HHS Public Access

Author manuscript

Neuron. Author manuscript; available in PMC 2017 July 21.

Published in final edited form as:

Neuron. 2015 October 21; 88(2): 367–377. doi:10.1016/j.neuron.2015.08.033.

Cortical and subcortical contributions to short-term memory for orienting movements

Charles D. Kopec[1,2], **Jeffrey C. Erlich**[1,2,3,4], **Bingni W. Brunton**[1,2,5], **Karl Deisseroth**[3,6], and **Carlos D. Brody**[1,2,3,*]

[1]Princeton Neuroscience Institute; Princeton University, Princeton, NJ 08544, USA

[2]Department of Molecular Biology; Princeton University, Princeton, NJ 08544, USA

[3]Howard Hughes Medical Institute

[4] NYU-ECNU Institute of Brain and Cognitive Science; New York University Shanghai, Shanghai 200122, China

[5]Department of Biology and Department of Applied Mathematics; University of Washington, Seattle, WA 98195, USA

[6]Department of Bioengineering; Neuroscience Program; Department of Psychiatry and Behavioral Sciences; CNC Program; Stanford University, Stanford CA 94305, USA

Abstract

Neural activity in frontal cortical areas has been causally linked to short-term memory (STM), but whether this activity is necessary for forming, maintaining, or reading out STM remains unclear. In rats performing a memory-guided orienting task, the frontal orienting fields in cortex (FOF) are considered critical for STM maintenance, and during each trial, display a monotonically increasing neural encoding for STM. Here, we transiently inactivated either the FOF or the superior colliculus, and found that the resulting impairments in memory-guided orienting performance followed a monotonically decreasing timecourse, surprisingly opposite to the neural encoding. A dynamical attractor model in which STM relies equally on cortical and subcortical regions reconciled the encoding and inactivation data. We confirmed key predictions of the model, including a time-dependent relationship between trial difficulty and perturbability, and substantial, supralinear, impairment following simultaneous inactivation of the FOF and superior colliculus during memory maintenance.

INTRODUCTION

Short-term memory, which relies on neural activity in frontal cortices (Fuster, 1997), is fundamental for successful interactions with a complex world. Memory-guided orienting

*To whom correspondence should be addressed; brody@princeton.edu.

CONTRIBUTIONS

CDK carried out the experiments, modeling, and analysis. JCE developed the original task and carried out the electrophysiological experiments. BWB and CDK carried out experiments to characterize eNpHR3.0 in our system. KD contributed the optogenetic reagents. CDK and CDB conceived the project and wrote the paper. CDB supervised all aspects of the project.

behaviors use a well-studied form of short-term memory (Funahashi et al., 1989; Goldman-Rakic, 1996). On each trial of a memory-guided orienting task, subjects are first presented with a sensory cue indicating an appropriate orienting direction; they must then withhold their motion throughout a memory delay period; and after a “go” cue that indicates the end of the delay period, they must read information out from short-term memory and perform the orienting motion.

Hours-long pharmacological silencing of cortical premotor areas such as the primate Frontal Eye Fields (FEF) and its suggested rat homologue, the Frontal Orienting Fields (FOF), produces an impairment that is substantially greater for memory-guided orienting than for sensory-guided orienting (Sommer & Tehovnik, 1997; Dias & Segraves, 1999; Erlich et al., 2011). As a result, neural activity in these areas, together with parietal and prefrontal cortices (Chafee & Goldman-Rakic, 1998; Harvey et al., 2012), is thought to be the neural substrate of the short-term memory for orientation. Correlations between firing rates during the memory delay and the subsequent orienting movement, in both rats (Erlich et al., 2011) and primates (Bruce et al., 1985; Schall & Thompson, 1999; DiCarlo & Maunsell, 2005; Wimmer et al., 2014) have suggested that these cortical areas are particularly necessary for a specific phase of short-term memory: maintenance. But no causal test of this commonly held hypothesis has yet been performed. (A different cortical region, the anterior lateral motor cortex (ALM) has recently been shown to be necessary in mice for short-term memory maintenance in a related task, memory-guided directional licking (Guo et al., 2014). The results presented below contrast with Guo et al.’s results, a comparison that we expand on in the discussion section.) Another potentially relevant brain region is the superior colliculus (SC) in the brain stem, which despite not being generally thought to subserve short-term memory maintenance, shares many properties, described in more detail below, with the rodent FOF and primate FEF during memory-guided orienting tasks.

