# Redundant, weakly connected prefrontal hemispheres balance precision and capacity in spatial working memory

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]Universitat de Barcelona, 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 Biomedical Engineering Institute, 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

How the prefrontal hemispheres coordinate to adapt to spatial working memory (WM) demands remains an open question. Recently, two models have been proposed: A specialized model, where each hemisphere governs contralateral behavior, and a redundant model, where both hemispheres equally guide behavior in the full visual space. To explore these alternatives, we analyzed simultaneous bilateral prefrontal cortex recordings from three macaque monkeys performing a visuo-spatial WM task. Each hemisphere represented targets across the full visual field and equally predicted behavioral imprecisions. Furthermore, memory errors were weakly correlated between hemispheres, suggesting that redundant, weakly coupled prefrontal hemispheres support spatial WM. Attractor model simulations showed that the hemispheric redundancy improved precision in simple tasks, whereas weak inter-hemispheric coupling allowed for specialized hemispheres in complex tasks. This interhemispheric architecture reconciles previous findings thought to support distinct models into a unified architecture, providing a versatile interhemispheric architecture that adapts to varying cognitive demands.

## Introduction

The world is spatially continuous, yet the inputs the primary visual cortex (V1) receives are divided at the vertical meridian, with each hemisphere processing only the contralateral visual field. These divided hemifield representations are stitched together by callosal projections between the two hemispheres, which selectively connect neurons with receptive fields along the vertical midline (Hubel & Wiesel, 1967; Peiker et al., 2013; Schmidt et al., 2010). However, while this division is well understood in early visual areas, it remains unclear how higher-order regions with less strict lateralization (Rainer et al., 1998), such as the prefrontal cortex (PFC), support complex cognitive functions like working memory (WM) across visual space (Svoboda & Li, 2018).

Unlike in V1, neurons in both prefrontal hemispheres have receptive fields across the full visual space (J. Chen et al., 2024; Panichello & Buschman, 2021; Xie et al., 2022), although a majority of neurons exhibit a contralateral preference (Funahashi et al., 1989). To explain how this organization can lead to unified responses, recent studies propose two potential models for the interhemispheric organization of spatial WM in the PFC: a specialized model in which each hemisphere is mostly responsible for contralateral information (Fig. 1a), and a redundant model where both prefrontal hemispheres represent the entire visual space and equally guide behavior for memories presented across space (Fig. 1b).

Behavioral studies of multi-item WM in human and non-human primates support the specialized model (Fig. 1a), since WM capacity is greater for items presented across than within visual hemifields (bilateral field advantage) (Alvarez & Cavanagh, 2005; Buschman et al., 2011; Delvenne, 2005; Delvenne et al., 2011; Dimond & Beaumont, 1971; Matsushima & Tanaka, 2014; Umemoto et al., 2010). This behavioral effect suggests independent storage of each visual hemifield. The independence of hemispheric resources is further reflected in neural activity recorded from PFC, where memory tuning is only reduced if several items are presented within the same, but not across hemifields (Buschman et al., 2011; Matsushima & Tanaka, 2014), suggesting reduced network competition for each hemisphere. Additionally, distractors have different effects on items presented in the same or the opposite hemifield (Gupta & Sridharan, 2023), and WM representations have been suggested to transfer across prefrontal hemispheres following a gaze shift across the vertical meridian (Brincat et al., 2021). Overall, these studies suggest specialized hemispheres with independent storage for each hemifield.

Conversely, optogenetic studies in mice suggest a redundant model, since disrupting short-term memory traces in one hemisphere alone does not impair behavior, whereas bilateral disruption does (Chen et al., 2021; Li et al., 2016a; Svoboda & Li, 2018). These findings suggest that each individual hemisphere is sufficient to store memories and guide

behavior across the full visual space (Fig. 1b), challenging the specialized model 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, consistent model. To further elucidate the interhemispheric organization of WM, we analyzed simultaneous recordings from the monkey bilateral prefrontal cortex during a visuo-spatial WM task. To understand the underlying network architecture, we modeled the data through bump attractor networks (Amari, 1977; Compte et al., 2000; Wilson & Cowan, 1972), a fundamental conceptual tool to interpret neural recordings in spatial WM. Such attractor models can successfully replicate WM features such as precision (Compte et al., 2000; Wimmer et al., 2014), history effects (Barbosa et al., 2020; Kilpatrick, 2018; Stein et al., 2020), location-dependent distraction (Compte et al., 2000; Gupta & Sridharan, 2023), feature-binding through synchronization (Barbosa et al., 2021), and limited multi-item WM capacity and interference based on within-network attractor competition (Almeida et al., 2015; Bouchacourt & Buschman, 2019; Buschman, 2021; Edin et al., 2009; Wei et al., 2012). This model thus constitutes a prime tool to address the computational basis of spatial WM. We here focused on two of these well-understood behavioral effects in WM, which act at different time scales and arise through different underlying neural mechanisms. First, within-trial memory diffusion, known to explain WM precision based on persistent activity (Compte et al., 2000; Wimmer et al., 2014). Second, across-trial history drift, explaining the attraction of current memories to the previous one (serial dependence) (Bliss et al., 2017; C. Fischer et al., 2024; J. Fischer & Whitney, 2014; Fritsche et al., 2017; Fritsche & de Lange, 2019; Kiyonaga et al., 2017; Liberman et al., 2014; Sheehan & Serences, 2022) based on synaptic plasticity (Barbosa et al., 2020; Bliss & D'Esposito, 2017; Kilpatrick, 2018; Stein et al., 2020, 2024).

Our neural analyses showed that WM across the visual space was stably and simultaneously represented by both prefrontal hemispheres in a lateralized manner (J. Chen

et al., 2024; Funahashi et al., 1989; Panichello & Buschman, 2021; Xie et al., 2022). Importantly, activity from both hemispheres equally guided behavior consistent with redundant memories in both hemispheres (Fig. 1b).

Furthermore, errors were weakly, but significantly correlated across hemispheres, consistent with weak inter-hemispheric connections. Using two redundant, weakly connected bump attractor networks of spiking neurons, we modeled this architecture, building on recent multi-network frameworks (Ardid et al., 2007; Barbosa et al., 2021; Froudist-Walsh et al., 2020; Gupta & Sridharan, 2023; Mejías & Wang, 2022) and thus replicated findings of memory diffusion and history drift. Importantly, simulations revealed complementary advantages of redundancy and weak coupling 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 WM by analyzing the behavior and bilateral population recordings from area 8Ar of three macaque monkeys performing a visuo-spatial oculomotor delayed response task (Umakantha et al., 2020). While the monkeys were fixating, the 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: WM precision and serial dependence. These behavioral effects have been linked, respectively, to within-trial diffusion and across-trial drift of bump attractor dynamics (Barbosa et al., 2020; Compte et al., 2000; Wimmer et al., 2014) (Fig. 1c, bottom) and can provide us with information on WM maintenance across different timescales. Monkeys exhibited good accuracy overall, but precision errors significantly

increased in longer delays (Fig. 1d,e), as previously reported (Papadimitriou et al., 2017; Ploner et al., 1998; White et al., 1994).

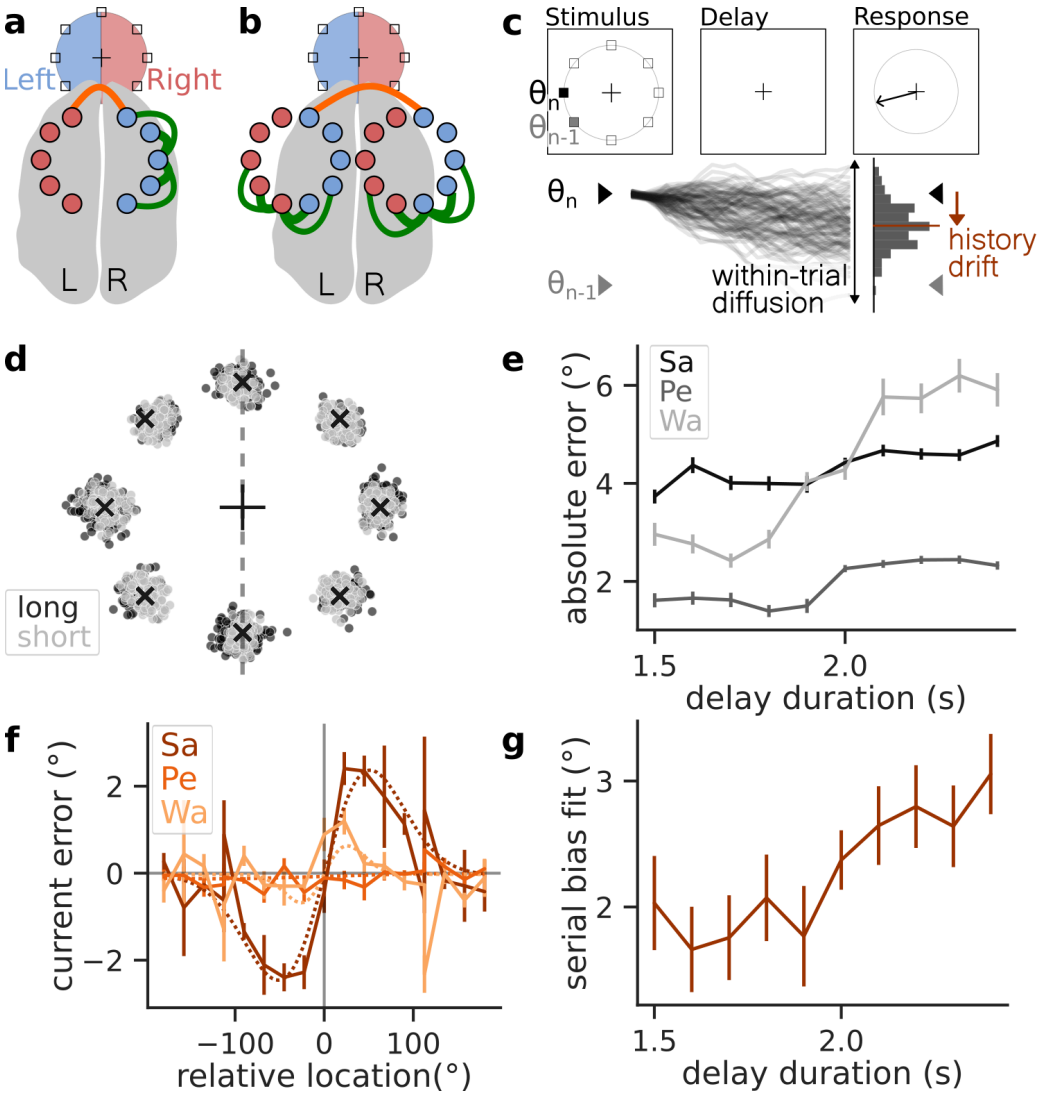



***Figure 1: Delay dependence of working memory within and across trials.***

*a) A specialized model of WM where each hemisphere represents the contralateral visual field with connections only along the vertical meridian, similar to V1 (Hubel & Wiesel, 1967; Peiker et al., 2013; Schmidt et al., 2010). Circles indicate neurons with preferred WM towards the colored hemifield with within-hemisphere (green) or across-hemisphere (orange) connections. b) A redundant model with a network covering both hemifields in each hemisphere. c) Schematic of an example trial in the behavioral task (top) illustrating two neural mechanisms (bottom) underlying memory inaccuracies. The memory of a target $\theta_n$ randomly diffused within a trial throughout the delay period and also drifted towards the recent history (previous trial = $\theta_{n-1}$). d) Responses (dots) were clustered around the target locations (crosses) similarly across the entire space, but error size depended on*

*delay length (here: highest vs. lowest delay quartile). e) Absolute response errors increased for all monkeys with delay length (OLS model: Sa: □ = 1.02 °/s, $p < 10^{-14}$. Pe: □ = 0.99 °/s, $p < 10^{-16}$. Wa: □ = 4.5 °/s, $p < 10^{-48}$, across monkeys: $p < 10^{-14}$). f) Serial dependence for the three monkeys illustrated strong idiosyncratic differences between animals, with significant serial dependence for Sa ($p < 10^{-70}$) and Wa ($p = 0.05$) but not Pe ($p = 0.76$). Model fits are shown as dashed lines (OLS model: Sa: σ = 0.9 rad: □ = 2.92º; Pe: σ = 2.15 rad: □ = 0.04º; Wa: σ = 0.45 rad: □ = 0.73º). g) Serial dependence increased with delay length for monkey Sa (OLS model: Delay-interaction: Sa: □ = 1.03 °/s, $p = 0.001$; Pe: □ = 0.21 °/s, $p = 0.2$; Wa: □ = 0.05 °/s, $p = 0.97$). Error bars are +/- 1 s.e.m..*

We also observed serial dependence, an attraction of the current trial memory report towards the previous target (Fig. 1f), replicating findings in both humans (Barbosa et al., 2020; Barbosa & Compte, 2020; Bliss et al., 2017; Cicchini et al., 2018; de Azevedo Neto & Bartels, 2021; C. Fischer et al., 2024; J. Fischer & Whitney, 2014; Fornaciai & Park, 2018, 2019; Fritsche et al., 2017; Fritsche & de Lange, 2019; Kiyonaga et al., 2017; Liberman et al., 2014; Manassi & Whitney, 2024; Markov et al., 2024; Sheehan & Serences, 2022; St John-Saaltink et al., 2016; Stein et al., 2020, 2024) and monkeys (Barbosa et al., 2020; Papadimitriou et al., 2015, 2017). We measured the effect size by fitting the response error using a first-order derivative of Gaussian (DoG1; for hyperparameter fit see 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 delay dependency (Fig. 1g), as previously reported (Bliss et al., 2017; Fritsche et al., 2017; Papadimitriou et al., 2015; Stein et al., 2020).

