OXFORD

Cerebral Cortex, 2021;00: 1–14

doi: 10.1093/cercor/bhab079
Original Article

ORIGINAL ARTICLE

Primate Spatial Memory Cells Become Tuned Early and Lose Tuning at Cell-Specific Times

Charalampos Papadimitriou[1,†], Charles D. Holmes[1,2,†] and Lawrence H. Snyder[1,2]

[1]Department of Neuroscience, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA and [2]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

Address correspondence to Charalampos Papadimitriou. Email: papadimitriou.c@gmail.com.

[†]C.P. and C.D.H. are co-first authors

Abstract

Working memory, the ability to maintain and transform information, is critical for cognition. Spatial working memory is particularly well studied. The premier model for spatial memory is the continuous attractor network, which posits that cells maintain constant activity over memory periods. Alternative models propose complex dynamics that result in a variety of cell activity time courses. We recorded from neurons in the frontal eye fields and dorsolateral prefrontal cortex of 2 macaques during long (5–15 s) memory periods. We found that memory cells turn on early after stimulus presentation, sustain activity for distinct and fixed lengths of time, then turn off and stay off for the remainder of the memory period. These dynamics are more complex than the dynamics of a canonical bump attractor network model (either decaying or nondecaying) but more constrained than the dynamics of fully heterogeneous memory models. We speculate that memory may be supported by multiple attractor networks working in parallel, with each network having its own characteristic mean turn-off time such that mnemonic resources are gradually freed up over time.

Key words: frontal eye fields, macaque, prefrontal cortex, working memory

Introduction

Working memory is the ability to actively maintain and transform information on the order of seconds. Most cognitive tasks rely on working memory. Many studies address the neural circuits that support working memory and have implicated prefrontal cortex (PFC) as a key locus. In this study, we focus on spatial working memory because spatial location is fundamental, continuous, and easily quantified. In monkeys, firing rates of neurons in the frontal eye fields (FEF) and dorsolateral prefrontal cortex (dlPFC), 2 prefrontal areas, elevate while subjects hold a spatial location in memory (Fuster and Alexander 1971; Kojima and Goldman-Rakic 1982; Bruce and Goldberg 1985; Funahashi et al. 1989, 1993; Pellegrino and Wise 1993; Chafee and Goldman-Rakic 1998; Ferrera et al. 1999; Constantinidis et al. 2001; Sommer and Wurtz 2001; Umeno and Goldberg 2001; Takeda and Funahashi 2002, 2004).

Much of the electrophysiology literature on spatial working memory is based on tasks with memory periods of ∼1–3 s. Elevated responses observed in these tasks are sustained, often with little or no decay, for the entirety of the memory period. These results have inspired models of neural bump attractor networks, the premier framework for working memory circuits (Amit 1992; Brunel 1996; Amit and Brunel 1997; Compte et al. 2000; Wang 2009). Bump attractor networks model spatial memory as a topographic map of nodes with local recurrent excitation and global recurrent inhibition. A memorized location is represented by the center of a "bump" of elevated activity that is distributed across nodes. The bump can be maintained indefinitely, even after the original stimulus is removed, due to a balance between the excitatory and inhibitory connections between nodes. Once formed, the amplitude and shape of the bump do not change over time unless and until the entire circuit is reset, at which point the bump disappears entirely. However,

the bump can drift slowly and randomly over time, maintaining its shape but contributing to a slow decay in memory accuracy (Compte et al. 2000).

Other studies have demonstrated more complex dynamics (Barak et al. 2013; Stokes et al. 2013; Murray et al. 2017; for a review, see Lundqvist et al. 2018). Rather than a single steady, sustained response that is repeated across all cells, these studies argue that the activities of individual cells can differ widely from one cell to the next. Some studies describe cells with upward or downward ramps of activity (Brody et al. 2003; Jun et al. 2010). Other describe more extreme dynamics in which individual cells activate only briefly, for example, 10s to 100s of milliseconds (Baeg et al. 2003; Harvey et al. 2012). When properly read out, these complex dynamics can nonetheless provide a continuous and robust memory trace (Goldman 2009; Murray et al. 2017). This is reminiscent of a liquid state machine, in which the responses of individual spiking neurons can be completely dissociated from the time course of the memory itself.

In this study, we investigated the dynamics of spatial working memory responses in PFC during memory periods of up to 15 s. Multiple memory locations (targets) continuously distributed in space were tested to ensure optimal excitation. Electrodes were continuously monitored and adjusted to maintain good isolation. We find that most memory cells lose their tuned memory activity before the end of a 15-s memory period. Surprisingly, the times at which tuning was lost (turn-off times) show greater consistency within individual cells compared with across cells. In other words, individual cells have characteristic mean turn-off times. Furthermore, once turned off, cells do not turn back on. Thus, the dynamics of memory activity over periods greater than a few seconds are more complex than those predicted by a canonical bump attractor network or even a decaying bump network but less complex than those predicted by fully heterogeneous models of memory such as liquid state machines.

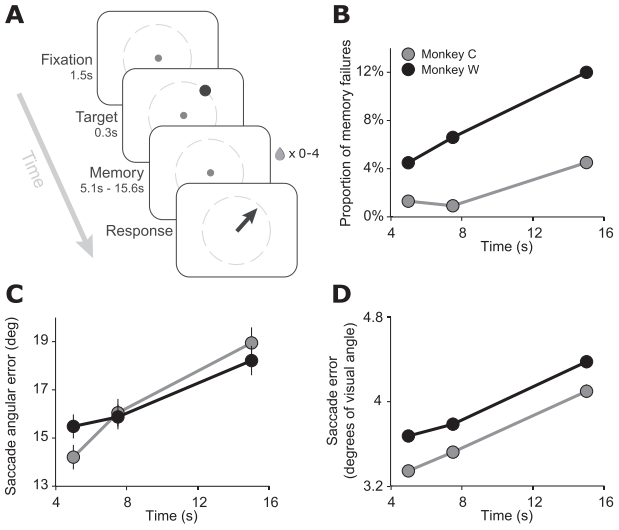

Figure 1. Memory task and performance. (*A*) The task begins with 1.5 s of central fixation. A peripheral stimulus turns on for 300 ms and is then extinguished. Memory targets can appear anywhere along a circle with a radius of 12 and 15 deg for monkeys C and W, respectively. After a memory period of 5.1–15.6 s, the subject makes a saccadic response to the remembered location. To encourage the animals to fixate through a long delay, up to 4 mid-trial rewards were delivered during the memory period (see Materials and Methods). (*B*) Proportion of memory failures, that is, memory-guided saccades directed >80 degrees of arc from the target, divided by the number of trials in which fixation was maintained up until the go cue. Memory failures are plotted as a function of memory period length for each monkey (Monkey C—gray, Monkey W—black). (*C*) The mean angular error of saccadic responses as a function of memory period length, excluding memory failures. Error bars are standard error. (*D*) The mean Euclidean error of saccadic responses as a function of memory period length. Standard errors in (*B*) and (*D*) are smaller than the data points themselves. Trials with endpoints >80 deg from the target (memory failure trials) are excluded from (*C*) and (*D*). The sets of 3 data points in (*B*–*D*) are not cumulative, but instead represent results from just the 5, 7.5, and 15-s trials, respectively.

Materials and Methods

Two cynomolgus macaques (*Macaca fascicularis*), C and W, were trained on a center-out memory-guided saccade task. During the task subjects sat in a completely dark room in a primate chair, head-fixed in a straight-ahead position and facing a screen located 30 cm away. Visual stimuli were controlled with custom software and projected using a CRT projector. Eye position was recorded using infrared video eye-tracking system (ISCAN, MA). The data and code used for analysis are available upon request. All procedures conformed to the Guide for the Care and Use of Laboratory Animals and were approved by the Washington University Institutional Animal Care and Use Committee.

Behavioral Task

In the memory-guided saccade task monkeys were required to remember a peripheral spatial location (Fig. 1A). Each trial began with the presentation of a fixation point on which monkeys had to fixate within 3.3 degrees of visual angle (dva). After 1.5 s, a peripheral memory target was flashed for 300 ms at a random location on a circle with a radius of 12 and 15 dva, for monkeys C and W, respectively, centered on the fovea. Stimulus presentation was followed by a memory period that lasted for 5.1–5.6 (5-s trials), 7.6–8.1 (7.5-s trials), or 15.6–16.1 s (15-s trials), during which time the subject maintained fixation while remembering the location of the flashed stimulus. Subjects received up to 4 small rewards during the memory period (Supplementary Fig. S1). The number of rewards changed over the course of data collection but was constant during any one day. At the end of the memory period, the fixation point disappeared, cueing the subject to shift their gaze to the remembered location. If the initial memory-guided saccade landed within 5.5 dva of the target, the subject received an immediate reward. This reward encouraged, but did not necessitate, precise mnemonic behavior. The memory target reappeared 300 ms after the initial saccade. The subject was then required to make a corrective saccade to within 3.5 dva of the visible target in order to receive a final large reward. Animals were allowed to blink throughout the task, including during the memory period. Blinks were detected when the pupil was at least 80% occluded. Eye positions detected during blinks were removed and interpolated based on eye positions just before and after the blink.

Behavioral Performance

We computed several measures of performance. Trials were immediately aborted and labeled as a "fixation break" if the animal's eye position moved >5 dva from the fixation point before the go cue. Of the remaining trials, those in which the memory-guided saccade landed >80 degrees of arc from the memory target were classified as "memory failures." Both fixation breaks

and memory failures were excluded from all subsequent behavioral and unit analyses, with the exception of the analysis shown in Figure 1B and Supplementary Figure S1.

We operationalized the animals' behavioral precision by computing angular error of the memory-guided saccades. Specifically, angular error was computed by subtracting the angle of the memory target from the angle of the memory-guided saccade. Positive and negative angular errors demonstrate the saccade was counter-clockwise and clockwise from the memory target position, respectively. To quantify precision, we computed the mean of the absolute value of saccadic angular error (Fig. 1C). We used a similar process to with Euclidean error, that is, the error in 2 dimensions (Fig. 1D).

