Cerebral Cortex, April 2017;27:2513–2527

doi:10.1093/cercor/bhw106
Advance Access Publication Date: 24 April 2016
Original Article

ORIGINAL ARTICLE

Ghosts in the Machine II: Neural Correlates of Memory Interference from the Previous Trial

Charalampos Papadimitriou[1], Robert L. White III[2] and Lawrence H. Snyder[1]

[1]Department of Anatomy and Neurobiology, Washington University in St. Louis, St. Louis, MO 63116, USA and
[2]Department of Psychology, University of California Berkeley, Berkeley, CA 94720, USA

Address correspondence to Charalampos Papadimitriou, Department of Anatomy and Neurobiology, Washington University School of Medicine, Box 8108, 660 South Euclid Avenue, Saint Louis, MO 63110, USA. Email: papadimitriou.c@gmail.com

Abstract

Previous memoranda interfere with working memory. For example, spatial memories are biased toward locations memorized on the previous trial. We predicted, based on attractor network models of memory, that activity in the frontal eye fields (FEFs) encoding a previous target location can persist into the subsequent trial and that this ghost will then bias the readout of the current target. Contrary to this prediction, we find that FEF memory representations appear biased away from (not toward) the previous target location. The behavioral and neural data can be reconciled by a model in which receptive fields of memory neurons converge toward remembered locations, much as receptive fields converge toward attended locations. Convergence increases the resources available to encode the relevant memoranda and decreases overall error in the network, but the residual convergence from the previous trial can give rise to an attractive behavioral bias on the next trial.

Key words: attractor network models, frontal eye fields, proactive interference, receptive field remapping, spatial working memory

Introduction

Working memory, the ability to actively maintain and transform information, is necessary for performing a wide range of cognitive tasks. Spatial working memory is of particular interest. In a spatial working memory task the memoranda and responses are locations in space and are naturally continuous, not categorical, allowing investigators to finely probe the properties of the working memory circuits. Spatial working memory tasks can be easily performed by animals, allowing animal neurophysiology to be compared with human functional magnetic resonance imaging (fMRI) data.

Neurophysiology studies in monkeys identify the prefrontal cortex (PFC) as one of the key regions involved in maintenance of spatial information. Dorsolateral PFC and frontal eye fields (FEFs) show a sustained increase in firing rates during spatial working memory tasks (Fuster and Alexander 1971; Kojima and Goldman-Rakic 1982; Bruce and Goldberg 1985; Funahashi et al. 1989, 1993; Di Pellegrino and Wise 1993; Chafee and Goldman-Rakic 1998; Ferrera et al. 1999; Umeno and Goldberg 2001; Constantinidis et al. 2001; Sommer and Wurtz 2001; Takeda and Funahashi 2002, 2004). FEF is also involved in transforming visual signals into saccadic commands (Bruce and Goldberg 1985; Schall 1991; Sommer and Wurtz 2000). During memory tasks that allow for saccade plan generation early in the memory period, fMRI signals in human FEF show increased coherence with a network of oculomotor areas (supplementary eye fields, dorsal anterior cingulate) involved in maintaining saccade goals. In contrast, during memory tasks that prevent saccade planning until late in the trial, FEF instead shows increased coherence with a different network of areas (dorsolateral PFC, superior frontal sulcus, posterior parietal cortex). These areas are thought to be involved in sustaining covert attention at a particular spatial location (Corbetta et al. 2002; Curtis et al. 2005). These results suggest that FEF plays an important role both in maintaining the perceived position of a stimulus and in transforming that information into a saccade plan that can be maintained over time.

© The Author 2016. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Working memory is susceptible to interference from previous stimuli (proactive interference: Jarvik et al. 1969; Moise 1976; Edhouse and White 1988; Dunnett and Martel 1990; see also Jonides and Nee 2006). Papadimitriou et al. (2015) identified proactive interference in a memory-guided saccade paradigm. The interference produces an attractive bias of saccadic responses toward previous memory targets with well-defined spatial and temporal characteristics. In this study, we look for neural correlates of this bias in the spiking patterns of FEF neurons. We find two possible candidates: residual activity encoding the previous target, and a shift in the activity encoding the current target. The shift in activity provides a better temporal match to the behavioral bias, yet is in a direction that is opposite to that which we would have predicted.

To resolve this inconsistency we suggest that the shift in target-encoding activity may arise from a shift in receptive field positions. Previous reports suggest that receptive fields in FEF move toward the target of an upcoming saccade (Zirnsak et al. 2014). Receptive field changes have also been shown in spatial attention tasks in V4 (Connor et al. 1997; Tolias et al. 2001). We present a model in which convergence of mnemonic fields toward memory targets can reconcile our neuronal and behavioral data.

Materials and Methods

Subjects

Two *Macaca mulatta* (M1, M3) and one *Macaca fascicularis* (M2) were used as subjects. Monkeys were fitted with a prosthetic device to stabilize the head, a single scleral search coil for eye movement recording (Robinson 1963; Judge et al. 1980), and a recording chamber over either the left or right arcuate sulcus. Sterile surgery was performed under inhalation anesthesia (isoflurane, 0.5–2.0%). Postoperative analgesics were provided as necessary. All surgical and behavioral procedures conformed to National Institutes of Health guidelines and were approved by the Washington University Institutional Animal Care and Use Committee.

Recording Procedures

During experiments, the monkey was seated in a Lexan box (Crist Instruments). Eye movements were monitored using earth-mounted 4′ rectangular field coils (CNC Engineering). Visual stimuli were projected (Electrohome, Model ECP 4100) onto a 100 × 80 cm screen placed 58 cm from the animal. The room was otherwise completely dark, as confirmed by a dark-adapted human observer. All aspects of the experiment were computer-controlled (custom software). Eye position was logged every 2 ms. Visual stimulus presentation times were accurate to within one video refresh (17 ms).

Electrophysiological recording and stimulation were performed with tungsten microelectrodes (FHC or Alpha Omega; 0.2–2.0 MΩ) Extracellular potentials were amplified (FHC) and filtered (band pass 400–5000 Hz; Krohn-Hite). Single units were isolated with a dual time–amplitude window discriminator (BAK Electronics).

Memory-guided Saccade Task

We trained 3 macaque monkeys to perform a memory-guided saccade task in which each animal first fixated a central fixation point. A peripheral target was then presented for 150 ms, followed by a 1.4, 2.8, or 5.6 s (randomly interleaved) memory period. Each animal was required to maintain fixation within 1.5° for 400 ms until the fixation point was extinguished, cueing the animal to saccade to within 1.5°–3.5° of the location of the target, depending on its eccentricity. Targets were presented at a single eccentricity, adjusted to the preferred eccentricity of the unit (range 5°–20°, mean ~13.5°). Targets were presented at up to 16 possible locations along the circumference of a virtual circle (angular separation of 22.5°). On average, we presented 12 targets per unit and collected 8 trials of each memory period length per target.

Memory Screening Task

This task was used to screen single units for further study during an experimental session. The task was identical to the memory-guided saccade task, except that the delay could vary randomly from 1000–2000 ms. Targets were presented at up to 16 possible locations at either 10° or 20° eccentricity (angular separation of 45°).

FEF Screening Task

FEF sites were defined as those at which electrical microstimulation with current <50 μA evoked consistent saccadic eye movements (bipolar stimulation pulses, negative leading, 250 μs per phase, 333 Hz, 70 ms duration, applied with a software-controlled stimulus isolation unit [FHC] (Bruce et al. 1985)). In the screening task, animals began by fixating a central target for 400 ms. The target was then extinguished, and in half of the trials stimulation began 100 ms later. The fixation point reappeared 300 ms after the initial offset. The animal was rewarded on all stimulation trials and also on control trials in which the eyes remained at the fixation target.

Behavior Analysis

Analysis of behavioral error was similar to the study by Papadimitriou et al. 2015. Briefly, within each memory-guided saccade trial we projected saccade response vectors to the unit circle. That is, we removed the radial component of each response and considered only the angular component. Using the interval 100–300 ms after the saccade, we calculated saccade error as the difference between the saccade angular direction and the target direction. For each target location we then subtracted the mean error across all trials (systematic error) from the error in each individual trial. Since the previous behavioral study revealed a pattern of errors that was very well described by a Gabor function, we used a similar function to fit response error as a function of relative target location:

$$y(x) = \text{height} \times \sin(\text{width} \times x)\, e^{-(\text{width} \times x)^2}, \qquad (1)$$

where *y* is the response error and *x* is relative direction of the previous trial's target (previous minus current target angles). The 400-ms fixation period that separates the end of one trial from the start of the next was short relative to the expected duration of the behavioral bias, which has been shown to persist for over 4 s (Papadimitriou et al. 2015).

Population-averaged Tuning Curves

We wished to know how neural tuning curves might be altered as a function of the previous trial's target. A tuning curve describes how a cell responds to a range of target locations. In the population-averaged tuning curve the receptive field centers of all cells are moved to a common location (0°) and cell responses are averaged. The receptive field center of each neuron was determined by fitting a Von Mises function to firing rate as a function of target direction in the interval 50–300 ms after target onset. Next, target directions were expressed relative to each cell's receptive field, that is, each cell's receptive field center was subtracted from each target direction. We then generated population-averaged tuning curves as a function of the current target (Fig. 2*b*), or population-averaged tuning surfaces as a function of both previous and current target directions (Fig. 5*b*).

Population Response Curves

For some analyses (Fig. 6*a*,*b*) we determine the "population response curve". Whereas a tuning curve describes how one cell responds to a range of different target locations, the population response curve describes how each cell in the population responds to one particular target location. Rather than constructing these curves based on the location of the target in 2 or even 3 spatial dimensions, we reduced the dimensionality of the problem by considering only targets arrayed in a circle centered on the fixation point. Cells are ordered by the location of their receptive field centers. As with the target location, we considered only a single dimension of receptive field centers, arrayed in a circle about the fixation point. The simplest form of a population response is a curve composed of points (*x*,y), in which *x* is the receptive field center of a cell, and *y* is the firing rate of that cell in response to a target at location C. C is held fixed for all cells in a given response curve.

