

Robustness of Working Memory to Prefrontal Cortex Microstimulation

 Joana Soldado-Magraner,^{1,2,3} Yuki Minai,^{2,3,4}  Byron M. Yu,^{1,2,3,5*} and  Matthew A. Smith^{2,3,5*}

¹Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, ²Center for the Neural Basis of Cognition, Carnegie Mellon University and University of Pittsburgh, Pittsburgh, Pennsylvania 15213, ³Neuroscience Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, ⁴Machine Learning Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, and ⁵Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Delay period activity in the dorsolateral prefrontal cortex (dlPFC) has been linked to the maintenance and control of sensory information in working memory. The stability of working memory-related signals found in such delay period activity is believed to support robust memory-guided behavior during sensory perturbations, such as distractors. Here, we directly probed dlPFC's delay period activity with a diverse set of activity perturbations and measured their consequences on neural activity and behavior. We applied patterned microstimulation to the dlPFC of two male rhesus macaques implanted with multielectrode arrays by electrically stimulating different electrodes in the array while they performed a memory-guided saccade task. We found that the microstimulation perturbations affected spatial working memory-related signals in individual dlPFC neurons. However, task performance remained largely unaffected. These apparently contradictory observations could be understood by examining different dimensions of the dlPFC population activity. In dimensions where working memory-related signals naturally evolved over time, microstimulation impacted neural activity. In contrast, in dimensions containing working memory-related signals that were stable over time, microstimulation minimally impacted neural activity. This dissociation could explain how working memory-related information may be stably maintained in dlPFC despite the activity changes induced by microstimulation. Thus, working memory processes are robust to a variety of activity perturbations in the dlPFC.

Key words: dimensionality reduction; dorsolateral prefrontal cortex; neural perturbations; population activity

Significance Statement

Memory-guided behavior is remarkably robust to sensory perturbations, such as distractors. The dorsolateral prefrontal cortex (dlPFC) is believed to underlie this robustness, given that it stably maintains working memory-related information in the presence of distractors. Here, we sought to understand the extent to which dlPFC circuits can robustly maintain working memory information during memory-guided behavior. We found that behavior was robust to electrical microstimulation perturbations in dlPFC and working memory signals were stably maintained in dlPFC despite widespread changes in the neural activity caused by the perturbations. Our findings indicate that working memory is robust to direct activity perturbations in the dlPFC, an ability that may be due to the processes that mediate similar robustness in the face of distractors.

Received June 21, 2024; revised May 2, 2025; accepted May 13, 2025.

Author contributions: J.S.-M., Y.M., B.M.Y., and M.A.S. designed research; J.S.-M. and Y.M. performed research; J.S.-M. analyzed data; J.S.-M., B.M.Y., and M.A.S. wrote the paper.

We thank the members of the Yu, Smith, Chase, and Batista labs for valuable discussions, Samantha Schmitt for assistance with data collection, and our animal care staff.

This work was supported by Japan Student Services Organization (Y.M.), National Institutes of Health CRCNS R01 MH118929 (B.M.Y. and M.A.S.), National Science Foundation NCS DRL 2124066 (B.M.Y. and M.A.S.), Simons Foundation 543065 and NC-GB-CULM-0003241-05 (B.M.Y.), National Institutes of Health NS129584 (B.M.Y.), and National Institutes of Health MH128393 and EY029250 (M.A.S.).

*B.M.Y. and M.A.S. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Joana Soldado-Magraner at jsoldadomagraner@cmu.edu.

<https://doi.org/10.1523/JNEUROSCI.2197-24.2025>

Copyright © 2025 the authors

Introduction

When animals are engaged in delayed response tasks that require the maintenance and control of sensory-related information to guide future actions, the brain must store this information in working memory and protect it from interference from other external and internal signals (Katsuki and Constantinidis, 2012; Lorenc et al., 2021; Wang, 2021). For example, in tasks that require remembering the location of a visual stimulus, the brain must maintain this information and avoid confounding it with other incoming visual signals, such as information about the location of a distracting visual stimulus (Katsuki and

Constantinidis, 2012; Suzuki and Gottlieb, 2013). The mechanism that grants working memory robustness to such disturbances is not well understood (Lorenc et al., 2021; Wang, 2021).

Delay period activity in the dorsolateral prefrontal cortex (dlPFC) has been linked to working memory-related computations (Fuster and Alexander, 1971; Goldman-Rakic, 1995). These signals may represent the content of working memory or ongoing attentional and executive control processes (Lebedev et al., 2004; Sreenivasan et al., 2014; Lara and Wallis, 2015) and are central to perform working memory tasks (Bauer and Fuster, 1976; Funahashi et al., 1993; Buckley et al., 2009). The rich and heterogeneous nature of delay period activity in dlPFC has motivated different models of working memory maintenance, including persistent activity (Wang, 2001; Constantinidis et al., 2018), dynamic representations (Druckmann and Chklovskii, 2012; Lundqvist et al., 2018), and activity-silent mechanisms (Mongillo et al., 2008; Stokes, 2015).

To guide behavior, working memory information must be maintained in the presence of perturbations such as distractors. Some theories propose that PFC can robustly and stably maintain working memory information in the presence of other evolving signals by dissociating dynamic and persistent working memory representations (Druckmann and Chklovskii, 2012; Murray et al., 2017). This dissociation might protect working memory information from interference from other incoming sensory inputs which might act as distractors (Parthasarathy et al., 2019). Many studies have shown that working memory behavior is robust to perturbations from sensory distractors (Katsuki and Constantinidis, 2012; Suzuki and Gottlieb, 2013; Parthasarathy et al., 2017; Cavanagh et al., 2018). Delay period activity in PFC is affected by distractors, but working memory-related signals are found to be largely preserved in the neural population activity, which aligns with behavior (Parthasarathy et al., 2017; Cavanagh et al., 2018).

Previous studies have sought to directly perturb delay period activity in dlPFC to influence memory-guided behavior. Long-lasting activity manipulations, for example, by microstimulating dlPFC throughout the entire delay period (Opris et al., 2005a) or pharmacologically inactivating dlPFC (Sawaguchi and Iba, 2001), affect memory-guided behavior. However, transient optogenetic inactivation of PFC minimally impacts memory-guided behavior in monkeys (Mendoza-Halliday et al., 2023). Relatedly, transient optogenetic inactivation of premotor cortex in mice has no impact on prepared movements (Li et al., 2016; Inagaki et al., 2019). These studies also found that task-relevant signals in the delay period activity were not affected or quickly recovered from such transient manipulations.

Here, we tested whether dlPFC can robustly maintain memory-guided behavior under a diverse set of electrical microstimulation perturbations. We implanted monkeys with multi-electrode arrays in the dlPFC, which allowed us to transiently stimulate the area with a variety of microstimulation spatial patterns while they performed a memory-guided saccade task. We then simultaneously recorded the effect of microstimulation on dozens of neurons in the dlPFC. We found that patterned microstimulation broadly affected dlPFC neural population activity and caused strong changes in working memory-related signals. However, the monkeys' behavior was minimally impacted. These observations could be reconciled when characterizing the effect of microstimulation at the population level. Microstimulation impacted activity in dimensions that reflected the natural time course of working memory representations. However, microstimulation minimally impacted dimensions

that contained stable working memory signals, and activity in these dimensions quickly recovered from the perturbation. Our findings indicate that working memory signals in dlPFC are robust to a wide range of microstimulation perturbations, making memory-guided behavior robust to direct perturbations of brain activity.

Materials and Methods

Subjects and surgical procedures. We implanted a 96-electrode "Utah" Array (Blackrock Microsystems) in the dlPFC of two adult, male rhesus macaques (*Macaca mulatta*) using sterile surgical techniques under isoflurane anesthesia. We implanted one array in the left 8Ar for Monkey W and dual arrays in the left and right 8Ar for Monkey S (on the prearcuate gyrus, immediately anterior to the arcuate sulcus). The head was immobilized for recordings with a titanium headpost attached to the skull with titanium screws, implanted in a separate procedure prior to the array implants. Experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of Carnegie Mellon University and complied with guidelines set forth in the National Institute of Health's Guide for the Care and Use of Laboratory Animals.

Behavioral task. In each experimental session, the monkeys performed a memory-guided saccade task. On each trial, the monkeys first fixated on a dot at the center of the screen. After establishing fixation (for 100 ms, Monkey W; for 200 ms, Monkey S), one of four peripheral targets (45, 135, 225, 315°) appeared on the screen for a brief period of time (100 ms, Monkey W; 200 ms, Monkey S). This was followed by a delay period, after which the center dot turned off (go cue) and the monkeys performed a saccade to the remembered target location to receive a liquid reward. In Monkey W, the delay period was either 1.25 or 1.55 s in duration (with probabilities 0.8 and 0.2). In Monkey S, it was 1.5 or 2 s in duration (with probabilities 0.5 and 0.5). The delay had different lengths so that the monkeys could not anticipate the go cue timing with certainty. Upon initiation of the saccade, the monkey's eye position had to reach the peripheral target location within 200 ms and maintain gaze within 2.1° (Monkey W) or 2.4° (Monkey S) of the target center for 150 ms to receive a liquid reward. Stimuli were displayed on a 21" cathode ray tube monitor with a resolution of 1024 × 768 pixels and a refresh rate of 100 Hz at a viewing distance of 59 cm. Stimuli were generated using custom software written in MATLAB (MathWorks) with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). Eye position was tracked monocularly using an infrared system at 1,000 Hz resolution (EyeLink 1000, SR Research).

Microstimulation experiments. We electrically microstimulated dlPFC on a subset of the trials in each session while the monkeys performed the memory-guided saccade task. On the rest of the trials, the monkey performed the task without microstimulation, which served as a control. On each trial, we stimulated on different electrodes of the array using single electrodes in Monkey W and pairs of electrodes simultaneously in Monkey S. In both cases, stimulation was monopolar. The array pedestal served as the ground, located a few centimeters from the implanted array. By choosing different electrodes, we changed the spatial location of the stimulation on each trial, so we refer to each stimulation condition as a different "microstimulation pattern." In each session, we stimulated with three (Monkey W) or four (Monkey S) different microstimulation patterns. We chose each pattern at random, but in some sessions we repeated some of the patterns used in previous days. Across sessions, we applied a total of 21 unique microstimulation patterns in Monkey W and four unique microstimulation patterns in Monkey S. Each session was organized in blocks. In each block, the experimental system performed for each trial a pseudorandomized selection of the different microstimulation conditions (three or four microstimulation patterns and no microstimulation) and the target angle conditions (four target angles). We ran several blocks per session to ensure sufficient amounts of trials were collected under each target angle and microstimulation condition (~20 trials for each target angle and microstimulation condition combination in Monkey W and 30 trials in Monkey S). We

refer to the set of trials collected under each microstimulation pattern as a “microstimulation experiment.” This includes trials across all target angle conditions. We performed 30 microstimulation experiments over the course of 10 sessions in Monkey W (three microstimulation patterns per session, four target angle conditions) and four microstimulation experiments within one session in Monkey S (four microstimulation patterns, four target angle conditions).

Microstimulation parameters. We used a Grapevine system and xipmex software to control microstimulation delivered using nano2 + stim headstages (Ripple). The microstimulation consisted of a 150 ms pulse train of 0.25 ms biphasic square pulses, with a frequency of 350 Hz and current of 50 or 125 μ A per stimulated electrode. We kept these parameter values fixed for all sessions. The stimulation parameters were chosen based on previous microstimulation studies in dlPFC (Wegener et al., 2008), and we set the current amplitude low enough not to induce any eye movements. In Monkey W, we stimulated on single electrodes with lower currents (50 μ A) because larger currents led to a substantial increase in trials in which fixation was aborted (though not consistent saccades). We stimulated using the array implanted in the left dlPFC and measured the effect of microstimulation on the same array. The microstimulation was applied during the delay period at 500 ms after peripheral target offset. Even though the monkey could predict the time of the microstimulation, we stimulated at a different location in the array (or did not stimulate) randomly on each trial. In Monkey S, we stimulated only using the right dlPFC array and measured the effect on both the right and the left dlPFC arrays. We stimulated with higher currents in this monkey, using pairs of electrodes each at 125 μ A. The reason for this was to induce sufficient activity modulation on the contralateral array. We could evoke saccades with some electrode pairs at microstimulation currents >100 μ A per electrode. In this monkey, we randomized the stimulation onset time (500, 600, or 700 ms after peripheral target offset).