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

While the monkeys performed this task, neural activity was recorded from bilateral 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 (Barbosa et al., 2020; Constantinidis

et al., 2018; Funahashi et al., 1989; Wimmer et al., 2014). We leverage the large population recordings to obtain a single-trial view of population activity by showing the evolving firing rates of multi-units sorted by their memory fields (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 end of the delay period (example trial in Fig. 2b, for 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 when averaging out noise across trials of the example session (Fig. 2c).

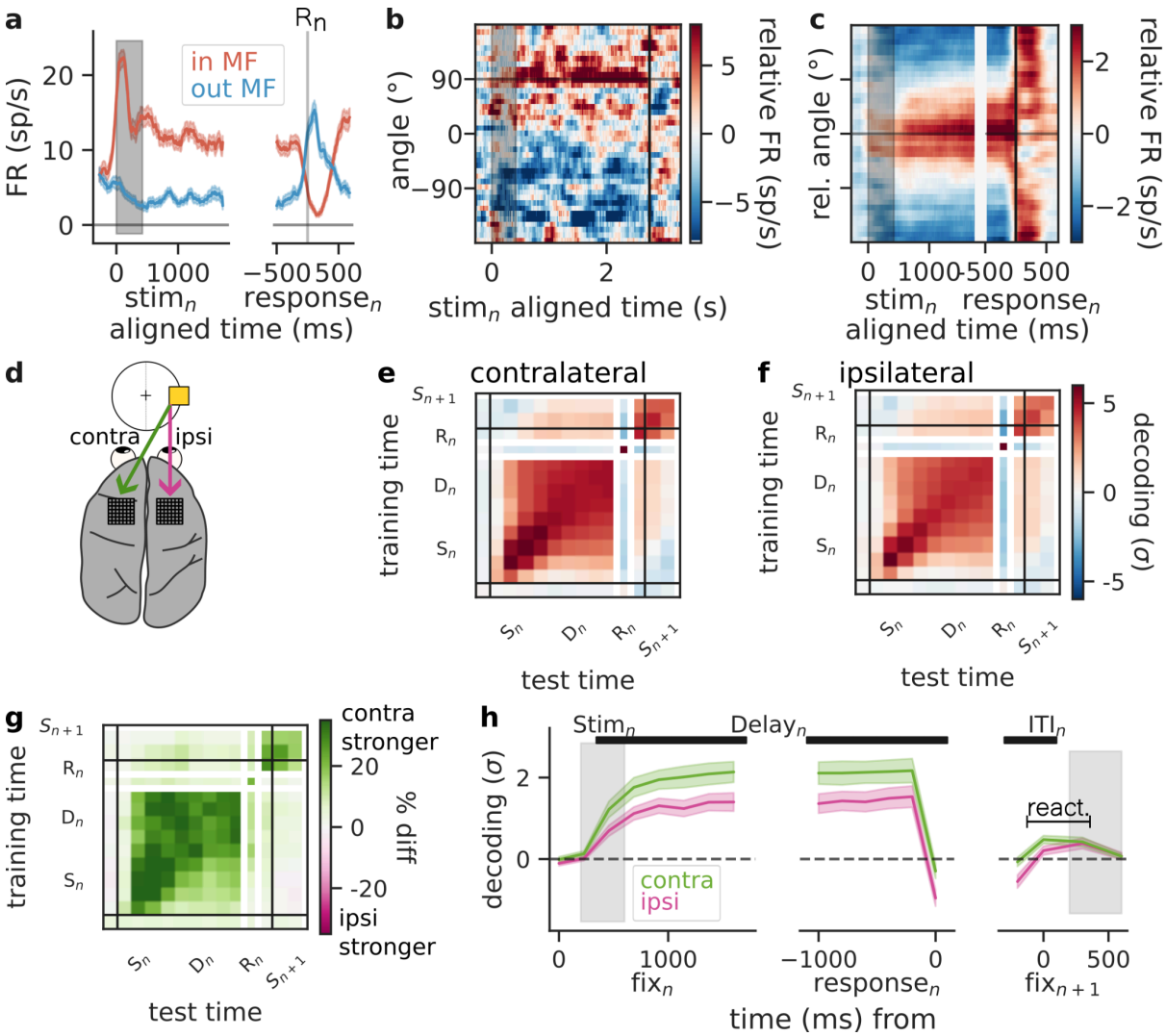



***Figure 2: Hemispheres represent the full visual space with a contralateral preference.*** *a) Average firing rate of an example multi-unit across trials where cues were inside (red) or outside (blue) of its memory field (calculated on the average delay activity). In total, 3180 out of 5708 multi-units were selective during the delay*

*(Sa: 86.58%, Pe: 23.63%, Wa: 68.95%, bootstrapped test (see Methods)). b) Tuned persistent population activity in a single example trial from stimulus start (gray background; stimulus location at 90°) to response (vertical black line). Multi-units are sorted by preferred location (calculated on left-out trials, Methods). c) Same as panel b but for an example session average. The average population activity is recentered around the stimulus location in each trial. d) Sketch of the bilateral arrays in area 8Ar in an example trial with the stimulus in the right hemifield, making the left hemisphere contralateral and the right hemisphere ipsilateral. e) Temporal generalization of the decoder accuracy, measured as the distance from a shuffled representation in units of their standard deviation (σ) for contralateral trials (average of left/right hemisphere). White lines indicate a discontinuity because of timing differences across trials; black lines indicate stimulus presentation. Same colorbar as f. f) same for ipsilateral trials. g) Lateralization percentage between contralateral and ipsilateral decoding ((contra-ipsi)/contra*100), indicating stronger contralateral code (average of 16.59% stronger during the delay). h) A Delay decoder (trained on the average delay activity) tested on contralateral and ipsilateral trials separately could disentangle the memory and response code showing emerging reactivations ($fix_{n+1}$) in both hemispheres (t-test: ipsilateral: $t = 2.54\sigma$, $p = 0.016$, contralateral: $t = 4.86\sigma$, $p < 10^{-4}$, paired t-test: difference contra-ipsi: $t = 1.8\sigma$, $p = 0.08$). Error bars are +/- 1 s.e.m..*

To investigate the extent of hemispheric lateralization across sessions (single session firing rate view in Extended Data Fig. 3), we computed linear decoders to predict the stimulus location in each trial based on the population activity in each hemisphere (see Methods). We trained a decoder on 80% of trials in each hemisphere at each time step and computed the mean-squared error (MSE) on contralateral and ipsilateral left-out trials separately (schematic Fig. 2d). We compared the MSE to a stimulus-shuffled representation (inverted z-score to shuffle, “decoding” in Figures; see Methods) and found a stable delay code (dashed square) in both contralateral (Fig. 2e) and ipsilateral (Fig. 2f) trials. Although each hemisphere reliably stored memories across the full visual field, contralateral decoding outperformed ipsilateral decoding by 16.59% (averaged decoding performance difference during the delay, Fig. 2g), leaving it unclear which hemisphere predominantly guided overall memory and behavior.

In what follows, we used a decoder trained on the average delay code (Fig. 2h, *Delay Decoder*) to focus on the encoding of memory contents rather than other task parameters,

such as motor response (Calangiu et al., 2022). This revealed a reemergence of memory-related activity during fixation of the upcoming trial, a reactivation (line at $fix_{n+1}$, Fig. 2h), which has previously been shown to correlate with serial dependence strength (Barbosa et al., 2020; Luo et al., 2024; Ranieri et al., 2022; Stein et al., 2024; H. Zhang & Luo, 2023). Notably, we found reactivations in both hemispheres (Fig. 2h), indicating little lateralization in the occurrence of reactivations.

These analyses showed that while both hemispheres stored full-field memories in a stable manner, contralateral memories were stored more accurately. This imbalance between the contralateral and ipsilateral memory code opens the question of how the brain reconciles the higher contralateral precision with the need for a unified behavioral response.

*Both hemispheres reflected trial-by-trial behavioral variability equally and had weakly correlated memory representations.*

To address how the lateralized memory representations are combined to provide a unified behavioral response, we correlated the neural decoder errors on each trial with the corresponding behavioral response errors (Papadimitriou et al., 2017; Wimmer et al., 2014). A positive correlation during the delay period indicates that the multi-units can predict which behavioral errors the monkeys will make later in the trial. Furthermore, a steady increase in correlation during the delay suggests that memories slowly diffuse away from the original memory location as proposed by attractor dynamics (Papadimitriou et al., 2017; Wimmer et al., 2014). We found steadily increasing correlations when pooling multi-units across both hemispheres (Fig. 3a).

To understand how each hemisphere contributes to the behavior, we trained decoders separately on each hemisphere and performed the same analysis. The left and right hemispheres showed no difference in correlation with behavior (Fig. 3c, Extended Data Fig. 4b). Surprisingly, when testing on ipsilateral vs. contralateral trials, both conditions showed

the same steady increase in correlations of neural prediction errors and behavioral errors (Fig. 3b, Extended Data Fig. 4a), even though the data showed a more precise memory code in the contralateral hemisphere (Fig. 2g). This analysis showed that both hemispheres equally predicted small behavioral errors in the full visual field, so both were equally weighted in guiding behavior. Therefore, our analyses ruled out the specialized model (Fig. 1a) and rather pointed towards a redundant architecture (Fig. 1b), in which both hemispheres encoded stimuli in the full visual field and equally guided behavioral responses.

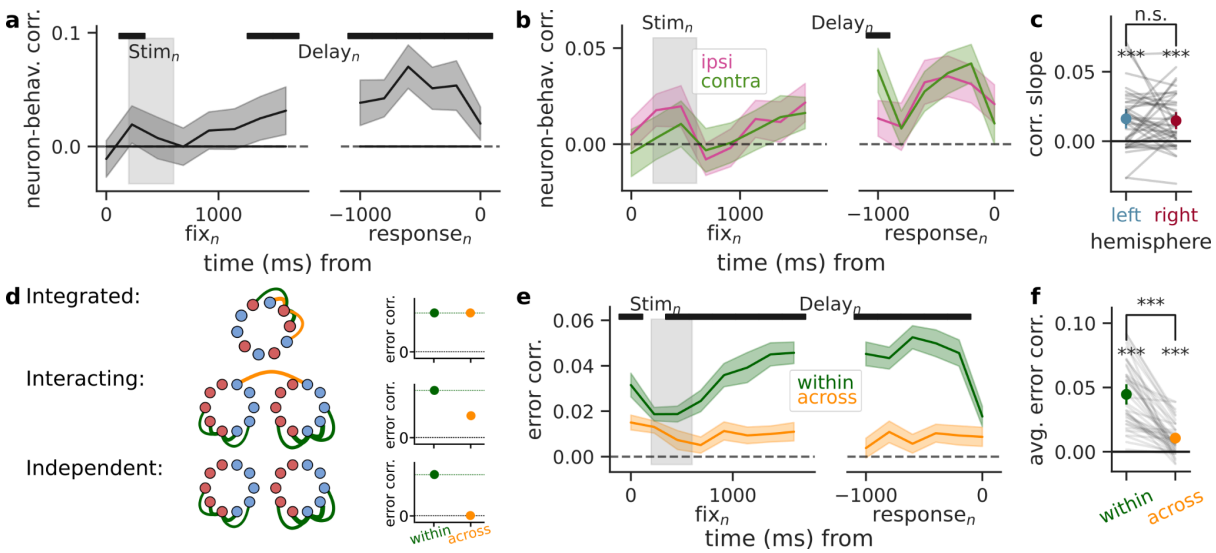



***Figure 3: Memories were stored redundantly in both hemispheres with interacting representations.*** *a) Memory diffusion is computed through correlating decoder and behavioral response errors. Neural prediction errors became more aligned with behavioral errors throughout the delay (t-test of slope: $t = 4.95$ 1/s, $p < 10^{-4}$). Error bars are 95% C.I. b) Same as a, but computed for ipsilateral and contralateral trials separately. By the end of the delay period, both neural representations significantly correlate with the responses but don't significantly differ from each other. (paired t-test of slopes: $t = -0.38$ 1/s, $p = 0.7$) c) Correlation increase does not significantly differ for the left vs. right hemisphere (paired t-test: $t = 0.375$ 1/s, $p = 0.71$). d) Schematic of different network architectures predicting different across-hemisphere error correlations. e) Error correlations across time shown 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 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.*

Having discarded the specialized model, we aimed to specify the hemispheric organization of the prefrontal cortex that supports the redundant model. A redundant architecture can range from tightly integrated hemispheres keeping a mirror copy of each other (Fig. 3d, top) to two uncoupled, independently diffusing hemispheres (Fig. 3d, bottom). These alternatives can be tested through analyzing how memories diffuse in each hemisphere. If both hemispheres maintained mirror copies of each item, memory errors would be identical in both hemispheres (Fig. 3d, top, right). On the other hand, independent hemispheres would lead to two separate sets of prediction errors as memories diffused independently (Fig. 3d, bottom, right). To distinguish between these conditions, we used the decoder prediction errors as a proxy for the differences in maintained memories. We correlated these neural errors either within a hemisphere (to get a baseline stability of correlations) or across hemispheres (to see how much code was shared). Specifically, we repeatedly trained four decoders on random halves of the multi-units recorded from each hemisphere (sub-hemispheres: $left_A$, $left_B$, $right_A$, $right_B$). The prediction errors of the decoders were then correlated within (*within*: $left_A$ vs. $left_B$, $right_A$ vs. $right_B$) and across hemispheres (*across*: $left_{A/B}$ vs. $right_{A/B}$, respectively). Error correlations were much stronger within than across hemispheres (Fig. 3e,f), suggesting a separate functional network for each hemisphere.