Because of variability in individual cell responses (e.g., different tuning amplitudes, widths, firing rates), generating population response curves from individual cell tuning curves requires recording a large number of cells for each possible receptive field center location. That is, in order to get a good estimate of the population response at each spatial location we would have to record from enough cells with receptive field centers at that location so that cell-by-cell response variability would be averaged out, leaving no differences in the average responses between cells with different receptive field centers. Instead, we record from 88 cells (with preferred directions that are biased toward the contralateral visual field) and apply a simplifying assumption. We assume that receptive fields of all cells in the memory circuit have the same shape, and we construct a single (population-averaged) tuning curve from all 88 cells. In other words, we assume that any cell's response depends only on the distance of the current target from that cell's receptive field center. Next, we find the response of a cell with a receptive field center at location D to a target at location C. This is just the firing rate of the population-averaged tuning curve at *x* = C–D. To generate the population response curve for a target C, we repeat this last step, iterating through all possible values of D.

In order to take into account not just the current target C but also the previous target P, we extended this method to an additional dimension. We first build the population-averaged tuning surface (Fig. 5*b*) over the domain of previous and current target locations (P and C, respectively). To relate the population-averaged tuning surface to the population response curve, we again make the simplifying assumption that each cell in the memory circuit has a tuning surface for previous and current target locations that is identical to the population-averaged tuning surface. Different receptive field center locations differ relative to both current and previous target locations. Therefore, on average, cells with a receptive field center at D degrees in visual space would respond in trials with a current target at C and previous target at P with a firing rate determined by the distance between their receptive field center and both the previous and current target locations. This point is defined by the coordinates *x* = (C – D), y = (P – D) on the population-averaged tuning surface. To construct the population response curve for a current target C and a previous target P, we repeat this last step, iterating through all possible values of C at 1° intervals. This is equivalent to generating responses from 360 cells with preferred directions that uniformly cover the visual space. This set of responses defines a slice with a slope of +1 through the surface of Figure 5*b*. Figures 6*a*,*b* show such slices, representing population response curves for particular combinations of current and previous targets. See Supplementary Figure 1 for additional details.

Neural Effects of Previous Target

We investigate two possible (and nonexclusive) effects of the previous target on the current trial–residual memory activity, which we call a ghost, and a shift in the center of neuronal activity representing the current target, which we call a shift. To compute the normalized ghost amplitude at any particular point in time, we calculate the firing rate on trials in which the previous target was close to (within 22.5°) the receptive field, and subtract as a baseline the firing rate on trials in which the previous target was far away (>112.5°) from the receptive field. In addition, to avoid contamination from the shift effect, only trials in which the current target was more than 90° from the previous target were used. The ghost response at each point in time was normalized to the maximum response to a current target (i.e., the response to a target centered in the receptive field) at that same time interval (e.g., Fig. 7*a*). We do not show normalized data prior to 500 ms after target presentation, since normalization of cells with late responses results in unstable results in this period.

We also calculated the shift in the response to the current target. When the population response curve shifts "away" from a previous target location, neuronal tuning curves shift "toward" that location (see Supplementary Fig. 2*a*). For a clockwise (counterclockwise) tuning curve shift, firing rate clockwise (counterclockwise) from the preferred direction will be elevated compared with firing rate counterclockwise (clockwise) from the preferred direction. To compute the amount of shift at a particular point in time we computed the firing rate difference 20–70° from the receptive field center on the same side as and opposite side of the previous target. The firing rate difference was then converted into a shift amount in degrees. To accomplish this, we shift all trials by +S and −S degrees relative to the preferred direction. We then calculate the firing rate difference on the same and opposite sides of the tuning curve flanks for each amount of shift, 2S (the distance between +S and −S). This gives the expected firing rate difference for a shift of 2S degrees. We used this procedure to calculate the expected firing rate difference for each S between 0° and 90° (in 0.01° steps). We matched each observed firing rate difference to the closest expected firing rate difference in the generated range. We then assigned the shift for that observed firing rate difference to be the corresponding S for the matched expected firing rate difference. In this way, we mapped observed firing rate differences to corresponding amounts of shift in degrees. Positive values indicate a population response curve shift away from the previous target location. We computed the shift over time by calculating the shift quantity at 1-ms intervals from 500 ms after target onset until the end of the memory delay. This is shown for trials in which previous and current targets were close together (Fig. 7*c*) or far apart (Fig. 7*b*).

Population Vector Readout

To decode neural memory activity, we used a population vector readout (Georgopoulos 1988) of population activity bumps, as described by the equation:

$$P = \sum_i a_i R_i,$$

where a_i is the normalized activity of a cell with receptive field center at R_i and P is the decoded remembered location.

Converging Receptive Fields Model

We modeled a network of neurons that uniformly cover a visual space 100 × 100 units. Receptive fields were modeled as 2-dimensional Gaussians of the form

$$R(x, y) = e^{-[(x-\mu_x)^2+(y-\mu_y)^2]/2\sigma^2} , \tag{2}$$

where μ_x and μ_y are the coordinates of the receptive field center and σ is the standard deviation. A sigma of 8 was used for our simulations. To avoid possible edge effects, the size of the visual space, the receptive field sigma, and the location of target presentations were chosen so that neural activity near the edges of the space was always at baseline levels.

Zirnsak et al. (2014) showed that receptive fields of neurons in FEF converge toward saccade targets. To simulate this, receptive fields in the model converged toward target locations by a fraction of their distance, c, from the target location. This quantity was scaled so that cells with receptive fields close to the target location converged by a larger proportion of their distance than cells that were far away. We defined the sigmoid function by which c was scaled as

$$s(d) = \frac{1}{1 + e^{-a(d-b)}},$$

where a defines the slope of the sigmoid, b is the value of x at the function's half-height, and d is the distance between the receptive field center at coordinates (R_x, R_y) and the target location at coordinates (T_x, T_y):

$$d = \sqrt{(T_x - R_x)^2 + (T_y - R_y)^2} .$$

Therefore, the total movement of each receptive field R toward a target location is

$$\vec{M} = M_x + M_y, \tag{3}$$

where the x-component M_x is defined as

$$M_x = \frac{c(T_x - R_x)}{1 + e^{-a(d-b)}} \tag{4}$$

and the y component is similarly defined.

Qualitatively, receptive fields near the target move closer to it while receptive fields far from the target do not move (Fig. 8*a*). In our simulations we set $a = -0.05$ and $b = 20$.

We found that convergence to the current target alone did not explain our neuronal data. However, when we also include a small amount of persistent convergence toward the previous trial's target, we find that the model precisely replicates our behavioral and neuronal data. With both previous and current target convergence the total receptive field movement is given by

$$\vec{M}_T = \vec{M}_P + \vec{M}_C, \tag{5}$$

where $\vec{M}_P$ and $\vec{M}_C$ are the movement vectors toward the previous and current target (respectively) and determined from equations (3) and (4). In particular, we set the convergence amount c to 0.6 for convergence toward current target and 0.2 for convergence toward the previous target.

In our model, receptive fields in both the memory circuit and the readout circuit converge toward the target location in a similar way.

To simulate the task we presented previous and current target combinations on a circle with radius of 15 units in our visual space. We then used a population vector readout that accounted for receptive field changes to determine the behavioral response location in the 2-dimensional space that is predicted by the activity of the circuit. Finally, we calculated behavioral angular bias as a function of the relative distance between the previous and current target locations predicted by the model (Fig. 8*d*) using the same analysis as the actual behavioral data (Fig. 1*d*).

We measured shifts in neuronal responses when receptive field changes are not accounted for (Fig. 8*b*,*c*). We first generated the population response curves for network neurons with the same procedure we used analyzing the data. To reproduce our experimental paradigm in which we selected only neurons that showed tuned responses, only network neurons that could be driven to 33% or higher of their maximal tuning amplitude were included in population response curves. This criterion does not change the behavior of the model or results, and only affects the signal to noise ratio of "recorded" neurons (removing it is equivalent to recording unresponsive neurons in a dataset).

Results

Response Bias Toward the Previous Target

Spatial memory responses are biased toward previously memorized locations in a memory-guided saccade task (Papadimitriou et al. 2015). In this study, we look for neural correlates of this bias in frontal memory circuits. We first replicated the basic behavioral finding. Three macaques made saccades to memorized locations after delays of 1.4, 2.8, or 5.6 s (Fig. 1*a*). Saccade endpoint error increases with delay (Fig. 1*b*). We defined trial-by-trial response error as the total error minus the mean of the error obtained for that particular target position (see Materials and Methods). A plot of response error as a function of the distance between the previous and current target location reveals a systematic behavioral bias toward the memory location of the previous trial (Fig. 1*d*). The bias can be well fit by a Gabor function (peak-to-peak height = 1.13°, fit $P < 0.005$) and was significant in each of the 3 individual animals (peak-to-peak height = 1.8, 1.3, and 0.95° for monkey H, J, and P, respectively; $P < 0.005$ for all fits). The fit accounts for 0.4% of the variance of the raw data (the total behavioral error) and 76% of the variance of the binned data (the effect of the previous target).

Neuronal Responses in FEFs

We looked for neural correlates of the behavioral response bias in the FEF. We recorded from 88 neurons in FEF while monkeys performed the memory-guided saccade task. Cells with sustained memory period activity were selected using a memory screening task (see Materials and Methods). Figure 2*a* shows population-averaged firing rate as a function of time. Activity was higher when a memory target was presented at the center (red trace) or flank (orange) of each receptive field, as compared with when the target was presented outside the receptive fields (green). The elevated activity persisted for the duration of the memory period and is well-fit by a Von Mises function (Fig. 2*b*).