Neural recordings. Electrophysiological recordings were performed with a Grapevine acquisition system (Ripple). Extracellular activity was recorded from the array, bandpass filtered (0.3–7,500 Hz), digitized at 30 kHz, and amplified by the system. Waveforms that exceeded a threshold were saved and stored for offline waveform classification. Thresholds were set by taking a multiple (4–5) of the root mean squared noise of the voltage measured on each electrode. Waveforms were automatically classified as either noise or spikes using an artificial neural network (Issar et al., 2020). All spiking waveforms which survived this classification on a given electrode were grouped together and treated as multiunit activity.

Data preprocessing. We excluded electrodes that in a given session had low spiking activity (firing rate <1 sp/s) and high Fano factors (>8). To compute those criteria, we used activity during no-microstimulation trials. We also removed electrodes with a high percentage of coincident spikes, which is an indication that they could be electrically shorted (pairs with >20% coincident spiking within all 0.5 ms windows over the entire session; we removed one electrode only from each electrically shorted pair). In Monkey W, we included all electrodes from the left dlPFC array that passed the criteria (50–70 electrodes per session). In Monkey S, we pooled the electrodes across the left and the right dlPFC arrays that passed the criteria (50 electrodes in the left array and 37 electrodes in the right array). We combined recordings across hemispheres because the recordings were from matched regions of cortex across the hemispheres and individual electrodes in each hemisphere showed microstimulation effects. Analysis of each hemisphere separately showed qualitatively similar results. Spike counts were binned in nonoverlapping 50 ms windows. In our analysis of both microstimulation and no-microstimulation experiments, we focus on the period –300 ms before to 750 ms after microstimulation onset. To avoid microstimulation artifacts, we did not analyze activity from microstimulation onset to 50 ms after microstimulation offset. Given that the delay period could have different lengths on each trial, the end of our analysis period (which we mark as t_{end} in our analyses) did not always correspond to the end of the delay period (at go cue). However, since the delay times were randomized, the monkeys would not know whether the end of the delay period

would occur by the time the 750 ms mark was reached. In Monkey S, given that we introduced variability in the microstimulation onset times, the period from microstimulation offset to go cue was shorter than 750 ms in some trials. This resulted in less trials contributing to some analyses for certain time periods after microstimulation, but we had a minimum of 16 trials per time step.

Single-unit analysis. We computed a measure of target angle selectivity to estimate the tuning strength of activity measured from each individual electrode (referred to as a “neural unit”) at various times during the delay period (Figs. 1C,D, 2A,C). It was computed as the difference between the maximum and minimum firing rate across the four target angle conditions at a given point in time (e.g., at t_{post}). This measure can account for relative changes in tuning strength between target angle conditions, but it is not sensitive to changes in direction preference. We chose this measure given that the most prominent effect of microstimulation involved strong changes in modulation strength. We did not observe systematic changes in direction preference with microstimulation across units, although we did observe changes in direction preference in some units.

To estimate the prevalence of microstimulation-induced early excitation, early inhibition, and rebound excitation effects in the neural population, we used the following procedure. For early excitation and inhibition, we computed the percentage of units whose firing rates were significantly higher/lower than in no-microstimulation conditions for at least two consecutive time bins following microstimulation offset (when most units show significant modulation; Fig. 2B). For rebound excitation effects, we calculated the percentage of units in each microstimulation experiment that exhibited early inhibition followed by rebound excitation. We considered that a unit presented rebound excitation if, following inhibition, it exhibited a significant increase in firing rates compared with no-microstimulation conditions for at least two consecutive time bins at any point later in the trial. We labeled a unit as experiencing early excitation, early inhibition, or rebound excitation if such microstimulation-induced activity pattern was detected in at least one of the four target angle conditions. To test for significant changes in firing rates induced by microstimulation, we used a two-tailed paired Wilcoxon rank sum test at a 5% significance level.

Microstimulation electrode tuning analysis. To test whether behavioral performance and classification accuracy could be influenced by the spatial tuning of the stimulation site, we first estimated the spatial tuning of each stimulation site using a vector average. We computed this vector average based on the neural responses to the four target angle conditions measured at the stimulation site (by averaging across no-microstimulation trials of each target angle condition, considering a 200 ms in the middle of the delay period, when microstimulation would be delivered in microstimulation trials). We excluded from the spatial tuning calculation stimulation electrodes that did not pass our data preprocessing criteria (see above, Data preprocessing). In the case of having two stimulation sites (Monkey S), we computed the vector average independently for each site and took the average of the two vectors. We then estimated the angle between the spatial tuning vector of the stimulation site and the vector representing the target direction presented to the monkey on a given trial. Next, we recomputed the behavioral metrics (precision, RT and fraction of misses; see below, Behavioral analysis) for each target angle (averaged across trials of a given target angle condition). We then computed the change in behavioral metric induced by microstimulation (the difference between the quantities along the horizontal and vertical axes in Fig. 3B–D, but estimated per target angle). We did the same to compute changes in classification accuracy induced by microstimulation (Fig. 4B; see below, Classification analysis). Finally, we tested whether there was a statistically significant correlation (Pearson’s correlation) between the change in behavioral metric or classification accuracy and the estimated angular distance (between the vector representing the spatial tuning of the stimulated site and the vector representing the target angle presented to the monkey).

Behavioral analysis. Saccade precision was estimated as the distance between the saccade endpoint to the target (in degrees of visual angle; Fig. 3B). Reaction times were computed as the time between go cue

(i.e., fixation dot offset) and the time the monkey's eye crossed the boundary of the virtual fixation window (Monkey W, 1.4° in diameter around the fixation point; Monkey S, 2.7° ; Fig. 3C). To compute the fraction of target misses [target misses / (target misses + correct trials); Fig. 3D], we considered all trials in which the monkey maintained fixation throughout the delay until the go cue. We excluded fixation breaks where the monkey left the fixation window before the go cue. In correct trials, the monkey reached the target window and maintained their eyes within that window for 150 ms. All other trials constituted "target misses," which included cases where monkeys initiated a saccade toward the target location but failed to reach the target window (which was the vast majority of the errors), as well as a very small number of cases where the monkey failed to initiate a saccade or performed a saccade toward an incorrect location. The monkeys were able to correctly report the memorized target location in this task equally well with and without microstimulation. Because of this, we focused primarily on analyzing neural activity from correct trials.

Classification analysis. We trained a Poisson Naive Bayes classifier to predict target angle identity on single trials at different time points during the delay period (Fig. 4). We trained the classifier separately for microstimulation and no-microstimulation experiments. A different classifier was trained at each time point during the delay period for each experiment. We used leave-one-trial-out cross-validation (LOOCV) to train and test the classifier and used activity from all units. To calculate confidence intervals for LOOCV classification accuracies in Figure 4A, we followed the Bayesian approach taken in Sadtler et al. (2015). We modeled the LOOCV binary classification results for each trial as a Bernoulli process with probability p (indicating the probability of correct classification). Since p was unknown, we set a uniform prior distribution on p between 0 and 1. Given this prior distribution and the distribution of observations (i.e., the binary string of classification results across trials), the posterior distribution on p could be computed, which is a beta distribution. Figure 4A shows the mean and the 95% confidence interval of the posterior distribution on p , estimated independently for each time point on each microstimulation condition.

To predict target angle identity based on activity within the memory subspace alone, we used a different classification procedure. We used fivefold cross-validation to estimate the memory subspace using both microstimulation and no-microstimulation trials (see below, Population analysis). On each fold, we projected the activity of no-microstimulation trials into the identified memory subspace and computed class means across trials and time for each target angle condition. Next, we projected the activity of held-out microstimulation trials and classified each trial based on the Euclidean distance to the closest class mean. Importantly, the held-out microstimulation trials had neither been used to estimate the memory subspace nor to compute the class means. Even though class means were computed from no-microstimulation trials, and that the means were based on averages across time, we could accurately and stably classify microstimulation trials at various times during the delay period (see Results, dlPFC's neural activity was minimally impacted by microstimulation in a memory subspace of the neural population activity).

Population analysis. We used two different dimensionality reduction methods, factor analysis (FA) and demixed PCA (dPCA), to estimate different subspaces of the neural population activity. FA is a statistical method whose objective is to find the "dominant" dimensions that capture the greatest shared variance among the neural units. We sought to estimate the dominant dimensions of the neural activity under regular task conditions, so we fit FA to no-microstimulation trials exclusively. In each session, we fit FA to the binned spike counts (50 ms bins) from each unit during the delay period in all no-microstimulation trials (including all target angle conditions). The dimensionality of the "dominant subspace" (4D) was estimated by computing the optimal dimensionality separately for each session based on cross-validated data likelihoods, and taking the rounded average of the estimates across all sessions. Low-dimensional activity was inferred by the model based on posterior mean estimates (Santhanam et al., 2009).

To find the "memory subspace," we used demixed dPCA (Kobak et al., 2016). dPCA's objective is to find dimensions of the neural activity

that are predominantly related to specific behavioral and task variables. We used dPCA to find memory dimensions that maximized target angle variance and minimized the effects of microstimulation, and also that contained target angle signals that were stable over time. To do this, the dPCA model was fit to both microstimulation and no-microstimulation trials, and different subspaces were found that "demixed" target angle, microstimulation, and time-related signals in the population, as well as signals due to their interaction (Fig. 6). We refer to the target angle dimensions as the "memory subspace." We fitted dPCA to the firing rates from each unit estimated in 50 ms bins during the delay period, obtained by averaging across trials of a given target angle and microstimulation condition (with no temporal smoothing). We set the dimensionality of the memory subspace to be 4D to match the dimensionality of the dominant subspace. This was necessary to be able to fairly compare activity across the two subspaces. We used fivefold cross-validation to estimate the memory dimensions. We confirmed that the four memory dimensions explained a large percentage (70%) of the target angle variance in the trial-averaged neural activity of held-out trials across sessions. This subspace contained only a small fraction of time-related and stimulation-related variance, and thus, target angle signals in this subspace were largely stable over time and were minimally influenced by microstimulation.

The distance metric in Figure 7 was computed separately for each target angle and microstimulation pattern and at each time point. First, we calculated the difference between the mean activity across trials in microstimulation and no-microstimulation conditions along each dimension in the subspace (four dimensions). Second, we standardized the differences by the 95% confidence interval of the activity across trials in no-microstimulation conditions along each dimension. Third, we computed the vector norm of the differences across all dimensions. In this way, we could compare distances across sessions and monkeys. This metric was computed based on held-out microstimulation trials that were not used in the estimation of the subspaces.

Statistical tests. To test whether microstimulation and no-microstimulation activity were statistically different, we used a two-tailed two-sample Wilcoxon rank sum test at a 5% significance level (Fig. 2A). In Figure 2B, we applied this test considering activity across all four target angle conditions. To test for significance of microstimulation-induced decreases in tuning strength (Fig. 2C), we used a one-tailed two-sample t test. To test for significance of microstimulation-induced changes in saccade precision (Fig. 3B) and reaction times (Fig. 3C), we used a two-tailed two-sample t test. To test for microstimulation-induced changes in target misses (Fig. 3D) and classification accuracies (Fig. 4B), we used a two-tailed two-sample Wilcoxon rank sum test. To assess whether we could classify above chance, we used the binomial test at a 5% significance level to compare the binary string used to compute the classification accuracy across trials (a value between zero and one) against a 0.25 chance level probability (since there are four target angle conditions). To assess whether a given microstimulation experiment induced a significant increase or decrease in classification probabilities with respect to the no-microstimulation condition (Fig. 4B), we used a binomial test at a 5% significance level to compare the binary string used to compute the classification accuracy across trials in the microstimulation condition against the classification accuracy obtained in the no-microstimulation condition. To test for the overall significance of linear regression models (Figs. 3B–D, 4B), we used an F test. To compare microstimulation and no-microstimulation activity within the dominant and memory subspaces (Fig. 7B), we used a multivariate two-sample test at a 5% significance level based on the statistical energy of the samples (Aslan and Zech, 2005), which compares the two-sample distributions in a multidimensional space.