We also found that across-hemisphere error correlations were significantly above zero (Fig. 3e,f, orange), suggesting either connections across hemispheres or a common input to both areas. To address a plausible source of common input, we considered uninstructed eye movements during fixation. Indeed, eye gaze locations around the fixation point correlated with the memorized stimulus (Extended Data Fig. 5a-b) (Gu et al., 2024; Linde-Domingo & Spitzer, 2024; Mostert et al., 2018; Willeke et al., 2019), so an afferent feedback could coordinate small errors to the left or the right of the stimulus in the two hemispheres. Additionally, we can account for brain-wide arousal signals, another type of common input, through taking uninstructed eye movements into consideration (Gu et al., 2024; Johnston et

al., 2022). However, we found that these signals do not confound the observed error correlations, since no correlation between gaze location and neural prediction errors was observed (Extended Data Fig. 5b). Furthermore, even after removing uninstructed eye movements from the decoder analysis (see Methods), we still qualitatively observed the same across-hemisphere prediction error correlations (t-test across: $t = 4.2$, $p = 0.00015$, Extended Data Fig. 5c, d) as reported previously (Fig. 3e, f). Based on these results, we propose that inter-hemispheric interactions could explain the observed error correlations across hemispheres.

In sum, we found evidence that the redundant prefrontal hemispheres are neither fully integrated (Fig. 3d, top) nor independent (Fig. 3d, bottom) but have separate but interacting memory representations (Fig. 3d, middle).

*Tuned inter-area connections led to shared memory diffusion in a WM model.*

To understand the neural dynamics supporting redundant but interacting hemispheres in WM, we implemented the proposed architecture (Fig. 3d, middle) in a computational network model (Fig. 4a). We used the bump-attractor framework, which can model memory diffusion consistent with the observed neural-behavior correlations (Fig. 3a) (Wimmer et al., 2014). Specifically, we modeled each hemisphere as a bump-attractor network of spiking excitatory and inhibitory neurons covering the full visual space. Brief external stimuli were delivered to each ring at the beginning of each trial (Fig. 4b, $S_n$, $S_{n+1}$), allowing both hemispheres to store presented items through tuned persistent firing during the delay (Fig. 4b, $D_n$, $D_{n+1}$).

We studied across-hemispheric interactions by manipulating the strength of excitatory connections between hemispheres, expressed as a percentage of the within-area connectivity (see Methods). Regardless of across-area connectivity strength, all models showed higher memory precision when averaging activity across both areas (system) instead of in a single area (area) (Fig. 4c). This highlights that redundancy between hemispheres improved memory performance.

Interestingly, stronger interhemispheric connections only stabilized the readout of each single area but did not affect system-wide performance (Fig. 4c). This occurred because, on the one hand, stronger across-area connectivity stabilized the network activity in each area, but on the other hand, this led to correlated errors across areas, which led to larger system-wide errors than averaging independent errors. However, while system-wide precision was not affected, memory stability (tuning strength) was enhanced with increased tuned interhemispheric connections (Fig. 4d).

Furthermore, tuned interhemispheric connections (i.e., greater efficacy between neurons with similar stimulus selectivity) were necessary to reproduce the observed across-area error correlations (Fig. 4e) observed in the data (Fig. 3e). Neither uncoupled (Fig. 4e, yellow, *0.0%*) nor 1.0%-coupled but untuned (Fig. 4e, gray, *untuned*) interhemispheric connections led to error correlations across areas. Thus, our model simulations suggest tuned coupling between left and right prefrontal hemispheres in spatial WM.

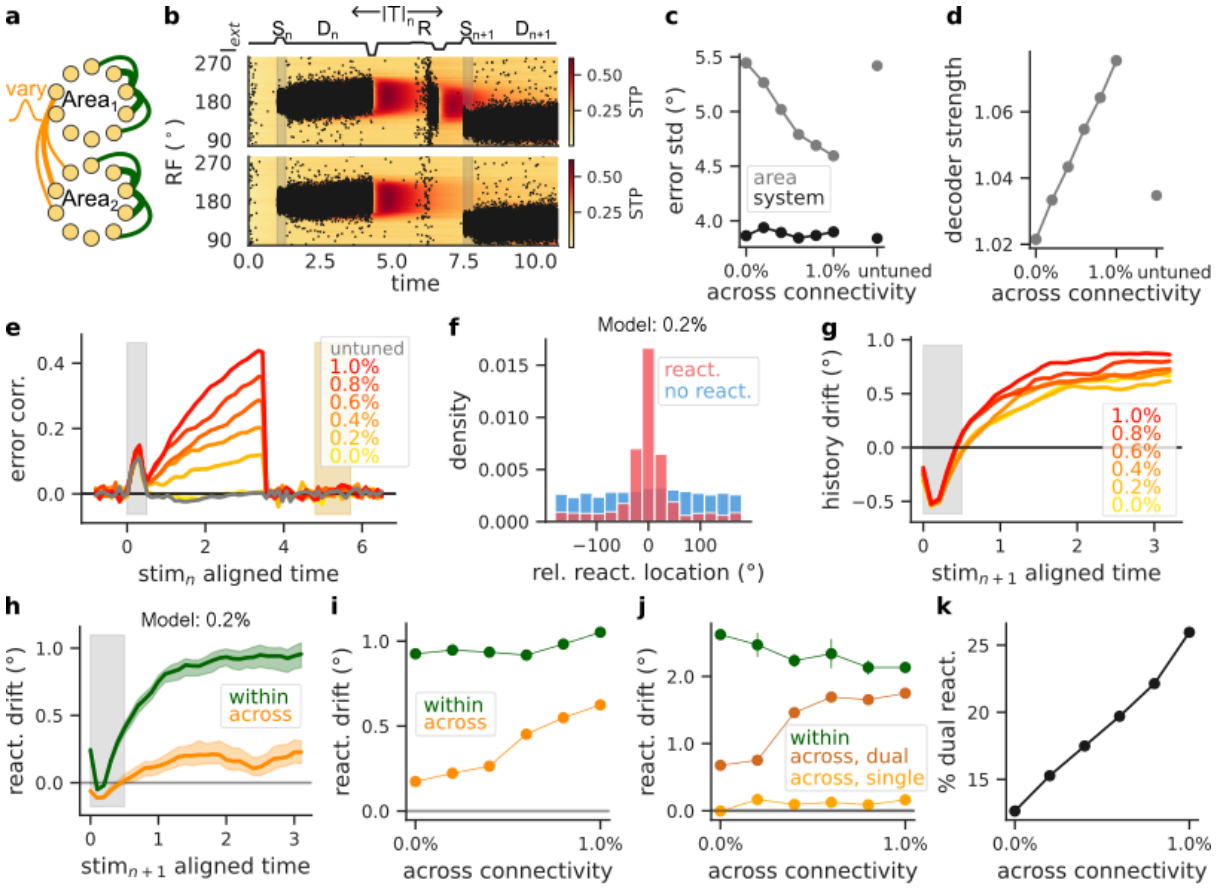


***Figure 4: Across-area connectivity increased coordination of within-trial memory diffusion and across-trial memory drift.*** *a) Model schematic: Each area was modeled as a bump-attractor network with short-term plasticity (STP). The strength and tuning of across-area connectivity varied across different model implementations. b) Example simulation of two consecutive trials, where only Area$_1$ but not Area$_2$ showed a reactivation. Top: Strength of external stimulus (stimulus width = 60°; see details in Methods). 'S' , 'D' and 'R' mark stimulus onset, delay, and reactivation, respectively. Middle: Spiking activity (dots) and STP dynamics (heatmap) for Area$_1$, with neurons sorted by receptive field preference. Bottom: Same for Area$_2$. c) Standard deviation of response errors by model type for the response of a single area (area) or the response computed on the average FR of both areas (system). Area precision improved with increasing tuned across-area connectivity strength while the system-wide precision remained constant at a minimal level. d) Tuned connectivity between models increased decoding strength (bump stability). e) Across-area error correlations for models with different connectivity profiles. Strong, but untuned (untuned, gray) and unconnected (0.0%, yellow) models failed to produce error correlations, while tuned models had increased correlations with increased across-area connectivity strength. f) Reactivations were rare (38% of trials) and tuned to the previous stimulus location. g) History drift (attraction to the previous stimulus location, see Methods) increased through the delay and weakly increased with stronger connectivity. h) Delay activity in each hemisphere drifted stronger towards reactivations within the same hemisphere (green) than the opposite hemisphere (orange) (exemplary for the 0.2% model). i) Increased across-area connectivity increased the amount of shared history drift (final time point of panel h, orange). j) Same as panel i, but split by reactivation types. Across-area attraction to reactivations increased only when reactivations occurred in both areas (dual reactivations, brown) but remained private to each network for single reactivation (orange) trials. k) Percentage of dual reactivations increased with connectivity strength.*

*Increasing inter-area connectivity in a WM model increased shared across-trial history drift.*

We further used this model to investigate how varying the inter-area connectivity affected across-trial memory drift leading to serial dependence in behavior (Fig. 1c). We were able to analyze this in the model, since excitatory connections within each area were facilitating (short-term plasticity, see Methods), leading to increased synaptic efficacy following synaptic activation. This mechanism has been recently proposed to underlie serial dependence in this task (Barbosa et al., 2020; Bliss & D'Esposito, 2017; Kilpatrick, 2018; Stein et al., 2020, 2024).

As done before for a single network (Barbosa et al., 2020; Kilpatrick, 2018; Stein et al., 2020, 2024), we simulated sequential trials separated by an “activity-silent” inter-trial interval (Fig. 4b, $ITI_n$). Towards the end of the inter-trial interval, we applied a nonspecific excitatory current (a “reactivation” current (*R*), Fig. 4b) to both networks, which was strong enough to generate a new bump representation in 38% of trials. Due to the underlying synaptic plasticity, reactivations were on average centered at the location of the previous stimulus (Fig. 4f). These reactivations in turn 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 (Barbosa et al., 2020). We modeled repulsive history effects (adaptation) in lower-level visual areas (Dragoi et al., 2000; Gibson & Radner, 1937; Hajonides et al., 2021; Sheehan & Serences, 2022) by shifting stimulus inputs away from the previous trial’s location (Bliss & D’Esposito, 2017; Stein et al., 2020) (see Methods). We found a system-wide delay-dependent increase in serial dependence from an initial repulsion to a later attraction as found in human experimental data (Bliss et al., 2017; Fritsche et al., 2017; Stein et al., 2020) for all connectivity strengths (Fig. 4g).

To clarify the impact of connectivity strengths on across-trial history drift, we analyzed how the mnemonic activity (Fig. 4b, $D_{n+1}$) in each area drifted towards the previous reactivation (Fig. 4b, R) occurring within the same area or in the other area (across). In the weakest-tuned model, the attraction towards the reactivation within the same area (green) was much stronger during the delay than the attraction to the other area’s reactivation (across, orange, Fig. 4h). Increasing the inter-area connectivity strength increased across-area history drift (Fig. 4i). This effect was driven by trials in which reactivations occurred in both hemispheres equally (*dual* reactivation trials, Fig. 4j, brown). In trials where a reactivation only occurred in a single area (*single* reactivation trials), the history drift remained private to each area across all connectivity strengths (Fig. 4j, orange). Therefore, shared history drift was driven by *dual* reactivations (Fig. 4k).

In sum, we discovered that reactivations, and with them across-trial history drift, were affected by increasing across-area connectivity strengths. Notably, history effects could be transferred across areas only if inter-area connectivity strength was high and reactivations occurred in both areas simultaneously. Based on these insights from the computational model, we could now further specify inter-hemispheric coupling strength in the neural data.

*PFC hemispheres simultaneously encode the same item during memory maintenance but have private reactivations.*

We first confirmed that the neural representations in the monkey PFC exhibited across-trial history drift, reproducing findings of a neural correlate of serial dependence in non-invasive recordings from humans (C. Fischer et al., 2024; Hajonides et al., 2021; Sheehan & Serences, 2022; St John-Saaltink et al., 2016). As predicted by the model, relating decoder prediction errors to the relative distance between the previous and the current target (Fig. 5a,b) revealed a neural correlate of behavioral serial dependence (Fig. 1f). During stimulus presentation, the across-trial history drift was repelled from the previous trial's stimulus (Fig. 5a, blue), which was overcome by an attraction at the end of the delay period (Fig. 5a, red). When fitting the neural history drift by a DoG1 curve similar to in behavior (fits are dashed lines in Fig. 5a; hyperparameter fit in Extended Data Fig 6d-f; see Methods), we observed a smooth change from repulsion into attraction throughout the delay (Fig. 5b). Notably, behavioral serial dependence seemed to scale with neural history drift at the end of the delay period across monkeys (Extended Data Fig. 7a), but more subjects would be necessary to establish a correlation at the population level.