We estimated the proportion of trials in which the animal completely forgot the target and was forced to guess (guess rate) by fitting a mixed probability model (Bays et al. 2009; Zhang and Luck 2009). The model describes the probability density function of the saccadic angular error, θ_{err}, as:

$$p\left(\theta_{err}\right) = (1-\gamma) f\left(\theta_{err}|\kappa\right) + \gamma \frac{1}{2\pi}$$

where $f\left(\theta|\kappa\right)$ is a von Mises distribution with a mean of zero and a concentration parameter κ, and γ is the guess rate. The model was fit to empirical saccadic angular errors with a maximum likelihood estimation paradigm optimized with a bounded limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm (Virtanen et al. 2020). A guess rate estimate was made for each session. For our purposes, only data from 15-s trials were considered.

We assessed whether the animals simplified spatial representations, for example, remembering which quadrant the target was in rather than its precise position. Such simplification might lead to systematic drift of spatial representations, that is, representations shift from target position to the simplified spatial direction. To test for such an effect, we binned trials by target direction (32 bins, each 11.25 deg). For each bin, we computed the direction for the mean saccadic endpoint (Supplementary Fig. S4). For this analysis, we separately considered each animal and each delay length (5, 7.5, or 15 s).

Electrophysiology

In each experimental session, 1 to 4 electrodes (AlphaOmega) were lowered into the FEF and/or dlPFC. Some time was dedicated to manually isolating as many single units as possible across all electrodes. Once a sufficient number of electrodes had isolated single units, neuronal memory tuning was assessed with a memory-guided saccade task with a 1.5-s delay. If at least one cell exhibited spatial tuning during the delay, all the isolated units (including untuned units) were selected for recording. Otherwise, all electrodes were lowered further and these search procedures were restarted. If no spatially tuned units could be found before the animal performed too many search trials (inferring the animal would perform too few long-duration trials) and one or more untuned cells were isolated, those units were selected for recording. Untuned cells are only included in this report if they were recorded from a track from which at least one cell with mnemonic spatial tuning (as verified by offline analysis) was recorded. In total, 161 cells were recorded and included in our analyses.

Sites were considered FEF sites if they lay within 250 microns of a site at which saccades could be evoked with 50 microamps of current or less (Bruce and Goldberg 1985). All other sites anterior to FEF were classified as dlPFC (see Fig. 8 and Supplementary Fig. S6 for locations of recording tracks).

Memory Tuning

We classified our cells into those that showed memory tuning in the early memory period and those that did not. We first identified the preferred direction by fitting a cosine function to the data. Next, we pooled trials into 2 groups—preferred direction trials, in which the target appeared within ±33.75 deg of the preferred direction, and null direction trials, in which the target appeared within ±33.75 deg of the direction diametrically opposite the preferred direction. Trials in which the target appeared outside these 2 ranges were excluded from this analysis. Cells were considered "tuned" if their firing rate was greater for preferred compared with null direction trials (*t*-test, $P < 0.05$) in either of 2 early memory periods (0.5–1.5 or 2–4 s after target onset). We found 70 cells that were significantly tuned in at least the first interval and 23 cells that were significantly tuned in the second interval but not the first interval, for a total of 93 tuned cells. The 68 cells with $P > 0.05$ in both intervals were classified as untuned, though 15 of these "untuned" cells (in the memory period) were tuned during stimulus presentation (50–300 ms after target onset). Using different or additional early memory intervals had only minor effects on these classifications.

We assessed the evolution of directional tuning over 15 s of memory at the population level by computing mean tuning curves (Fig. 3A–E) at an array of intervals. Across all cells, preferred directions were aligned to 0 deg. Target locations were grouped into 45 deg bins. Population-averaged tuning curves were estimated by fitting von Mises functions to the mean binned firing rates over particular intervals (the 2 early memory intervals—0.5–1.5 and 2–4 s—and 3 intervals corresponding to the end of our 3 delay lengths—3–5, 6–7.5, and 12–15 s). Confidence intervals (CIs) for fit statistics were computed with a bias-corrected and accelerated bootstrap (Sheppard et al. 2020).

Tuning Changes and Decay

We modeled how the random drift of a bump in an attractor circuit that does not decay would affect firing rates in single-unit recordings, and in particular, if random drift might account for the drop in tuning we observed in our memory cells (see Results). We assumed that the observed behavioral error was an exact readout of bump drift, that is, the amount of error on any one trial indicates the amount of drift on that trial. Note that this assumption likely overestimates the actual drift, but since we ultimately conclude that drift is too small to account for the observed change in neuronal activity (see Results), the assumption is conservative, assuming that other sources also contribute to the behavioral error would only strengthen our conclusion. We separately considered 5, 7.5, and 15-s trials. For each trial we shifted the early memory tuning curve (Fig. 3A) by an amount equal to the saccadic angular error of that trial and then sampled the firing rate at a random target location. As with our observed data, we fit our simulated data with von Mises functions. The amplitude of these fits provides a prediction of how random drift over 5, 7.5, or 15 s might be expected to diminish tuning. In order to determine whether random drift was sufficient to account for the decrease in tuning that we

observed, we then compared this prediction with our observed data fits. We performed a similar analysis in 2 dimensions for 15-s data. Fit statistics were assessed with bootstrapping.

We estimate mean turn-off times—the times at which cells on average become untuned—for each cell. We first computed tuning amplitude in each cell for the 0.5–1.5 and 2–4-s memory period intervals. The larger of the 2 values was taken as the maximal tuning of the cell. A cell's mean turn-off time was then defined as the time when the trial averaged tuning of the cell first dropped below 25% of its maximal value and thereafter remained below that level for at least 1.5 s. Variations of the criterion value did not change our conclusions. Because some cells did not show tuning until over 2 s into the memory period and could thus not lose tuning before then, the histogram of mean turn-off times (Fig. 4A) was truncated on the left at 2.5 s. Nine cells reached peak tuning early and turned off before 2.5 s had elapsed; these cells were placed in the first histogram bin. We compared our 25% criterion method to a statistical method and found that the methods agreed for cells with sufficiently large initial tuning and sufficiently many trials (Supplementary Fig. S9). The statistical method tended to estimate earlier turn-off times when either initial tuning or trial counts were low.

We assessed whether turn-off times were cell-specific in 2 ways. In the first analysis, for each cell, we split trials randomly into 2 groups and computed the mean turn-off time of each group (Fig. 4D). Subsequently, we computed a correlation across cells. This procedure was carried out 10 000 times and an average correlation was produced. A positive correlation would support that turn-off times are more consistent within cells compared with across cells. For the second analysis, we included trials in which the cell does not turn off. We correlated turn-off time with the proportion of sustained trials (Fig. 4E). If turn-off time is consistent within cells, then cells with later mean turn-off times should have a greater number of sustained trials. Importantly, these 2 measures are independent—mean turn-off time is recomputed excluding sustained trials. Sustained trials for each cell were identified as follows: Preferred direction firing rate time courses were averaged across trials and the baseline firing rate was subtracted from this average. A threshold was defined as 50% of the maximum value during an interval from 0.5 to 4 s after the target onset. Trials for which the firing rate rose above this threshold and remained above this threshold were considered sustained. Trials during which the firing rate rose above and then dropped below this threshold were considered turned off. All other trials were excluded.

We compared cell turn-off time to intrinsic time scales measured during fixation. Our analysis and inclusion criteria was identical to those described by Wasmuht et al. (2018), with the exception that our fixation interval was 1.5 s instead of 0.5 s. In accord with the description by Wasmuht and colleagues, we excluded 2 cells with a mean firing rate < 1 sp/s, 6 cells with at least one time bin with no spikes across all trials, and 34 cells with quasi-linear fits. Additionally, we excluded 1 cell for which an optimal fit could not be found and 1 cell with an outlier time constant less than 1 ms. This left 48/93 cells (52%) to be included in the analysis.

Sustained Cells

We wished to know whether there are cells that sustain memory activity indefinitely, but we were unable to employ indefinitely long memory periods and thus could not distinguish true sustained cells from cells that would eventually turn off. To address this, we modeled cell turn-off as an exponentially distributed random variable and fit a curve to turn-off times between 2.5 and 15 s of memory. We then extrapolated this curve to estimate the number of cells that we would expect to maintain memory beyond 15 s. The 9 cells that turned off before 2.5 s were not included in the fitted data, because we may have undersampled cells with short memory periods, though the results were similar without this exclusion. An exponential decay fit the data better than a linear decay. Extrapolating the fit beyond 15 s and taking its area (light gray region in Fig. 8A) provides an estimate of the number of cells expected to turn off at some time after 15 s. We then tested the hypothesis that all of our cells fit this pattern, that is, that there are no true sustained cells, by subtracting this estimate from the observed number of cells still holding a memory after 15 s and asking if the difference was greater than that which would be predicted by chance given uncertainties in the fit and the data.

Untuned Cells

We wanted to know if cells that are initially untuned might become tuned later in the memory period. We recorded 68 cells without tuning in the first 4 s of the memory period and tested whether they developed tuning at any point later in the memory period. To accomplish this, we split the data for each cell into 500 ms time bins, fit a cosine function to the firing rates in each bin, and tested for a significant fit. We conducted a bootstrap analysis to compare the observed proportion of tuned intervals with the proportion expected by chance. We simulated 1000 shuffled cell populations with 68 cells each. Each trial for a given cell was replaced by a randomly sampled (with replacement) trial of the same duration (5, 7.5, 15 s) from the set of all trials in the population. We calculated the proportion of significantly tuned intervals for each of the shuffled cell populations to generate a distribution for the proportion of tuned intervals expected by chance, to which we compared the observed proportion of tuned intervals (Table 1). We also repeated the entire analysis for time bins of 2000 ms (Table 2).

