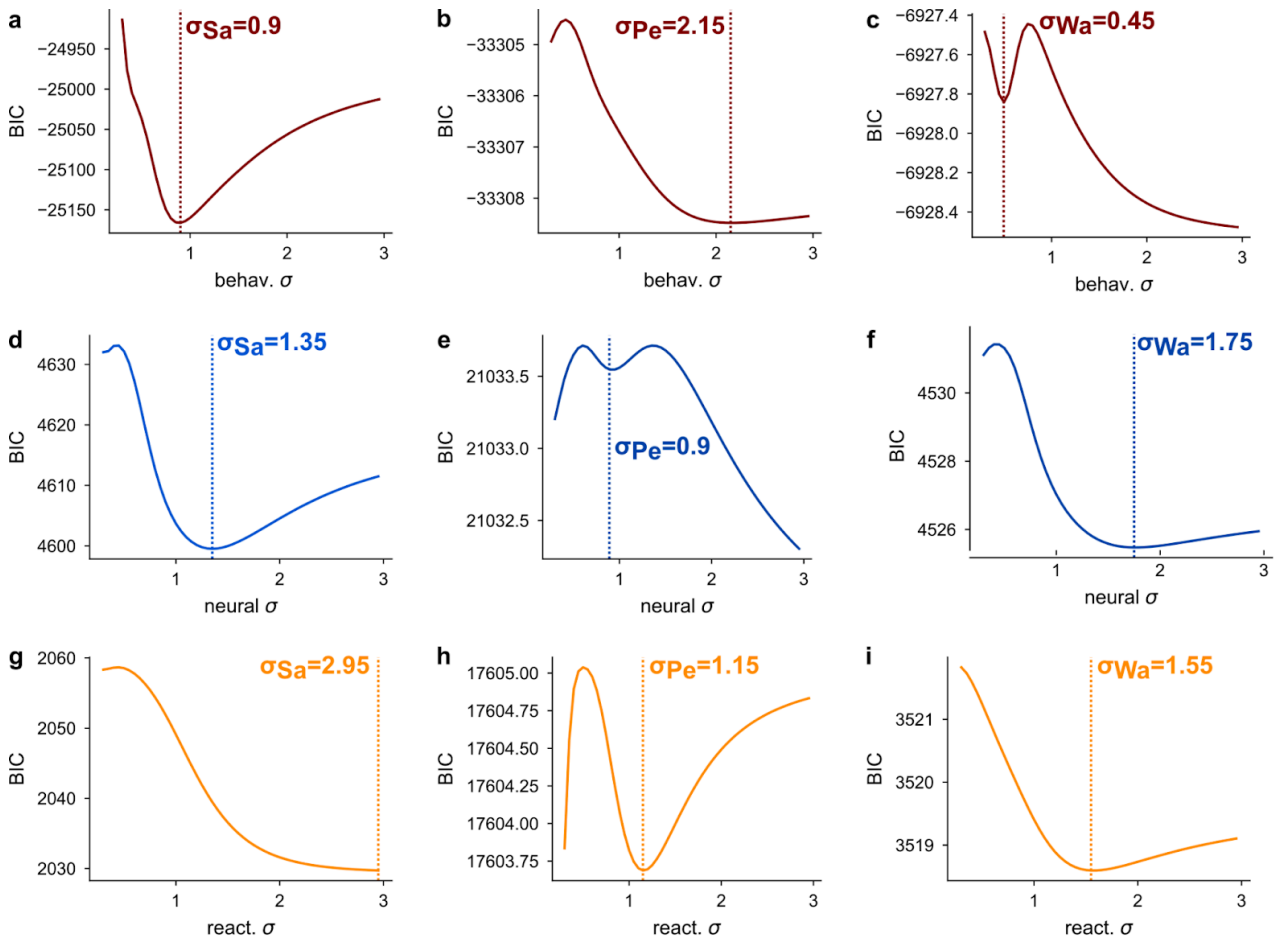
Extended Data Fig. 3: Neural firing averaged across trials for a) contralateral and b) ipsilateral stimuli in an example session (monkey: Sa, session: 0).