FEF reflects saccade endpoints as well as target locations. This can be seen by contrasting the population-averaged activity across all recorded cells when the memory-guided saccade lands either counter-clockwise (>12.5°; mean = 16.2°) or clockwise (< −12.5°; mean = −16.1°) of the memory target location (Fig. 3*a*, red versus blue trace). The tuning curves are constructed

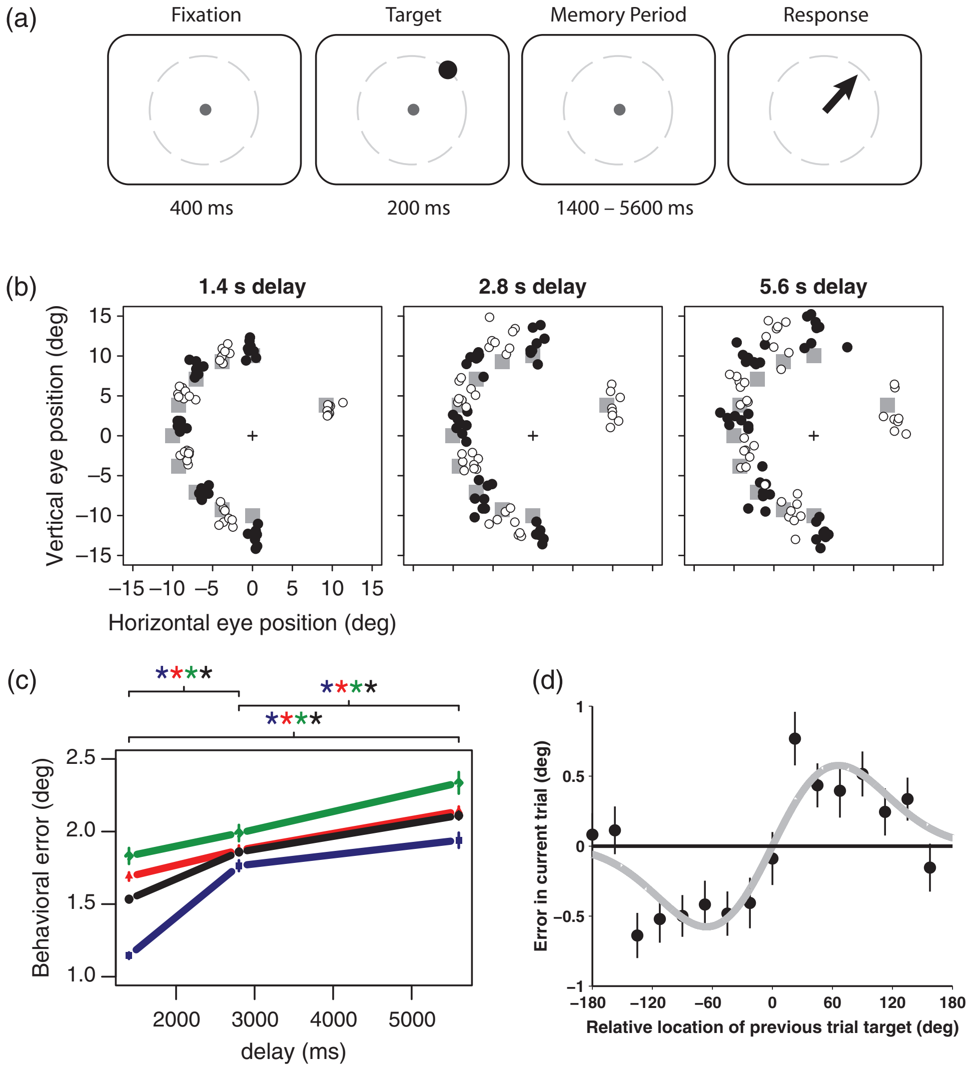

Figure 1. Behavioral task and responses. (*a*) Memory-guided saccade task. Subjects fixated on a central target presented at the center of the screen. After a fixation period of 400 ms a memory target was displayed for 150 ms at one of 16 possible peripheral locations at a fixed eccentricity. Target presentation was followed by a memory period between 1.4 and 5.6s in duration during which the subject continued to fixate. After the memory period the fixation target disappeared and the subject responded by making a saccade to the remembered location. The dashed gray line indicates where targets might appear; it was not visible to the animal. (*b*) Saccade endpoints in a subset of representative trials. Gray squares represent target locations. Responses have been colored black or white to more easily identify the associated memory target. (*c*) Euclidian distance in degrees of visual angle between the mean endpoint of saccades to the target location for each of the 3 subjects (red, green, blue) and for all subjects pooled (black). Significant *t*-tests ($P < 0.05$) of the difference between delay lengths are indicated by "*" of the corresponding color. (*d*) Error in current trial response as a function of previous target location relative to current target location. When the previous target was clockwise from the current target (negative *x*-axis) the saccadic response was biased clockwise from the current target (negative y-axis) and when the previous target was counter-clockwise from the current target the saccadic response was biased counter-clockwise from the current target. The gray line is the Gabor fit to the raw data (peak-to-peak height = 1.13, fit $P < 0.005$).

based on activity recorded in the 500 ms immediately prior to the go cue. If FEF encoded only the target at that time interval, the two curves would perfectly overlap. If FEF encoded only the saccade endpoint, the curves would be separated from one another by 32.3°, the difference in the means of the saccade endpoints used to construct each curve. In fact, they are separated by 25.3°. Figure 3*b* shows similar data from 7 different bins of saccade error. In each case, the vector sum readout of the data (a prediction of saccade error if saccade endpoint is encoded; see Materials and Methods) is plotted as a function of the actual error in the saccade endpoints. A linear fit with a slope of 0 would indicate that FEF encodes only the target location. A slope of 1 (dashed line) would indicate coding of only the saccade endpoint. The actual slope (solid line) is intermediate (0.70 ± 0.24 ° per deg, $P < 0.005$), indicating that, in the final 500 ms prior to the go cue, FEF neurons encode a location that is closer to the saccade endpoint than to the target.

Figure 3*c* shows how this measure changes over time. At the start of the trial (50–300 ms after target onset), the neural activity encodes target location independent of saccade error (slope = 0.01 ± 0.17 ° per deg, $P = 0.97$). Early memory period activity (350–750 ms after target onset) is influenced by (or influences) the saccade endpoint (slope = 0.30 ± 0.14 ° per deg, $P < 0.02$), and in the final 500 ms, the effect is twice as strong (slope = 0.70 ± 0.24 ° per deg, $P < 0.005$). The early and late slopes are significantly different from each other ($P < 0.05$). Thus, while FEF activity initially encodes the location of the memory target, it becomes more closely linked to the endpoint of the upcoming memory-guided saccade as the trial progresses. This suggests that if the bias related to the target position from the previous trial (Fig. 1*d*) is produced either within or upstream of FEF, then this bias can be read out from the FEF neurons (see also Wimmer et al. 2014).

Residual Memory Trace in the ITI and Subsequent Trial

Papadimitriou et al. (2015) modeled the bias from the previous target using a combination of a long-term and a short-term store. Only the long-term store, modeled as a bump attractor, is biased by previous target position. A simple way to produce this bias is for a remnant of activity encoding the previous target to persist into the subsequent trial. The attractor dynamics may then merge this remnant from the previous trial's target with the "bump" encoding the current target. The merger would result in a single bump of activity, encoding a location between the current and previous target. This would manifest in the behavior as a bias toward the location encoded in the previous trial. To test this hypothesis, we looked for evidence of a remnant or "ghost" of the previous target during the fixation period, after the animal had successfully completed the previous memory trial and returned to the fixation point, but before the target for the current trial had appeared.

We plotted firing rate during the fixation period, prior to target onset, as a function of the previous trial's target position relative to the receptive field. The elevation in firing rate seen on the previous trial when the target was in the receptive field (Fig. 2*a*) persisted, in an attenuated form, into the subsequent trial's fixation interval. Figure 4*a* shows data from an example cell. The firing rate is normalized to the activity recorded 50–300 ms after target presentation. The ghost activity in the fixation period is about one-quarter as large as the previous visually evoked activity from that same target. Even though the previous trial has ended, the cell shows clear tuning to the previous target location and is well fit by a Von Mises function ($P < 0.0001$). Of 88 cells, 49 showed significant ($P < 0.05$) tuning to the previous target location during the fixation interval and only 8 showed significant tuning for a location opposite to the previous target (Fig. 4*b*). Figure 4*c* shows population-averaged firing rates, similar to Figure 2*a* but sorted by the target location from the previous trial. The

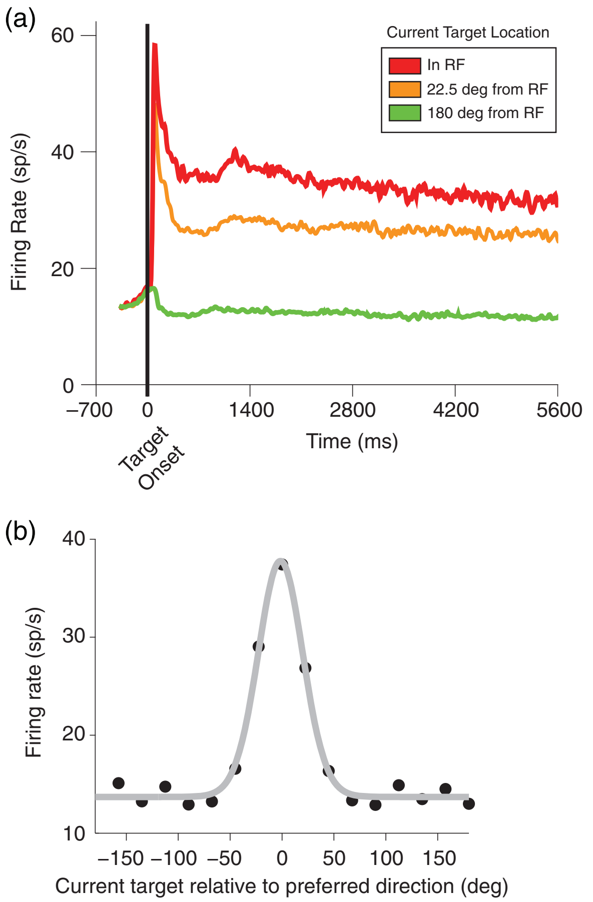

Figure 2. Tuned and sustained memory responses in FEF neurons. (*a*) FEF population response when the memory location was presented at the center of cells' receptive fields (red trace), 22.5° from the receptive field centers (orange trace) or 180° away from the receptive fields (green trace). Firing rates when the target was presented in the receptive field stay high after target offset (150 ms) and for the duration of the memory period. (*b*) Population firing rate as a function of target location relative to the receptive field is well fit by a Von Mises function (time interval 500–1500 ms, adjusted $R^2 = 0.99$, $P < 0.005$).