Results

We designed an experimental protocol using electrical microstimulation to probe delay period activity in dlPFC during working memory. We implanted two monkeys with multielectrode arrays in dlPFC, area 8Ar (Fig. 1A), and applied different microstimulation (uStim) patterns while they performed a memory-guided

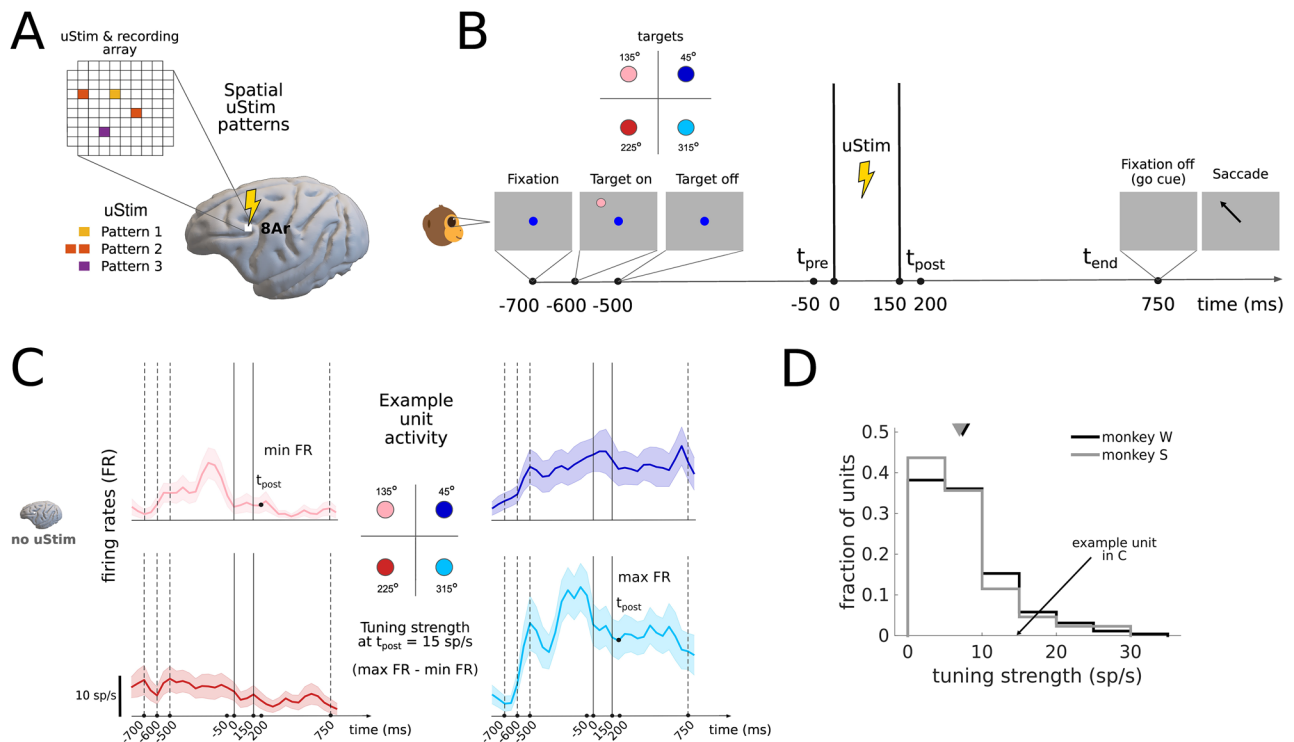


Figure 1. Multielectrode electrical microstimulation to probe dlPFC delay period activity during working memory. **A**, Electrical microstimulation (uStim) protocol. A Utah array was implanted in the dlPFC, area 8Ar. Subthreshold stimulation was applied to either individual electrodes or pairs of electrodes in the array, creating a variety of spatial microstimulation patterns (see Materials and Methods). Neural population activity was simultaneously recorded from all electrodes in the array after microstimulation. **B**, Memory-guided saccade task. After the monkey acquired fixation, a target was briefly presented (for 100 ms) at one of four possible spatial locations. This was followed by a variable delay period (1.25–2 s, see Materials and Methods), after which the go cue (fixation off) signaled the monkey to saccade to the remembered target location. Microstimulation was applied during the delay period, using 3–4 different spatial patterns on each session (see Materials and Methods). t_{pre} , 50 ms before uStim onset; t_{post} , 50 ms after uStim offset; t_{end} , end of delay period (time of go cue). **C**, Trial-averaged activity (PSTHs) of one representative unit for each of the four target angle conditions. Shown are firing rates (FR) computed from no-microstimulation trials ($n = 22$) in an example session. This unit is spatially tuned during the delay period—activity is higher for the 315° target than for the other targets. Tuning strength at $t_{\text{post}} = 15$ spikes per second (sp/s), computed as max FR (target 315°)–min FR (target 225°). Vertical lines mark the trial epochs indicated in panel **B**. In no-microstimulation trials, t_{post} marks the same time as in microstimulation trials, but no stimulation was delivered in this case. Trials are aligned to the time at which microstimulation would be delivered during microstimulation trials (see Materials and Methods). PSTHs have been smoothed for visualization with a Gaussian filter with a standard deviation of 40 ms. Monkey W, session 20220810, unit 30. **D**, Tuning strengths across all units at t_{post} (Monkey W, $n = 10$ sessions; Monkey S, $n = 1$ session). Triangles indicate mean tuning strength across all units and sessions (7 ± 6 sp/s, mean $\pm$ SD, Monkey W; 7 ± 6 sp/s, Monkey S).

saccade task (Fig. 1B). We used microstimulation patterns below the threshold that evoked saccades, consisting of either single electrodes or pairs of electrodes at different spatial locations in the array (Fig. 1A; see Materials and Methods). In area 8Ar, the current thresholds required to evoke saccades tend to be higher than the nearby frontal eye fields (Bruce et al., 1985), though subthreshold stimulation can still impact behavior (Opris et al., 2005a). We focused on area 8Ar in the dlPFC since it is implicated in spatial working memory and contains rich and heterogeneous signals related to the transformation of sensory inputs to oculomotor actions in memory-guided saccade tasks (Funahashi et al., 1989; Opris et al., 2005a; Leavitt et al., 2018; Khanna et al., 2020).

At the beginning of the task, a target was briefly shown at one out of four possible spatial locations (Fig. 1B; Materials and Methods). This was followed by a variable delay period (1.5–2 s; from “target off” to “fixation off” in Fig. 1B), after which the animal reported the location of the target with a saccadic eye movement to the remembered location. We stimulated during the delay period of the task when only a fixation spot was present on the screen and the animal had to hold in memory the location of the presented target. On each trial, we stimulated with a different microstimulation pattern (Fig. 1A), with 3–4 patterns selected for each experimental session (see Materials and Methods). Interleaved with the

microstimulation trials, we also ran trials where no microstimulation was applied. We consider a “microstimulation experiment” as the collection of trials collected under a given microstimulation pattern within a session, which includes trials for all target angle conditions (see Materials and Methods). Within a session, we performed multiple microstimulation experiments. Similarly, we consider a “no-microstimulation experiment” as the collection of trials collected with no microstimulation in a given session, for all target angle conditions.

Recordings from different electrodes in the array revealed delay period activity that was tuned to the different spatial locations (Fig. 1C, example unit), as previously reported in this area (Funahashi et al., 1989; Leavitt et al., 2018; Khanna et al., 2020). Some of these signals were persistent throughout the delay (Fig. 1C, target 45° activity), a canonical feature of delay period activity in the prefrontal cortex (Fuster and Alexander, 1971; Constantinidis et al., 2018). Other neural responses were dynamic (Fig. 1C, target 135° activity), consistent with more recent descriptions of working memory signals in PFC (Murray et al., 2017; Lundqvist et al., 2018; Khanna et al., 2020; Wang, 2021). Across units and sessions, tuning strength (a measure of selectivity to different targets; see Materials and Methods) ranged from a few spikes per second (sp/s) to over 30 sp/s (Fig. 1D). Thus, the dlPFC neural population contained

working memory-related signals with target angle location information of the type needed to solve the task.

Microstimulation broadly affected dlPFC delay period activity during working memory

Having identified working memory-related signals in the dlPFC population during the delay period, we sought to understand how different microstimulation perturbations impacted these signals in individual neural units. For this, we measured the effect of microstimulation on the activity across all recorded units in the array and quantified the impact on tuning strength.

We found that microstimulation disrupted working memory-related signals in dlPFC. Neural activity was strongly modulated by microstimulation, and this modulation induced changes in tuning strength, even on correct trials (Fig. 2*A*, example units). We observed that microstimulation often suppressed activity (Fig. 2*A*, top panel, no microstimulation; bottom panel, microstimulation). This suppression was observed in all target angle conditions for many units, resulting in reductions of tuning strength (e.g., from 15 sp/s to 0 sp/s at microstimulation offset, t_{post} ; Fig. 2*A*, top vs bottom left panels). In some cases, microstimulation also caused increases in activity for some conditions (Fig. 2*A*, top vs bottom right panels) and concomitant increases of tuning strength (e.g., from 15 sp/s to 40 sp/s at t_{post} ; Fig. 2*A*, top vs bottom right panels).

The microstimulation effect on neural activity was widespread across the dlPFC population, with over 50% of the recorded units significantly modulated for one or more of the four target angle

conditions immediately after microstimulation offset (at t_{post} ; Fig. 2*B*; Monkey W: $58 \pm 16\%$ of units, mean $\pm$ SD, $n = 30$ microstimulation experiments; Monkey S: $50 \pm 7\%$ of units, mean $\pm$ SD, $n = 4$ microstimulation experiments). The most common effect of microstimulation was to suppress activity (with $26 \pm 12\%$ of units significantly suppressed right after microstimulation offset, mean $\pm$ SD, $n = 34$ microstimulation experiments; see Materials and Methods). Few units were excited after microstimulation offset ($4 \pm 3\%$ of units, mean $\pm$ SD, $n = 34$ microstimulation experiments) or exhibited inhibition followed by rebound excitation ($4 \pm 3\%$ of units, mean $\pm$ SD, $n = 34$ microstimulation experiments; Kumaravelu and Grill, 2024), as seen in example unit 1 (Fig. 2*A*, left panels). Some units remained modulated by the end of the delay period (at t_{end} ; Fig. 2*B*; Monkey W: $13 \pm 5\%$, mean $\pm$ SD, Monkey S: $14 \pm 3\%$, mean $\pm$ SD). This indicates that activity did not always recover from the microstimulation perturbation by the time the monkeys were instructed to perform the saccade.