Here we used temporally-precise optogenetic inactivation to dissect the role of the FOF and SC in short-term memory. We selectively and unilaterally silenced each brain region during specific time periods within trials of a memory-guided orienting task. Using the same task, Erlich et al. 2011 found that neural encoding in the FOF for the upcoming orienting movement is weakest during the cue stimulus and grows monotonically throughout the memory delay period. This led us to predict that the behavioral effect of inactivating the FOF would be weakest during the cue stimulus (when there is little encoding to disrupt), and would grow monotonically as it occurred later in the delay period (when there is more substantial encoding to disrupt). Directly contrary to these expectations, we found that unilateral inactivation of either the FOF or the SC had its biggest behavioral effect during cue presentation, and the effect magnitude decreased monotonically during the delay period. In other words, the stronger the neural encoding that was silenced, the weaker the effect produced by that silencing.

One potential explanation for these results is that the observed neural encoding is largely an epiphenomenon, and that neural activity causal to memory maintenance during the delay period is gradually transferred out of the FOF and SC and mainly held in some other brain region(s). However, we found that an alternative explanation, which would fully reconcile the electrophysiological and optogenetic findings, was provided by a dynamical attractor

model previously used to jointly describe decision-making and short-term memory (Wong & Wang, 2006; Machens et al., 2005). In our adaptation of this model, the FOF and the SC play fully causal and equivalent roles within a mutual-inhibition attractor network that maintains the memory over the delay period. During each trial, the model transitions from a relatively perturbable decision-making phase with little encoding (during the stimulus cue), into a relatively unperturbable maintenance phase with strong encoding (during the delay period, when the system has entered a dynamical attractor that maintains the memory). The model thus displays monotonically increasing encoding together with a monotonically decreasing effect of silencing.

We probed two key predictions of the model. First, the model predicts that silencing both the FOF and the SC simultaneously at the end of the delay period should have a substantial effect, greater than the sum of inactivating either region alone. Second, the model predicts that during the cue period, difficult trials should be more perturbable than easy trials, but that by the end of the delay period, perturbability should be independent of trial difficulty. This is because by the end of the delay period, the system's fall into a dynamical attractor will have erased any information about how it got there. We confirmed both predictions. Taken together, our data support the idea that the encoding and maintenance of short-term memory for orientation is distributed across both cortical and subcortical regions, and is well described by a dynamical attractor mechanism.

Results

We trained rats to perform a memory-guided orienting task (Fig. 1A; (Erlich et al., 2011)). In these tasks, inhibitory perturbations of frontal premotor cortices are expected to be more informative than excitatory perturbations: low-amplitude electrical microstimulation of the primate FEF readily evokes involuntary short-latency orienting motions (Robinson & Fuchs, 1969; Bruce & Goldberg, 1985) that can break memory delay period fixation (i.e., break motion withholding), and thus preclude completion and analysis of normal behavioral trials. We found similar effects when Channelrhodopsin (ChR2) in rat FOF was used to cause an excitatory perturbation, even when the perturbation was very short (Fig. S1A–C). In contrast, even long inhibitory perturbations did not hinder normal completion of behavioral trials (Fig. S1D,E). Inhibitory perturbations are therefore the focus of our study. Unilateral injections into the FOF of each rat were made with a virus engineered to drive expression of the light-activated chloride pump eNpHR3.0 (Gradinaru et al., 2010; Tye et al., 2011), under the control of the pyramidal neuron-specific calcium calmodulin protein kinase IIα promoter (AAV-CaMKIIα-eNpHR3.0-eYFP; Fig. S2A). We made five injection tracts per animal, arranged as a cross with a 500 μm spacing. Together with the injections, a chemically-sharpened 125 μm-diameter optical fiber was implanted, 1 mm deep, into the center of the virus injection area (Lambelet et al., 1998; Hanks et al., 2015). During behavioral sessions, the optical fiber was connected to a computer-controlled 532 nm laser. Simultaneous laser stimulation and electrophysiological recordings in an anesthetized preparation indicated that our procedures produced robust neuronal inhibition within a sphere of ~750 μm radius from the fiber tip (Hanks et al., 2015). Both anesthetized and further awake behaving recordings indicated onset and offset of inhibition in the tens of milliseconds (Fig. S3).