Having demonstrated that neural history drift could be reliably computed from population activity in the monkey prefrontal cortex, we determined whether this drift was private or shared across hemispheres using the drift of memories towards reactivations in the same or opposite hemisphere as in the model. We confirmed that reactivations could be successfully extracted by a Delay decoder (see $fix_{n+1}$ in Fig. 2h) and were tuned to the previous stimulus

location (Fig. 5c) for strongly reactivated trials (Extended Data Fig. 7b; see Methods). When determining the drift of the neural traces towards the previous reactivations (hyperparameter fits in Extended Data Fig 6g-i; see Methods), our analysis revealed an attraction towards reactivations in the same, but not the opposite hemisphere (Fig. 5d). Based on our model simulations, this finding suggests weak interactions between the hemispheres (Fig. 4h,i).

Additionally, we investigated how different reactivation strengths (decoder strength during reactivations) affected the reactivation bias. We split the data into *none*, *single* and *dual* reactivation trials based on a 30% reactivation strength threshold for each hemisphere (Extended Data Fig. 7b) similar to Fig. 4j and observed that in *none* 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 were in-line with a model with very weak inter-hemispheric interactions (Fig. 4j, e.g. 0.2%).

The model further predicted that the observed hemisphere-specific, private reactivation drift, requires most trials to exhibit *single*-hemisphere reactivations (Fig. 4j). To test this, we analyzed the relationship of reactivation decoder strengths across hemispheres. Using a Gaussian Mixture Model (GMM; see Methods), we fitted the data to four hypothesized classes (*none,* $single_{left}$, $single_{right}$, *dual*). Notably, across 250 runs with different initial parameters, the GMM established eight different class means that were all centered off-diagonally (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 (both hemispheres strongly reactivate; diagonal) were rare. When we reduced the GMM to three classes, the model consistently identified the same off-diagonal class means *(none,* $single_{left}$, $single_{right}$; Extended Data Fig. 7 d-f) across all 250 re-runs, suggesting that three (and not four) classes underlied the distribution, with no evidence for a *dual* 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). These analyses suggest that while hemispheres covary in their representation strengths during the delay period, reactivations are predominantly independent and hemisphere-specific.

We conclude that the prefrontal hemispheres interacted through tuned but weak inter-hemispheric coupling, which led 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 strength correlations, Fig. 5f) but perform separately when inputs only triggered events in one hemisphere (reactivations remaining private, Fig. 5d,e).

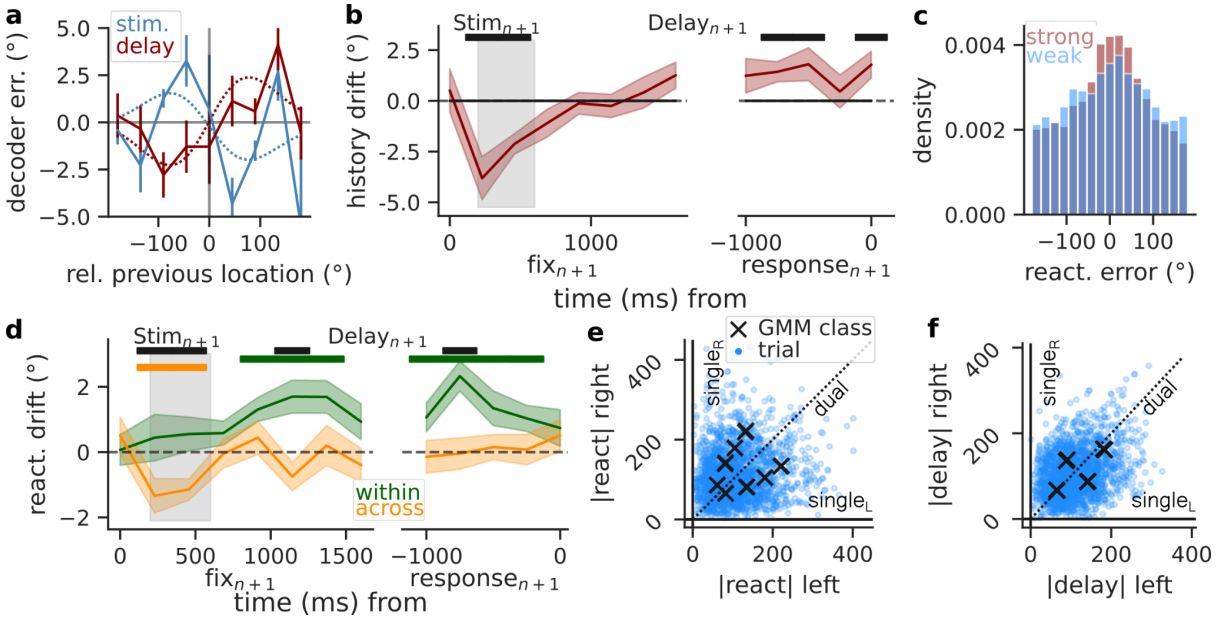



***Figure 5: Private across-trial history drift in monkey PFC indicates weak but specific inter-hemispheric coupling.*** *a) Decoder error plotted over the relative location of the previous stimulus location showed neural repulsion during the current stimulus presentation but attraction at the end of the delay. Dashed lines indicate DoG1 fit. b) Neural history drift (DoG1 fit from panel a) at each time point showed an initial repulsion from the previous stimulus location which continuously grew into an attraction throughout the delay. c) Precision of reactivation location to the previous stimulus location was higher in stronger reactivated trials (cut-off set at 30% of each session's reactivation strength; see Methods). d) Attraction towards the previous reactivation location*

*remained private within each hemisphere (green, t-test: $t = 6.63°$, $p < 10^{-7}$) while remaining at zero across hemispheres (orange, t-test: $t = 0.15°$, $p = 0.88$). e) Distribution of decoder reactivation prediction strengths in the left and right hemispheres. Class means (crosses) were determined using a Gaussian Mixture Model (250 repeats). Notably, clusters did not appear along the diagonal, indicating few dual reactivations. f) Delay decoder prediction strengths did co-fluctuate across hemispheres. Error bars represent +/- 1 s.e.m.; black top bar indicates significance from 0.*

*Reconciling competing models: Weakly connected hemispheres allow for hemispheric specialization in multi-item and redundancy in single-item tasks.*

While we showed that the weakly connected, redundant hemispheric architecture is consistent with memory diffusion and history drift in single-item WM, we did not explain the findings in the literature that favor a specialized architecture. Specifically, findings of reduced memory interference when several items are presented across hemifields rather than in the same hemifield (bilateral field advantage) in behavior and neurons seem to argue for a strong lateralization of the hemispheres (Alvarez & Cavanagh, 2005; Buschman et al., 2011; Delvenne, 2005; Delvenne et al., 2011; Dimond & Beaumont, 1971; Matsushima & Tanaka, 2014; Umemoto et al., 2010).

However, even in the proposed redundant architecture, we found that some neural code (reactivations) remained private to each hemisphere, suggesting that the prefrontal hemispheres were capable of maintaining separate memory representations, provided they were independently activated. We reasoned that the lateralization of inputs to the PFC from upstream areas (Pasternak et al., 2015) could lead to the preferential activation of each hemisphere for contralateral stimuli. This stimulus input lateralization could lead to a hemispheric specialization when multiple, competing items are presented (Umemoto et al., 2010; Y. Zhang et al., 2018). We tested this in our network model to analyze if the seemingly contradicting findings could be unified through the redundant network model.

We simulated a simple multi-item WM task in the model: Two stimuli were simultaneously presented to the network and were read out at the end of the delay. The only change to the network from Fig. 4 was that each area was assigned a hemisphere ($Area_1$ = left, $Area_2$ = right) and received a 20% weaker input stimulus when stimuli were presented ipsilaterally rather than contralaterally. We presented equidistant stimuli either across (Fig. 6a) or within hemifields simultaneously (Fig. 6b). In across-hemifield trials, each area stored its stronger contralateral input, leading to both items being remembered (Fig. 6c, example trial).

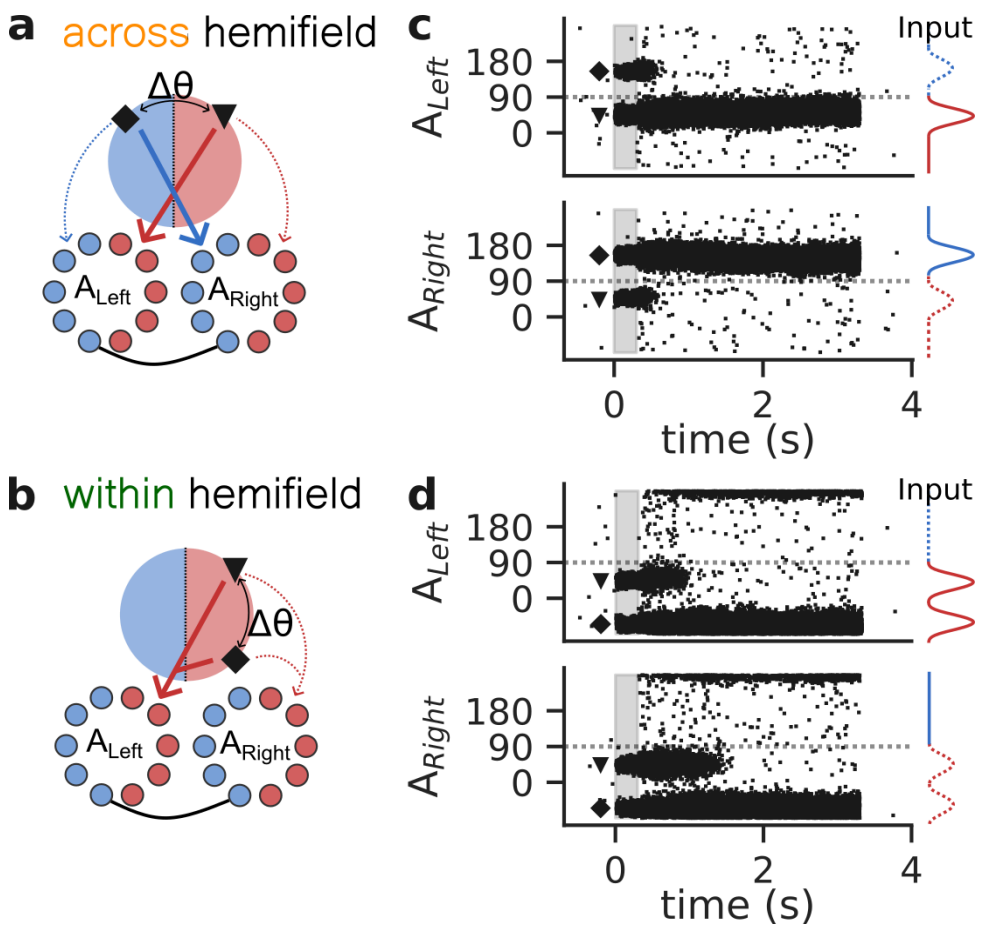



***Figure 6: Stimulus lateralization in a redundant network model led to hemispheric specialization in a multi-item task.*** *a) The same network model as previously (Fig. 4, 0.2% connected) was simulated for two simultaneously presented items across hemifields (stimuli only vary in location, shapes only for illustration purposes). Stimuli were projected stronger into the contralateral hemisphere (thickness of arrows). b) Same as a) for two stimuli presented within the same hemifield. Since both stimuli were presented in the same hemifield, the same area was contralateral (ipsilateral) for both items, and both items were therefore presented with equal*

*strength. c) Example trial for stimuli presented across hemifields (top = left, bottom = right hemisphere). Triangle and diamond markers indicate the stimulus locations (see a); dashed horizontal lines indicate hemifield separation (at 90°); Gaussians on the right represent the stimulus strength (height) by hemifield (color). Both areas retained their respective contralateral item until the end of the delay ($A_{Left}$ retains triangle, $A_{Right}$ retains diamond). d) Same as c) but with stimuli presented within the same hemifield. Both stimuli were presented with equal strength to each area. Both areas retained the same item (diamond), causing a forgetting error of the other item (triangle).*

However, in within-hemifield trials (Fig. 6d), both stimuli were contralateral (ipsilateral) to the same hemisphere, and therefore both stimuli impinged equally with 100% (80%) stimulus strength in each area. This led to a competition between items, since the capacity limit of each area was one item and therefore caused each hemisphere to randomly forget one item. This could result in the same item being forgotten in both areas (example trial, Fig. 6d) and therefore to a lower system-wide WM capacity when targets were presented within the same rather than across hemifields (capacity: within = 1.75, across = 2; percent correct of one randomly asked item: within = 75%, across = 100%).