Cells With Opposite Early and Late Tuning

Of the 93 memory cells, 15 showed early tuning that was opposite in sign compared with memory tuning. To identify these cells, we tested the following time intervals in order, stopping once we found significant tuning: 200–600, 300–700, 400–800, 500–900, and 600–1000 ms (each relative to target onset). If the polarity of tuning in the first significantly tuned early interval opposed that in the 2–4-s memory period interval, then that cell was classified as a cell with opposite early and late tuning. Of the 15 cells identified, 11 showed opposite tuning in the 100–300-ms interval, 3 in the 200–300-ms interval, and 1 in the 300–700-ms interval (shown as an example in Fig. 9A). We found no cells with opposite early and late tuning that did not show significant early tuning by the 300–700-ms interval.

Active- and Inactive-Cell Tuning

We investigated how the subgroup of cells that remained on at any given time behaved over the course of the memory period (Fig. 5A). We estimated mean active-cell tuning by stepping through each time point, t, during the memory period

Table 1 Cells untuned in the early memory period show no tuning in later memory. Columns are the proportion of time intervals that show significant tuning at criterion *P* values of 0.01, 0.025, and 0.05; the *P* value that this is different from the proportion expected by chance, and the maximum proportion compatible with the *P* value. The analysis was done using 500-ms time intervals. In no case is the percentage of significantly tuned cells higher than that expected by chance, as determined by repeating the analysis on 1000 simulated cell populations in which trials have been randomly shuffled and taking the 2.5th percentile. {#tbl:cells-500 ms}

Significance criterion	Proportion tuned	*P* value	Chance bound
$P < 0.01$	0.005	0.89	0.013
$P < 0.025$	0.018	0.82	0.029
$P < 0.05$	0.038	0.90	0.053

Table 2 Same as Table 1, but with 2000-ms time intervals. {#tbl:cells-2000 ms}

Significance criterion	Proportion tuned	*P* value	Chance bound
$P < 0.01$	0.007	0.56	0.017
$P < 0.025$	0.017	0.76	0.037
$P < 0.05$	0.037	0.85	0.065

and computing the mean tuning of all cells with a mean turn-off time greater than *t*. Likewise, we estimated mean inactive-cell tuning by stepping through each time point, *t*, during the memory period and computing the mean tuning of all cells with a mean turn-off time less than or equal to *t*.

We compared our experimental data to cells from a simulated decaying bump network. We modified code written by Wimmer et al. (2014, Supplementary Code 3) to model our cells. In brief, a network consisting of 512 topographically connected nodes (analagous to neurons) is stimulated so as to introduce a "bump" of activity across the network. The bump of activity is maintained by the network but slowly decays over the delay. We iteratively modified parameters so that the network would decay to 0–50% of initial tuning over 15 s, matching the data we recorded. For each simulation, we extracted the firing rate of a cell at the initial center of the bump and the firing rate of a cell opposite the initial center of the bump. We then estimated tuning by computing the difference between these 2 rates. We carried out 740 network simulations as well as 190 simulations of a similar network that did not decay. This procedure yielded 930 independent tuning trajectories. We grouped these trajectories into sets of 10, averaged them and smoothed them with a Gaussian filter ($\sigma = 25$ ms) to produce 93 mean tuning trajectories: 74 from the decaying network and 19 from the nondecaying network. This set of procedures produced a data set similar to our experimental data: 19 sustained cells and 74 nonsustained cells. We then compared active- and inactive-cell tuning in the recorded and simulated data sets (Fig. 5B). Similar to our experimental data, turn-off times were estimated as the time when mean tuning dropped below 25% of initial tuning (mean during 1–2 s after stimulus onset) and thereafter remained below that level for at least 1.5 s.

Comparison With Previous Literature

It is critical to know whether FEF and dlPFC cells that we identify as memory cells are similar to memory cells identified in previous studies of FEF and dlPFC. To establish this, we analyzed our data using approaches from those studies. Typically, investigators used memory periods of no more than 3 s (e.g., Chafee and Goldman-Rakic 1998, 3 s; Clark et al. 2012, 1 s; Leavitt et al. 2018, 0.5–1.5 s; Markowitz et al. 2015, 1.0–1.5 s; Mendoza-Halliday et al. 2014, 1.2–2.0 s; Funahashi et al. 1989, 3 s; Wimmer et al. 2014, 3 s; Romo et al. 1999, 3–6 s; for a review, see Constantinidis et al. 2001). We first replicated the memory tuning plots of Funahashi et al. (1989, Fig. 3); Chafee and Goldman-Rakic (1998, Figs 11 and 12); Clark et al. (2012, Figs 1–4), comparing unit activity over just the first 3 s of the memory period for preferred and null direction targets (Fig. 7*A–D*). Next, we replicated the receiver operating curves (ROC) analysis of Clark et al. (2012) (Fig. 7*E–H*). We computed a single AUC value for each cell and considered the distribution (Fig. 7*I*). The interval used for that computation depended on cell identity. We used 0.5–1.5 s after target onset for the 70 cells that were significantly tuned during that interval and 2–4 s for the remaining 23 cells. Additionally, for single cells, ROC areas under the curve (AUC) were calculated at each time point over an interval with a sliding 500-ms window (Fig. 7*E–H*).

Results

We used long memory periods of up to 15 s to systematically examine the time course of spatial working memory activity in prefrontal memory circuits. Two cynomolgus macaques (*M. fascicularis*), C and W, performed a memory-guided saccade task (Fig. 1A). Subjects were shown a brief peripheral stimulus and required to remember its spatial location for a period of 5.1–5.6, 7.6–8.1, or 15.1–15.6 s. For simplicity, we nominally refer to these as 5, 7.5, and 15-s trials, respectively. After the delay, the animals were rewarded for making a saccade to the remembered location. Mid-trial rewards were used to encourage the animals to stay on task through the longer delays (see Materials and Methods).

Behavioral performance deteriorates with longer memory periods. As the memory period extends from 5 to 15 s, the incidence of grossly inaccurate trials, that is, angular errors >80 degrees of arc from the target, increases from 1% to 5% in Monkey C and from 5% to 12% in Monkey W ($P < 0.0001$, *t*-test, for each monkey; Fig 1B). The average error in the remaining saccadic responses increases with delay time—the angular error increases from ∼15 to ∼18 deg (Fig. 1C; $P < 0.001$ for each monkey) and Euclidean error increased from ∼3.5 to ∼4.3 dva (Fig. 1D; $P < 0.001$ for each monkey). See Holmes et al. (2018) for additional details. Trials in which the subject broke fixation prior to receiving a go cue (Supplementary Fig. S1) were excluded from this and subsequent analyses.

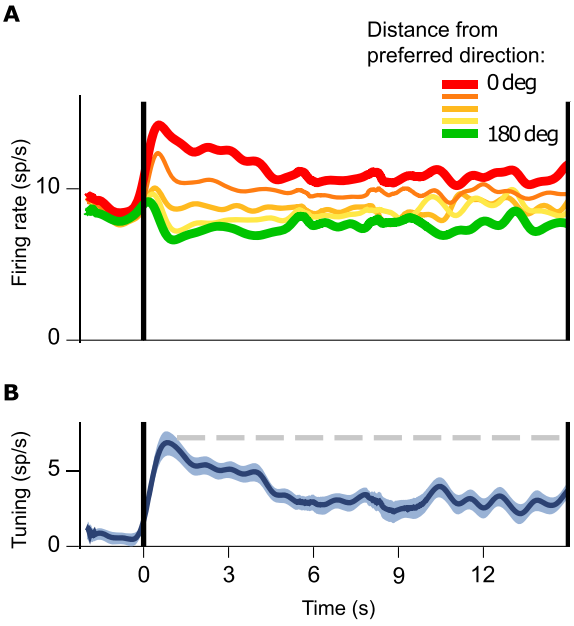

Figure 2. Tuning throughout the memory period for 93 memory cells. (*A*) Population neural activity when the memory targets were in the cells' preferred directions (0 deg; red trace), null directions (180 deg; green trace), or at various points between (orange and yellow traces). (*B*) Memory tuning (difference between the red and green traces of *A*). Shading indicates the ±1 standard error of the mean.

We recorded from 88 cells in the FEF and 73 cells in the dlPFC, totaling 161 cells. We found no substantial differences in our analyses between these areas and thus pooled the data across areas. Of the 161 recorded cells, 93 (41 FEF, 52 dlPFC) showed significant tuning in an early memory interval (either 0.5–1.5 or 2–4 s after target onset; see Materials and Methods) and were thus classified as memory neurons. When a memory target was in a memory cell's preferred direction, population activity was elevated compared to when a target was in the null (opposite) direction (Fig. 2*A*). This difference was sustained for the entire duration of the trial, but waned over the course of memory, particularly over the first 6 s (Fig. 2*B*).

Population Tuning Decay Is Not Explained by Network Drift

Bump attractor networks are susceptible to random drift. On any one trial, the network may transition across a continuum of stable states in a random walk. Consequently, trial-averaged tuning curves become shallower and wider over time (Compte et al. 2000; see also Supplementary Fig. S2). Indeed, our data demonstrate that tuning curves, computed as von Mises fits to population averaged activity, attenuate over the course of 15 s (Fig. 3*A*–*E*). Fits decrease in amplitude (peak-to-trough) with time, decaying from 5.2 sp/s in early memory (bootstrap 95% CI, [4.5, 6.1] sp/s) to 2.1 sp/s by late delay (−60%; [1.5, 2.6] sp/s).