We first asked whether the effects seen under hours-long pharmacological inactivation (Sommer & Tehovnik, 1997; Dias & Segraves, 1999; Erlich et al., 2011) could be reproduced when inactivations lasted only a few seconds. In a randomly chosen 25% of trials, we turned the laser on at the start of the auditory orienting cue, and kept the laser on throughout the entire trial (2 s), until the animal had completed its orienting motion (Fig. 1A, "Laser On"). As expected, this produced a robust ipsilateral bias specific to the Laser On trials (15.5%, $p<0.01$, n=4 rats, Fig. 1B). Whole-trial inactivation in a sensory-guided orienting task that lacks a short-term memory component (Fig. 1C) produced no detectable bias (1.1 +/− 3.7%, p=0.29, n=3 rats, Fig. 1D; see also S1F,G), providing further evidence for the involvement of the FOF in orienting guided specifically by short-term memory. These data also demonstrate that inhibition of the FOF does not impair the ability to produce normal orienting motions (consistent with Erlich et al., 2011; see also Fig. S4).

We next tested the effect of inactivating only during the 1 s fixation period, encompassing decision formation, short-term memory encoding, and memory maintenance, but prior to initiating the motor response. This again resulted in a significant ipsilateral response bias (17.9%, $p<0.01$, n=7 rats, Fig. 2B), confirming that FOF activity preceding the motor response is causal to the subject's choice.

As shown in Fig. 2A, neural encoding of the direction of the upcoming orienting movement is weakest in the FOF during the cue presentation, and gradually increases, with the strongest predictive neural signal found immediately before the end of the memory delay period (reanalysis of data from Erlich et al., 2011). If disrupting a weak neural code produces a weaker behavioral effect than disrupting a strong neural code, then we would expected the smallest behavioral effect from FOF inactivation during the sensory cue, and the largest effect from inactivation immediately before the end of the memory delay period. We tested this prediction using short inactivation windows at different timepoints, with one window per trial in a randomly chosen 25% of trials (Fig. 2C,D).

The data produced a pattern precisely opposite to the prediction. Inactivation windows of both 500 ms (Fig. 2C) and 250 ms (Fig. 2D) produced the largest effects when placed during the sensory cue (500 ms Cue period: 11.6%, n=7 rats, $p<0.01$; 250 ms Cue first half: 12.1%, n=3 rats, $p<0.01$), and the magnitude of the effect decreased monotonically thereafter (Fig. 2C,D; Tables S1,S2). Immediately before the end of the delay period, inactivation of the FOF produced the smallest behavioral effect (2.5%, n=7 rats, p=0.012, Fig. 2D, open arrow), even though at that time the neural signal is the strongest (Fig. 2A, open arrows), and eNpHR3.0-mediated inhibition produces robust silencing (Fig. S3). To assess whether the small bias magnitude during the memory delay could be a consequence of incompletely silencing some parts of the FOF, in a second group of rats we infected a larger area of frontal cortex and implanted two optical fibers, one 500 μm anterior and the other 500 μm posterior from our original coordinates (Fig. S5A,B). We estimate that we thus doubled the volume of cortical tissue that we were silencing. Despite this doubling, we observed no additional response bias (p=0.226, n=2 rats dual fiber, n=5 rats single fiber, Fig. S5C,D), indicating that our single-fiber experiments had already saturated the size of the effect that could be produced by unilateral silencing of the FOF. Inactivation at a later timepoint, during the subject's orienting response, had no discernible effect on side choice (−0.1 +/− 3.5%,

p=0.53, n=6 rats, Fig. 2C, solid arrow). Nor was there an effect on movement times (Fig. S4).