Next, we sought to quantify the impact of microstimulation on working memory-related signals in units that were significantly modulated by microstimulation. In many units, the microstimulation perturbation caused strong changes in tuning strength (Fig. 2*C*). The modulated units exhibited both decreases and increases in tuning strength, but overall there was a significant reduction in tuning strength right after microstimulation offset (Fig. 2*C*, left panel). The changes in tuning strength were often substantial (>10 spikes per second). By the end of the delay period, tuning strength remained reduced on average across units and

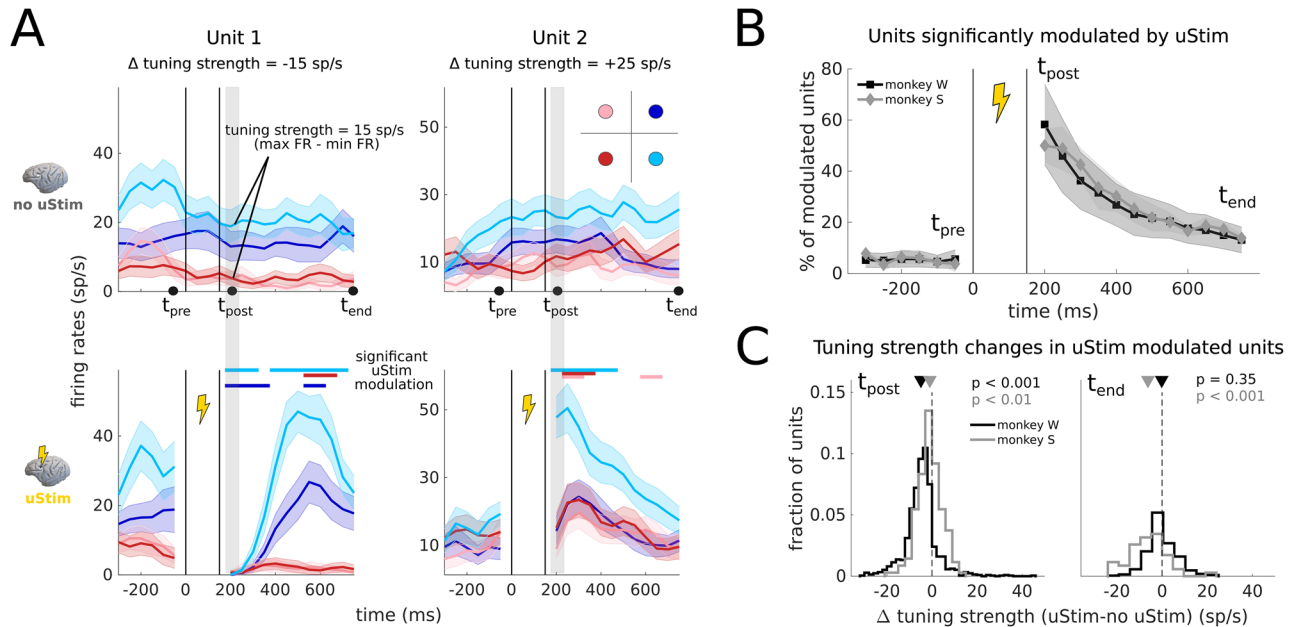


Figure 2. Microstimulation impacts working memory-related signals in dlPFC. **A**, Effect of microstimulation on neural activity and target angle tuning for two example units (left panel, same unit as in Fig. 1*C*). Top, PSTHs for trials without microstimulation (no-uStim). Bottom, PSTHs for trials with microstimulation (uStim). Microstimulation can induce both decreases (left panels) or increases (right panels) in activity. Furthermore, the tuning strength can decrease (-15 sp/s at t_{post} , left panels) or increase ($+25$ sp/s at t_{post} , right panels). Colored bars above the bottom panels indicate significant microstimulation-induced activity modulation for each target angle compared with no-microstimulation activity for the same target angle ($p < 0.05$, two-tailed paired Wilcoxon rank sum test, $n = 22$ trials per condition). In microstimulation trials, PSTHs are aligned to the time of microstimulation onset (bottom panels). In no-microstimulation trials, the PSTHs are aligned to the time at which microstimulation would be delivered during microstimulation trials (top panels; see Materials and Methods). PSTHs have been smoothed for visualization with a Gaussian filter with a standard deviation of 40 ms. Monkey W, session 20220810, units 30 and 56. **B**, Percentage of units that are significantly modulated by microstimulation at different times during the delay period (mean $\pm$ SD across all microstimulation experiments; Monkey W, $n = 30$; Monkey S, $n = 4$; $p < 0.05$, two-tailed paired Wilcoxon rank sum test). **C**, Tuning strength changes induced by microstimulation at t_{post} and t_{end} in units significantly modulated by microstimulation. Tuning strength was predominantly reduced by microstimulation at t_{post} ($p < 0.001$, Monkey W; $p < 0.01$, Monkey S; one-tailed paired t test; all units across all microstimulation experiments). Triangles indicate mean tuning strength across all units and sessions for each monkey. By t_{end} , tuning strength remained reduced on average across units and experiments for Monkey S, but not for Monkey W ($p = 0.35$, Monkey W; $p < 0.001$, Monkey S; one-tailed paired t test), although some units remained affected in monkey W.

experiments for one of the monkeys (Fig. 2C, right panel), though some units remained affected in both monkeys. Thus, we found that microstimulation broadly impacted activity across the dlPFC neural population and produced strong and long-lasting changes in tuning strength, affecting working memory-related signals that may be crucial for the task.

Microstimulation minimally impacted memory-guided behavior

Having found that there was a pronounced effect of microstimulation on neural activity (Fig. 2), we next sought to determine whether microstimulation impacted working memory behavior. We asked whether microstimulation had an impact on the eye movements to remembered targets. We observed that saccade trajectories were qualitatively similar in microstimulation versus no-microstimulation experiments (Fig. 3A). However, since microstimulation can have subtle impacts on behavior (Opris et al., 2005a, b; Murphey and Maunsell, 2008), we quantified the impact of microstimulation on several eye movement metrics.

First, the precision of saccades (distance of the saccade endpoint to the presented target) was not impacted by microstimulation (Fig. 3B). Across all microstimulation experiments in both animals, we found no significant difference in saccade precision between microstimulation and no-microstimulation experiments.

Second, we asked whether microstimulation had an impact on the readiness of the animals to report the remembered target location by examining the saccadic reaction time (RT; Fig. 3C).

The overall RT across experiments was ~ 3 ms faster in the no-microstimulation versus microstimulation experiments, which was small compared with the typical range of RTs within a given experiment (no-microstimulation, 181 ± 7 ms; microstimulation, 184 ± 10 ms; mean $\pm$ SD). Nonetheless, this difference was statistically significant. This small but significant difference could indicate that the microstimulation had a slight tendency to disrupt the memory signal or the saccadic preparation (Opris et al., 2005a; Churchland and Shenoy, 2007).

Third, microstimulation did not increase the rate of errors (Fig. 3D). We considered target misses, which applied to cases where monkeys failed to report the correct target location after go cue presentation (see Materials and Methods). This included cases where monkeys initiated a saccade toward the target location but failed to reach the target window (which was the vast majority of the errors), as well as a small number of cases where the monkey failed to initiate a saccade or performed a saccade toward an incorrect location. We found no statistically significant difference in the fraction of target misses in microstimulation versus no-microstimulation trials (Fig. 3D).

Finally, we considered the possibility that the impact of microstimulation on performance could depend on the spatial tuning of the stimulation site. In particular, microstimulation could have a stronger impact in cases where the spatial tuning of the stimulation site matched the target location presented on a given trial. Overall, we found no clear relationship between the spatial tuning of the stimulation site and the changes in behavior (see Materials and

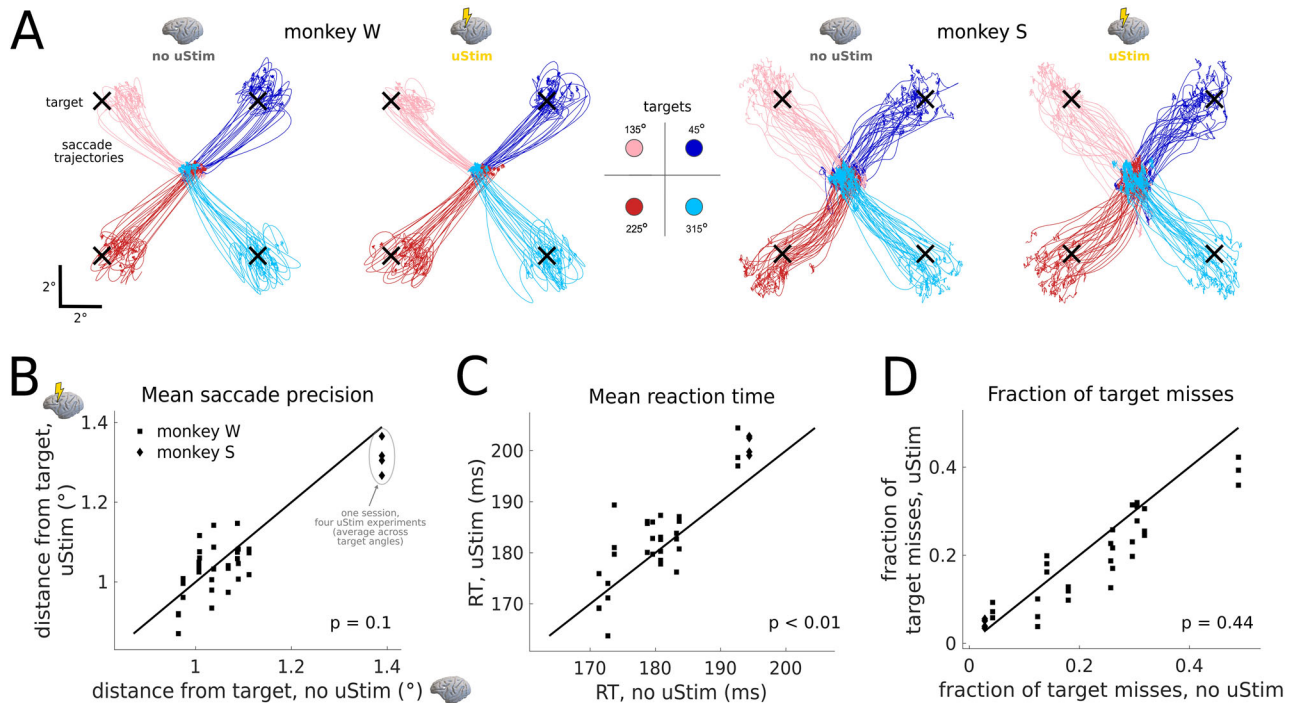


Figure 3. Microstimulation in dlPFC minimally impacts memory-guided behavior. **A**, Saccade trajectories in microstimulation (uStim) and no-microstimulation (no-uStim) trials for an example session of each monkey. Monkey W, session 20220801; Monkey S, session 20210311. Trajectories are shown from go cue to target acquisition, for all target angle conditions (trials per condition, $n = 22$, Monkey W; $n = 33$, Monkey S). Trajectories have been smoothed for visualization with a Gaussian filter with a standard deviation of 2 ms. **B**, Mean saccade precision in microstimulation versus no-microstimulation trials, measured as the trial-averaged distance from the saccade endpoint to the target. Distances on microstimulation trials are not significantly different from distances on no-microstimulation trials ($p = 0.1$, two-tailed paired t test; all microstimulation experiments across both monkeys, $n = 34$). **C**, Mean reaction time (RT) on microstimulation versus no-microstimulation trials. RTs on microstimulation trials are significantly higher than on no-microstimulation trials ($p < 0.01$, two-tailed paired t test; $n = 34$), but only by 3 ms on average. **D**, Fraction of target misses (see Materials and Methods) on microstimulation versus no-microstimulation trials. The fraction of target misses on microstimulation trials is not significantly different than on no-microstimulation trials ($p = 0.44$, two-tailed paired Wilcoxon rank sum test; $n = 34$). In panels **B–D**, there was a significant linear relationship between the behavioral metrics measured in microstimulation versus no-microstimulation trials ($p < 0.001$, F test). Each individual square/diamond indicates a different microstimulation experiment (which includes all trials for a given microstimulation pattern within a session, averaged across all target angle conditions). The squares/diamonds appear organized in columns because each microstimulation experiment within a session was compared against the same no-microstimulation experiment of that session.

Methods). However, we found a small but significant effect on saccade precision. In particular, microstimulation seemed to slightly increase saccade precision (albeit by a very small amount, $<0.3^\circ$) when the spatial tuning of the stimulated site was close to the target direction presented (Pearson's correlation = 0.3, $p = 0.02$). Thus, the observed microstimulation-induced neural activity changes had minimal consequences for the behavioral performance of the monkeys.

dIPFC preserved working memory-related information after microstimulation

How can we reconcile the pronounced microstimulation effects on neural activity (Fig. 2) with the little or no impact of microstimulation on behavior (Fig. 3)? One possible explanation is that the area we are perturbing, despite containing signals relevant to the task, is not causally implicated in the generation of the behavior. However, the large body of literature involving lesions, cooling, and inactivation in the dIPFC points at a critical and necessary role of this region in this type of working memory behavior (Bauer and Fuster, 1976; Funahashi et al., 1993; Buckley et al., 2009). An alternative explanation is that, similar to what has been previously reported with sensory distractors (Parthasarathy et al., 2017; Cavanagh et al., 2018), the induced

neural changes from microstimulation do not entirely disrupt working memory information.