To quantitatively test if random drift of the network's bump of activity could sufficiently account for the observed drop in activity, we compared the observed end-of-delay tuning curves to simulated curves obtained as follows: For each trial, we shifted the early population tuning curve (Fig. 3*A*) by an amount equal to the saccadic angular error of the trial. The shape of the tuning curve was always the same—only the center differed, that is, drifted, from trial to trial. We then sampled the firing rate at a single random direction, analogous to our actual data collection. We binned and averaged the data by direction and fit a von Mises function to these data. This procedure was carried out separately for 5, 7.5, and 15-s trials. As expected, the simulated end-of-delay tuning curves' amplitudes were attenuated compare to the early tuning curve (differences from initial tuning—5-s trials, −1.8 sp/s, 95% bootstrap CI [−2.6, −0.8] sp/s; 7.5-s trials, −3.1 sp/s, CI [−4.4, −2.1] sp/s; 15-s trials, −3.2 sp/s, CI [−4.4, −2.0] sp/s; Figs 3*C*–*E*). However, these analyses underestimate the observed attenuation (differences between simulated and observed end-of-delay amplitudes—5-s trials, +1.2 sp/s, CI [+0.3, +2.1] sp/s; 7.5-s trials, +2.4 sp/s, CI [+1.5, +3.7] sp/s; 15-s trials, +2.4 sp/s, CI [+1.3, +3.4] sp/s; Fig. 3*F*). We also simulated random drift in 2 dimensions for 15-s trials, allowing activity to drift on the 2-dimensional plane representing the screen on which the targets were presented. Like the 1-dimensional drift, 2-dimensional drift predicted less attenuation than what was observed. These simulations formalize the intuition that the large drop in tuning cannot be explained by bump attractor drift alone. While drift may occur and may contribute to tuning loss, our results suggest that some other mechanism must be the primary driver of the drop in tuning.

Cell Dynamics Are Heterogeneous But Systematic

We next compared the loss of tuning in the network to activity profiles of single cells and asked whether the tuning time courses were correlated across cells. A synchronized loss in tuning would be consistent with an imperfect, or decaying, bump network. A decaying bump network is similar to a bump attractor network, except that inhibition slightly dominates excitation. Thus, activity in a decaying bump network gradually falls back to baseline over time (Wimmer et al. 2014). In a decaying bump network, cells should lose tuning with time constants that are similar to each other and to the dynamics of the network as a whole.

We operationalized tuning loss by calculating a "mean turn-off time"—the time when a cell switches from being tuned to untuned as calculated across trials. We calculated tuning as a function of time for each cell and defined the mean turn-off time as the time when the cell dropped to 25% of its maximum tuning and remained below that level for at least 1.5 s thereafter. Mean turn-off times were broadly distributed throughout the entire memory period (Fig. 4*A*,*B*; see also Supplementary Fig. S3). Only 19 of the 93 cells did not turn off over 15 s of memory. With a homogeneous process, we would have expected few if any cells to drop to 25% of maximum tuning, since the observed population activity only drops to 45% after 15 s. These results indicate a heterogeneous decay mechanism rather than a homogeneous one.

We next asked whether cells that lose memory tuning may regain it before the end of the memory period. We excluded the 19 sustained cells from this analysis. We additionally excluded another 25 cells with a mean turn-off time greater than 9 s, as these cells would only have ∼5 s to turn back on, potentially biasing our results. We found that only 1 of the remaining 49 cells showed a recovery of significant memory tuning later in the memory period (criterion for significance, $P < 0.05$, no multiple comparisons correction). Including the 25 cells that turned off after 9 s of memory did not change this result.

We next asked whether the broad distribution of turn-off times was times were due to systematic differences in dynamics

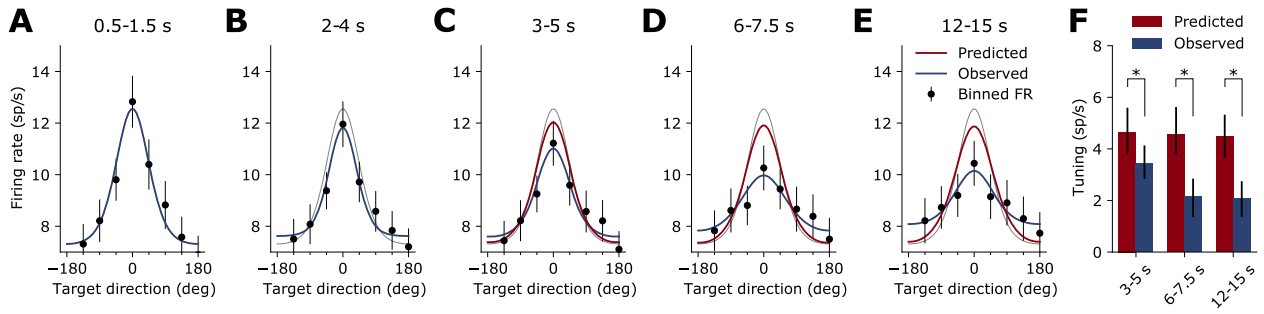

Figure 3. (*A–E*) Population activity as a function of memory target location at different times in the memory period (*A*, 0.5–1.5 s; *B*, 2–4 s; *C*, 3–5 s; *D*, 6–7.5 s; *E*, 12–15 s). Data points and their error bars indicate observed firing rate means and their standard errors, respectively. Blue lines depict von Mises fits to the data. Thin gray lines depict fits to the 0.5–1.5 s (early) data, for comparison. Red curves (*C–E*) depict the tuning curve predicted from drift simulation. (*F*) Tuning amplitudes predicted from drift simulation (red) and amplitude actually observed (blue) at the end of delays for 5, 7.5, and 15 s. Amplitude is computed as the peak-to-trough difference of the von Mises fit to the data. Error bars indicate bootstrap 95% CIs. The difference between predicted and observed amplitude is significant ($P < 0.05$, 2-sided bootstrap test) for all 3 trial lengths.

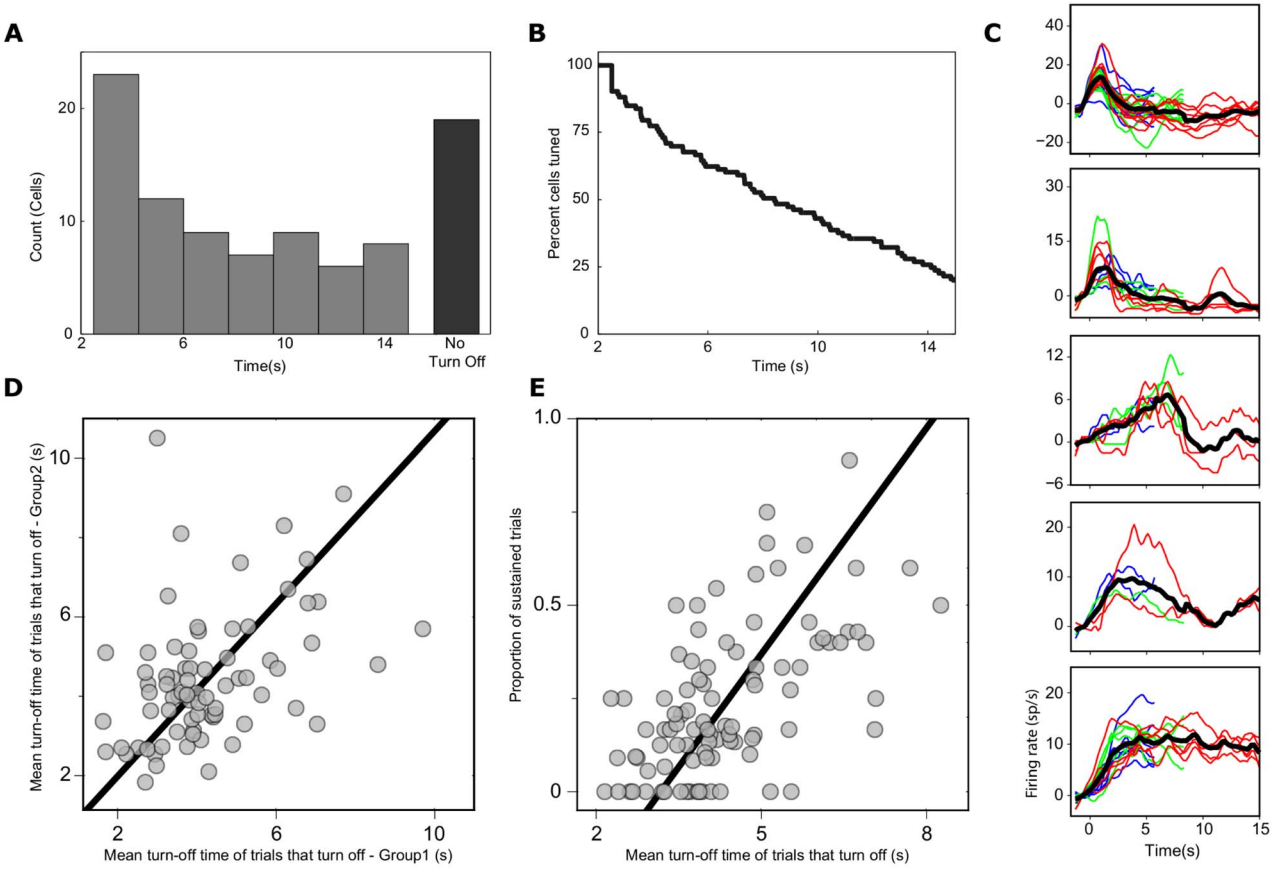

Figure 4. Tuning properties of individual cells. (*A*) Distribution of mean turn-off times, that is, when cells trial-averaged tuning drop to 25% of its early memory magnitude. (*B*) Survival curve of mean turn-off times showing the percentage of cells that remain on (tuned) throughout the memory period. Of the 93 cells, 19 (20%) do not turn off even after 15 s of memory. (*C*) Firing rates from 5 example cells in individual 5 s (blue), 7.5 s (green), and 15 s (red) trials when the memory target was in the cell's preferred direction. The black trace is the mean response. (*D*) Mean turn-off times estimated from 2 randomly selected subsets of trials for each cell are correlated ($r = 0.41$, $P < 0.001$). Each point represents data from one cell; data from all but the 19 persistent cells are included. The line represents a type II regression. (*E*) Correlation of mean turn-off time of trials in which the cell turned off versus the proportion of trials in which tuning persisted for the entire memory period ($r = 0.57$, $P < 0.0001$). Each point represents data from 1 cell; data from all 93 cells are included. The line represents a type II regression.

across cells or simply due to obtaining too few trials from each cell. At the extreme, if we obtained only 1 trial per cell, turn-off times would vary greatly from cell to cell. Therefore, we asked if the distribution of turn-off times within each individual cell was more limited than the distribution of turn-off times found across the entire population. Figure 4C shows 5 example cells: 2 that consistently turn off early, 2 that consistently turn off late, and 1 with a consistent sustained response. Generally, activity

patterns were consistent within each cell: any given trial was more similar in time course to the other trials from that same cell, compared with trials from other cells.