Therefore, the proposed redundant, weakly connected architecture could not only account for findings of diffusion and history effects in single-item WM but could also replicate the reported bilateral field advantage observed in multi-item WM tasks (Alvarez & Cavanagh, 2005; Buschman et al., 2011; Delvenne, 2005; Delvenne et al., 2011; Dimond & Beaumont, 1971; Matsushima & Tanaka, 2014; Umemoto et al., 2010).

## Discussion

In this study, we explored the prefrontal inter-hemispheric architecture that supports spatial WM. We showed that while mnemonic spatial representations exhibited a contralateral bias in the 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). 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 allowed for shared memory diffusion (Fig. 4e) but independent history effects such as reactivations and drift (Fig. 4f-k), consistent with the experimental data (Fig. 5). Finally, we demonstrate that this architecture can represent WM items both redundantly and privately, depending on WM demands (Fig 6).

Importantly, this study addresses a longstanding debate about interhemispheric WM organization. Multi-item WM studies demonstrate a bilateral field advantage, where memories for items presented within the same hemifield have reduced precision (Alvarez & Cavanagh, 2005; Buschman et al., 2011; Delvenne, 2005; Delvenne et al., 2011; Dimond & Beaumont, 1971; Matsushima & Tanaka, 2014; Umemoto et al., 2010) and reduced encoding in the neural activity (Buschman et al., 2011; Matsushima & Tanaka, 2014), while memories across hemispheres do not. This suggests independent contralateral storage of each hemifield (Fig. 1a). In contrast, optogenetic studies in mice indicate that the disruption of *both* hemispheres is necessary to impair behavior, implying mirrored memory representations (G. Chen et al., 2021; Li et al., 2016; Svoboda & Li, 2018). Our model reconciles these findings by positing a redundant, weakly connected network with input lateralization. This network mirrors representations across hemispheres if a single item is presented (Fig. 4), but lateralizes representations when multiple items compete for lateralized resources, replicating the bilateral field advantage in multi-item WM (Fig. 6), consistent with available behavioral and neural data for this phenomenon (Alvarez & Cavanagh, 2005; Buschman et al., 2011; Delvenne, 2005; Delvenne et al., 2011; Dimond & Beaumont, 1971; Matsushima & Tanaka, 2014; Umemoto et al., 2010). Crucially, this flexibility arises naturally from the weak lateral bias of inputs from upstream areas into a network with fixed architecture, without the need for abrupt changes in connectivity for different task demands as suggested by previous studies (Matsushima & Tanaka, 2014; Pinotsis et al., 2019).

Remarkably, the redundant yet weakly coupled prefrontal architecture offers distinct advantages tied to specific WM demands. When demands are low (below capacity, here: single-item tasks), redundancy of memories across hemispheres boosts WM precision (Fig. 4c) and stability (Fig. 4d). However, when demands are high (above capacity, here: multi-item WM), the risk of catastrophic forgetting outweighs small precision errors. The network can adapt to this and favors hemisphere-specific representations enabled by stimulus lateralization and low interhemispheric connectivity (Fig. 6). Taken together, this architecture flexibly adapts to varying cognitive demands with a fixed, parsimonious architecture.

As it stands, our model cannot yet explain recent findings of memory transfer across hemispheres during mental rotations across the vertical meridian (Brincat et al., 2021). In our framework, mental rotations would occur independently in the representations held in each hemisphere, without a transfer of information between hemispheres. However, we speculate that including the known contralateral preference of PFC neural memory fields (Funahashi et al., 1989) in our model could explain the findings of memory transfer. Specifically, the rotation of the memory within each hemisphere would increase (decrease) firing rates and code accuracy in the post-rotation contralateral (ipsilateral) hemisphere, mimicking experimental results (Brincat et al., 2021) without the need for a direct memory transfer. Future studies should test this prediction to validate and refine the model.

The model contains weak interhemispheric connections, based on the weak but significant encoding error correlations observed in the experimental data (Fig. 3e). While we account for uninstructed eye movements as a potential source of common input (Extended Data Fig. 5), other uninstructed movements can also elicit shared neural variability (Talluri et al., 2023). However, our analysis concerns shared neural information underlying memory imprecisions, a very specific correlation that can only be conceivably induced by movements that cause the consistent rotation of mnemonic representations in the two hemispheres. In this respect, we argue that eye gaze is the main confound to address, as it has been shown

to change spatial encoding during the delay period (Gu et al., 2024; Linde-Domingo & Spitzer, 2024; Liu et al., 2023; Mostert et al., 2018; Willeke et al., 2019; Willett & Mayo, 2023). We therefore argue that the identified correlations point towards a task-related communication subspace between the prefrontal hemispheres, which could be realized through tuned inter-hemispheric connections anatomically supported by the corpus callosum.

While our analyses support the interpretation of weak inter-hemispheric connections, a recent study suggests stronger inter-hemispheric connections. Specifically, it was shown that after optogenetic disruption of the neural activity of one hemisphere in the mouse ALM, stimulus-tuned activity reemerges (G. Chen et al., 2021; Inagaki et al., 2019; Li et al., 2016). This effect is due to inter-hemispheric interactions, since no reemergence occurs when both hemispheres are silenced or the corpus callosum is cut (Li et al., 2016). In our current simulations, weak inter-hemispheric connections are not sufficient to allow for memory transfer if one area is silenced (simulations not shown). However, it has been proposed that during the task, the frontal cortex has an increased background activity relayed from other areas (Inagaki et al., 2019). This could allow for a quicker memory transfer during the task than in idle periods (reactivations), thus allowing the recovery of memory transfer across hemispheres with weak connections that don’t cause dual reactivations.

Another possible explanation for this discrepancy is that while the optogenetic disruptions are performed in a premotor area in the mouse (G. Chen et al., 2021; Inagaki et al., 2019; Li et al., 2016), we focus on the prefrontal area 8Ar in the monkey. Premotor areas have more and stronger callosal connections than prefrontal areas (Schwartz, 1995), which could also explain a stronger connectivity underlying information transfer. It would therefore be insightful to study the lateralization of WM along the callosal hierarchy and test whether a relative hemispheric disconnection characterizes association in comparison with sensory and (pre)motor cortices. This would contribute a lateral dimension not yet considered in studies that have emphasized the role of distributed brain processing in WM maintenance

(Christophel et al., 2017; Dotson et al., 2018; Edin et al., 2009; Froudist-Walsh et al., 2020; Mejías & Wang, 2022), serial dependence (Tschiersch, 2022), and multi-item WM (Buschman et al., 2011; Gupta & Sridharan, 2023; Panichello & Buschman, 2021).

Finally, our findings can have implications for several neuropsychiatric disorders, such as schizophrenia and autism, which show a reduction in callosal connectivity (Arnone et al., 2008; Patel et al., 2011) and deficits in WM performance (Forbes et al., 2009; Stein et al., 2020, 2021). Based on our model, we would predict that the bilateral field advantage as a percentage of within- vs. across-hemifield memory should remain or even be increased in patients due to reduced callosal connections. On the other hand, the lack of callosal connections could lead to a decrease in WM precision in single-item tasks since the prefrontal hemispheres cannot stabilize each other's memory representations.

In conclusion, our study bridges paradoxical findings from single- and multi-item WM 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 WM.

## Acknowledgements

We are grateful to Megan McDonnell for helpful advice and discussion. 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


Almeida, R., Barbosa, J., & Compte, A. (2015). Neural circuit basis of visuo-spatial working memory precision: A computational and behavioral study. *Journal of Neurophysiology*, *114*(3), 1806–1818. https://doi.org/10.1152/jn.00362.2015

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643. https://doi.org/10.1111/j.1467-9280.2005.01587.x

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, *27*(2), 77–87. https://doi.org/10.1007/BF00337259

Ardid, S., Wang, X.-J., & Compte, A. (2007). An Integrated Microcircuit Model of Attentional Processing in the Neocortex. *Journal of Neuroscience*, *27*(32), 8486–8495. https://doi.org/10.1523/JNEUROSCI.1145-07.2007

Arnone, D., McIntosh, A. M., Tan, G. M. Y., & Ebmeier, K. P. (2008). Meta-analysis of magnetic resonance imaging studies of the corpus callosum in schizophrenia. *Schizophrenia Research*, *101*(1), 124–132. https://doi.org/10.1016/j.schres.2008.01.005

Barbosa, J., Babushkin, V., Temudo, A., Sreenivasan, K. K., & Compte, A. (2021). Across-Area Synchronization Supports Feature Integration in a Biophysical Network Model of Working Memory. *Frontiers in Neural Circuits*, *15*. https://doi.org/10.3389/fncir.2021.716965

Barbosa, J., & Compte, A. (2020). Build-up of serial dependence in color working memory. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-67861-2

Barbosa, J., Stein, H., Martinez, R. L., Galan-Gadea, A., Li, S., Dalmau, J., Adam, K. C. S., Valls-Solé, J., Constantinidis, C., & Compte, A. (2020). Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. *Nature Neuroscience*, *23*(8), Article 8. https://doi.org/10.1038/s41593-020-0644-4

Bliss, D. P., & D'Esposito, M. (2017). Synaptic augmentation in a cortical circuit model reproduces serial dependence in visual working memory. *PLoS ONE*, *12*(12), e0188927. https://doi.org/10.1371/journal.pone.0188927

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-15199-7

Bouchacourt, F., & Buschman, T. J. (2019). A Flexible Model of Working Memory. *Neuron*, *103*(1), 147-160.e8. https://doi.org/10.1016/j.neuron.2019.04.020

Brincat, S. L., Donoghue, J. A., Mahnke, M. K., Kornblith, S., Lundqvist, M., & Miller, E. K. (2021). Interhemispheric transfer of working memories. *Neuron*, *109*(6), 1055-1066.e4. https://doi.org/10.1016/j.neuron.2021.01.016

Buschman, T. J. (2021). Balancing Flexibility and Interference in Working Memory. *Annual Review of Vision Science*, *7*, 367–388. https://doi.org/10.1146/annurev-vision-100419-104831

Buschman, T. J., Siegel, M., Roy, J. E., & Miller, E. K. (2011). Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of Sciences*, *108*(27), 11252–11255. https://doi.org/10.1073/pnas.1104666108

Calangiu, I., Kollmorgen, S., Reppas, J., & Mante, V. (2022). *Primate pre-arcuate cortex actively maintains persistent representations of saccades from plans to outcomes* [Preprint]. Neuroscience. https://doi.org/10.1101/2022.09.26.509463

Chen, G., Kang, B., Lindsey, J., Druckmann, S., & Li, N. (2021). Modularity and robustness of frontal cortical networks. *Cell*, *184*(14), 3717-3730.e24. https://doi.org/10.1016/j.cell.2021.05.026

Chen, J., Zhang, C., Hu, P., Min, B., & Wang, L. (2024). Flexible control of sequence working memory in the macaque frontal cortex. *Neuron*, *112*(20), 3502-3514.e6. https://doi.org/10.1016/j.neuron.2024.07.024

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, *21*(2),

111–124. https://doi.org/10.1016/j.tics.2016.12.007

Cicchini, G. M., Mikellidou, K., & Burr, D. C. (2018). The functional role of serial dependence. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1890), 20181722. https://doi.org/10.1098/rspb.2018.1722

Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X.-J. (2000). Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model. *Cerebral Cortex*, *10*(9), 910–923. https://doi.org/10.1093/cercor/10.9.910

Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X.-L., Wang, M., & Arnsten, A. F. T. (2018). Persistent Spiking Activity Underlies Working Memory. *Journal of Neuroscience*, *38*(32), 7020–7028. https://doi.org/10.1523/JNEUROSCI.2486-17.2018

de Azevedo Neto, R. M., & Bartels, A. (2021). Disrupting Short-Term Memory Maintenance in Premotor Cortex Affects Serial Dependence in Visuomotor Integration. *Journal of Neuroscience*, *41*(45), 9392–9402. https://doi.org/10.1523/JNEUROSCI.0380-21.2021

Delvenne, J.-F. (2005). The capacity of visual short-term memory within and between hemifields. *Cognition*, *96*(3), B79–B88. https://doi.org/10.1016/j.cognition.2004.12.007

Delvenne, J.-F., Castronovo, J., Demeyere, N., & Humphreys, G. W. (2011). Bilateral Field Advantage in Visual Enumeration. *PLoS ONE*, *6*(3), e17743. https://doi.org/10.1371/journal.pone.0017743

Dimond, S., & Beaumont, G. (1971). Use of two cerebral hemispheres to increase brain capacity. *Nature*, *232*(5308), 270–271. https://doi.org/10.1038/232270a0

Dotson, N. M., Hoffman, S. J., Goodell, B., & Gray, C. M. (2018). Feature-Based Visual Short-Term Memory Is Widely Distributed and Hierarchically Organized. *Neuron*, *99*(1), 215-226.e4. https://doi.org/10.1016/j.neuron.2018.05.026

Dragoi, V., Sharma, J., & Sur, M. (2000). Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron*, *28*(1), 287–298.

https://doi.org/10.1016/s0896-6273(00)00103-3

Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegnér, J., & Compte, A. (2009). Mechanism for top-down control of working memory capacity. *Proceedings of the National Academy of Sciences*, *106*(16), 6802–6807. https://doi.org/10.1073/pnas.0901894106

Fischer, C., Kaiser, J., & Bledowski, C. (2024). A direct neural signature of serial dependence in working memory. *eLife*, *13*. https://doi.org/10.7554/eLife.99478.1