To test this, we analyzed the dIPFC neural population activity to determine if working memory-related information was preserved in the population after microstimulation. To extract this information from the neural population, we trained target angle classifiers using delay period activity separately for microstimulation and no-microstimulation experiments. Furthermore, a separate classifier was trained at each time step and used neural activity from all units (see Materials and Methods). In this way, we could assess whether microstimulation might have reduced or eliminated the target angle information present in the dIPFC population. We found that we could classify well above chance and with similar accuracy in both microstimulation and no-microstimulation experiments (Fig. 4A, example session with three different microstimulation experiments). Importantly, we could classify similarly well throughout the entire delay period. Across all experiments we performed, classification accuracy was not significantly different in microstimulation versus no-microstimulation experiments (Fig. 4B, shown at t_{pre} , t_{post} and t_{end}). There was variability in classification accuracy across microstimulation experiments (individual squares and diamonds), but in all cases we could classify above chance (dashed lines), even

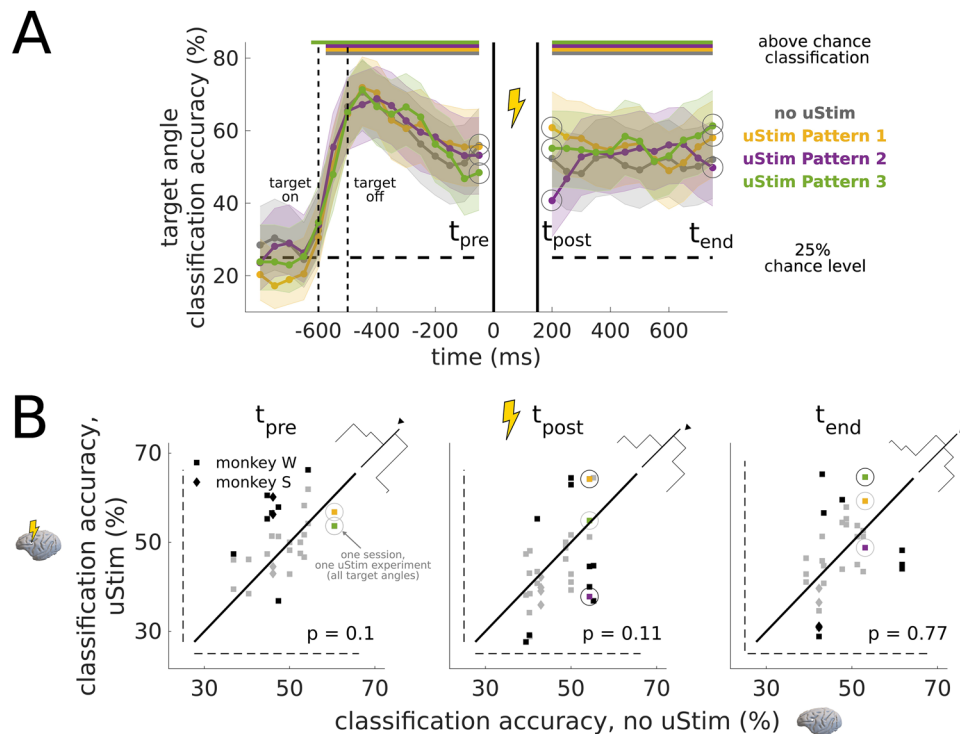


Figure 4. Working memory-related information in dIPFC is preserved after microstimulation. **A**, Cross-validated target angle classification accuracy for microstimulation (uStim) and no-microstimulation (no-uStim) experiments. Example session with three microstimulation experiments (colored lines) and one no-microstimulation experiment (gray line). A separate classifier was trained on each time step for each experiment. Classification was well above chance (dashed line, chance probability: 0.25) throughout the delay period (indicated by colored bars; $p < 0.05$, binomial test), even at t_{post} . Classification accuracies were similar in microstimulation and no-microstimulation experiments. Shading indicates 95% confidence interval (see Materials and Methods). Lines have been smoothed for visualization with a Gaussian filter with a standard deviation of 40 ms. Monkey W, session 20220802. **B**, Classification accuracies for all microstimulation experiments at different time points during the delay period. Classification accuracy was not significantly different in microstimulation versus no-microstimulation experiments (left to right panels for t_{pre} , t_{post} , and t_{end} ; $p = 0.1$, $p = 0.11$ and $p = 0.77$, two-tailed paired Wilcoxon rank sum test; $n = 34$). Classification accuracy difference (no-microstimulation - microstimulation) averaged across all stimulation patterns: $t_{pre} = -2.5\%$, $t_{post} = 2.3\%$, and $t_{end} = 0.6\%$. In all cases, classification was above chance (dashed lines), even at t_{post} (middle panel). There was a significant linear relationship between the accuracies estimated in microstimulation versus no-microstimulation trials at each of the three time points ($p < 0.05$, F test). Black symbols indicate microstimulation patterns that exhibit significant changes in classification accuracy with respect to no-microstimulation conditions ($p < 0.05$, binomial test). Gray symbols indicate no significant changes. One exception is that, for the symbols corresponding to the example session shown in panel **A**, the symbols are colored by the uStim pattern (3 colored symbols in each panel). To indicate significance for the colored symbols, we use a black or gray circle around the symbol to indicate significant or not significant changes, respectively. The purple square in the left panel overlaps with the green square and is therefore not visible.

immediately after microstimulation (at t_{post} , middle panel). The variability in classification accuracy could not be accounted for based on the spatial tuning of the stimulated sites (see Materials and Methods). This analysis demonstrates that dlPFC robustly encodes working memory-related information after microstimulation. The presence of working memory-related signals in the neural population activity may explain why behavior is not disrupted by perturbations that have strong effects on the activity of individual units.

dlPFC's neural activity was minimally impacted by microstimulation in a memory subspace of the neural population activity

Having found that working memory information is preserved in the dlPFC population after microstimulation, we next sought to find which dimensions of the population activity were perturbed by microstimulation. We considered two subspaces. First, we applied FA to extract the dimensions of the activity that captured the greatest shared variance among the neurons during no-microstimulation conditions (Fig. 5A, the “dominant subspace”; see Materials and Methods). Second, we used demixed PCA (dPCA; Kobak et al., 2016) to specifically look for dimensions of the activity that were not affected by microstimulation

and which could contain stable working memory-related signals during the delay (Fig. 5B, the “memory subspace”; see Materials and Methods). The existence of these dimensions might explain the robustness of behavior, since working memory-related signals could be stably read out throughout the delay period (Parthasarathy et al., 2019).

The dominant subspace captured two prominent features of the population activity: working memory tuning and the time evolution of responses (Fig. 5A, middle panel). These dimensions together captured the dominant dimensions of covariation among the units and not necessarily variance due to specific task variables (such as target angle and time in the trial; see Materials and Methods). However, target angle tuning, as well as changes in activity over time, tend to be shared among neurons. This shared variance is precisely what FA seeks to capture. This is why we observed target angle tuning in this subspace. In particular, in no-microstimulation experiments, the activity occupied different locations in the space depending on target angle condition, reflecting target tuning. Additionally, neural activity during the delay period evolved from its initial state, moving from one location (at t_{pre}) to a different location (at t_{end}). This explains why the dominant subspace contained working memory-related signals that evolved over time.

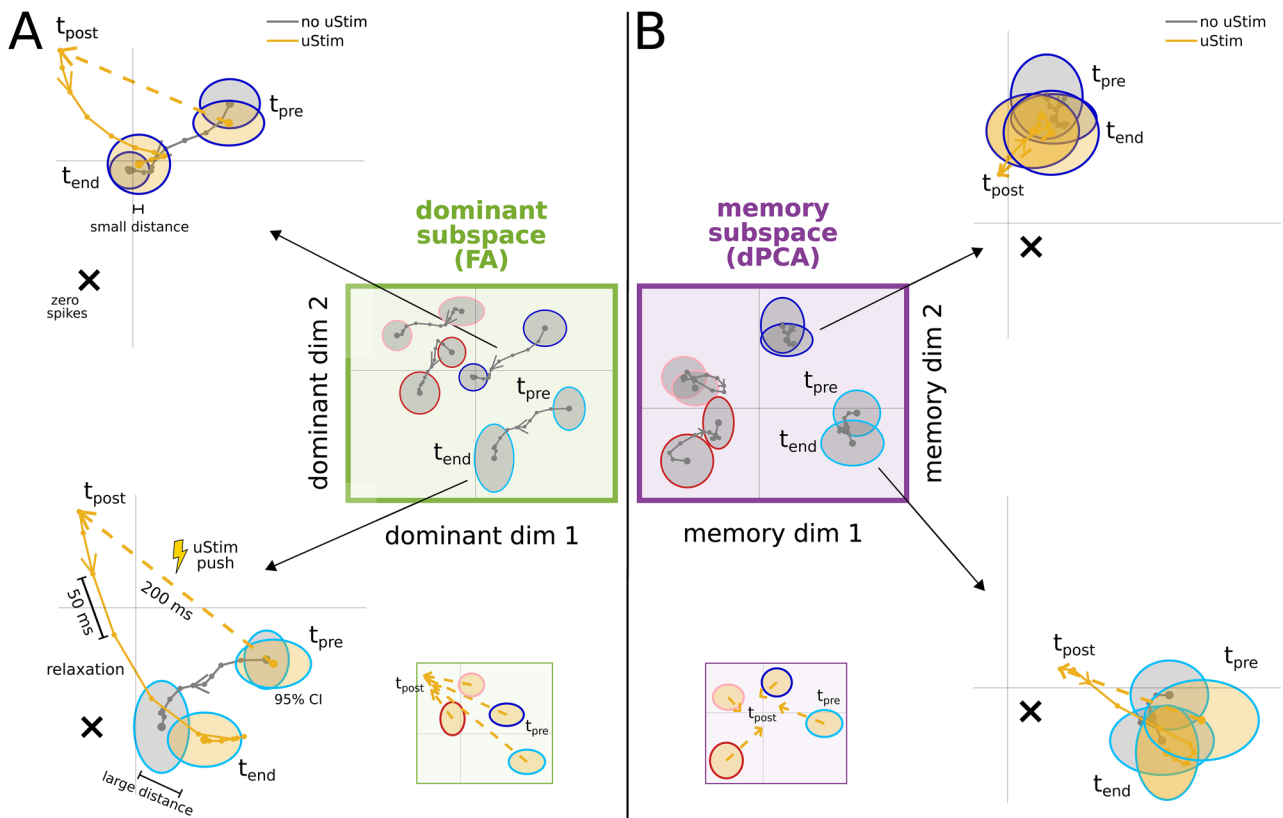


Figure 5. Working memory representations in dominant and memory subspaces of dlPFC population activity. **A**, Middle panel, trial-averaged neural trajectories in the dominant subspace during example no-microstimulation trials (gray lines) for all target angle conditions. Top and bottom left panels, trial-averaged neural trajectories in the dominant subspace during example microstimulation (uStim, yellow lines) and no-microstimulation (no-uStim, gray lines) trials, for two example target angle conditions (45 and 315°). For no-microstimulation trials, trajectories are shown from t_{pre} to t_{end} . For microstimulation trials, trajectories are shown from t_{post} to t_{end} . Data were excluded from t_{pre} to t_{post} due to microstimulation artifacts. The change in activity induced by microstimulation from t_{pre} to t_{post} is indicated by yellow dashed arrows. Inset at the bottom right corner shows activity at t_{pre} and t_{post} during microstimulation trials for all target angle conditions. 95% confidence intervals (CI) across all trials ($n = 22$) are shown at t_{pre} and t_{end} . Microstimulation strongly modulates activity in the dominant subspace (yellow dashed arrows are long), and the activity does not always recover to its natural state by t_{end} (distance between microstimulation and no-microstimulation activity distributions at t_{end}). **B**, Trial-averaged neural trajectories in the memory subspace. Same conventions as in panel **A**. Microstimulation modulation is weak in this subspace (yellow dashed arrows are short), and the activity remains in the same location throughout the delay period. The X markers in panels **A** and **B** indicate the location in state space corresponding to 0 spike counts across all units. In panels **A** and **B**, we show population activity in the two leading dimensions of the dominant and the memory subspaces. Trajectories in panels **A** and **B** have been smoothed with a Gaussian filter with a standard deviation of 60 ms. Monkey W, session 20220810.