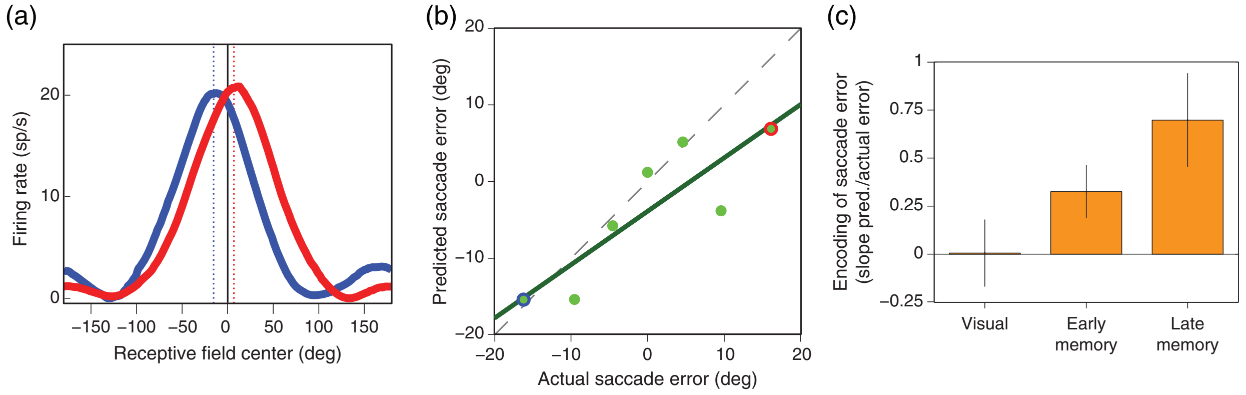

Figure 3. Neural activity reflects behavioral responses. (*a*) Population activity for trials when response error was greater than 12.5° (mean = 16.2; red trace) and less than −12.5° (mean = −16.1°; blue trace). The dotted lines show the encoded location determined with population vector decoding (red trace, 6.9°; blue trace, −15.4°). (*b*) Linear regression of saccade error predicted by neural activity and response error observed behaviorally for the time interval 500–0 ms prior to the go-cue. Trials are binned by observed response error (−10° to 10°, steps of 5°, bin width of 5°). We also included 2 bins with response error >12.5° (mean = 16.2) and less than −12.5° (mean = −16.1°). The data points outlined in blue and red correspond to the curves in *a*. The regression line (green) has a slope of 0.70 ° per deg (regression $P < 0.015$). The dashed line shows a slope of one. (*c*) Linear regression slopes for the visual period (50–300 ms after target onset; slope = 0.006 ° per deg; $P = 0.97$), early memory (350–750 ms after target onset; slope = 0.33 ° per deg; $P < 0.02$), and late memory (−500 to 0 ms prior to go-cue; slope = 0.70 ° per deg $P < 0.005$).

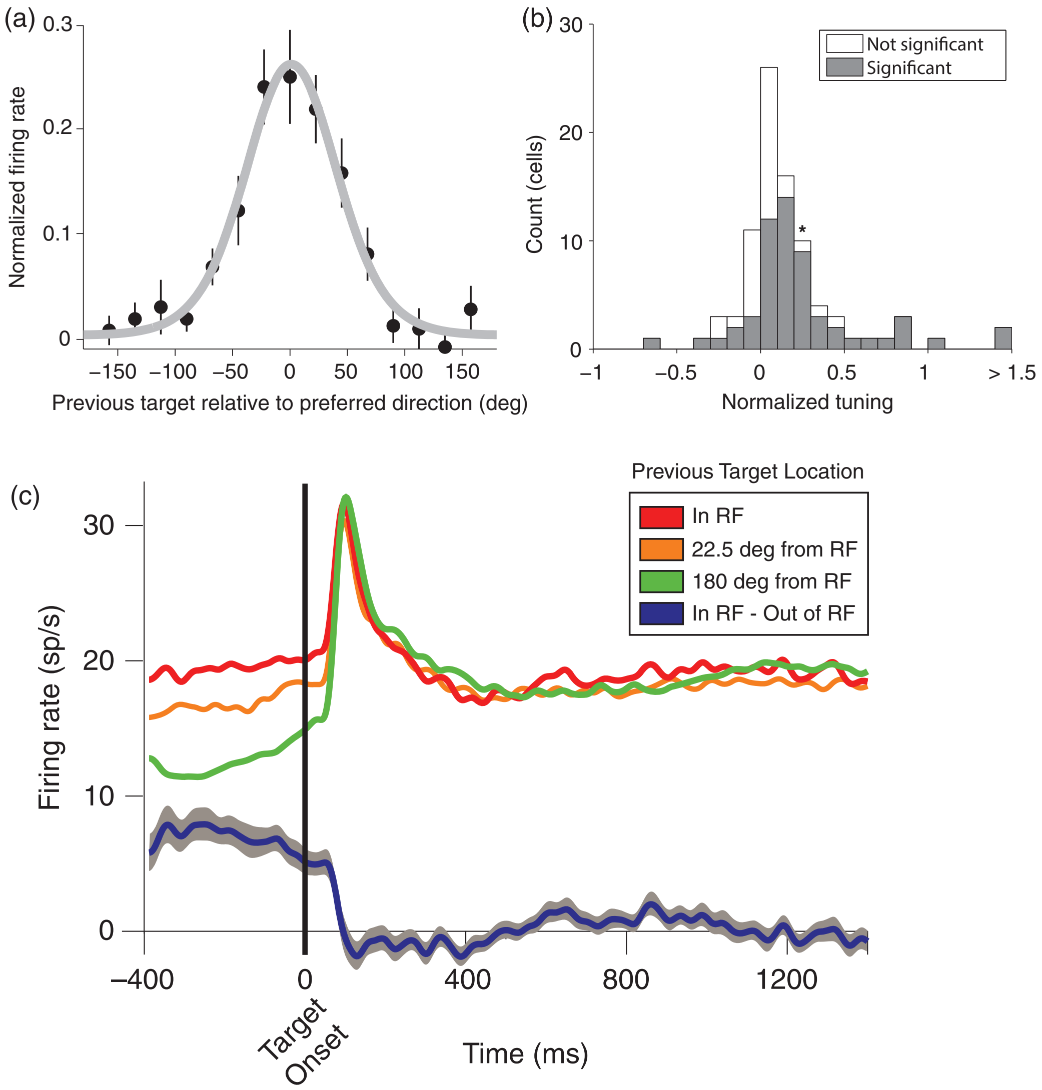

Figure 4. Residual memory tuning from the previous trial. (*a*) Firing rate of an example cell during fixation as a function of target location on the previous trial. Firing rate is scaled to the tuning amplitude 50–300 ms after target onset and the baseline is removed. (*b*) Histogram of normalized tuning to the previous target in the current trial fixation period (as in *a*) for the cell population. Gray bars indicate cells that show a significant difference in firing rates for previous targets in versus out of their receptive fields ($P < 0.05$) and white bars indicate cells that did not show a significant difference. The asterisk indicates the bin that includes the example cell in *a*. (*c*) Population firing rate when the previous target was presented at the center (red trace), 22.5° away (orange trace) or 180° away from the cells' receptive fields (green trace). The blue trace shows the difference between the red and green traces. Note that these traces are sorted by previous target position, not by current target position or by the relative current versus previous target position (compare Figs 2*a* and 5).

figure confirms that, within the fixation period that separates the end of one trial from the start of the next, there is a ghost of the previous trial's memory activity. The population-averaged effect across all cells is 5.41 ± 0.95 sp/s ($P < 0.0001$), which is 32% of the activity 500–0 ms before the end of the previous memory period (Fig. 2*a*, far right). The ghost disappears abruptly once the next target is presented. The response to target onset shows no tuning, that is, the red, orange, and green traces overlay one another shortly after the vertical line at time zero in Figure 4c. To some extent this is to be expected, since the traces are sorted on the previous target position, and previous and current target positions are completely independent of one another. However, it is contrary to the model of Papadimitriou et al. (2015). This model predicted that the residual ghost would merge with and shift the current trial's bump, preserving a small bias in firing related to the previous trial's target position such that the red trace would remain slightly higher than the green trace. There is no evidence for this in Figure 4c; the firing rate difference between the red and green trace is −0.21 ± 0.41 sp/s ($P = 0.613$) in the interval from 200 to 1400 ms after target presentation.

Influence of Previous Target on Neural Activity

The influence of the previous trial on behavior is weak for previous targets far from the current target, and strongest when the previous and current memory targets are about 60° from one another (Fig. 1*d*). This is consistent with attractor models, in which broad inhibition quickly quashes activity that is far from the dominant bump, with little effect on the dominant bump itself. Only residual activity near the dominant bump would be expected to exert an influence. The manifestation of the ghost of the previous trial might therefore depend on how far away it is from the current target. To test this idea, we constructed a population-averaged tuning surface as a function of current and previous target locations. In order to combine data across all recorded cells, we expressed target locations relative to the center of each cell's receptive field.

Figure 5 shows the resulting population tuning surfaces for the fixation and early memory periods. The activity ghost appears in the fixation period (panel a) as a horizontal band at $y = 0$, that is, on trials in which the previous target is aligned with the receptive field. During the memory period (panel b), activity is dominated by the current target. This is indicated by the vertical band at $x = 0$, reflecting trials in which the current target is aligned with the receptive field and therefore evokes a large response. There is also a faint but persistent ghost of the previous target in the memory period. The ghost is smaller in amplitude than in the fixation period, and appears only when the previous and current targets are more than about 100° apart, that is, points defined by the locus of $y = 0°$ and $x < 100$ and $x > -100°$. These loci are indicated by the green circles. The pattern is precisely the opposite of what we predicted from the behavioral data. Instead of the ghost being most obvious in cases in which the previous and current targets are close together ($x = \pm 60°$, magenta ovals), the ghost is instead most obvious when the previous and current targets are far apart (green circles).