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), Article 5. https://doi.org/10.1038/nn.3689

Forbes, N. F., Carrick, L. A., McIntosh, A. M., & Lawrie, S. M. (2009). Working memory in schizophrenia: A meta-analysis. *Psychological Medicine*, *39*(6), 889–905. https://doi.org/10.1017/S0033291708004558

Fornaciai, M., & Park, J. (2018). Attractive Serial Dependence in the Absence of an Explicit Task. *Psychological Science*, *29*(3), 437–446. https://doi.org/10.1177/0956797617737385

Fornaciai, M., & Park, J. (2019). Serial dependence generalizes across different stimulus formats, but not different sensory modalities. *Vision Research*, *160*, 108–115. https://doi.org/10.1016/j.visres.2019.04.011

Fritsche, M., & de Lange, F. P. (2019). The role of feature-based attention in visual serial dependence. *Journal of Vision*, *19*(13), 21. https://doi.org/10.1167/19.13.21

Fritsche, M., Mostert, P., & Lange, F. P. de. (2017). Opposite Effects of Recent History on Perception and Decision. *Current Biology*, *27*(4), 590–595. https://doi.org/10.1016/j.cub.2017.01.006

Froudist-Walsh, S., Bliss, D. P., Ding, X., Jankovic-Rapan, L., Niu, M., Knoblauch, K., Zilles, K., Kennedy, H., Palomero-Gallagher, N., & Wang, X.-J. (2020). *A dopamine gradient controls access to distributed working memory in monkey cortex* (p. 2020.09.07.286500). bioRxiv. https://doi.org/10.1101/2020.09.07.286500

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual

space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61*(2), 331–349. https://doi.org/10.1152/jn.1989.61.2.331

Gerstner, W., Kistler, W. M., Naud, R., & Paninski, L. (2014). *Neuronal Dynamics—From single neurons to networks and models of cognition and beyond*. https://neuronaldynamics.epfl.ch/

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology*, *20*(5), 453–467. https://doi.org/10.1037/h0059826

Gu, Q., Zhang, Q., Han, Y., Li, P., Gao, Z., & Shen, M. (2024). Microsaccades reflect attention shifts: A mini review of 20 years of microsaccade research. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1364939

Gupta, S., & Sridharan, D. (2023). *Distractors induce space-specific neural biases in visual working memory* (p. 2023.12.24.573161). bioRxiv. https://doi.org/10.1101/2023.12.24.573161

Hajonides, J. E., van Ede, F., Stokes, M. G., Nobre, A. C., & Myers, N. E. (2021). *Multiple and Dissociable Effects of Sensory History on Working-Memory Performance* (p. 2021.10.31.466639). bioRxiv. https://doi.org/10.1101/2021.10.31.466639

Hikosaka, O., & Wurtz, R. H. (1983). Visual and oculomotor functions of monkey substantia nigra pars reticulata. IV. Relation of substantia nigra to superior colliculus. *Journal of Neurophysiology*, *49*(5), 1285–1301. https://doi.org/10.1152/jn.1983.49.5.1285

Hubel, D. H., & Wiesel, T. N. (1967). Cortical and callosal connections concerned with the vertical meridian of visual fields in the cat. *Journal of Neurophysiology*, *30*(6), 1561–1573. https://doi.org/10.1152/jn.1967.30.6.1561

Inagaki, H. K., Fontolan, L., Romani, S., & Svoboda, K. (2019). Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature*, *566*(7743), 212–217. https://doi.org/10.1038/s41586-019-0919-7

Johnston, R., Snyder, A. C., Khanna, S. B., Issar, D., & Smith, M. A. (2022). The eyes reflect an internal cognitive state hidden in the population activity of cortical neurons.

*Cerebral Cortex*, *32*(15), 3331–3346. https://doi.org/10.1093/cercor/bhab418

Kilpatrick, Z. P. (2018). Synaptic mechanisms of interference in working memory. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-25958-9

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial Dependence across Perception, Attention, and Memory. *Trends in Cognitive Sciences*, *21*(7), 493–497. https://doi.org/10.1016/j.tics.2017.04.011

Li, N., Daie, K., Svoboda, K., & Druckmann, S. (2016). Robust neuronal dynamics in premotor cortex during motor planning. *Nature*, *532*(7600), Article 7600. https://doi.org/10.1038/nature17643

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial Dependence in the Perception of Faces. *Current Biology : CB*, *24*(21), 2569–2574. https://doi.org/10.1016/j.cub.2014.09.025

Linde-Domingo, J., & Spitzer, B. (2024). Geometry of visuospatial working memory information in miniature gaze patterns. *Nature Human Behaviour*, *8*(2), 336–348. https://doi.org/10.1038/s41562-023-01737-z

Liu, B., Nobre, A. C., & van Ede, F. (2023). Microsaccades transiently lateralise EEG alpha activity. *Progress in Neurobiology*, *224*, 102433. https://doi.org/10.1016/j.pneurobio.2023.102433

Luo, M., Zhang, H., & Luo, H. (2024). *Reactivated past decisions repel early sensory processing and attract late decision-making* (p. 2024.02.26.582221). bioRxiv. https://doi.org/10.1101/2024.02.26.582221

Manassi, M., & Whitney, D. (2024). Continuity fields enhance visual perception through positive serial dependence. *Nature Reviews Psychology*, *3*(5), 352–366. https://doi.org/10.1038/s44159-024-00297-x

Markov, Y. A., Tiurina, N. A., & Pascucci, D. (2024). Serial dependence: A matter of memory load. *Heliyon*, *10*(13), e33977. https://doi.org/10.1016/j.heliyon.2024.e33977

Matsushima, A., & Tanaka, M. (2014). Different Neuronal Computations of Spatial Working Memory for Multiple Locations within versus across Visual Hemifields. *The Journal of*

*Neuroscience*, *34*(16), 5621. https://doi.org/10.1523/JNEUROSCI.0295-14.2014

Mejías, J. F., & Wang, X.-J. (2022). Mechanisms of distributed working memory in a large-scale network of macaque neocortex. *eLife*, *11*, e72136. https://doi.org/10.7554/eLife.72136

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science (New York, N.Y.)*, *319*(5869), 1543–1546. https://doi.org/10.1126/science.1150769

Mostert, P., Albers, A. M., Brinkman, L., Todorova, L., Kok, P., & Lange, F. P. de. (2018). Eye Movement-Related Confounds in Neural Decoding of Visual Working Memory Representations. *eNeuro*, *5*(4). https://doi.org/10.1523/ENEURO.0401-17.2018

Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, *592*(7855), 601. https://doi.org/10.1038/s41586-021-03390-w

Papadimitriou, C., Ferdoash, A., & Snyder, L. H. (2015). Ghosts in the machine: Memory interference from the previous trial. *Journal of Neurophysiology*, *113*(2), 567–577. https://doi.org/10.1152/jn.00402.2014

Papadimitriou, C., White, R. L., & Snyder, L. H. (2017). Ghosts in the Machine II: Neural Correlates of Memory Interference from the Previous Trial. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(4), 2513–2527. https://doi.org/10.1093/cercor/bhw106

Pasternak, T., Lui, L. L., & Spinelli, P. M. (2015). Unilateral Prefrontal Lesions Impair Memory-Guided Comparisons of Contralateral Visual Motion. *The Journal of Neuroscience*, *35*(18), 7095–7105. https://doi.org/10.1523/JNEUROSCI.5265-14.2015

Patel, S., Mahon, K., Wellington, R., Zhang, J., Chaplin, W., & Szeszko, P. R. (2011). A meta-analysis of diffusion tensor imaging studies of the corpus callosum in schizophrenia. *Schizophrenia Research*, *129*(2), 149–155. https://doi.org/10.1016/j.schres.2011.03.014

Peiker, C., Wunderle, T., Eriksson, D., Schmidt, A., & Schmidt, K. E. (2013). An Updated Midline Rule: Visual Callosal Connections Anticipate Shape and Motion in Ongoing

Activity across the Hemispheres. *Journal of Neuroscience*, *33*(46), 18036–18046. https://doi.org/10.1523/JNEUROSCI.1181-13.2013

Pinotsis, D. A., Buschman, T. J., & Miller, E. K. (2019). Working Memory Load Modulates Neuronal Coupling. *Cerebral Cortex*, *29*(4), 1670–1681. https://doi.org/10.1093/cercor/bhy065

Ploner, C. J., Gaymard, B., Rivaud, S., Agid, Y., & Pierrot-Deseilligny, C. (1998). Temporal limits of spatial working memory in humans. *European Journal of Neuroscience*, *10*(2), 794–797. https://doi.org/10.1046/j.1460-9568.1998.00101.x

Rainer, G., Asaad, W. F., & Miller, E. K. (1998). Memory fields of neurons in the primate prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(25), 15008–15013. https://doi.org/10.1073/pnas.95.25.15008

Ranieri, G., Benedetto, A., Ho, H. T., Burr, D. C., & Morrone, M. C. (2022). Evidence of Serial Dependence from Decoding of Visual Evoked Potentials. *The Journal of Neuroscience*, JN-RM-1879-21. https://doi.org/10.1523/JNEUROSCI.1879-21.2022

Schmidt, K. E., Lomber, S. G., & Innocenti, G. M. (2010). Specificity of Neuronal Responses in Primary Visual Cortex Is Modulated by Interhemispheric CorticoCortical Input. *Cerebral Cortex*, *20*(12), 2776–2786. https://doi.org/10.1093/cercor/bhq024

Schwartz, M. L. (1995). Organization and Development of Interhemispheric Connections of the Prefrontal Cortex in Rhesus Monkey. In A. G. Reeves & D. W. Roberts (Eds.), *Epilepsy and the Corpus Callosum 2* (pp. 29–47). Springer US. https://doi.org/10.1007/978-1-4899-1427-9_4

Sheehan, T. C., & Serences, J. T. (2022). Attractive serial dependence overcomes repulsive neuronal adaptation. *PLOS Biology*, *20*(9), e3001711. https://doi.org/10.1371/journal.pbio.3001711

St John-Saaltink, E., Kok, P., Lau, H. C., & de Lange, F. P. (2016). Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(23), 6186–6192. https://doi.org/10.1523/JNEUROSCI.4390-15.2016

Stein, H., Barbosa, J., & Compte, A. (2021). Towards biologically constrained attractor models of schizophrenia. *Current Opinion in Neurobiology*, *70*, 171–181. https://doi.org/10.1016/j.conb.2021.10.013

Stein, H., Barbosa, J., Rosa-Justicia, M., Prades, L., Morató, A., Galan-Gadea, A., Ariño, H., Martinez-Hernandez, E., Castro-Fornieles, J., Dalmau, J., & Compte, A. (2020). Reduced serial dependence suggests deficits in synaptic potentiation in anti-NMDAR encephalitis and schizophrenia. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-18033-3

Stein, H., Barbosa, J., Soldevilla, D. L., Rosa-Justicia, M., Morató, A., Galan-Gadea, A., Prades, L., Muñoz-Lopetegui, A., Ariño, H., Martinez-Hernandez, E., Guasp, M., Castro-Fornieles, J., Dalmau, J., Santamaria, J., & Compte, A. (2024). *Neural signatures of reduced serial dependence in anti-NMDAR encephalitis and schizophrenia*. OSF. https://doi.org/10.31234/osf.io/kz73s

Stimberg, M., Brette, R., & Goodman, D. F. (2019). Brian 2, an intuitive and efficient neural simulator. *eLife*, *8*, e47314. https://doi.org/10.7554/eLife.47314

Svoboda, K., & Li, N. (2018). Neural mechanisms of movement planning: Motor cortex and beyond. *Current Opinion in Neurobiology*, *49*, 33–41. https://doi.org/10.1016/j.conb.2017.10.023

Talluri, B. C., Kang, I., Lazere, A., Quinn, K. R., Kaliss, N., Yates, J. L., Butts, D. A., & Nienborg, H. (2023). Activity in primate visual cortex is minimally driven by spontaneous movements. *Nature Neuroscience*, *26*(11), 1953–1959. https://doi.org/10.1038/s41593-023-01459-5

Tschiersch, M. (2022). Causal evidence for the higher-order origin of serial dependence suggests a multi-area account. *Journal of Neurophysiology*, *128*(2), 336–338. https://doi.org/10.1152/jn.00048.2022

Umakantha, A., Williamson, R. C., Yu, B., & Smith, M. A. (2020). Slow coordinated activity across hemispheres in PFC predicts pupil size. *Computational and Systems Neuroscience (COSYNE) 2020*, 123–124.