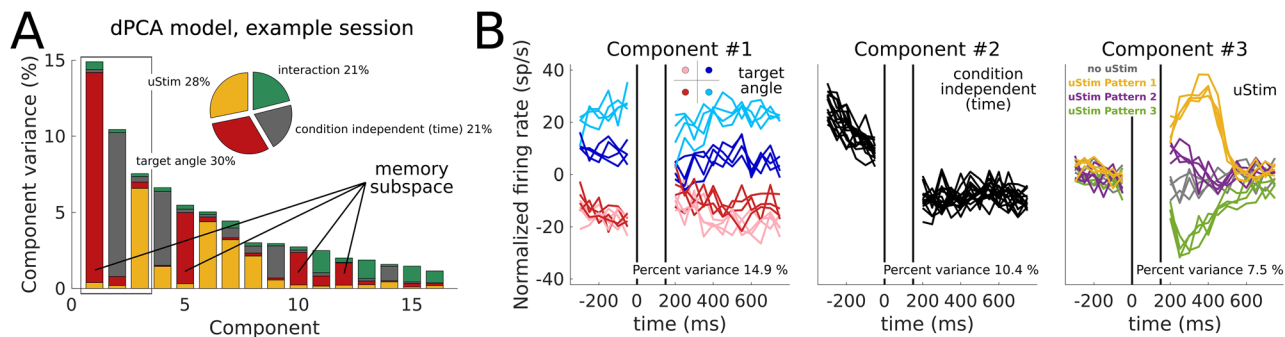


Figure 6. Memory subspace estimation using dPCA. **A**, Percentage of variance of the trial-averaged neural population activity captured by each dPCA dimension in an example session (Monkey W, session 20220810). The dPCA model was fit to find dimensions that primarily captured target angle, microstimulation or time-related variance, as well as variance due to their interaction (red, yellow, gray, and green colors, respectively). The top four dPCA target angle-related dimensions defined the memory subspace (dPCA dimensions 1, 5, 10, and 12). The memory dimensions captured ~30% of the total variance of the trial-averaged activity (inset, red portion of pie chart). **B**, Trial-averaged neural activity projected onto the first target angle, time, and microstimulation dPCs (left to right panels). These dimensions corresponded in this case to the top three dPCA dimensions, which capture most of the variance of the trial-averaged neural activity (panel **A**, large rectangle; 14.9%, 10.4%, and 7.5% of the variance). Each trace represents the trial-averaged neural activity of one target angle and microstimulation condition (each panel contains 16 traces in total, given 4 target angle conditions and 4 microstimulation conditions). In the left panel, the traces are colored based on target angle condition. In the right panel, they are colored based on microstimulation condition.

The memory subspace, on the contrary, captured working memory tuning that was stable over time (Fig. 5*B*, middle panel). The dPCA objective explicitly decorrelated (or “demixed”) dimensions with target angle variance from dimensions with microstimulation variance, as well as from dimensions with time-related variance (see Materials and Methods). The memory subspace was defined by the top dPCA target angle-related dimensions, which captured ~30% of the total variance of the trial-averaged activity (Fig. 6*A*, red bars; example session). Variance related to microstimulation and the passage of time was primarily captured by other dPCA dimensions (Fig. 6*A*, yellow and gray bars, respectively). The memory dimensions reflected target angle-related signals that were stable over time (Fig. 6*B*, left panel, first memory dimension; Fig. 5*B*, middle panel, memory dimensions 1 and 2). The other dPCA dimensions captured signals related to the passage of time (Fig. 6*B*, middle panel, first time dimension) and the effect of microstimulation (Fig. 6*B*, right panel, first microstimulation dimension).

We found that microstimulation strongly impacted activity in the dominant subspace (Fig. 5*A*, top and bottom left panels, target angles 45 and 315°; bottom right panel, all target angles). It need not be the case that the activity along the dominant dimensions would be modulated by the microstimulation, since these dimensions were found using no-microstimulation trials. The effect of microstimulation could be visualized as a push of the neural activity along specific directions (dashed yellow arrows) and away from its natural time course (gray trajectories, no-microstimulation). At microstimulation offset (t_{post}), the activity relaxed toward the end location of the trajectories under no-microstimulation conditions (yellow trajectories; t_{end} in no-microstimulation). In this dominant subspace, the effect of microstimulation was evident (dashed yellow arrows), and the activity did not always recover to its natural state by the end of the trial (top left panel, recovery, versus bottom left panel, incomplete recovery; t_{end} locations). These population-level effects match the observations made at the individual unit level, where we found strong microstimulation modulation at t_{post} and a persistent modulatory effect until the end of the trial for a subset of the units (Fig. 2). The dominant subspace captured ~50% of the target angle variance of the trial-averaged activity in held-out microstimulation trials (Monkey W: $50 \pm 13\%$, mean with 95% CI, $n = 10$ sessions; Monkey S: 53%, $n = 1$), between 5–20% of

microstimulation variance (Monkey W: $19 \pm 10\%$; Monkey S: 6%) and 25–40% of time-related variance (Monkey W: $44 \pm 16\%$; Monkey S: 25%).

Contrary to what was found in the dominant subspace (Fig. 5*A*), the activity in the memory subspace remained in roughly the same location throughout the entire delay period, in both microstimulation and no-microstimulation experiments (Fig. 5*B*). This location depended on the target angle condition (top vs bottom right panels, angles 45 and 315°; bottom left panel, all angles), indicating preserved target tuning over time. Modulation between t_{pre} and t_{post} was small (length of dashed arrows) and the activity recovered to its natural state by the end of the delay period (t_{end} , no-microstimulation). The memory subspace captured ~70% of the target angle variance of the trial-averaged activity in held-out microstimulation trials (Monkey W: $68 \pm 8\%$, mean with 95% CI, $n = 10$ sessions; Monkey S: 68%, $n = 1$; Materials and Methods), only ~3% of microstimulation variance (Monkey W: $4 \pm 2\%$; Monkey S: 2%), and 2% of time-related variance (Monkey W: $4 \pm 2\%$; Monkey S: 1%). The small amount of time-related and microstimulation-related variance confirms that the memory subspace contained stable working memory information throughout the delay that was minimally impacted by microstimulation.

Consistently across microstimulation experiments, population activity was strongly modulated by microstimulation in the dominant subspace, whereas it was minimally impacted by microstimulation in the memory subspace (Fig. 7). To see this, we quantified the distance between microstimulation and no-microstimulation activity at various times during the delay period. Activity after microstimulation deviated from no-microstimulation activity in the dominant subspace, particularly at t_{post} (Fig. 7*A*, green, same example experiment as in Fig. 5; Fig. 7*B*, top panels, green, mean with 95% CI across all experiments). This indicated that microstimulation pushed activity away from its natural state. After this push, the activity did not always recover by the end of the trial (Fig. 7*B*, bottom panels, green, % of experiments with significant activity modulation due to microstimulation in the dominant subspace; Monkey W, at t_{post} : 73%, at t_{end} : 15%, $n = 120$ microstimulation and target angle conditions; Monkey S, at t_{post} : 100%, at t_{end} : 19%, $n = 16$). In contrast, activity in the memory subspace remained minimally impacted by microstimulation consistently across

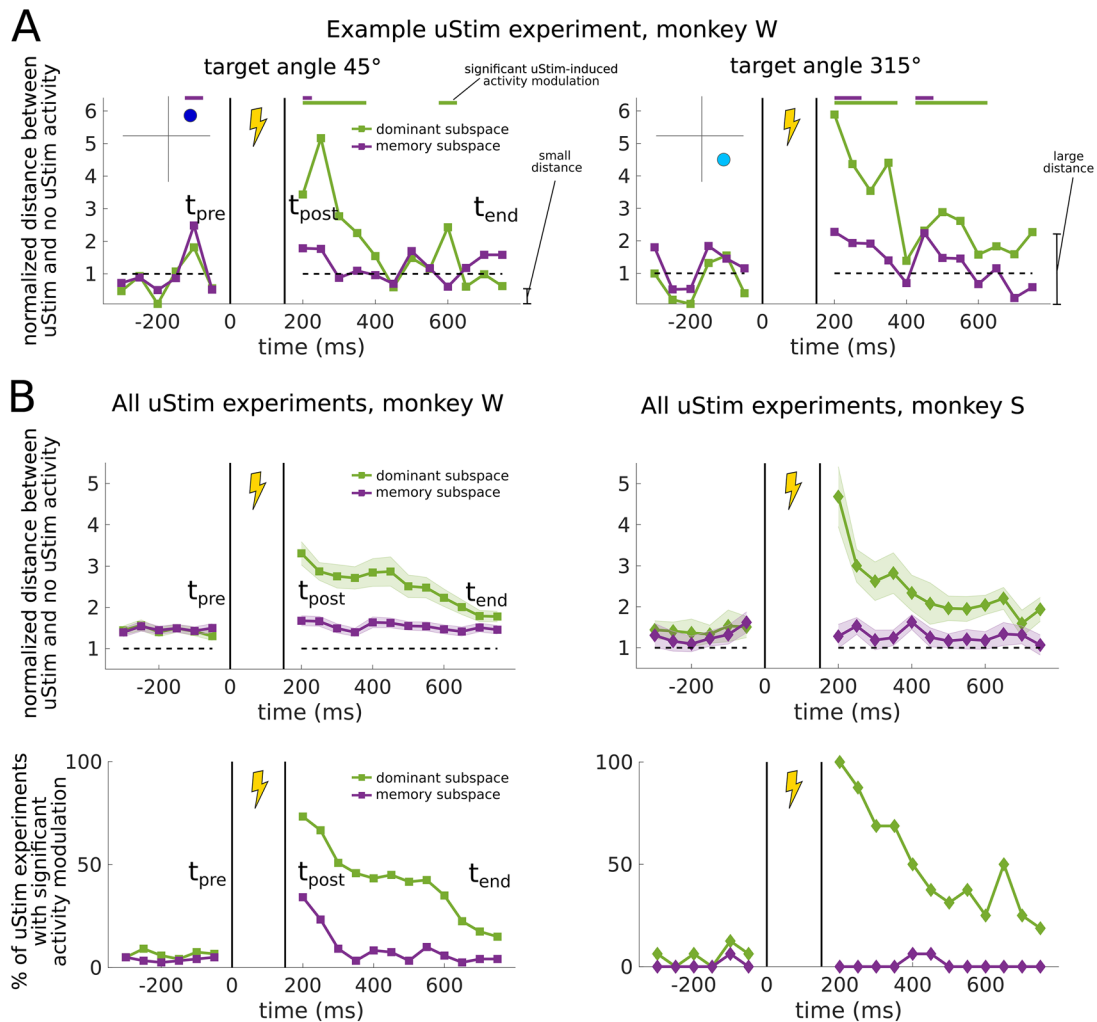


Figure 7. Microstimulation minimally impacts activity in the memory subspace. **A**, For an example microstimulation experiment, distance between microstimulation and no-microstimulation trajectories throughout the delay period in the dominant (green lines) and memory (purple lines) subspaces. Distances correspond to the trajectories shown in Figure 5 for target angles 45° (left panel) and 315° (right panel). Distances are normalized to the 95% confidence interval of the no-microstimulation trials at each time step (see Materials and Methods). Colored bars above panels indicate significant microstimulation-induced activity modulation in the dominant (green) and memory (purple) subspaces. For target angle 45°, activity in the dominant subspace recovers by t_{end} (distance is small). For target angle 315°, activity in the dominant subspace does not fully recover by t_{end} (distance is large). Monkey W, session 20220810. **B**, Top panels, Mean distance between microstimulation and no-microstimulation trajectories across all microstimulation experiments in the dominant (green lines) and memory (purple lines) subspaces. Error bars indicate 95% CI across all experiments (Monkey W, $n = 120$ microstimulation and target angle conditions; Monkey S, $n = 16$). Bottom panels, Percent of microstimulation experiments with significant microstimulation-induced activity modulation in the dominant (green lines) and memory (purple lines) subspaces ($p < 0.05$, two-tailed two-sample multivariate test). Activity in the dominant subspace is significantly modulated by microstimulation after microstimulation offset in many experiments, and in some experiments it remains modulated by t_{end} . In contrast, activity in the memory subspace quickly recovers from the perturbation in most experiments. In panel **A**, the quantifications were computed based on neural activity in the two leading dimensions of the dominant and the memory subspaces (to match the examples shown in Fig. 5). The quantifications in panel **B** were computed using all four dimensions for each of the subspaces.

microstimulation experiments (Fig. 7A, purple, same example experiment as in Fig. 5; Fig. 7B, top panels, purple, mean with 95% CI across all experiments), and the modulation was largely absent by the end of the trial (Fig. 7B, bottom panels, purple: % of experiments with significant activity modulation due to microstimulation in the memory subspace; Monkey W, 34% at t_{post} , 4% at t_{end} , $n = 120$; Monkey S, 0% at t_{post} , 0% at t_{end} , $n = 16$). This might explain the robustness of behavior after microstimulation, since working memory-related signals could be stably read out with the same decoder throughout the delay period.