We can use the tuning surface of Figure 5*b* to determine how each cell in the population will respond for any combination of previous and current target locations. To capture cells with receptive field centers at all possible locations for a specific combination of current and previous target position, we must take a slice through the surface with slope of +1. As an example, consider a trial in which the (current) target is 130° counterclockwise to the previous target. The relevant points on the surface are those for which the previous target direction (expressed relative to the direction of the receptive field center of each cell, or preferred direction, which ranges from −180 to + 180°) equals the current target direction (expressed relative to the preferred direction) minus 130°, or $y = x - 130$. Thus the population response to any combination of current and previous target locations is described by a line with a slope of +1. When the current and previous targets coincide, this line runs from the bottom left to the top right. For all other cases, the line starts on the far left, ascends to the top of the plot, wraps around to the bottom, and then continues on up again. This results in two parallel line segments. For the particular example of a current target 130° counterclockwise to the previous target, the locus of points forms two line segments, one from ($x = -180$, $y = -50$) to ($x = +50$, $y = 180$) and the other from ($x = +50$, $y = -180$) to ($x = 180$, $y = -50$). See Materials and Methods and Supplementary Figures 1 and 3 for details.

Figure 6*a* contrasts two slices through the tuning surface of Figure 5*b*. The slices represent the conditions when the previous target was 130° away from the current target in either a clockwise (blue) or counterclockwise (red) direction. As in Figure 5*b*, the most prominent feature is a large bump at 0°, representing the response to the current target. A much smaller bump is present in each curve at the location of the previous target, corresponding to the slightly elevated firing previously noted in the tuning surface of Figure 5*b* (green circles), close to the $y = 0$ line—the ghost of the previous trial's bump. To quantify this effect we took all trials in which the previous and current targets were separated by 90–170°, clockwise or counterclockwise, and measured the difference in firing at the previous target location, that is, the separation between the blue and red lines at the dashed lines. (For target separations <90°, the effect is different—see next paragraph. For separations approaching 180°, the red and blue lines must converge.) The mean separation, that is, the height of the ghost, was 3.04 ± 1.2 sp/s ($P < 0.01$).

In Figure 6*b*, we show a similar plot as in panel A, but now representing the condition in which the previous target was close to the current target—40° clockwise (blue) or counterclockwise (red). Once again, the prominent feature is a large bump at 0°, representing the current target position. The behavioral data and

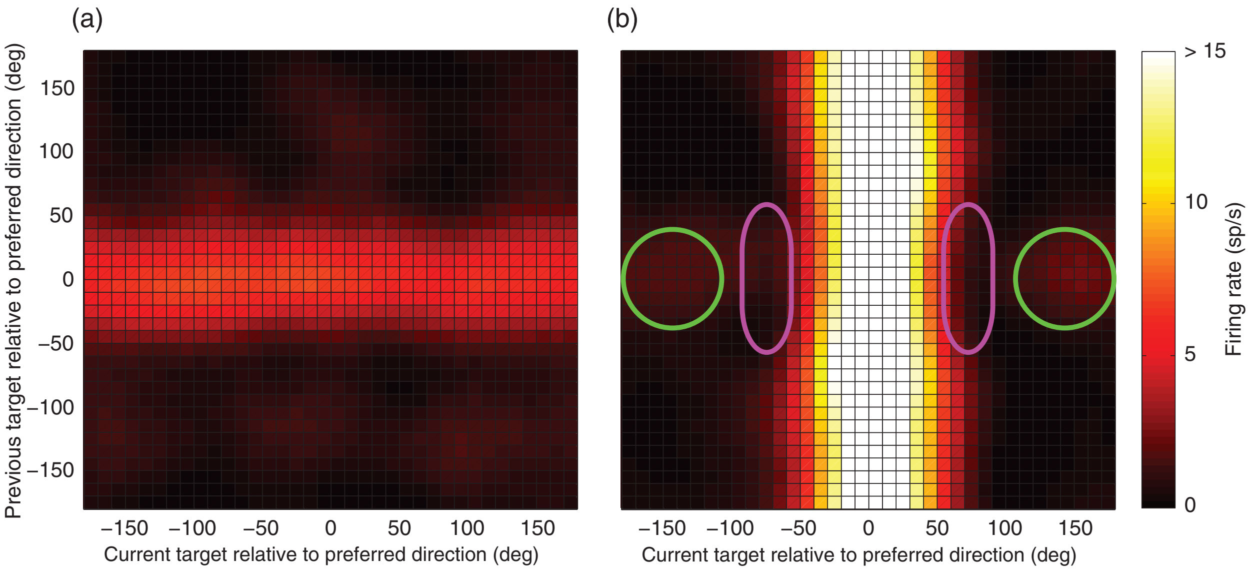

Figure 5. Two-dimensional population tuning curve of firing rate as a function of previous and current target location. In both panels, preferred direction of each unit has been rotated to 0°. (*a*) Neural activity as a function of previous and current target location during the fixation period −375 to −175 ms prior to target onset. Fixation period activity is elevated when the previous target was in the preferred direction ($y = 0°$). (*b*) Activity during the memory period 1000–1500 ms after target onset. Activity is high when the current target is in the receptive field ($x = 0°$). Smaller but clear activity elevation is evident when the previous target was in the preferred direction ($y = 0°$) and the current target is away from the preferred direction ($x > 90°$). In these figures, baseline activity of each neuron has been subtracted, but tuning amplitude has not been normalized. Amplitude normalization and baseline subtraction do not affect our results.

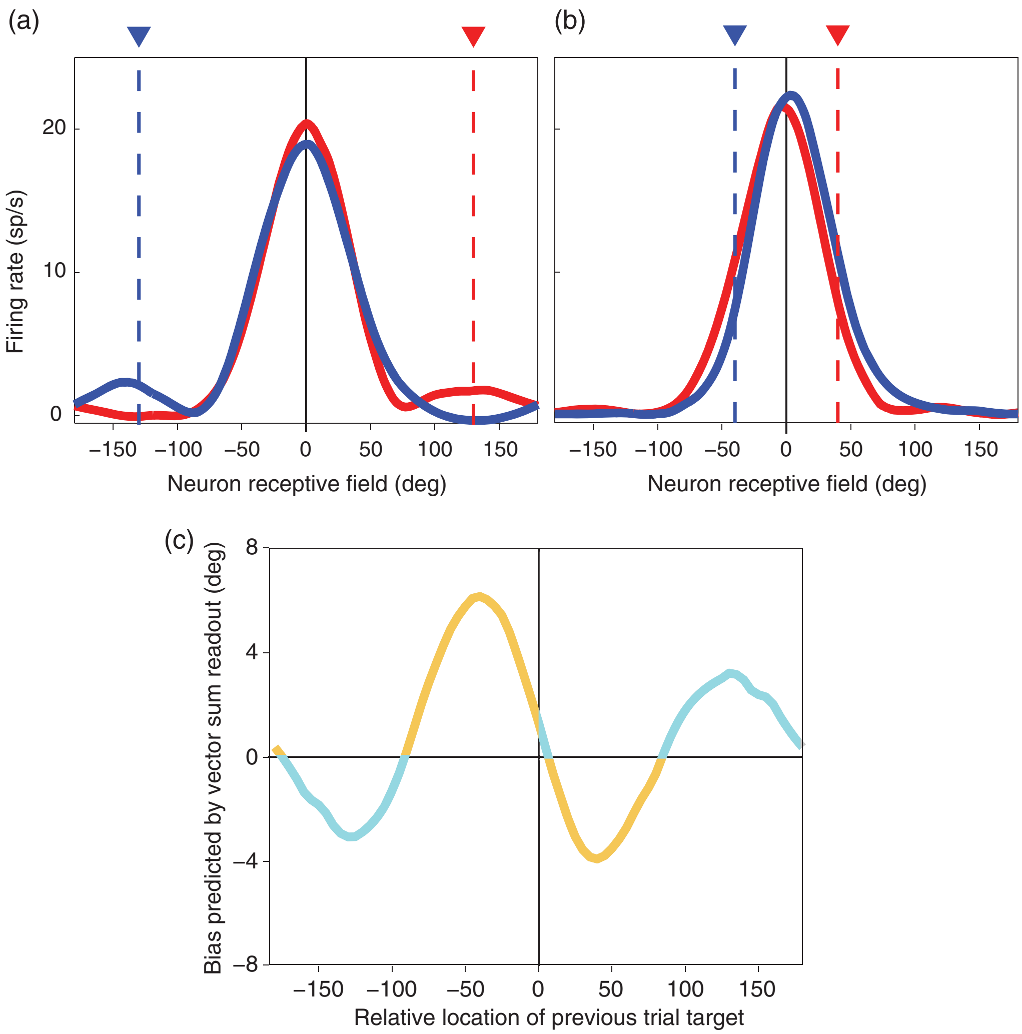

Figure 6. Population response curves and behavioral readout. (*a*) When the previous target was at 130° or −130° (red and blue triangle, respectively), the activity in neurons with a preferred direction near 130° or −130° (red and blue traces respectively) is elevated in the current trial. (*b*) When the previous target was at 40° or −40° (red and blue triangle, respectively), the activity in neurons with a preferred direction near 40° or −40° (red and blue traces respectively) is reduced in the current trial. In *a* and *b* baseline activity of each neuron has been subtracted, but tuning amplitude has not been normalized. Amplitude normalization and baseline subtraction do not affect our results. (*c*) Population vector readout of FEF activity in the interval 1000–1500 ms after target onset. When the previous and current targets are close together (e.g., *b*), the readout predicts a repulsive bias (orange) away from the previous target location in the behavioral response. When the previous and current targets are far apart (e.g., *a*), the readout predicts an attractive bias (blue) toward the previous target location.

the attractor model both suggest that attractive bias will be strongest when the previous and current targets are close together. This leads to a prediction of a large ghost. Instead, at the location corresponding to the previous target (the dashed vertical lines), the blue trace is above the red on the right and below the red on the left. This is exactly the opposite of the pattern in Figure 6*a*. We quantified this by taking all trials in which the previous and current targets were separated by up to ±80° and measured the difference in firing. The mean difference was −6.05 ± 1.0 sp/s ($P < 0.0001$). The negative sign means that the effect of a nearby previous target was to lower firing rate in the subsequent trial.

This effect can either be a suppression at the location of the previous target or a shift of the current target representation in a direction away from the previous target location. A suppression can produce a repulsive shift in the current target activity bump by suppressing the bump flank on the same side as the previous target more than the flank on the opposite side. A shift, in contrast, would be associated with a suppression of activity on the near flank and an increase in activity on the far flank. We analyzed this and found that the effect we observe is best described as a shift away from the previous target location, or a combination of a shift and a suppression, and not a pure suppression (see Supplementary Fig. 4). Yet both the behavior and the model attractor network predict a shift toward the previous target location. Thus these results are inconsistent with our predictions.