Umemoto, A., Drew, T., Ester, E. F., & Awh, E. (2010). A bilateral advantage for storage in visual working memory. *Cognition*, *117*(1), 69–79. https://doi.org/10.1016/j.cognition.2010.07.001

Vallat, R. (2018). Pingouin: Statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026

Wei, Z., Wang, X.-J., & Wang, D.-H. (2012). From Distributed Resources to Limited Slots in Multiple-Item Working Memory: A Spiking Network Model with Normalization. *Journal of Neuroscience*, *32*(33), 11228–11240. https://doi.org/10.1523/JNEUROSCI.0735-12.2012

White, J. M., Sparks, D. L., & Stanford, T. R. (1994). Saccades to remembered target locations: An analysis of systematic and variable errors. *Vision Research*, *34*(1), 79–92. https://doi.org/10.1016/0042-6989(94)90259-3

Willeke, K. F., Tian, X., Buonocore, A., Bellet, J., Ramirez-Cardenas, A., & Hafed, Z. M. (2019). Memory-guided microsaccades. *Nature Communications*, *10*(1), 3710. https://doi.org/10.1038/s41467-019-11711-x

Willett, S. M., & Mayo, J. P. (2023). Microsaccades are directed toward the midpoint between targets in a variably cued attention task. *Proceedings of the National Academy of Sciences*, *120*(20), e2220552120. https://doi.org/10.1073/pnas.2220552120

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons. *Biophysical Journal*, *12*(1), 1–24.

Wimmer, K., Nykamp, D. Q., Constantinidis, C., & Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*, *17*(3), Article 3. https://doi.org/10.1038/nn.3645

Xie, Y., Hu, P., Li, J., Chen, J., Song, W., Wang, X.-J., Yang, T., Dehaene, S., Tang, S., Min, B., & Wang, L. (2022). Geometry of sequence working memory in macaque prefrontal cortex. *Science*, *375*(6581), 632–639. https://doi.org/10.1126/science.abm0204

Zhang, H., & Luo, H. (2023). Feature-specific reactivations of past information shift current

neural encoding thereby mediating serial bias behaviors. *PLOS Biology*, *21*(3), e3002056. https://doi.org/10.1371/journal.pbio.3002056

Zhang, Y., Ye, C., Roberson, D., Zhao, G., Xue, C., & Liu, Q. (2018). The bilateral field advantage effect in memory precision. *Quarterly Journal of Experimental Psychology (2006)*, *71*(3), 749–758. https://doi.org/10.1080/17470218.2016.1276943

## Methods

### Behavioral task and recordings

Three male adult rhesus monkeys (Macaca mulatta) performed an oculo-motor delayed response task (Hikosaka & Wurtz, 1983), 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 (3453 trials), 12 (5666 trials), 22 (9065 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 (Umakantha et al., 2020). 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 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 $\theta_r$ was considered for further analysis. In each trial, the angular error ($\theta_e$) was computed as the circular distance between the stimulus angle ($\theta_s$) and the response ($\theta_r$). We subtracted the mean of $\theta_e$ 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($\theta_e$) ~ delay) across sessions using the *statsmodels.formula.api* Python package.

### Serial dependence plots

To visualize serial dependence, we plotted the current trial's error ($\theta_e^{curr}$) 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 $\theta_e^{curr}$ with the $sign(\Delta_{prev,curr})$ and plotting it over $abs(\Delta_{prev,curr})$ (Barbosa & Compte, 2020).

### Serial dependence strength

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) (Stein et al., 2020):

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

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

For each monkey, we determined $F(x) \ = \ f'_{normalized}(\sigma_{opt},\ x)$ , where $\sigma_{opt}$ 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 (Wimmer et al., 2014). In complex-number notation:

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

The preferred orientation of each multi-unit $\theta_{pref}$ is given by the angle of the complex-valued $MF$ ($\theta_{pref} \ = \ arg(MF)$) 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, 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 $\theta_{pref}$. To avoid the circularity of generating a bump during the delay solely by sorting by $\theta_{pref}$, we estimated $\theta_{pref}$ 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 $\beta_t$ 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 $\beta_t$ 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):

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

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

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

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 $\theta_s$. To compute a decoding score, we z-scored the $MSE$ against the shuffled $MSE$s ($MSE_{shuffle}$) and provided the output as a measure of standard deviation σ. The decoding score is thus obtained as:

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

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 (Barbosa et al., 2020; Stein et al., 2024), 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 $\theta_e$ 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 of the multi-units in each hemisphere (independent analysis, left vs. right hemisphere, Fig. 3b, 3c). The obtained prediction errors (combined, left, right) were each circularly correlated 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. 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 (*circ_corrcc* from *pingouin* package (version 0.4.0) (Vallat, 2018); uniform correction) 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. We computed the decoder from the gaze in the same manner as from the multi-units (see section *Population decoding of stimulus location*) but replacing the neural matrix X by the complex eye locations. 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 tested a possible role of gaze location in driving decoder-decoder error correlations within/across hemispheres (see above): We added the complex 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 predictions 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).

**Neural history drift**

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

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

**Reactivations**

Fig. 2h shows that reactivations occur immediately before stimulus presentation. We therefore estimated the location and strength of reactivations in individual trials using a Single-trial 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**

This analysis was performed 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 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 in the opposite 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: *none,* $single_{left}$, $single_{right}$ 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.

**Bump-attractor model**

We adapted the previously proposed bump attractor model implementation from (Barbosa et al., 2020) based on (Gerstner et al., 2014) in Brian2 (Stimberg et al., 2019). 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 (Barbosa et al., 2020), 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 (Barbosa et al., 2020), following the dynamics of facilitation ($u$) and depression ($x$) proposed by (Mongillo et al., 2008):

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

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

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 (Barbosa et al., 2020).

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 (Bliss & D'Esposito, 2017; Stein et al., 2020).

*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 untuned 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 (Barbosa et al., 2021). 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 ($\theta_{bump}$) from each area of the model was determined by computing the circular mean of the excitatory neurons' preferred direction ($\theta_n$) weighted by their firing rates ($r_n$):

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

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 WM simulations (Fig. 4) was used with an across-area connectivity of 0.2% (which fits the data in Fig. 5). The only change to the network was the introduction of a slight stimulus input lateralization (Pasternak

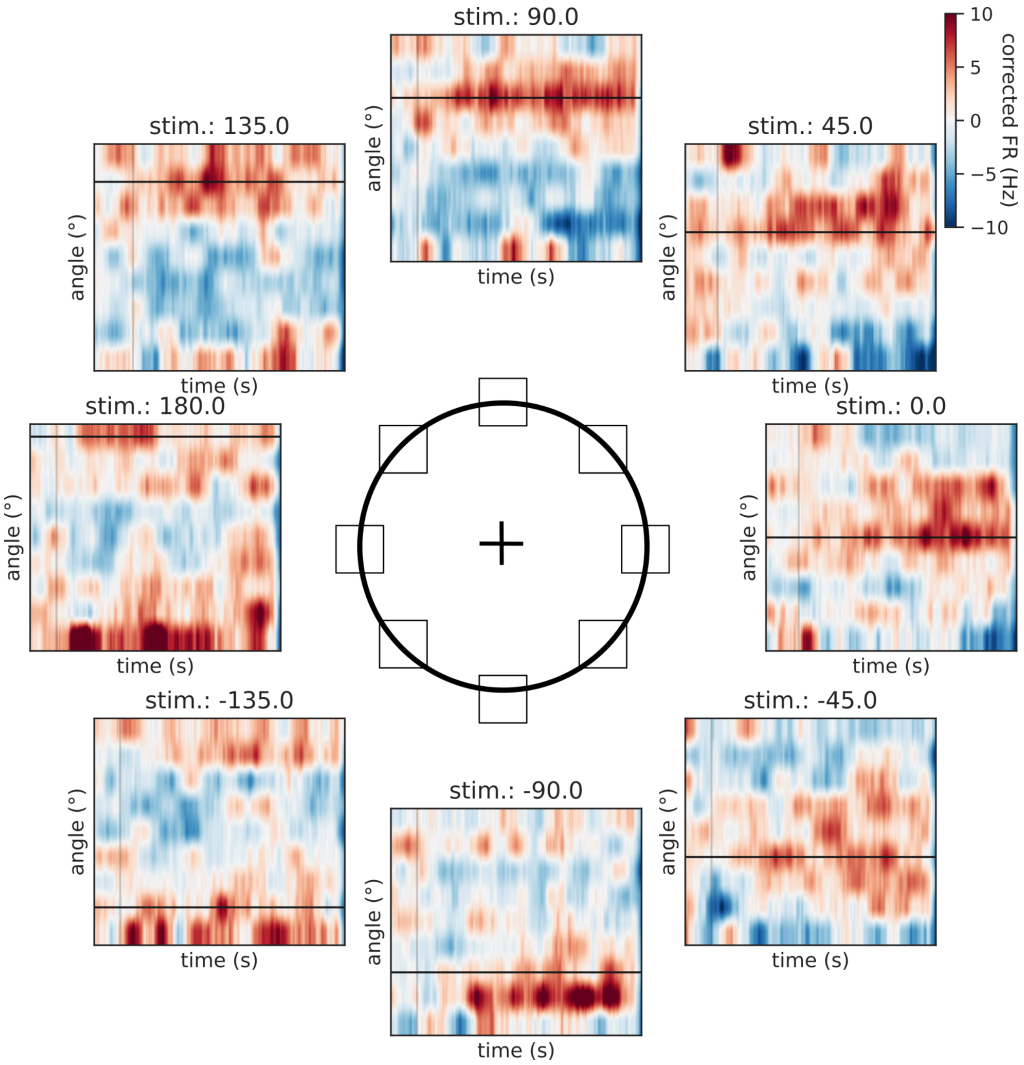
et al., 2015): We assigned each area to a hemisphere ($Area_1$ = left, $Area_2$ = right) and varied the stimulus strength according to the target's hemifield (90-270° = left; 0-90°, 270-360° = right). When stimuli were presented contralaterally to a hemisphere, the stimulus strength was set to $0.3\,nA$, as before, but if the stimuli were presented ipsilaterally the stimulus strength was reduced to $0.24\,nA$.

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.

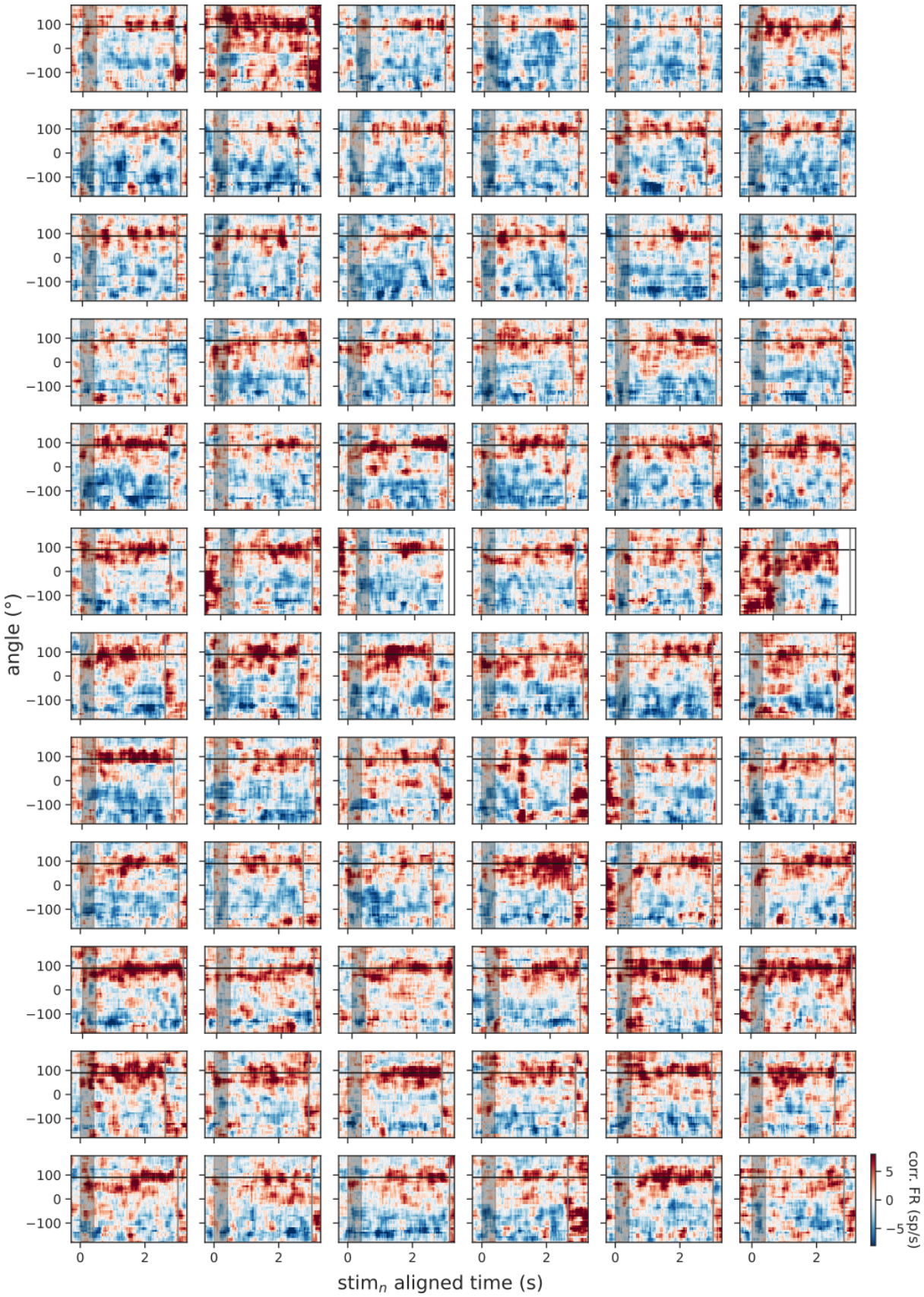
For this proof-of-concept we reused the model from the previous analysis, which has a default capacity of one item per ring. This led to a system-wide capacity of two-items since we have two attractor networks (Area1, Area2). However, fine-tuning the model with stronger and more tuned excitation (Wei et al., 2012) or inhibition (Edin et al., 2009) can increase the WM 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 led to a precise readout if the bump for the corresponding stimulus had been maintained and a random location readout in case the memory was not retained.