To test this possibility, we attempted to extract working memory-related information from this memory subspace before and after microstimulation. Using a static classification method based on no-microstimulation activity from the memory subspace alone (see Materials and Methods), we could stably and accurately read out target angle information throughout

the delay period during microstimulation trials (classification accuracy: Monkey W, t_{pre} : $56 \pm 2\%$, t_{post} : $46 \pm 3\%$, t_{end} : $52 \pm 2\%$, mean with 95% CI, $n = 30$ microstimulation experiments; Monkey S, t_{pre} : $57 \pm 5\%$, t_{post} : $37 \pm 3\%$, t_{end} : $39 \pm 2\%$, mean with 95% CI, $n = 4$; above chance classification). It was also possible to read out working memory-related information in the dominant subspace with similar accuracy. This is because the dominant and memory subspaces partially overlapped, and as a consequence both subspaces contained signals related to working memory (minimum subspace angle = $60 \pm 3^\circ$, mean $\pm$ SD, $n = 55$ sessions and cross-validation folds, Monkey W; $63 \pm 2^\circ$, $n = 5$, Monkey S; defined as the smallest angle between the two 4D subspaces in high-d space). Importantly, classification accuracies obtained with the memory subspace-based static classification method were similar to those obtained using classifiers with time-varying parameters and which relied on all dimensions of

the data (Fig. 4). This meant that working memory-related signals could be stably read out with the same decoder throughout the delay period, making this information readily available to downstream areas (Parthasarathy et al., 2019).

In summary, we found that dynamic and stable working memory representations coexist in dlPFC and that microstimulation differentially impacts them. In dimensions where working memory signals naturally evolved over time, microstimulation strongly modulated neural activity. In contrast, in dimensions containing working memory signals that were stable over time, microstimulation minimally impacted neural activity, and working memory information could be stably read out throughout the delay period.

Discussion

Here, we studied working memory computations by directly perturbing delay period activity in the dlPFC using electrical microstimulation. We implanted monkeys with multielectrode arrays in dlPFC, which allowed us to stimulate with a variety of spatial microstimulation patterns while monitoring the effect on the recorded neural population. We found that microstimulation broadly affected the activity of individual neurons in dlPFC, including changes to the tuning strength of individual neurons that displayed working memory-related activity. However, we found minimal impact on the ability of the animals to correctly perform the task. This apparent contradiction was reconciled at the population level, where we found that working memory information was preserved in dlPFC after microstimulation and that the information could be stably read out from a specific “memory subspace” of the neural population activity. Our findings indicate that working memory exhibits robustness to microstimulation perturbations in the dlPFC.

PFC might be endowed with circuit properties that grant working memory signals stability to diverse activity perturbations. For example, a previous study also found a stable memory subspace that was not affected by perturbations from sensory

distractors (Parthasarathy et al., 2019). Cognitive-related signals that are not disrupted by external stimuli or contextual events have also been found in PFC during executive control functions other than working memory, such as categorical reasoning (Freedman et al., 2003; Cromer et al., 2011), rule-based decision-making (Mante et al., 2013; Siegel et al., 2015), and selective attention (Snyder et al., 2021). This feature might be an intrinsic and general property of high order areas such as PFC, which is required to form and maintain stable cognitive states to guide behavior (Snyder et al., 2021).

Working memory information could be robustly and stably maintained within PFC in the presence of other evolving signals by representing stable and dynamic information in different subspaces of the neural population activity (Druckmann and Chklovskii, 2012; Murray et al., 2017). We found such subspace dissociation of stable and dynamic working memory signals in the dlPFC population (Fig. 8*A*, signals in memory vs dominant subspaces). Importantly, while the “stable” memory subspace was minimally affected by microstimulation, the “dynamic” dominant subspace was strongly modulated by microstimulation (Fig. 8*B*). The separation of stable and dynamic variables in different subspaces of dlPFC’s neural activity might underlie working memory robustness to sensory perturbations (Parthasarathy et al., 2019), as well as to microstimulation perturbations. However, the dominant and memory subspaces overlapped along some dimensions (Fig. 8), which indicates that the two subspaces shared some signals. This could reflect the interaction of time-varying attentional and motor preparation processes with stable working memory representations (Parthasarathy et al., 2019). Alternative subspace dissociation mechanisms involve completely segregating different signals in orthogonal subspaces. This has been proposed in the motor cortex as a mechanism to separate preparatory activity from movement activity so that movements are not prematurely generated during motor planning (Kaufman et al., 2014; Elsayed et al., 2016). Related orthogonal mechanisms might be in place to prevent cognitive and arousal-related signals

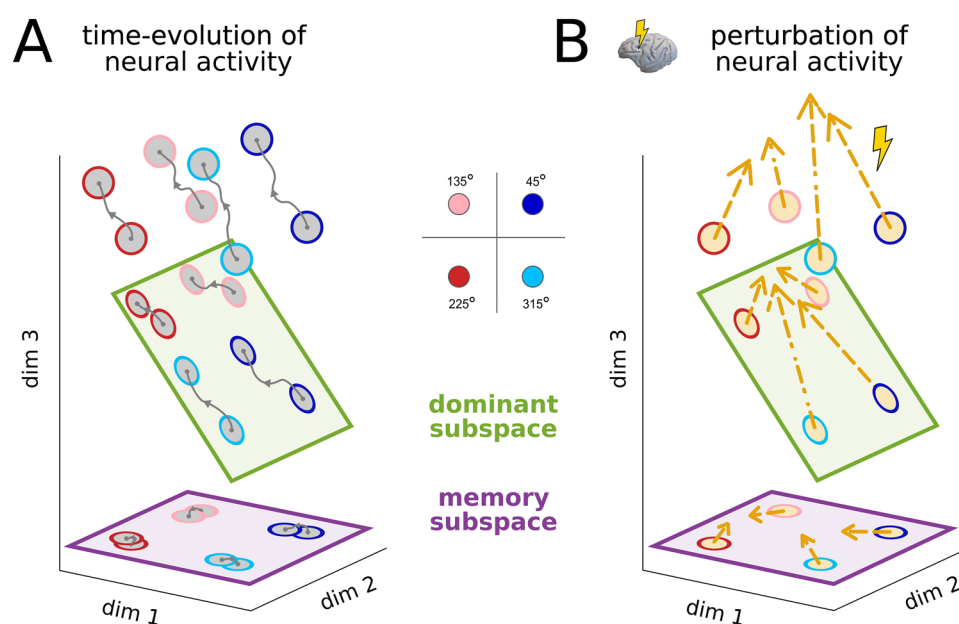


Figure 8. Population-level working memory representations in dlPFC for robustness to perturbations. **A**, Organization of the dominant and memory subspaces in dlPFC’s population activity space. The dominant subspace (green) captures the time evolution of working memory-related signals, whereas the memory subspace (purple) reflects the stable component of the working memory signals. **B**, Microstimulation perturbations push activity along different directions in neural population activity space. The perturbations are strongly reflected in the dominant subspace (i.e., the perturbations have a substantial projection onto this subspace), but they are minimally reflected in the memory subspace (i.e., the perturbations are largely orthogonal to this subspace).

from influencing motor responses (Johnston and Smith, 2024) or value-related signals from prematurely driving choice (Yoo and Hayden, 2020).

Several factors might have contributed to the lack of impact of stimulation on behavior (Fig. 3). First, the time of stimulation during the delay period might play a role. Our stimulation was in the middle of the delay period. In PFC, stimulation during delayed response tasks might be most effective at the beginning of a delay period, when the memory has just been encoded and motor preparation is minimal (Stamm, 1969). In the motor cortex, the opposite is found: stimulation is more effective later in the delay period, closer to movement execution (Churchland and Shenoy, 2007). Second, increasing the task difficulty (e.g., by adding more targets or by increasing the delay duration) could make it easier for microstimulation to impact the memory network and, consequently, behavior. Third, PFC is part of a larger network responsible for various executive control functions, including working memory, that comprises multiple brain areas (Fuster, 2015). The stimulation was local within this network, which might not be enough to disrupt this large-scale and resilient recurrent network (Li et al., 2016). Finally, it is also possible that PFC might be responsible for executive functions supporting working memory, but not storing the memory per se (Lebedev et al., 2004; Sreenivasan et al., 2014; Lara and Wallis, 2015; Mackey et al., 2016). Regardless of whether the working memory-related representations found in dlPFC reflect a memory or a process that supports working memory, our results indicate that such representations are robust to microstimulation perturbations within the dlPFC network.

Several studies that perturbed neural activity in dlPFC during the delay period of working memory tasks were able to induce stronger effects on behavior than the effects reported here. These studies involved longer-lasting activity manipulations than the type we performed, e.g., microstimulation throughout the entire delay period (Opris et al., 2005a), large-scale lesion (Funahashi et al., 1993), cooling (Bauer and Fuster, 1976), or pharmacological inactivations (Sawaguchi and Iba, 2001). Another related study that showed effects on behavior during working memory applied microstimulation in the frontal eye fields but also throughout the whole delay period (Opris et al., 2005b). In visually guided tasks, brief microstimulation perturbations during saccade planning have been shown to impact eye movements when applied to the lateral PFC (Wegener et al., 2008) and the dorsomedial PFC (Yang et al., 2008).

In contrast, recent studies that performed transient inactivations of PFC and motor areas during delayed response tasks minimally impacted behavior. In a working memory task, brief optogenetic inactivation of the lateral PFC in monkeys did not impact working memory-related signals and behavior (Mendoza-Halliday et al., 2023). In a delayed motor task, transient optogenetic inactivation of mouse premotor cortex (ALM) did not impact behavior (Li et al., 2016; Inagaki et al., 2019). Similarly, in a delayed motor task, transient optogenetic stimulation of monkey motor and premotor cortices produced no behavioral impact, though electrical microstimulation minimally influenced behavior (O'Shea et al., 2022). These studies found that task-relevant signals during the delay period were not affected (Mendoza-Halliday et al., 2023) or quickly recovered (Li et al., 2016; Inagaki et al., 2019) from such transient manipulations, even though activity was strongly modulated by the stimulation. Similarly, we found task-relevant signals that were either not affected by the microstimulation or that tended to recover from it (Figs. 2, 5, and 7).

Taken together, this body of work points to possible explanations for how robustness to perturbations is maintained in neural circuits. One possibility is that task-relevant information is affected by the stimulation but that compensatory mechanisms are in place that restore this information, potentially through redundancy (e.g., across hemispheres as in Li et al., 2016). A second possibility is that task-relevant information is affected, but not completely disrupted, and that the information is still accessible in certain subspaces of the same neural population (Murray et al., 2017; Parthasarathy et al., 2019). A third possibility, perhaps related to the second, is that task-relevant information is affected but that the dimensions in which this information resides do not align with the dimensions of the activity that are truly consequential for behavior (O'Shea et al., 2022). Future work involving additional perturbations and monitoring of neural activity more broadly across cortical areas will be necessary to resolve these potential explanations.

A way by which the brain may retain robustness in the face of perturbation is through an attractor network. Attractors are robust to brief and modest perturbations, such as noise, sensory distractors, or direct activity manipulations (Wang, 2021). There are various attractor networks that might underlie the maintenance of working memory-related information in dlPFC and that could implement the possibilities discussed above. First, attractor networks can be instantiated by recurrent dynamics supported by specific patterns of network connectivity (Compte et al., 2000; Li et al., 2016; Murray et al., 2017; Inagaki et al., 2019; Parthasarathy et al., 2019; Zhou et al., 2023; Stroud et al., 2024). Second, attractor networks can incorporate an “activity-silent” mechanism (Mongillo et al., 2008; Stokes, 2015) via short-term synaptic plasticity to maintain a short-term memory even in the absence of changes in neural activity (Barbosa et al., 2020). Synaptic weights might be less prone to interference from sensory distractors (Miller et al., 2018; but see Wang, 2021), so the same plasticity-based mechanism could mediate the robustness to microstimulation perturbations. Third, attractor networks can exhibit oscillatory activity (Compte et al., 2000), which has been observed during working memory (Pesaran et al., 2002) and might contribute in a frequency-specific manner to gating of sensory information (Miller et al., 2018). This mechanism might mediate the suppression of distractors (Bonnefond and Jensen, 2012) and electrical microstimulation.