Readout

We generalized these results across a full range of target positions. We generated a family of population activity curves like those of Figure 6*a*,*b*, for all relative target positions and for several different time points, and used a population vector method to read out the location encoded by the activity. We hypothesized that a systematic error or bias in the readout would match the behavioral bias seen in Figure 1*d*. In order to test this hypothesis, we plotted the predicted bias in saccade endpoint (actual target location minus the neuronal readout of activity 1000–1500 ms into the memory period) as a function of the distance between the previous and current target locations (Fig. 6c). The plot provides a prediction of the behavioral bias we might expect to see in memory-guided saccades after a 1000–1500-ms memory period, based only on FEF activity. (Given that the influence of FEF activity on saccade endpoint is only 70% complete (Fig. 3b), this plot may overestimate the magnitude of the effect, but correctly captures the sign—attraction versus repulsion.)

The results do not match our expectations. When the previous and current target are far apart (>90°, as in Fig. 6a), the ghost from the previous trial biases the population readout in an attractive direction, such that the predicted bias has the same sign as the relative location of the previous target (Fig. 6c, blue sections of the trace). This attractive bias matches the attractive bias that is observed in the behavior (Fig. 1d). However, the predicted peak attraction occurs at 130°, whereas the peak attraction in the behavior occurs at 60°. When the previous and current targets are close together (less than 90°, as in Fig. 6b), the shift in FEF activity away from the location of the previous target biases the population readout in a repulsive direction, such that the bias and previous target locations have opposite signs (Fig. 6c, orange section of the trace). This repulsion is opposite to the attractive bias that is observed in the behavior (Fig. 1d). Thus, although FEF memory circuits show clear previous trial effects, a straightforward readout of the activity does not match the observed behavior.

Previous Target Effects Over Time

We now turn from the spatial pattern of the previous target effect to the temporal pattern. The magnitude of the attractive behavioral bias grows over the first several seconds of the delay period (Papadimitriou et al. 2015). Papadimitriou et al. modeled this by proposing that the behavior is driven by two independent stores working in parallel: a rapidly decaying but veridical visual sensory store and a sustained but distorted working memory store. The sustained store has a constant bias, present from the very start of the trial. This store has no information about the veridical target location and therefore has no way to correct its bias. The behavior relies on a weighted average of the two stores. Initially the unbiased visual store has a high amplitude and so early responses are nearly veridical. However, the visual store decays rapidly. After several seconds the output is driven almost entirely by the sustained store, and so becomes biased. Thus the model predicts that FEF, the putative sustained store, will be biased from the very start of each trial and that this bias will persist over time.

Figure 7a shows the time course of the normalized height of the residual ghost. The ghost is present at the start of trials in which the previous and current targets are far apart, with a normalized amplitude (relative to the response to a visual target) of ~20%. However, the ghost disappears rapidly. This does not match the time course of the behavioral bias, which persists for

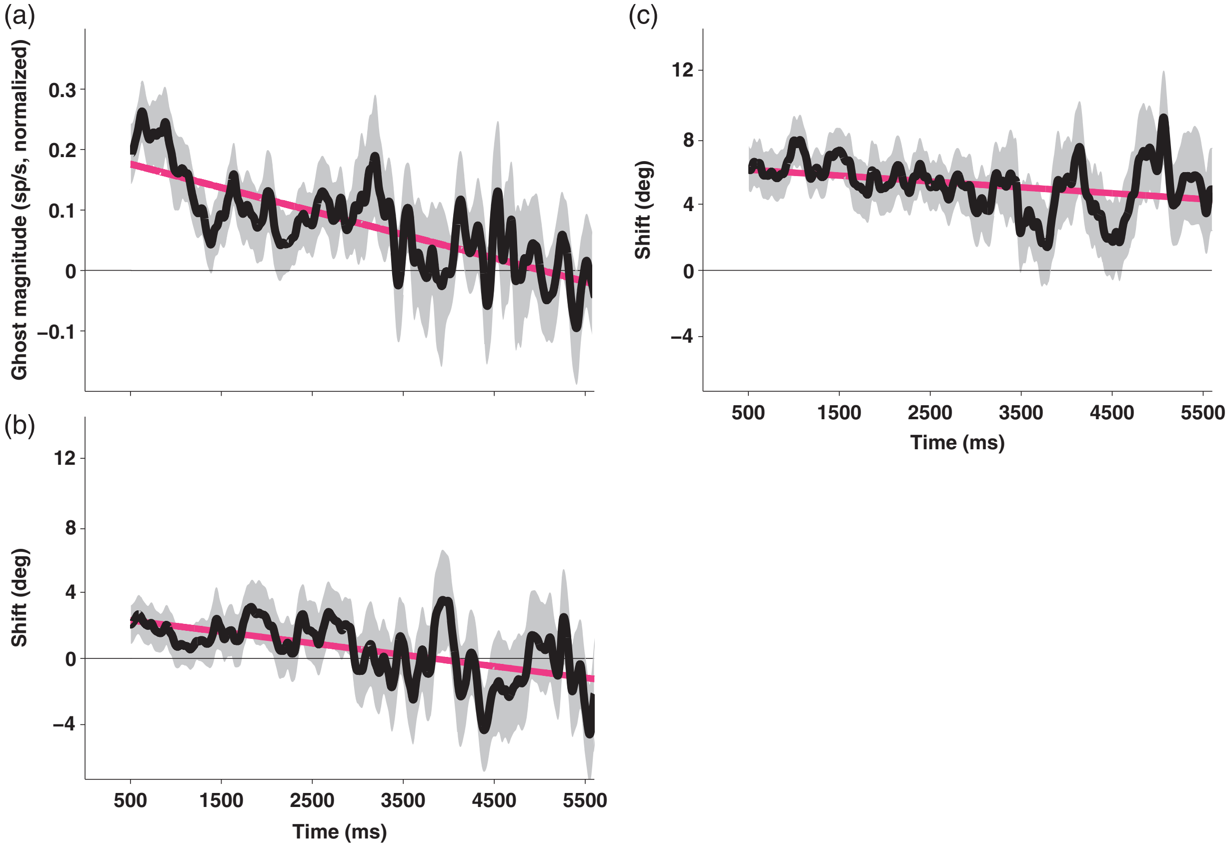

Figure 7. Ghosts of distant targets decay, but shifts resulting from near targets persist. (a) The ghost (the residual activity encoding the previous target) is not sustained through the delay period. Ghost amplitude, measured for previous and current targets that are far from one another (>90° apart), is initially ~20% as large as the visually evoked response, but decreases with a slope of 4% per second. It disappears entirely after 3.5 s (mean effect 3.5–5.6 s after target onset = 0.02 ± 0.05%, $P = 0.736$). (b) The disappearance of the ghost in a is not accompanied by a shift of the activity bump (mean shift = −0.57 ± 1.8°, $P = 0.737$). (c) The shift in the current target representation, measured for previous and current targets that are close together (<90° apart), is largely sustained (slope = −0.24°/s). It remains highly significant even at the end of the delay period (mean effect 3.5–5.6 s after target onset = 4.38 ± 1.26°, $P < 0.002$). In all panels, red lines are linear fits.

over 5 s without attenuation (Papadimitriou et al. 2015; Supplementary Figure 5*a*). Thus neither the temporal nor the spatial aspects of the ghost match the observed behavioral bias.

Attractor network models predict that a residual ghost of activity from the previous trial will merge with the representation of the current trial's target, shifting the current target representation toward the previous target location. This could conceivably explain the rapid disappearance of the ghost. In this case, the bias would manifest as an attractive shift of the current target representation, starting as the ghost disappears and persisting to the end of the trial. Figure 7*b* shows that this was not the case; the early disappearance of the ghost was not accompanied by an attractive shift of the target representation.

In contrast, Figure 7*c* shows that, when the previous and current targets were close together, a strong repulsive shift was present throughout the entire delay period. As previously noted, this shift appears to be in the wrong direction to produce an attractive bias. While the sign of the effect is reversed, the spatial profile (the relative locations of previous and current target at which the maximum effect occurs) and the temporal aspects of the shift are consistent with the observed behavior.

In summary, we observe two distinct neuronal effects of the previous target. Neither effect provides a good match to the behavior. Ghost activity has the right sign to produce an attractive bias. However, neither its spatial nor temporal properties match those of the behavior. In particular, we find the maximum behavioral effect when the current and previous targets are separated by 60°, but under these conditions the ghost disappears as soon as the target appears. When the current and previous targets are far apart, we find no behavioral effect, yet the ghost persists for 3.5 s after the target appears. Thus the ghost activity does not match the spatial and temporal patterns of the behavioral bias. In contrast, the neuronal shift effect shows a better, but still incomplete, match to the behavior. In particular, both the behavioral bias and the neuronal shift occur only when a target appears close to the location of the previous target, and both persist for the entire delay period. However, the neuronal shift predicts a repulsive bias, while the behavior shows an attractive bias. Thus, neither the ghost nor the shift can explain the behavioral bias.

One way to reconcile the repulsive shift predicted by neural activity with the attractive bias observed behaviorally would be if subjects fixate at a location biased toward the previous target. FEF encodes the relative change in eye position (the saccade vector) and not the absolute saccade endpoint. A large enough displacement in initial fixation position could result in a saccade vector that is biased away from the previous target, even while the saccade endpoint is biased toward the previous target (see Supplementary Fig. 6*a*). To address this possibility we recomputed the behavioral effect, taking into account the subject's eye position immediately prior to saccade onset (−200 to 0 ms). We found that saccade vectors calculated in this way still were still biased toward the previous target (see Supplementary Fig. 6*b*), indicating that small differences in fixation location cannot account for the discrepancy between neural activity and behavior.