We quantified turn-off time consistency by assessing mean turn-off times across subsets of trials. For each cell, we randomly split all the trials in which the cell turned off into 2 groups and computed a mean turn-off time for each group. We then tested the correlation between group mean turn-off times across cells. If turn-off times reflect idiosyncratic properties of individual cells, then the mean turn-off times will be correlated. If turn-off times are not cell specific, then mean turn-off times would not be correlated. This analysis is analogous to a cross-validation test. Figure 4D illustrates the results of one example random split. Mean turn-off times are clearly correlated across cells ($r = 0.41$, $P < 0.001$). We repeated this analysis 10 000 times, each time with a new random split. Across these 10 000 repetitions, we consistently found clear correlation (mean $r = 0.35$; $P < 0.0001$, permutation test).

This analysis necessarily excludes trials in which the firing rate was sustained. In a second analysis, we used sustained trials to further assess whether mean turn-off time is cell specific. We reasoned that if mean turn-off time is a cell-specific property, then mean turn-off time should correlate with the proportion of sustained trials—cells with an early mean turn-off time should be less likely to stay on for the entire duration of a trial, and likewise, cells with a later mean turn-off time should be relatively more likely to stay on for the entire duration of the trial. Alternatively, if turn-off times reflect noise in a homogeneous network, then mean turn-off time and the proportion of sustained trials should be uncorrelated. For this analysis, we recomputed turn-off times only with trials in which the cell turned off (see Materials and Methods). Doing so enforced that turn-off time and the number of sustained trials were independent. We find a clear correlation between each cell's mean turn-off time and the proportion of sustained trials for that cell (Fig. 4E, $r = 0.57$, $P < 0.0001$).

Individual cells may appear to have characteristic mean turn-off times if turn-off time reflects memory decay and memory decay varies from session to session. To rule out this possibility, we compared mean turn-off time with an array of measures of performance: hit rate, median saccadic angular error, median saccadic absolute error, and a guess rate (see Materials and Methods). In particular, if differences in turn-off times across sessions reflect or underlie differences in behavioral performance, then turn-off times and behavior should be correlated. However, none of the behavioral measures exhibited a significant correlation (Spearman's R) with turn-off time (hit rate, $R = 0.12$, $P = 0.24$; median saccadic angular error, $R = -0.041$, $P = 0.70$; median saccadic absolute error, $R = 0.020$, $P = 0.85$; guess rate, $R = -0.18$, $P = 0.086$; no corrections for multiple comparisons). We also leveraged sustained cells by asking whether behavioral performance differed when a cell's tuning was sustained as compared with when tuning was lost. Again, no relationship was found (rank-sums tests; hit rate, $P = 0.48$; median saccadic angular error, $P = 0.86$; median saccadic absolute error, $P = 0.78$; guess rate, $P = 0.16$).

Though variation in behavior could not be linked to cell turn-off times, some other latent variable that varies from session-to-session may explain the apparent cell-specificity of turn-off time. If this were true, then we would expect simultaneously recorded cells to have similar mean turn-off times. We asked if this was the case by computing intraclass correlation of turn-off times across 15 pairs of simultaneously recorded cells (Vallat 2018). Correlation was not significant ($R = 0.15$, 95% CI [−0.36, 0.6]). Taken together, our analyses indicate that turn-off times are cell specific, varying less across trials for individual cells than across cells, and suggest that the mechanism underlying the loss of tuning is at least partly cell-specific in nature.

Previous studies have described intrinsic time scales of memory neurons (Wasmuht et al. 2018). These time scales predict tuning dynamics during a memory task. We asked whether the intrinsic timescales of our neurons could predict mean turn-off time. We found no correlation between these values ($R = 0.094$, $P = 0.52$).

Finally, we considered whether animals might, over time within a single long trial, shift from remembering a precise location to more a generic representation (e.g., a quadrant). Such behavior would invalidate our interpretation of the neural data. A shift of memory from the specific location to a generalized representation could move the memory out of a cell's mnemonic field and thus cause the cell to lose its tuning. To test for such behavior, we asked if target end points systematically shift to cluster about a small number of directions and if the degree of this clustering increases over longer delays. Memory saccades were biased toward horizontal directions (Supplementary Fig. S4). However, for systematic shifts of this nature to cause cells to turn off, the representation would need to shift by a distance greater than or equal to the cell's tuning curve's half-width-at-quarter-height. Absolute directional shifts across bins had a maximum of 25.4 and 23.5 deg for Monkeys W and C, respectively—substantially less than the population average half-width-at-quarter-height (~75 deg). Furthermore, a drift sufficient to turn cells off in the middle of the memory period would cause other cells to turn on in the middle of the memory period. In our data set, cells either develop spatial mnemonic tuning in the first 2 s or remain untuned throughout even our longest memory interval (see Individual Cells Become Tuned Early and Consistently). Finally, this explanation would predict that the mnemonic fields of sustained cells would be clustered about the points toward which the memory representations drift, which was not the case (Supplementary Fig. S4).

Cells Transition From a Distinct Turned-On State to a Distinct Turned-Off State

We have contrasted a model in which all cells turn off from gradual decay with identical rates with a model in which cells turn off at different times from heterogeneous dynamics. We argue for the latter, but the nature of the heterogeneity is unclear. At one extreme, all cells may gradually decay but with different rates. At another extreme, cells may turn off abruptly at different times. To characterize the heterogeneity, we asked whether individual cells show a progressive drop in tuning prior to when they are estimated to turn off. At each millisecond, the population of cells was split into 2 groups: cells with a mean turn-off time after the current time point (on cells) and cells with a mean turn-off time at or before the current time point (off cells). As time progresses, more cells shift from on to off. We examined the mean tuning over time for on cells, off cells, and for the entire population (Fig. 5A). Although the tuning of the entire population decays over time, the average tuning strength of the on cells exhibits a modest increase in tuning over the 15-s memory period. Cells that have turned off are completely untuned (with the exception of an activity increase at the end of the longest memory period that may reflect anticipatory effects). These results suggest that the decay in population level

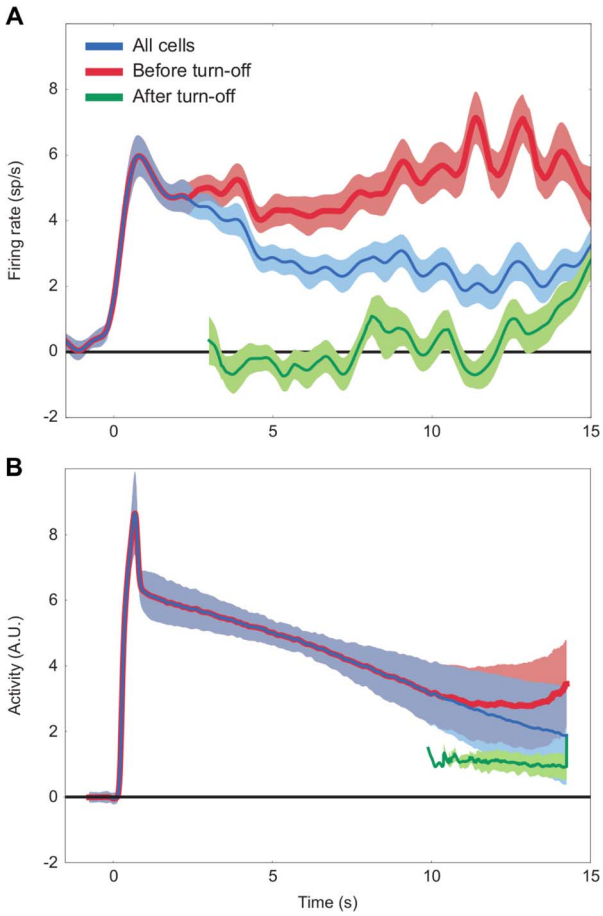

Figure 5. Tuning of cells conditioned on mean turn-off time. (*A*) Blue trace—tuning of the entire population of 93 cells. Red trace—tuning strength of cells with a mean turn-off time greater than current time (on cells). Green trace—tuning of cells with a mean turn-off time less than or equal to the current time (off cells). Note that the data in the red trace comprise progressively fewer cells with time (from 93 to 19), whereas the green trace comprises progressively more cells (from 9 to 74). (*B*) Tuning of 93 modeled bump attractor cells, of which 19 are sustained and 74 are nonsustained. Format as in (*A*).

tuning results from cells shifting from an on state to an off state and not from gradual decay of individual neurons.

For comparison, we modeled a decaying bump network of 74 nonsustained cells and 19 sustained cells, matching our recordings (see Materials and Methods). Tuning differed dramatically in the simulation (Fig. 5*B*) compared with our recorded data (Fig. 5*A*). Whereas some of our experimentally observed cells turned off as early as 2.5 s into the memory period, all of the modeled cells remained on for at least 10 s. Once off, modeled cells retained some tuning (recall that "turned-off" cells are classified by having dropped below 25% of maximal tuning and so may still be slightly tuned), unlike our recorded cells that were virtually untuned. Most striking is the fact that the on cells in our model show a clear drop in tuning, demonstrating that individual model cells decay gradually, in contrast to the population of recorded cells, which maintain constant tuning for as long as they are on.