The work presented here aimed to understand the response of PFC to microstimulation perturbations and found that working memory signals were robust to such perturbations. An important goal for future work is to design perturbations that can selectively impact working memory. One possible approach is to design stimulation experiments to specifically impact targeted neurons and behavior (Moore and Fallah, 2004). Another approach is to optimize stimulation patterns to produce customized effects on neural populations and behavior (Tafazoli et al., 2020; Nejatbakhsh et al., 2023; Minai et al., 2024). Understanding the ways in which populations of neurons maintain robust cognitive states, along with the development of tools to customize perturbations of neural populations to achieve specific targeted states, is one of the important goals of modern systems neuroscience.

Data and Code Availability

The data that support the findings of this study are publicly available in Zenodo at <https://doi.org/10.5281/zenodo.15640851>. The code to reproduce the analysis is openly available in KiltHub at

<https://doi.org/10.1184/R1/29293772.v1>. All versions of the code, including new releases, can be found in GitHub at https://github.com/jsoldadomagraner/PFC_uStim.

References

- Aslan B, Zech G (2005) New test for the multivariate two-sample problem based on the concept of minimum energy. *J Stat Comput Simul* 75:109–119.
- Barbosa J, Stein H, Martinez RL, Galan-Gadea A, Li S, Dalmau J, Adam KCS, 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. *Nat Neurosci* 23:1016–1024.
- Bauer RH, Fuster JM (1976) Delayed-matching and delayed-response deficit from cooling dorsolateral prefrontal cortex in monkeys. *J Comp Physiol Psychol* 90:293–302.
- Bonnefond M, Jensen O (2012) Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Curr Biol* 22:1969–1974.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Bruce CJ, Goldberg ME, Bushnell MC, Stanton GB (1985) Primate frontal eye fields II. Physiological and anatomical correlates of electrically evoked eye movements. *J Neurophysiol* 54:714–734.
- Buckley MJ, Mansouri FA, Hoda H, Mahboubi M, Browning PGF, Kwok SC, Phillips A, Tanaka K (2009) Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science* 325:52–58.
- Cavanagh SE, Towers JP, Wallis JD, Hunt LT, Kennerley SW (2018) Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nat Commun* 9:3498.
- Churchland MM, Shenoy KV (2007) Delay of movement caused by disruption of cortical preparatory activity. *J Neurophysiol* 97:348–359.
- Compte A, Brunel N, Goldman-Rakic PS, Wang X-J (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb Cortex* 10:910–923.
- Constantinidis C, Funahashi S, Lee D, Murray JD, Qi X-L, Wang M, Arnsten AFT (2018) Persistent spiking activity underlies working memory. *J Neurosci* 38:7020–7028.
- Cromer JA, Roy JE, Buschman TJ, Miller EK (2011) Comparison of primate prefrontal and premotor cortex neuronal activity during visual categorization. *J Cogn Neurosci* 23:3355–3365.
- Druckmann S, Chklovskii DB (2012) Neuronal circuits underlying persistent representations despite time varying activity. *Curr Biol* 22:2095–2103.
- Elsayed GF, Lara AH, Kaufman MT, Churchland MM, Cunningham JP (2016) Reorganization between preparatory and movement population responses in motor cortex. *Nat Commun* 7:13239.
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci* 23:5235–5246.
- Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol* 61:331–349.
- Funahashi S, Bruce CJ, Goldman-Rakic PS (1993) Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic “scotomas”. *J Neurosci* 13:1479–1497.
- Fuster J (2015) *The prefrontal cortex*. London: Elsevier.
- Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. *Science* 173:652–654.
- Goldman-Rakic PS (1995) Cellular basis of working memory. *Neuron* 14:477–485.
- Inagaki HK, Fontolan L, Romani S, Svoboda K (2019) Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* 566:212–217.
- Issar D, Williamson RC, Khanna SB, Smith MA (2020) A neural network for online spike classification that improves decoding accuracy. *J Neurophysiol* 123:1472–1485.
- Johnston R, Smith MA (2024) Brain-wide arousal signals are segregated from movement planning in the superior colliculus. *eLife* 13:RP99278.
- Katsuki F, Constantinidis C (2012) Unique and shared roles of the posterior parietal and dorsolateral prefrontal cortex in cognitive functions. *Front Integr Neurosci* 6:17.
- Kaufman MT, Churchland MM, Ryu SI, Shenoy KV (2014) Cortical activity in the null space: permitting preparation without movement. *Nat Neurosci* 17:440–448.
- Khanna SB, Scott JA, Smith MA (2020) Dynamic shifts of visual and saccadic signals in prefrontal cortical regions 8Ar and FEF. *J Neurophysiol* 124:1774–1791.
- Kleiner M, Brainard D, Pelli D (2007) What's new in Psychtoolbox-3? (Abstract). *Perception* 36 ECVP Abstract Supplement.
- Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi X-L, Romo R, Uchida N, Machens CK (2016) Demixed principal component analysis of neural population data. *eLife* 5:e10989.
- Kumaravelu K, Grill WM (2024) Neural mechanisms of the temporal response of cortical neurons to intracortical microstimulation. *Brain Stimulat* 17:365–381.
- Lara AH, Wallis JD (2015) The role of prefrontal cortex in working memory: a mini review. *Front Syst Neurosci* 9:173.
- Leavitt ML, Pieper F, Sachs AJ, Martinez-Trujillo JC (2018) A quadrantic bias in prefrontal representation of visual-mnemonic space. *Cereb Cortex* 28:2405–2421.
- Lebedev MA, Messinger A, Kralik JD, Wise SP (2004) Representation of attended versus remembered locations in prefrontal cortex. *PLoS Biol* 2:e365.
- Li N, Dae K, Svoboda K, Druckmann S (2016) Robust neuronal dynamics in premotor cortex during motor planning. *Nature* 532:459–464.
- Lorenc ES, Mallett R, Lewis-Peacock JA (2021) Distraction in visual working memory: resistance is not futile. *Trends Cogn Sci* 25:228–239.
- Lundqvist M, Herman P, Miller EK (2018) Working memory: delay activity, yes! Persistent activity? Maybe not. *J Neurosci* 38:7013–7019.
- Mackey WE, Devinsky O, Doyle WK, Meager MR, Curtis CE (2016) Human dorsolateral prefrontal cortex is not necessary for spatial working memory. *J Neurosci* 36:2847–2856.
- Mante V, Sussillo D, Shenoy KV, Newsome WT (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503:78–84.
- Mendoza-Halliday D, Xu H, Azevedo FAC, Desimone R (2023) Dissociable neuronal substrates of visual feature attention and working memory. *BioRxiv Prepr Serv Biol*:2023.03.01.530719.
- Miller EK, Lundqvist M, Bastos AM (2018) Working memory 2.0. *Neuron* 100:463–475.
- Minai Y, Soldado-Magraner J, Smith MA, Yu BM (2024) MiSO: optimizing brain stimulation to create neural activity states. Available at: <https://neurips.cc/virtual/2024/poster/95894> [Accessed Nov. 15, 2024].
- Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. *Science* 319:1543–1546.
- Moore T, Fallah M (2004) Microstimulation of the frontal eye field and its effects on covert spatial attention. *J Neurophysiol* 91:152–162.
- Murphy DK, Maunsell JHR (2008) Electrical microstimulation thresholds for behavioral detection and saccades in monkey frontal eye fields. *Proc Natl Acad Sci USA* 105:7315–7320.
- Murray JD, Bernacchia A, Roy NA, Constantinidis C, Romo R, Wang X-J (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc Natl Acad Sci USA* 114:394–399.
- Nejatbakhsh A, Fumarola F, Esteki S, Toyozumi T, Kiani R, Mazzucato L (2023) Predicting the effect of micro-stimulation on macaque prefrontal activity based on spontaneous circuit dynamics. *Phys Rev Res* 5:043211.
- Opris I, Barborica A, Ferrera VP (2005a) Microstimulation of the dorsolateral prefrontal cortex biases saccade target selection. *J Cogn Neurosci* 17:893–904.
- Opris I, Barborica A, Ferrera VP (2005b) Effects of electrical microstimulation in monkey frontal eye field on saccades to remembered targets. *Vision Res* 45:3414–3429.
- O'Shea DJ, et al. (2022) Direct neural perturbations reveal a dynamical mechanism for robust computation. :2022.12.16.520768. Available at: <https://www.biorxiv.org/content/10.1101/2022.12.16.520768v1> [Accessed Jan. 9, 2023].
- Parthasarathy A, Herikstad R, Bong JH, Medina FS, Libedinsky C, Yen S-C (2017) Mixed selectivity morphs population codes in prefrontal cortex. *Nat Neurosci* 20:1770–1779.
- Parthasarathy A, Tang C, Herikstad R, Cheong LF, Yen S-C, Libedinsky C (2019) Time-invariant working memory representations in the presence of code-morphing in the lateral prefrontal cortex. *Nat Commun* 10:4995.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442.

- Pesaran B, Pezaris JS, Sahani M, Mitra PP, Andersen RA (2002) Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat Neurosci* 5:805–811.
- Sadtler PT, Ryu SI, Tyler-Kabara EC, Yu BM, Batista AP (2015) Brain-computer interface control along instructed paths. *J Neural Eng* 12:016015.
- Santhanam G, Yu BM, Gilja V, Ryu SI, Afshar A, Sahani M, Shenoy KV (2009) Factor-analysis methods for higher-performance neural prostheses. *J Neurophysiol* 102:1315–1330.
- Sawaguchi T, Iba M (2001) Prefrontal cortical representation of visuospatial working memory in monkeys examined by local inactivation with muscimol. *J Neurophysiol* 86:2041–2053.
- Siegel M, Buschman TJ, Miller EK (2015) Cortical information flow during flexible sensorimotor decisions. *Science* 348:1352–1355.
- Snyder AC, Yu BM, Smith MA (2021) A stable population code for attention in prefrontal cortex leads a dynamic attention code in visual cortex. *J Neurosci* 41:9163–9176.
- Sreenivasan KK, Curtis CE, D'Esposito M (2014) Revisiting the role of persistent neural activity during working memory. *Trends Cogn Sci* 18:82–89.
- Stamm JS (1969) Electrical stimulation of monkeys' prefrontal cortex during delayed-response performance. *J Comp Physiol Psychol* 67:535–546.
- Stokes MG (2015) "Activity-silent" working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci* 19:394–405.
- Stroud JP, Duncan J, Lengyel M (2024) The computational foundations of dynamic coding in working memory. *Trends Cogn Sci* 28:614–627.
- Suzuki M, Gottlieb J (2013) Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. *Nat Neurosci* 16:98–104.
- Tafazoli S, MacDowell CJ, Che Z, Letai KC, Steinhardt CR, Buschman TJ (2020) Learning to control the brain through adaptive closed-loop patterned stimulation. *J Neural Eng* 17:056007.
- Wang X-J (2021) 50 years of mnemonic persistent activity: quo vadis? *Trends Neurosci* 44:888–902.
- Wang XJ (2001) Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci* 24:455–463.
- Wegener SP, Johnston K, Everling S (2008) Microstimulation of monkey dorsolateral prefrontal cortex impairs antisaccade performance. *Exp Brain Res* 190:463–473.
- Yang S, Heinen SJ, Missal M (2008) The effects of microstimulation of the dorsomedial frontal cortex on saccade latency. *J Neurophysiol* 99:1857–1870.
- Yoo SBM, Hayden BY (2020) The transition from evaluation to selection involves neural subspace reorganization in core reward regions. *Neuron* 105:712–724.e4.
- Zhou S, Seay M, Taxidis J, Golshani P, Buonomano DV (2023) Multiplexing working memory and time in the trajectories of neural networks. *Nat Hum Behav* 7:1170–1184.