A Proposed Model to Resolve Neuronal and Behavioral Manifestations of Bias

We next consider whether the phenomenon of shifting receptive fields might help explain the neuronal–behavioral discrepancy. Neurons in some areas involved in visual, oculomotor, and mnemonic processing appear to temporarily shift their receptive fields toward the goal of an upcoming saccade or attended location (Connor et al. 1997; Tolias et al. 2001; Zirnsak et al. 2014). The shifts we observe could reflect temporary shifts in receptive field locations. Previous studies have also revealed systematic mislocalizations of stimuli that occur under the same circumstances as receptive field shifts (Ross et al. 1997, 2001; Hamker et al. 2008). These earlier findings led us to hypothesize that the two effects that we observed—shifts in tuning curves during a memory period and behavioral mislocalizations of remembered targets—might be explained if receptive fields converge toward remembered locations and some fraction of that convergence persists across trials.

To test this idea, we simulated a network of memory neurons with receptive fields uniformly tiling visual space. When a memory target is presented to the network, neurons shift their receptive fields toward the target with an amplitude that is proportional to their distance from the target, multiplied by a sigmoid. The multiplication by a sigmoid confines the shifting to the vicinity of the target; receptive fields far from the target do not shift. Figure 8*a* shows the resulting shifts. The starting points of the depicted vectors represent the original receptive field centers, and the endpoints represent the final shifted position due to stimulus presentation. By construction, there is strong convergence toward the current target (red; both left and right panels) and a weak convergence toward the previous target (right panel; blue).

These receptive field shifts affect the network readout. Imagine a neuron with a receptive field centered 16° to the left of the fovea. If this field shifts 10° to the right, its new center will be 6° to the left of the fovea. In a vector sum readout, this cell would "vote" for a position 16° to the left. After the shift, the cell would respond most strongly to a target appearing at 6° left, not 16° left. Yet this strong response to a target at 6° left would be mistaken as a "vote" for the 16° leftward location; the cell would now bias the vector sum readout to the left. In general, a receptive field shift in one direction would shift a vector sum readout in the opposite direction. Note, however, that with just one target, the shifts of the receptive fields across the population are symmetric (Fig. 8*a*, left). As a result, the biases produced by individual cells will exactly cancel one another, producing no net bias in the vector sum readout. The addition of even a small residual shift from the previous trial will break the symmetry and result in a distorted readout (Fig. 8*a*, right). Since shifts bias the readout in the opposite direction from the shift, the distortion will result in a repulsion away from the location of the previous target.

Our quantitative simulations confirm this qualitative description. We replicated the structure of our task, presenting memory targets along a circle of radius 15° from the fixation point. The simulated neurons shifted their receptive field locations during the memory period, as described for area FEF (Zirnsak et al. 2014). In the model, there was no residual ghost activity from the previous trial. We asked what the effect of the receptive field shifts would be on the neuronal data. Population response curves from the simulated neurons shows repulsive neuronal shifts, just as in the data (compare Fig. 8*b*, with Fig. 6*b*). Vector sum readouts of the simulated network show repulsive biases, matching the repulsive bias in the recorded neuronal data (compare Fig. 8*c*, with Fig. 6*c*). (For simplicity, the simulation did not include residual ghosts when previous and current targets were far apart [Fig. 6*a*]; had this been included, then the small portions of the actual behavioral readout shown in blue in Fig. 6*c* would also have been replicated.) Critically, although the readouts of the simulated network generally match the vector sum readouts of the actual neuronal recordings, neither of these two readouts consistently match the

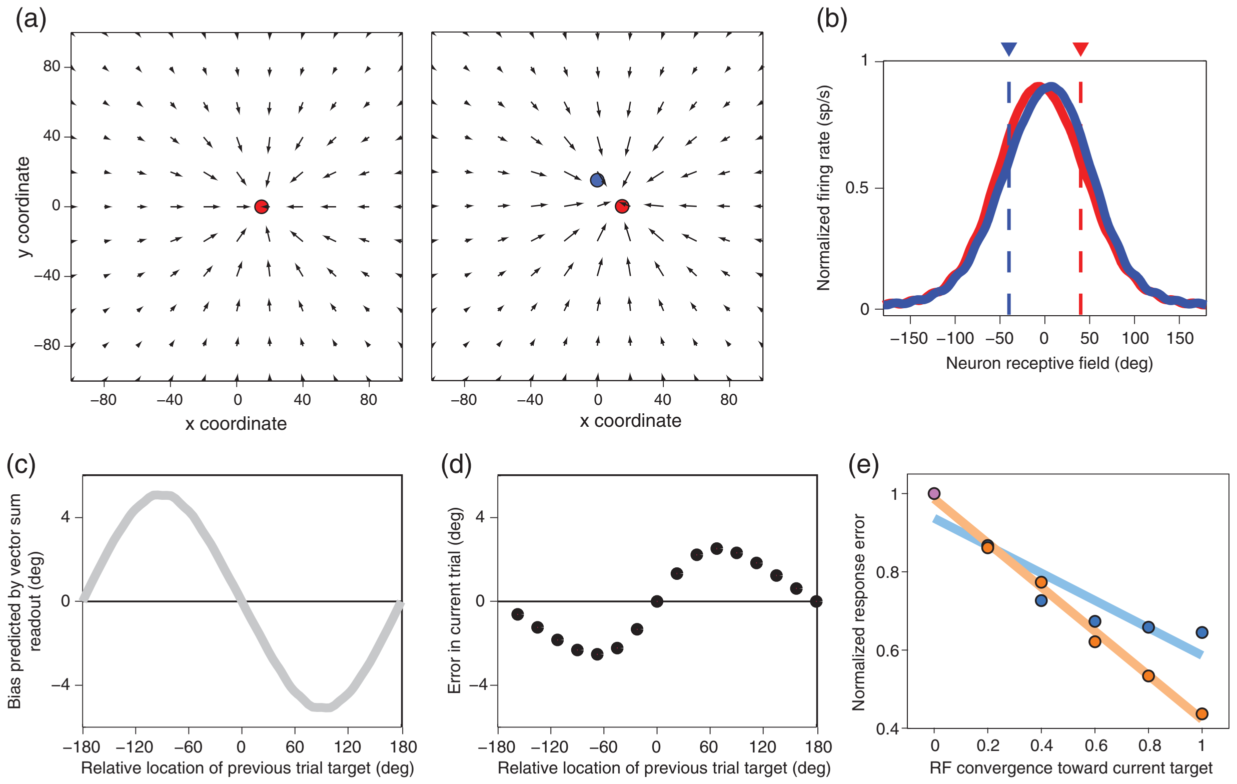

Figure 8. Convergence of receptive fields can explain both neuronal activity patterns and behavioral bias. (*a*) (Left) Example of receptive fields convergence toward the memory target ($c = 0.6$). Convergence amount is comparable to Zirnsak et al. (2014). (Right) Receptive fields converge toward the current memory target ($c = 0.6$) with some residual convergence toward the previous target ($c = 0.2$). (*b*) Population response curves appear to move away from the previous target location when receptive fields shift toward both the current and previous target (compare Fig. 6*b*). (*c*) The population vector sum of population response curves like those in *b* predict repulsive bias (compare Fig. 6*c*). (*d*) When the readout takes receptive field shifts into account (see text) the distorted distribution of receptive fields produces attractive bias (compare Fig. 1*d*). (*e*) Response error as a function of RF convergence toward current target (*x*-axis) in the presence (orange) or absence (blue) of convergence toward the previous target. When noise is added to cell firing rates convergence of receptive fields improves performance, even when some convergence persists into the subsequent trial. See text for additional details.

behavioral results, which show an attractive rather than a repulsive bias (Fig. 1*d*).

Next, we asked what the vector sum readout of the simulated network would look like if we assume that the network “knows” about the receptive fields shifts and takes them into account when generating the readout. In this case, a cell whose RF is normally at 16° left but which shifts over 10° to the right would “vote” for the 6° left position, not the 16° position. As a result, individual cells show no bias. However, because the population of cells no longer uniformly tiles space, the vector sum becomes biased. In particular, the vector sum is biased toward the location with the most dense accumulation of receptive fields, that is, the point in space toward which the receptive fields are converging. This results in a strong attractive bias toward the current target location, and a weak attractive bias toward the previous target location (Fig. 8*d*). The attractive bias in the simulation’s behavior is along the vector pointing from the current target toward the previous target, and therefore in addition to the angular component has a radial component as well. This is consistent with the behavioral error from our subjects who also show a radial component in behavioral error (see Supplementary Fig. 7).

A shift in the overall distribution of receptive fields, such that they cluster about a particular stimulus location, can be viewed as a shift in computational or representational resources to that location. For example, increasing the density of neurons that encode a memorized location could serve to make the memory trace more robust and resistant to noise compared with a network that lacked such a shifting mechanism. Our simulation confirms this intuition. We added random noise to each cell’s response, prior to computing a vector sum readout. We then calculated the response error for each trial under various conditions of receptive field convergence (Fig. 8*e*). Response error was normalized to the error when convergence to both the previous and current target was 0 (purple point in the top left corner). As we varied the degree of convergence toward the current target from no convergence ($x = 0$; no change in receptive field locations) to complete convergence ($x = 1$; all receptive fields are aligned with the target position), the error in the behavioral response decreased linearly (orange points and trace; 0.6% reduction in error per 1% of RF convergence, $P < 0.0001$). This improvement in the behavioral response was reduced but still present even when a fraction of the convergence (0.2) from the previous trial persists into the current trial, as in our model (blue trace, 0.4% reduction in error per 1% RF convergence, $P < 0.03$).

Discussion

Behavioral responses in spatial memory tasks are biased toward the memoranda of the previous trial (Fig. 1*d*). To identify neural correlates of this bias, we recorded from FEF during a memory-guided saccade task. We selected spatially tuned cells with sustained responses during a memory period (Fig. 2). These cells code target location early in the trial and saccade endpoint late in the trial (Fig. 3). A small amount of activity persisted after the end of each trial and could be seen in the subsequent fixation period prior to the appearance of the next target (Fig. 4). When averaged across all conditions, this residual or ghost activity disappeared as soon as a new target appeared. Given that the behavioral bias depends on the distance between the previous and current targets, we examined the ghost activity as a function of this distance (Fig. 5). We found that ghost activity persists during the memory period only when the current and previous targets are separated by >90° (Fig. 6). However, this activity cannot explain the behavioral bias, since the behavioral bias is strongest when the previous target is 60° from the current target (spatial mismatch). In addition, even when the target separation was large and the ghost did persist, the ghost lasted only about 3 s, whereas the behavioral bias persisted indefinitely (Fig. 7*a*).