Individual Cells Become Tuned Early and Consistently

We next asked if the times at which cells first became tuned showed similar variance to the variance of turn-off times. To address this, we assessed whether cells turned on at any time during in the memory period or just at the very start. For this purpose, we could not use the 93 memory cells we have previously described as they were selected specifically because they became tuned early in the memory period, and this would bias our results. However, during the same set of experiments, we also recorded from 68 cells without tuning in the first 4 s of the memory period. These cells were located in the same FEF or dlPFC tracks as the 93 cells with early tuning. Nineteen of these 68 cells exhibited significantly tuned visual responses ($P < 0.05$) during the 300 ms of stimulus presentation. We tested whether any of these 68 cells became tuned later in the memory period. We divided the memory period into 500-ms intervals and fit cosine functions of target location to the firing rates within each interval. We computed the proportion of significantly tuned intervals using 3 different significance criteria (Table 1) using a permutation test (see Materials and Methods). We also repeated this analysis with 2000-ms intervals (Table 2). We found that the number of significantly tuned intervals was not greater than that expected by chance for either 500 or 2000-ms intervals, for any criterion value. We did not correct for multiple comparisons across bin sizes and significance criteria. The lack of a correction will increase our false positive rate. However, this is a conservative choice since all 6 analyses yield negative results. Taken together, our results indicate that if a cell will become tuned, it will do so soon after the onset of working memory. Note that there was some heterogeneity in the time that cells became tuned, consistent with previous reports, but that this heterogeneity was limited—all of the tuned cells that were recorded developed their tuning within the first 2 s of the memory period (Supplementary Fig. S5). These previous studies used memory periods of just a few seconds, and so the finding that few cells turn on after the first several seconds of the memory period could easily be missed.

In contrast to our results, some previous studies find that memory cells can alternate between putative on and off states throughout the memory period (Baeg et al. 2003; Brody et al. 2003; Jun et al. 2010; Harvey et al. 2012). Many of these studies used only a small number of predetermined memoranda. In contrast, we collected data using a continuous circular array of targets and then analyzed memory activity in response to the targets that drove the cell most strongly. We asked whether memory responses are different for optimal versus nonoptimal target locations (Fig. 6). We reasoned that with a target in the flank of the preferred direction, even a modest drift could cause a large variation in firing rate (see Supplementary Fig. S2). Indeed, when strongly driven, cells turn on early and sustain their activity throughout the memory interval, but, when driven suboptimally by a target in the flanks of the preferred direction (45 deg off the optimal position), the cell in Figure 6*A* does not show tuning until relatively late in the trial and then appears to turn on and off multiple times throughout the memory period. The cell in Figure 6*B* turns on early, appears to turn off at 2.5 s, and appears to turn back on for the last half of the memory period. These examples are typical of our entire population; targets that maximally drive memory activity show sustained responses, whereas nonoptimal stimuli drive weaker, temporally inconsistent responses.

Memory Tuning Appears Robust During Early Memory

Though 80% of our cells lose their tuning during the memory period, we aimed to confirm that they are classical memory cells and would meet the criteria of prior studies of memory

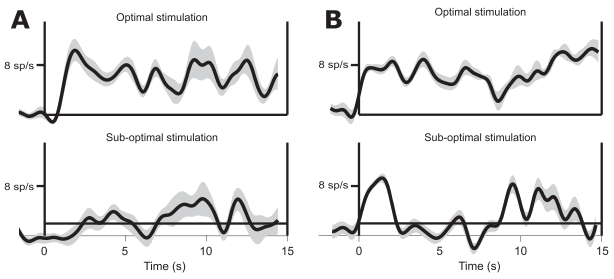

Figure 6. Memory responses appear to turn on and off over the course of the memory period when driven suboptimally. Two example cells (*A* and *B*) with well-behaved sustained memory responses are shown. Top traces show the mean firing rate when memory targets are in the cell's preferred direction, compared with 180 deg away (null direction). Bottom traces show the mean firing rate when the memory targets are at a flank 45 deg away from the preferred direction, compared with 180 deg away. Each cell exhibits sustained responses when driven optimally but fluctuates on and off when driven suboptimally.

activity. Memory activity is typically demonstrated by illustrating that directional tuning increases and remains above baseline throughout the delay period (e.g., Fig. 3 from Funahashi et al. 1989; Figs 11 and 12 from Chafee and Goldman-Rakic 1998; Figs 1–4 from Clark et al. 2012). We replicated these analyses with our population. As most other studies use substantially shorter delays, we only considered the first 3 s of our delay period. Four example cells with sustained tuning are shown in Figure 7*A*–*D*. The first 2 cells exhibit an additional transient response when the target appears. The next cell reaches a peak just as the target is withdrawn. The fourth cell becomes active only after the target disappears. Despite these differences, all 4 example cells show sustained tuning over the first 3 s of the memory period. Nonetheless, all 4 example cells turned off well before the end of the 15 s memory period (Fig. 7*A*, 4.3 s; Fig. 7*B*, 4.1 s; Fig. 7*C*, 3.0 s; Fig. 7*D*, 7.6 s). All of our cells showed similar responses. On average, activity is sustained over the first 3 s are clearly evident (Fig. 2). Thus, our memory cells qualitatively resemble those seen in previous studies.

We further validated that these neurons match the criteria used in the literature for memory cells by confirming that an ideal observer can use their activity to discriminate memory targets in the preferred and null directions using the responses from just a single trial. We computed time-resolved ROC AUC over the initial 3 s of memory (compare with Fig. 2*B* from Clark et al. 2012). All 4 example cells sharply increased discriminability around the time of the stimulus, and maintained above-chance discriminability over the initial 3 s of the memory period (Fig. 7*E*–*H*). We show a histogram of AUC values for each cell, computed 0.5–1.5 or 2–4 s after target onset (see Materials and Methods). Generally, discriminability was well above chance (median = 0.848; Fig. 7*I*). These analyses demonstrate that these memory cells are very similar to memory cells recorded in the many previous studies of memory in FEF and dlPFC.

Sustained Cells May Reflect a Distinct Population

We asked how the 19 sustained cells were related to the 74 nonsustained cells. We first determined the probability distribution that describes mean turn-off times for the 74 cells that turn off within 15 s. The distribution was best described by an exponentially decaying function with a time constant of 5.4 s (Fig. 8*A*). Mean turn-off times from 2.5 to 15 s account for 90% of the area under the full distribution curve (Fig. 8*B*). The remaining 10% of the area indicates the expected proportion of mean turn-off times in our sample that are part of this distribution but exceed 15 s. Based on this distribution we expect 8.5 cells in our sample of 93 to be sustained for at least 15 s (95% CI, 5–14 cells). This is significantly fewer than the number we observe ($n = 19$). This result suggests that at least some of the sustained cells may be part of a different population. However, the proportion of memory cells that may be part of a circuit with longer sustained memory responses is small compared with the total number of memory cells. Furthermore, sustained cells are indistinguishable from nonsustained cells during early memory—sustained

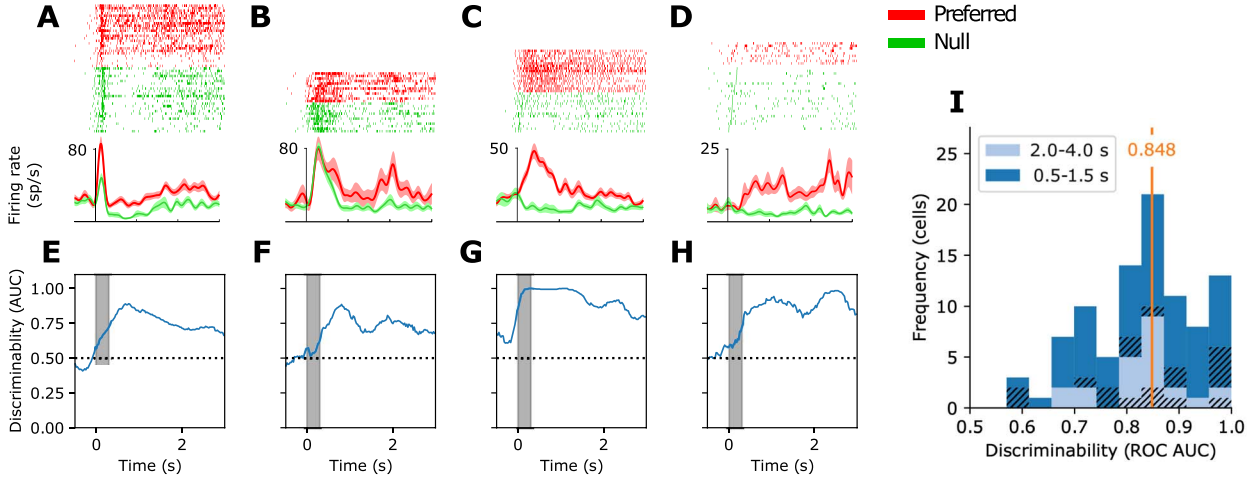

Figure 7. Memory cells are indistinguishable from those seen in previous studies. (*A*–*D*) Spike rasters and mean firing rate for 4 example cells in response to a target presented in the preferred ("red") or null ("green") directions. Data are shown over the first 3 s of memory to match or exceed the durations used in many previous studies. See Materials and Methods and Results for additional details and specific comparison studies. (*E*–*H*) Time resolved ROC AUC values for the 4 example cells in (*A*–*D*). AUC values are computed using a sliding 500-ms window. (*I*) Distribution of ROC AUC values for the 93 memory cells. ROC AUC values are computed from 0.5 to 1.5 or 2 to 4 s after the target first turns on, depending on when the cell first became tuned (see Materials and Methods). Hatching indicates sustained cells, that is, cells that did not turn off over the 15-s memory period (see Materials and Methods).

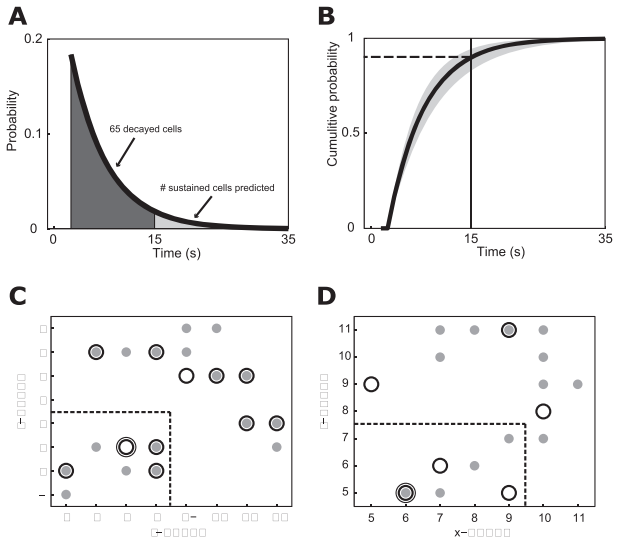

Figure 8. Properties of sustained cells. (*A*) Fitted probability density exponential distribution to mean turn-off times. The dark gray area represents cells that are not sustained, that is, cells with mean turn-off times prior to 15 s. The light gray area is the proportion of cells predicted to sustain activity for 15 s or more, based on the distribution of nonsustained cells. (*B*) Cumulative distribution of the function in (*A*). In total, 90% of the cells have mean turn-off times, indicating a prediction that 10% sampled cells should have mean turn-off times later than 15 s. Error bars show 95% CIs of the fit. (*C*) Anatomical distribution of sustained cells for Monkey C. Each gray closed circle represents a location in which one or more nonsustained memory cells were recorded. Black open circles indicate the locations of sustained cells, with double black lines when 2 cells were recorded at those coordinates. Dashed lines separate the region in and around the principle sulcus (bottom left), from the arcuate sulcus (top right). See Supplementary Figure S6. (*D*) Anatomical distribution of sustained cells for Monkey W. Format is identical to (*C*).