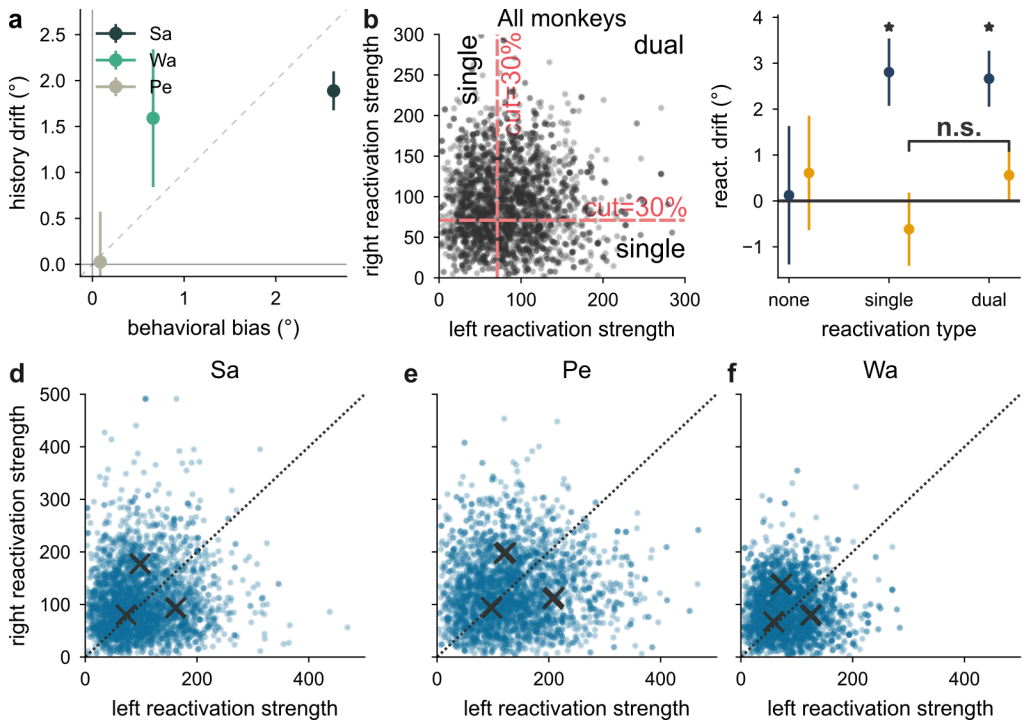
Extended Data Fig. 4: Both hemispheres contribute equally to the response. *a) Quantification of the slopes of ipsilateral vs. contralateral correlation with behavioral response of Fig. 3b. Ipsilateral (t-test: t = 3.07 1/s, p = 0.004) and contralateral (t-test: t = 3.34 1/s, p = 0.002) slopes differ significantly from zero, but do not differ from each other (paired t-test: t = -0.38 1/s, p = 0.7). b) Full time course of the correlation of left vs. right hemisphere prediction errors with behavioral errors (slopes shown in Fig. 3c). Error bars are s.e.m., top bars indicate significance of time points (black: difference between left, right).*



Extended Data Fig. 5: Error correlations are not due to common input. *a) A significant portion of the stimuli can be decoded from uninstructed eye movements even during the delay (while the monkey is fixating). However, this portion is smaller than the decoding from the multi-units. b) The angles of uninstructed eye-movements are significantly correlated to target angles during the task. However, gaze errors are not correlated to behavioral errors during the delay, confirming that uninstructed eye movements do not provide common inputs informative about the errors correlated in Fig. 3e. c) Analysis from Fig. 3e but when introducing a regressor for eye movements during the training of the decoder. During testing only the angles inferred by the multi-units are interpreted. Error correlations across areas are still significantly above 0, excluding common input as an explanation. d) Quantification of average delay correlations of panel c (t-test: within: $t = 11.75$, $p < 10^{-13}$, across: $t = 4.95$, $p = 0.00015$; paired t-test: within-across: $t = 9.18$, $p < 10^{-10}$).*



Extended Data Fig. 6: BIC-score evaluation of best hyperparameter σ (width of the curve) for DoG1 model fits. *BIC score for different hyperparameters σ computed for a-c) behavioral, d-f) neural and g-i) reactivation attraction for each monkey (Sa: a,d,g; Pe: b,e,h; Wa: c,f,i). Best DoG1 fit is determined as the first occurring BIC minimum.*



Extended Data Fig. 7: Reactivations occur mostly independently in each hemisphere, as predicted by the model a) The strength of the neural history drift (extracted from decoder errors) covaries with the strength of the behavioral serial dependence across monkeys. b) Reactivation strengths extracted from the left, right hemisphere decoders in 1000 individual example trials. The cut-off of 30% of reactivation strength in each session is approximately indicated through red lines, splitting trials into high and low reactivation trials for each hemisphere. c) Last time point of reactivation drift shown in Fig. 5d when splitting trials based on their reactivation strength (see panel b) into *null*, *single* or *dual* reactivation trials (cut-off = 30% in each session). No reactivations led to no reactivation drift. Both, single and dual reactivations led to a reactivation drift within the area but did not differ significantly from each other (t-test within: Null: t = 0.08°, p = 0.93, Single: t = 3.82°, p = 0.0005, Dual: t = 4.36°, $p < 10^{-4}$, t-test across: Null: t = 0.49°, p = 0.62, Single: t = -0.77°, p = 0.44, Dual: t = 0.95°, p = 0.34, Single vs. Dual: t = -1.19°, p = 0.24, paired t-test within-across: Null: t = -0.23°, p = 0.81, Single: t = 3.43°, p = 0.001, Dual: t = 2.28°, p = 0.028) suggesting that the dual class was produced by noise rather than a true difference in reactivation strengths. d,e,f) Gaussian mixture model of fitting three classes on left vs. right reactivations. The classes were located around zero and off-diagonal, indicating that reactivations most often occur not at all or in only one of the hemispheres. Results are similar for the three monkeys Sa (d), Pe (e), (Wa) (f) and stable across different initial conditions (250 repetitions) for all three monkeys.