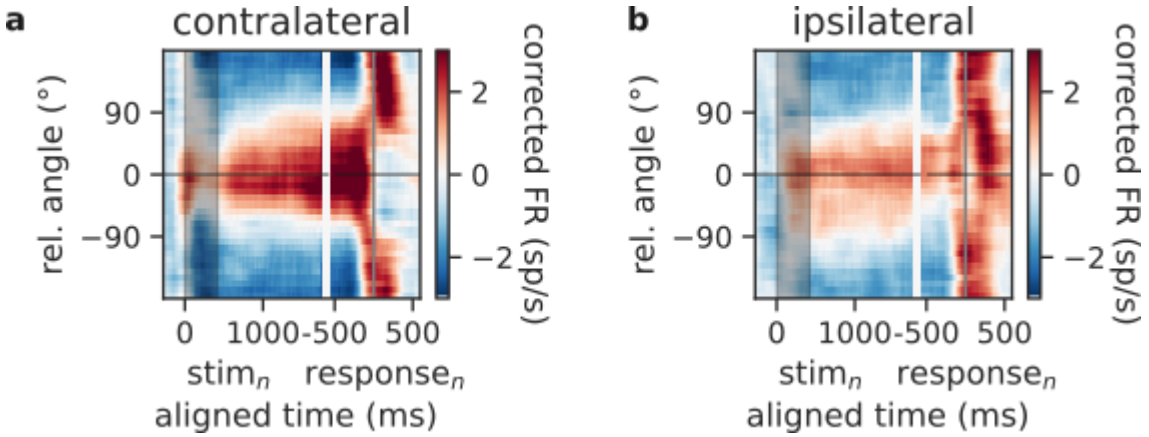
## Extended Data



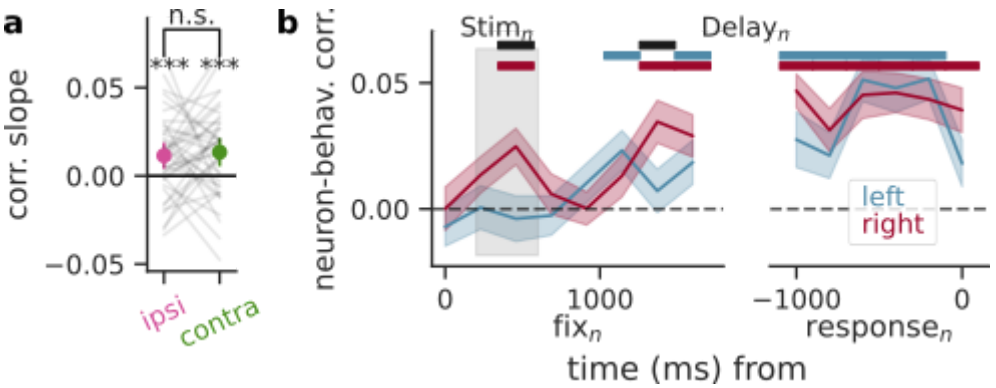

***Extended Data Fig. 1: Neural firing in example single trials with varying stimulus location.*** *Multi-units (y-axis) are grouped by their memory fields into 8 bins across circular space and the average of the multi-unit's mean-subtracted firing rates within each bin (win.: 200ms, step: 10ms) is shown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



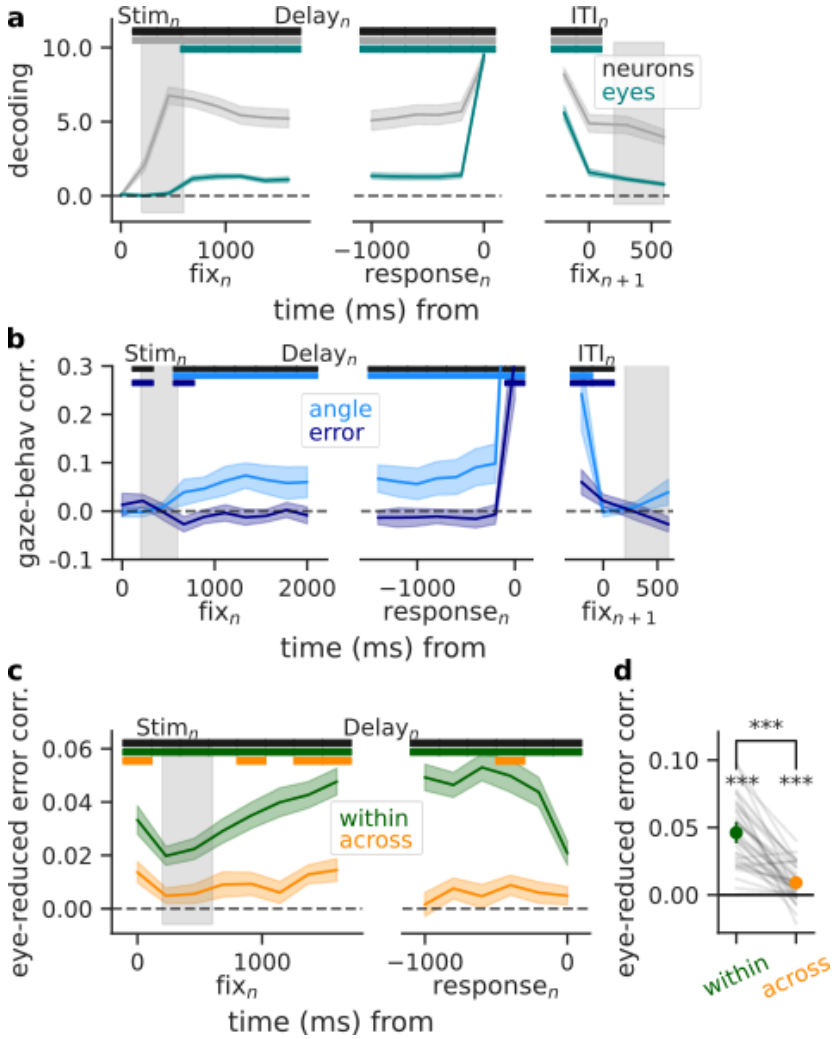

***Extended Data Fig. 2: Neural firing in all single trials with stimulus location at 90° in an example session (monkey: Sa, session: 0).***



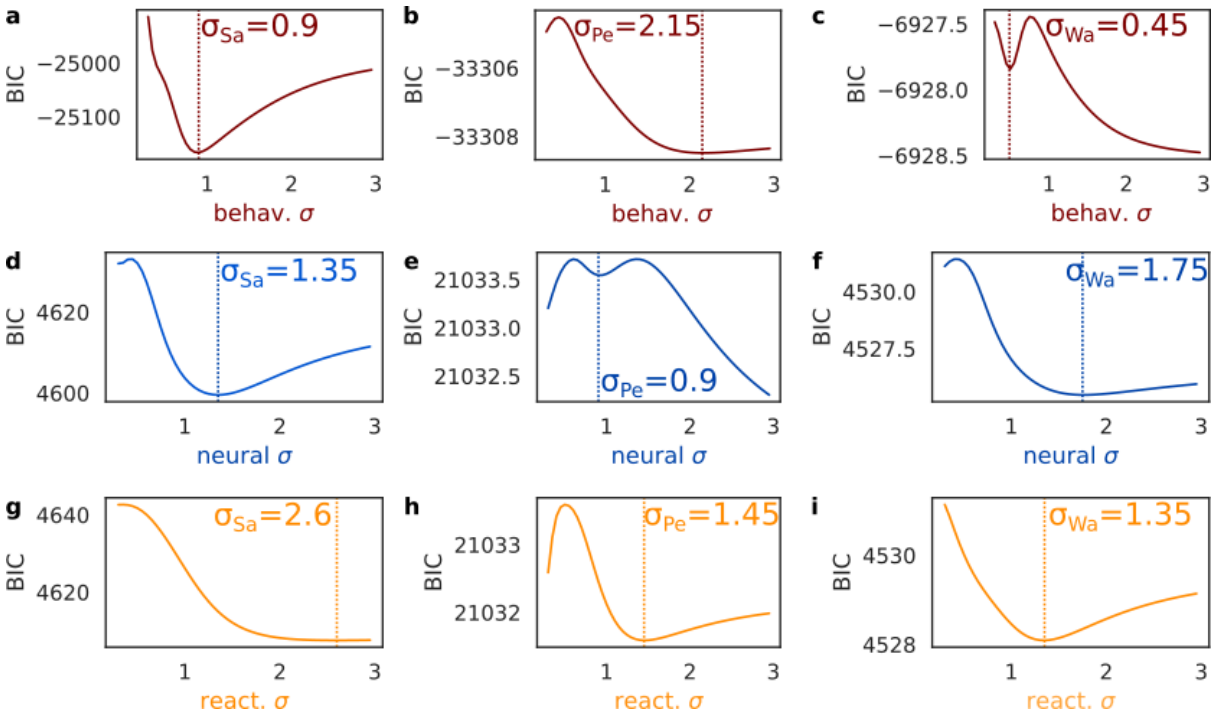

***Extended Data Fig. 3: Neural firing averaged across trials for a) contralateral and b) ipsilateral stimuli in an example session (monkey: Sa, session: 0).***



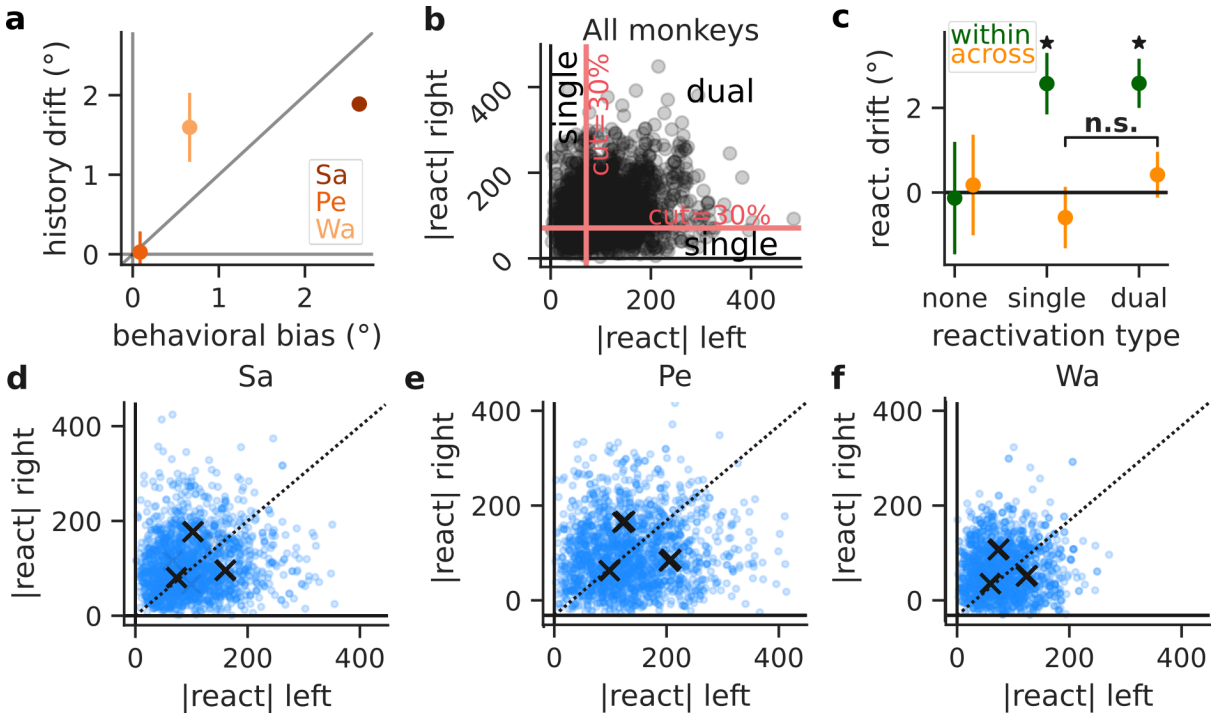

***Extended Data Fig. 4: Both hemispheres contribute equally to the response.*** *a) Quantification of the slopes of ipsilateral vs. contralateral correlation with behavioral response of Fig. 3b. Ipsilateral (t-test: t = 3.07 1/s, p = 0.004) and contralateral (t-test: t = 3.34 1/s, p = 0.002) slopes 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 none, 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: None: t = -0.1°, p = 0.92, Single: t = 3.54°, p = 0.001, Dual: t = 4.43°, p < $10^{-4}$, t-test across: None: t = 0.14°, p = 0.88, Single: t = -0.81°, p = 0.41, Dual: t = 0.77°, p = 0.44, Single vs. Dual: t = -1.1°, p = 0.27, paired t-test within-across: None: t = -0.17°, p = 0.85, Single: t = 3.54°, p = 0.001, Dual: t = 2.52°, p = 0.016) 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.