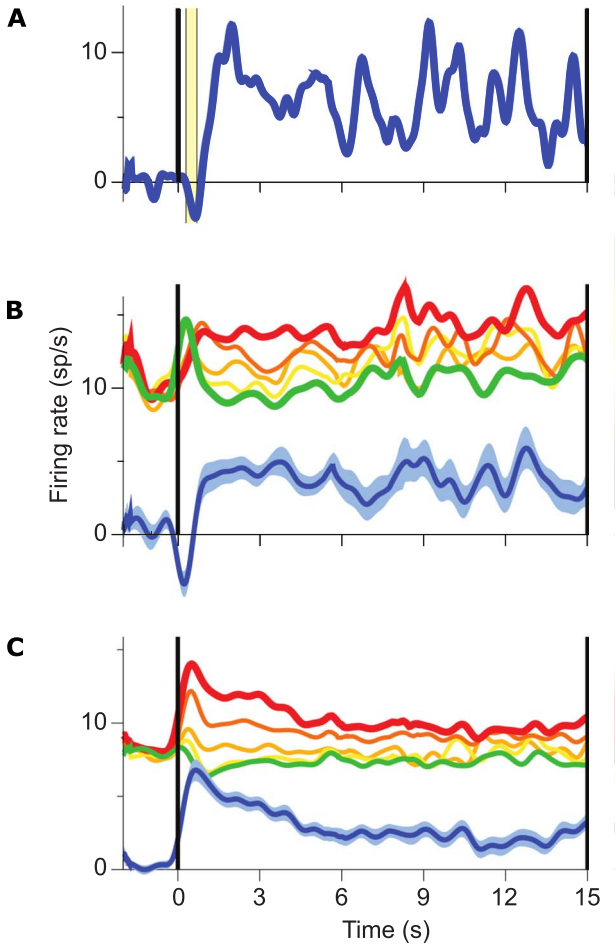

Figure 9. Cells with opposite early and late tuning. (*A*) Example cell with early tuning that is opposite in polarity from its late tuning. The highlighted time interval (yellow) is the first interval (300–700 ms) of significant tuning (−2.5 sp/s, $P < 0.006$). (*B*) Population activity of 15 cells, each of which shows significant early tuning that is opposite in polarity from its late tuning. Opposite tuning is generally driven by elevated activity for directions in the memory-interval null direction (an upward deflection of the green curve showing the response to a target in the null direction), rather than suppressed activity in the preferred direction (red trace). Population tuning in these cells shows little or no decay with time. (*C*) Population activity of cells that are not oppositely tuned in early versus late time periods. Formats in (*B*) and (*C*) are each identical to Figure 2.

cells have similar discriminability to nonsustained cells (Fig. 7*I*, rank sums test, $P = 0.66$), as well as preferred direction firing rate ($P = 0.81$), null direction firing rate ($P = 0.95$), and tuning strength ($P = 0.97$). Cells that turned on earlier (0.5–1.5 s) were neither more nor less likely to be sustained than cells that turned on later (2–4 s; χ^2 test, $P = 0.81$).

If some or all sustained cells form a separate population from nonsustained cells, then they may be anatomically clustered. To determine if this was the case, we examined the anatomical location of recorded cells for monkeys C and W (Fig. 8*C*,*D*, respectively; also see Supplementary Fig. S6). Sustained cells were not clustered but were instead distributed within and near the principal and arcuate sulci, similar to the distribution of cells that turned off.

Interestingly, 15 of the 93 cells (16%) showed visually evoked or early memory tuning that was opposite in sign from the tuning in the later intervals (Fig. 9; −6.1 sp/s ± 1.7 sp/s from 100 to 300 ms). These cells had similar tuning magnitudes as the other 78 cells later in the memory interval (oppositely tuned cells—4.6 ± 0.4 sp/s; nonoppositely tuned cell—4.2 ± 0.7 sp/s). Seven of these 15 cells (47%) were sustained and constitute a relatively large proportion of all sustained cells (37%). Firing rate patterns with opposite early and late tuning are consistent with certain types of inhibitory interneurons that may be involved in spatial memory circuits (Gabbott and Bacon 1996; Constantinidis et al. 2001; Constantinidis and Goldman-Rakic 2002; Wang et al. 2004; Zhou et al. 2012; see also Lawrence et al. 2005). This finding suggests that the inflated number of sustained cells compared with the number predicted by the analysis of Figure 8*A* could reflect a separate population of sustained inhibitory interneurons.

Discussion

Previous studies of working memoryusing primarily short (e.g., 1–3 s) memory periods found that mnemonic tuning is often sustained for the entire memory period (Fuster and Alexander 1971; Kojima and Goldman-Rakic 1982; Bruce and Goldberg 1985; Funahashi et al. 1989, 1993; Amit 1992; Pellegrino and Wise 1993; Chafee and Goldman-Rakic 1998; Ferrera et al. 1999; Sommer and Wurtz 2001; Umeno and Goldberg 2001; Takeda and Funahashi 2002, 2004). These findings inspired attractor network models with stable, nondecaying memory states (Amit 1992; Brunel 1996; Amit and Brunel 1997; Compte et al. 2000; Wang 2009). A subclass of attractor networks, a bump attractor network, represents continuous memoranda such as spatial locations as

a localized "bump" of elevated activity, whose amplitude and shape can be sustained indefinitely without decay. In this study, we sought to more thoroughly characterize spatial memory tuning over longer memory delays (5–15 s) with respect to the predictions of bump attractor network models.

We trained 2 nonhuman primate subjects to perform a spatial working memory task with 5–15-s delay periods and recorded single-unit activity from the FEF and the dlPFC while the animals performed the task. Because the task memoranda and behavioral responses were continuous, we could identify the optimal target direction for each cell. We actively monitored recorded cells to ensure good isolation over time. Our data indicate that memory cells turn on (i.e., exhibit spatial tuning) within the first few seconds of the memory period, remain active for a variable but cell-specific period of time, and then turn off (i.e., lose their tuning). Once turned off, cells do not regain their tuning. Most cells (80% in our data) have mean turn-off times shorter than our 15-s memory period.

Our conclusions stem from the observation that mean tuning decreases by 55% over a 15-s memory interval. We considered 2 possible mechanisms for this gradual drop in activity—random drift of the representation and gradual homogeneous decay. The first possibility, random drift, was that noise can perturb the state of a canonical bump attractor network, causing the bump to move in a random walk (Compte et al. 2000). This drift varies across trials, and as a result, we expect the spatial tuning curves, which are computed across multiple trials, to lose amplitude and widen over time (Supplementary Fig. S2). In our data, however, tuning amplitude decayed much faster than would be predicted by drift (55% decay by ~12–15 s observed versus 15% decay predicted after 15 s).

A second possible explanation for a drop in tuning is a bump network that homogeneously decays with time as a consequence of an imbalance between excitation and inhibition. We asked whether the decay we saw was consistent with a decaying bump network by modeling a network that decays at the same overall rate as what we observed in our recordings (Fig. 5B). In the model, the drop in tuning is progressive and similar across all cells, even cells that have not turned off. Most model cells turn off within a few seconds of each other. The decay is apparent even when excluding model cells that have turned off. Contrary to these predictions, we observed that individual cells transitioned relatively quickly between distinct on- and off-states (Fig. 5A) while mean turn-off times across cells were broadly distributed. Cells that remained on modestly increased in activity, which helps explain the difference between the drop in the population's tuning (55%) and the proportion of cells that turn off (80%). Thus, while random drift may occur, it cannot sufficiently account for the dynamics we observe.

Many studies show that single cells may ramp up or down during the memory period (Brody et al. 2003; Jun et al. 2010; Murray et al. 2017). Although not emphasized in our results, we see similar ramping in our own data. For example, the third panel of Figure 4C shows a cell with a clear upward ramp for ~6 s, and the fourth panel shows a cell with an upward ramp over ~3 s followed by a slow downward ramp over ~10 s. Our finding of cell-specific turn-off times compliments the prior finding that cells show heterogeneous activity profiles, for example, upward and downward ramps.

Other studies have suggested that individual cells may turn on and off multiple times during a single memory period or that individual cells involved in memory may turn on for brief periods (10s of milliseconds) in the beginning, middle, or end of a memory period (Baeg et al. 2003; Harvey et al. 2012). Some of these studies have used stimuli that do not maximally drive memory cells. Mouse studies, for example, typically do not precisely map mnemonic fields. We show that testing with suboptimal stimuli, that is, stimuli presented at positions on the flanks of the preferred direction, can produce responses with multiple transitions between on- and off-states. These turnings on and off may reflect bump drift coupled with a steep slope of the tuning curve on the flanks (Supplementary Fig. S2). Yet these same cells show sustained activity when driven by an optimal stimulus. We also found that once an optimally stimulated cell turns off, it does not regain tuning later in the memory period. Furthermore, cells that are not tuned in the first few seconds of the memory period do not acquire tuning later. Thus, while we see heterogeneous patterns of memory activity, for example, ramping, which are consistent with many previous studies using shorter memory periods, we also see consistent patterns that previous studies did not detect. In particular, we see consistent turn-off times ranging from 2 to 15 s, and we find that cells tend to turn on early, stay on for many seconds, and do not regain tuning once they have lost it (turned off).