When the previous target appeared within 90° of the current target, there was no ghost, but the population activity encoding the current target was shifted in position. However, this shift was directed away from the location of the previous target, that is, in a direction opposite that which would be predicted by the behavioral bias (Fig. 6*b*). Unlike the ghost but like the behavioral bias, this shift persisted throughout the duration of the trial (Fig. 7*c*). The fact that the shift is directed away from the previous target (Fig. 6*b*) does not contradict the fact that the firing rate in general codes the saccade direction (Fig. 3*c*). The previous target effect is small, accounting for only 0.4% of the variance of the total behavioral error. As a result, neural activity can reflect overall behavioral error (the direction of the saccade relative to the target, Fig. 3*c*) with one sign, and simultaneously reflect the previous target bias with an opposite sign. In summary, our data show that neural activity in FEF is influenced by prior memoranda, but a conventional readout of this activity (Fig. 6*c*) is not congruent with the observed behavior (Fig. 1*d*). Specifically, when the previous and current targets are close together, the neural activity (the shift in the representation of the current target) predicts that saccades will be repulsed away from the previous target, whereas the observed behavior shows a strong attractive bias. When the previous and current targets are far apart, the neural activity predicts a large attractive bias, whereas the observed behavior shows minimal bias.

To reconcile the neuronal data with the behavioral responses, we propose that receptive fields in FEF shift in response to memory targets, and that the fields do not completely revert back to their original locations after the end of a trial (Fig. 8). In a model, a small amount of residual shift exactly reproduces both behavioral and neuronal effects: the memory-guided saccades read out of the model show an attractive bias toward the location of the memoranda of the previous trials, and the activity in the simulated neurons show a repulsive shift.

Receptive fields may converge to over-represent a location in order to increase processing of that location. More specifically, in the case of spatial working memory, receptive field convergence may make the memory trace more robust to noise, since the effect of stochastic fluctuations in activity will drop as the number of neurons involved increases (Fig. 8*e*, orange trace). If a fraction of this convergence persists into the subsequent trial, then this will introduce a bias toward the previous trial's target location (Fig. 8*d*). However, as long as the residual convergence from the previous trial is small compared with the convergence toward the current trial's target, the total behavioral error will still be reduced as compared with the case of no convergence (Fig. 8*e*, blue trace).

Adaptation Versus Receptive Field Changes

An alternative explanation for the neuronal–behavioral discrepancy that we observe is a form of firing rate adaptation. Neurons with receptive field centers at the previous target location (adaptation) or some distance away (e.g., surround-suppression) may fire at reduced levels on the next trial, compared with the level they would have fired absent the effect of a previous target. This would produce a pattern of results similar to what we have shown, with a readout of neural activity that would be biased away from the previous target location. However, this explanation would account for the neural results but not the behavioral findings. In order for adaptation to account for the behavioral findings, downstream circuits closer to the motor output would have to show facilitation to counteract the suppression in FEF circuits, and this facilitation would need to overcompensate for the FEF adaptation in order to convert the repulsive FEF bias into an attractive behavioral bias. In addition, the fact that neurons show clear shifts in activity is further evidence against the adaptation model (e.g., see Supplementary Fig. 4).

Shift Mechanism

Receptive field shifts of the type we have posited could occur through Hebbian-type mechanisms. For example, the propagation of activity related to a mnemonic target into FEF might result in an increase in the efficacy of the synapses conveying that activity. On a subsequent trial, these strengthened connections would cause neurons in FEF to be more readily driven by inputs from the previously active location, broadening the point image in FEF and thereby effectively producing a receptive field shift. Connections between active FEF neurons and the output cells to which those FEF neurons project could undergo a related Hebbian-type strengthening, with strengthened connections causing FEF neurons to more readily drive output neurons. This would effectively reverse the changes that had occurred at the input level, resulting in an output that compensates for the earlier shift (see Supplementary Fig. 2 for details). We note, however, that we have no single unit data to either directly support or contradict this particular model.

We modeled the pattern of activity changes that we observe as shifts in receptive fields of FEF neurons. A strong prediction of the model is that behavioral responses evoked by direct stimulation (e.g., microstimulation) of FEF memory cells should be biased in the direction of the previous target location. Furthermore, direct stimulation of the projection targets of these cells would show no bias.

Template Matching Versus Vector Sum Readout

Supplementary Figures 2*a*,*b* show that when the receptive fields of cells move toward a location, the vector sum of the outputs from those cells will produce a repulsive bias. An alternative to the vector sum readout used in our model is the template-matching algorithm that was first described by Abbott (1994). Supplementary Figures 8*a*,*b* show that the template matching mechanism, like the vector sum, also results in a repulsive bias. However, when neuronal responses are amplitude modulated, such that cells with receptive fields close to a particular point in space respond more strongly than cells far from that point, then the readouts of both vector sum and template-matching mechanisms are shifted toward the selected point. This is shown in Supplementary Figures 2c,2d for the vector sum output, and in Hamker et al. 2008 for the template matching algorithm.

Our prefrontal data contain shifts in receptive field locations without systematic changes in response amplitude. Therefore, the read-outs of both vector-sum and template-matching mechanisms will move away from (not toward) the direction of the receptive field shift. Such a repulsive bias is not seen in the behavior.

Like the current study, Zirnsak et al. (2014) found receptive field convergence in FEF neurons and an attractive bias in the associated behavior during a saccade task. As described above, we find in our simulations that when receptive fields converge, both the vector sum and the template matching algorithms produce a repulsive bias in the associated behavior. There are at least two possible explanations for why Zirnsak et al. obtained an attractive rather than repulsive bias using the template-matching algorithm. One possibility is that the attractive bias is idiosyncratic to the particular way in which their simulation was set up (see Supplementary Fig. 9). A more interesting possibility is that the neurons that Zirnsak et al. recorded showed not just the receptive field shifts that were highlighted in their study, but also strong systematic modulations in response amplitude that were not described. While the receptive field shifts would produce repulsion, sufficiently strong amplitude modulation could drive an attractive behavioral bias. Since the authors do not comment on response amplitude, and instead present only normalized data, it is difficult to determine whether amplitude modulations were in fact present and drove their convergent results. One intriguing possibility, consistent with the results from both the current study and from Zirnsak et al. is that strong amplitude modulations might occur in movement neurons, a class of cells that are active mainly at the time of the saccade. Many movement neurons do not show strong memory responses in the preferred directions (Lawrence et al. 2005) and therefore would have been underrepresented in our study compared with Zirnsak et al. Movement cells are likely driven by cells with memory activity. Therefore amplitude modulations in movement neurons would be consistent with the downstream mechanisms we postulated in the previous paragraph.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding

This work is supported by National Eye Institute (grant number R01-EY012135 to L.H.S.), National Institute of Mental Health (grant number 5F31MH094076 to C.P.), and National Science Foundation (grant number T32-NS073547 to C.P.).

References

Abbott LF. 1994. Decoding neuronal firing and modelling neural networks. Q Rev Biophys. 27:291–331.

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

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.

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.

Connor CE, Preddie DC, Gallant JL, Van Essen DC. 1997. Spatial attention effects in macaque area V4. J Neurosci. 17: 3201–3214.

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

Corbetta M, Kincade JM, Shulman GL. 2002. Neural systems for visual orienting and their relationships to spatial working memory. J Cogn Neurosci. 14:508–523.

Curtis CE, Sun FT, Miller LM, D’Esposito M. 2005. Coherence between fMRI time-series distinguishes two spatial working memory networks. Neuroimage. 26:177–183.

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

Dunnett SB, Martel FL. 1990. Proactive interference effects on short-term memory in rats: I. Basic parameters and drug effects. Behav Neurosci. 104:655–665.

Edhouse WV, White KG. 1988. Cumulative proactive interference in animal memory. Anim Learn Behav. 16:461–467.

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.

Georgopoulos A. 1988. Neural integration of movement: role of motor cortex in reaching. FASEB J. 2:2849–2857.

Hamker FH, Zirnsak M, Calow D, Lappe M. 2008. The peri-saccadic perception of objects and space. PLoS Comput Biol. 4:e31.

Jarvik ME, Goldfarb TL, Carley JL. 1969. Influence of interference on delayed matching in monkeys. J Exp Psychol. 81:1.

Jonides J, Nee DE. 2006. Brain mechanisms of proactive interference in working memory. Neuroscience. 139:181–193.

Judge SJ, Richmond BJ, Chu FC. 1980. Implantation of magnetic search coils for measurement of eye position: An improved method. Vision Res. 20:535–538.

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, Snyder LH. 2005. Delay-period activity in visual, visuomovement, and movement neurons in the frontal eye field. J Neurophysiol. 94:1498–1508.

Moise SL. 1976. Proactive effects of stimuli, delays, and response position during delayed matching from sample. Anim Learn Behav. 4:37–40.

Papadimitriou C, Ferdoash A, Snyder LH. 2015. Ghosts in the machine: memory interference from the previous trial. J Neurophysiol. 113:567–577.

Robinson DA. 1963. A method of measuring eye movement using a scleral search coil in a magnetic field. IRE Trans Biomed Electron. 10:137–145.

Ross J, Morrone MC, Burr DC. 1997. Compression of visual space before saccades. Nature. 386:598–601.

Ross J, Morrone MC, Goldberg ME, Burr DC. 2001. Changes in visual perception at the time of saccades. Trends Neurosci. 24: 113–121.

Schall JD. 1991. Neuronal activity related to visually guided saccades in the frontal eye fields of rhesus monkeys: comparison with supplementary eye fields. J Neurophysiol. 66:559–579.

Sommer MA, Wurtz RH. 2000. Composition and topographic organization of signals sent from the frontal eye field to the superior colliculus. J Neurophysiol. 83:1979–2001.

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.

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

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.

Tolias AS, Moore T, Smirnakis SM, Tehovnik EJ, Siapas AG, Schiller PH. 2001. Eye movements modulate visual receptive fields of V4 neurons. Neuron. 29:757–767.

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

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.

Zirnsak M, Steinmetz NA, Noudoost B, Xu KZ, Moore T. 2014. Visual space is compressed in prefrontal cortex before eye movements. Nature. 507:504–507.