To keep the animals on task over memory period of up to 15 s, the animals received rewards mid-trial. A limitation of this design is that these rewards may perturb the time course of activity, perhaps even driving cells to turn off early. Indeed, animals were more likely to break fixation after a mid-trial reward (Supplementary Fig. S1), though a fixation break does not necessitate a lapse in memory. Most of our recording sessions did not have a mid-trial reward until 7.5 s into the memory period, yet cells from these sessions often turn off before that time. Furthermore, mean turn-off times do not appear to be clustered around the times of mid-trial reward, as would be expected if mid-trial rewards cause turn off (Supplementary Fig. S7), and a reward-triggered average of firing rates across cells also reveals no loss of turning specifically associated with the reward (Supplementary Fig. S8).

While our results are novel, they nonetheless replicate the main findings of earlier studies. As noted above, previous studies describe memory cells as having activity that persists over the course of a delay (Constantinidis et al. 2018). The fact that 80% of our cells lose their tuning during the delay may suggest that our cells represent a different population. However, the earlier studies used substantially shorter delays—as little as 1 s (see Introduction). When we consider only the first few seconds of the delay, our memory cells are indistinguishable from those described in the literature (Figs 2 and 7). It is only over much longer delays that the gradual decay of these cells becomes apparent. Furthermore, we do see cells with multiple on/off cycles and other complex dynamics similar to those reported previously (Baeg et al. 2003; Brody et al. 2003; Jun et al. 2010; Harvey et al. 2012), but only in response to suboptimal stimuli. Since the mnemonic fields of spatial memory networks appear to completely tile space, looking only at the responses to a small subset of stimuli (e.g., just 2 or 4 target locations) will fail to correctly characterize the behavior of the network. The complex dynamics that have been reported while using just a small number of stimulus locations may reflect responses from the flanks of the tuning curves and thereby mischaracterize the true network dynamics. Alternatively, since many of these responses were described in rodents, it is possible that this reflects a species difference.

Our results help refine the leading models of working memory circuits. Qualitatively, the dynamics we report are more

complex than bump attractor network models or decaying bump network models, yet more constrained than proposed alternatives such as reservoir network computational models (Maass et al. 2002; Verstraeten et al. 2007; Appeltant et al. 2011; Bernacchia et al. 2011) or some feed-forward networks (Goldman 2009). A parsimonious hypothesis would be that memory networks consist of many weakly coupled recurrent subnetworks, such as bump attractor networks, each with its own hard-wired time constant. Early in the memory period more of these attractor circuits may be active, but over time subnetworks progressively "turn off." Such a mechanism may help free up neural resources so that they become available to encode new information. Alternatively, the turned-off cells may still participate in coding the original information via activity–silent mechanisms (Stokes 2015). Future work should investigate the functional relevance of these turn-off times and whether turn-off times serve as a marker for the degree of coupling between neurons.

Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

Notes

We thank B. Acland, E. Mooshagian, J. Kang for helpful discussion and feedback regarding this work and J. Tucker for helping with technical issues. *Conflict of Interest*: None declared.

Funding

The National Eye Institute (grant number R01-EY012135 to L.H.S.); the National Institute of Mental Health (grant number 5F31MH094076 to C.P.); the National Science Foundation (grant number T32-NS073547 to C.D.H and C.P.). The National Institute of Mental Health and National Institute of Biomedical Imaging and Bioengineering (grant number R01EB028154)

References

Amit DJ. 1992. *Modeling brain function: the world of attractor neural networks*. Cambridge: Cambridge University Press.

Amit DJ, Brunel N. 1997. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb Cortex*. 7:237–252.

Appeltant L, Soriano MC, Van der Sande G, Danckaert J, Massar S, Dambre J, Schrauwen B, Mirasso CR, Fischer I. 2011. Information processing using a single dynamical node as complex system. *Nat Commun*. 2:1–6.

Baeg E, Kim Y, Huh K, Mook-Jung I, Kim H, Jung M. 2003. Dynamics of population code for working memory in the prefrontal cortex. *Neuron*. 40:177–188.

Barak O, Sussillo D, Romo R, Tsodyks M, Abbott L. 2013. From fixed points to chaos: three models of delayed discrimination. *Prog Neurobiol*. 103:214–222.

Bays PM, Catalao RFG, Husain M. 2009. The precision of visual working memory is set by allocation of a shared resource. *J Vis*. 9:7–7.

Bernacchia A, Seo H, Lee D, Wang X-J. 2011. A reservoir of time constants for memory traces in cortical neurons. *Nat Neurosci*. 14:366.

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

Bruce CJ, Goldberg ME. 1985. Primate frontal eye fields. I. Single neurons discharging before saccades. *J Neurophysiol*. 53:603–635.

Brunel N. 1996. Hebbian learning of context in recurrent neural networks. *Neural Comput*. 8:1677–1710.

Chafee MV, Goldman-Rakic PS. 1998. Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J Neurophysiol*. 79:2919–2940.

Clark KL, Noudoost B, Moore T. 2012. Persistent spatial information in the frontal eye field during object-based short-term memory. *J Neurosci*. 32:10907–10914.

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

Constantinidis C, Funahashi S, Lee D, Murray JD, Qi X-L, Wang M, Arnsten AF. 2018. Persistent spiking activity underlies working memory. *J Neurosci*. 38:7020–7028.

Constantinidis C, Goldman-Rakic PS. 2002. Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *J Neurophysiol*. 88:3487–3497.

Ferrera VP, Cohen JK, Lee BB. 1999. Activity of prefrontal neurons during location and color delayed matching tasks. *Neuroreport*. 10:1315–1322.

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

Fuster JM, Alexander GE. 1971. Neuron activity related to short-term memory. *Science*. 173:652–654.

Gabbott PL, Bacon SJ. 1996. Local circuit neurons in the medial prefrontal cortex (areas 24a, b, c, 25 and 32) in the monkey: I. Cell morphology and morphometrics. *J Comp Neurol*. 364:567–608.

Goldman MS. 2009. Memory without feedback in a neural network. *Neuron*. 61:621–634.

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

Holmes CD, Papadimitriou C, Snyder LH. 2018. Dissociation of LFP power and tuning in the frontal cortex during memory. *J Neurosci*. 38:8177–8186.

Jun JK, Miller P, Hernández A, Zainos A, Lemus L, Brody CD, Romo R. 2010. Heterogenous population coding of a short-term memory and decision task. *J Neurosci*. 30:916–929.

Kojima S, Goldman-Rakic PS. 1982. Delay-related activity of prefrontal neurons in rhesus monkeys performing delayed response. *Brain Res*. 248:43–50.

Lawrence BM, White RL 3rd, Snyder LH. 2005. Delay-period activity in visual, visuomovement, and movement neurons in the frontal eye field. *J Neurophysiol*. 94:1498–1508.

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.

Lundqvist M, Herman P, Miller EK. 2018. Working memory: delay activity, yes! Persistent activity? Maybe not. *J Neurosci*. 38:7013–7019.

Maass W, Natschläger T, Markram H. 2002. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput*. 14:2531–2560.

Markowitz DA, Curtis CE, Pesaran B. 2015. Multiple component networks support working memory in prefrontal cortex. *Proc Natl Acad Sci*. 112:11084–11089.

Mendoza-Halliday D, Torres S, Martinez-Trujillo JC. 2014. Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat Neurosci*. 17:1255.

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*. 114:394–399.

Pellegrino G, di Wise SP. 1993. Visuospatial versus visuomotor activity in the premotor and prefrontal cortex of a primate. *J Neurosci*. 13:1227–1243.

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

Sheppard K, Khrapov S, Lipták G, Mikedeltalima, Capellini R, Hugle, esvhd, Fortin A., JPN, Adams A, Jbrockmendel, M, Rabba, Rose ME, Rochette T, RENE-CORAIL X, Syncoding. 2020. *Bashtage/arch: Release 4.15*.

Sommer MA, Wurtz RH. 2001. Frontal eye field sends delay activity related to movement, memory, and vision to the superior colliculus. *J Neurophysiol*. 85:1673–1685.

Stokes MG. 2015. Activity-silent'working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci*. 19:394–405.

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

Takeda K, Funahashi S. 2002. Prefrontal task-related activity representing visual cue location or saccade direction in spatial working memory tasks. *J Neurophysiol*. 87:567–588.

Takeda K, Funahashi S. 2004. Population vector analysis of primate prefrontal activity during spatial working memory. *Cereb Cortex*. 14:1328–1339.

Umeno MM, Goldberg ME. 2001. Spatial processing in the monkey frontal eye field. II. Memory responses. *J Neurophysiol*. 86:2344–2352.

Vallat R. 2018. Pingouin: statistics in python. *J Open Source Software*. 3:1026.

Verstraeten D, Schrauwen B, d'Haene M, Stroobandt D. 2007. An experimental unification of reservoir computing methods. *Neural Netw*. 20:391–403.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J *et al*. 2020. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods*. 17:261–272.

Wang X-J. 2009. Attractor network models. In: *Encyclopedia of neuroscience*. Oxford: Elsevier Ltd, Oxford Academic Press, pp. 667–679.

Wang X-J, Tegnér J, Constantinidis C, Goldman-Rakic P. 2004. Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory. *Proc Natl Acad Sci*. 101:1368–1373.

Wasmuht DF, Spaak E, Buschman TJ, Miller EK, Stokes MG. 2018. Intrinsic neuronal dynamics predict distinct functional roles during working memory. *Nat Commun*. 9:1–13.

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

Zhang W, Luck SJ. 2009. Sudden death and gradual decay in visual working memory. *Psychol Sci*. 20:423–428.

Zhou X, Katsuki F, Qi X-L, Constantinidis C. 2012. Neurons with inverted tuning during the delay periods of working memory tasks in the dorsal prefrontal and posterior parietal cortex. *J Neurophysiol*. 108:31–38.