These data thus suggest that frontal cortical activity plays an important causal role during short-term memory encoding, a gradually diminishing role in maintenance during the delay period, –becoming only marginally important by the delay period's end–, and is not required for movement execution.

Subsequent to the formation of orienting motor plans (which, in our task, is not distinguished from making a Left vs. Right decision, and which could occur as soon as the first few hundred milliseconds of the sensory cue), a causal role in the maintenance of those plans in short-term memory could move to brain areas other than the FOF, possibly downstream in the motor pathway. One candidate for such a downstream region is the superior colliculus (SC) (Wurtz & Albano, 1980; Munoz & Wurtz, 1995b,a; Reep et al., 1987). Although this subcortical region has generally not been thought to subserve short-term memory maintenance, it shares many properties with the primate FEF and rodent FOF, including involvement in the control of orienting movements (Wurtz & Albano, 1980; Felsen & Mainen, 2008; Cavanaugh et al., 2012; Stubblefield et al., 2013), firing patterns during delayed orienting tasks that predict the upcoming orienting movement (Sommer & Wurtz, 2004; Felsen & Mainen, 2012), greater impairment of memory-guided orienting than sensory-guided orienting after pharmacological inactivation (Hikosaka & Wurtz, 1985), and involvement in other higher functions, such as attention and target selection (Port &Wurtz, 2009; Nummela & Krauzlis, 2010; Zenon & Krauzlis, 2012). Anatomically, the SC receives direct projections from the FEF in primates (Komatsu & Suzuki, 1985; Stanton et al., 1988) and from the FOF in rats (Reep et al., 1987), and in turn projects back to frontal premotor cortex via the mediodorsal nucleus of the thalamus (Sommer & Wurtz, 2008). We injected the same AAV-CaMKIIα-eNpHR3.0-eYFP unilaterally into the SC (Fig. S2B), and repeated our experiments, but now inactivating the SC. The pattern of results found in the SC closely mirrored those found in the FOF (Fig. 3), with no statistically significant differences between the two brain regions (r=0.98, p=0.271 that SC differs from FOF, n=3–5 rats for SC depending on inactivation period, Fig. 3, Table S2). Instead of revealing differential roles for the FOF and SC, these data suggested that the two regions may play closely related or parallel roles during memory-guided orienting.

One explanation for the two sets of inactivation results, then, is that a causal role in maintenance of orienting motor plans gradually moves away from both the FOF and the SC during the delay period. The monotonically increasing neural encoding of Fig. 2A could be mostly an efference copy, or perhaps largely an epiphenomenon. Below we test an alternative explanation. We found that assigning equal, fully causal roles in memory maintenance to the FOF and SC within a model previously used to jointly describe decision-making and short-term memory (Wong & Wang, 2006; Machens et al., 2005) was sufficient to reproduce the entire set of results. Below we first describe the model, then identify two key untested predictions it makes, and finally confirm both predictions experimentally.

In the proposed model, which is a simple adaptation of (Hopfield, 1984; Wong & Wang, 2006; Machens et al., 2005), mutual inhibition between two opposite-preference neuronal

populations produces dynamics with two stable neural activity attractors, representing a “go Right” and a “go Left” decision. The activity in each of two nodes (Left and Right) is represented by a variable U; the output of each node is a sigmoidal function of U and is represented by the variable V (Hopfield, 1984). There is mutual inhibition between the two nodes (of strength I), self-excitation within each node (of strength M), and an external input Ex to each node (Fig. 4A). In addition, the activity of each node is perturbed by a white noise Wiener process dW, scaled by σ, which represents overall noise in the system. Subscripts L and R indicate the Left and Right nodes, respectively:

$$\begin{aligned}\tau\frac{\mathrm{d}U_L}{\mathrm{d}t} &= -U_L + M\cdot V_L - I\cdot V_R + Ex_L + \sigma \mathrm{d}W_L\\ \tau\frac{\mathrm{d}U_R}{\mathrm{d}t} &= -U_R + M\cdot V_R - I\cdot V_L + Ex_R + \sigma \mathrm{d}W_R\\ V_L(t) &= \tfrac{1}{2}\tanh U_L(t) + \tfrac{1}{2}\\ V_R(t) &= \tfrac{1}{2}\tanh U_R(t) + \tfrac{1}{2}\end{aligned} \quad \text{(Equation 1)}$$

We set τ = 100 ms (Wang, 2008).

The external input Ex is composed of a background input B plus an additional term ϕ that depends on the identity of the auditory cue. Thus, letting E_{cue} be a positive constant, and letting ϕ range from 0 for the strongest Right cue to 1 for the strongest Left cue, during the Cue period we have:

$$\begin{aligned}Ex_L(i) &= B + \phi E_{\mathrm{cue}}\\ Ex_R(i) &= B + (1-\phi)E_c\end{aligned} \quad \text{(Equation 2)}$$

And during the Memory period:

$$\begin{aligned}Ex_L(i) &= 0\\ Ex_R(i) &= 0\end{aligned} \quad \text{(Equation 3)}$$

The action of halorhodopsin eNpHR3.0 mediated inhibition is modeled by reducing the output V of one node by the fraction:

$$h = \left(1 - \frac{1}{n}\right) \quad \text{(Equation 4)}$$

where n is a free parameter.

At the start of each trial, the state of the system begins on the unstable manifold that divides the two basins of attraction. As time evolves, the sensory stimulus biases which of the two attractors the state of the system is likely to move towards. Eventually the network settles into one of the attractors, corresponding to a categorical decision, which persists as a memory because attractors are stable points of the dynamics (Fig. 4B) (Wong & Wang, 2006). The system is most sensitive to perturbations during the formation of the memory

during cue presentation (equivalent here to the formation of the Left vs. Right decision (Wang, 2002)), where the energy landscape is flattest (Fig. 4C).

In our adaptation, each of the two nodes comprise neurons from both the FOF and the SC (Fig. 4A). We found that perturbing the model network by silencing approximately one third ($n = 3.03$) of either the “go Right” or the “go Left” group during the sensory cue presentation significantly affects the Left vs. Right outcome (14.4% ipsilateral choice bias). However, after the system has been fully captured by one of the attractors, the same magnitude of perturbation, applied now during the short-term memory delay period, produces a significantly smaller bias (5.0%, Fig. 4C,E). This simple model provides a very good quantitative match to the data (Fig. 4D,E), for both FOF and SC, with a monotonic decrease in the magnitude of the behavioral effect (Fig. 4E), even while the strength of the neural encoding in the model shows a monotonic increase through each trial (increasing separation, as time unfolds, between red and blue trajectories in Fig. 4B). Put another way, point attractors are by definition stable points, and the further the system falls into one, the harder it is to perturb. Good fits to the data can be achieved if the FOF and SC each comprise a fraction of the total ranging from $n = 2.15$ to 5.10 (Fig. S6, 95% confidence interval), demonstrating that the features of interest are robust properties of this distributed dynamical attractor model, and indicating that the model is compatible with there being other brain regions (such as posterior parietal (Harvey et al., 2012) or prefrontal (Chafee & Goldman-Rakic, 1998) cortices) that, like the FOF and SC, may be contributors to the circuit.

TESTING MODEL PREDICTIONS

The dynamic attractor network makes four clear predictions. First, as the system evolves, the information encoded in the network will increase (Fig. 2A and Fig. 4B). Second, concomitant with the increase in encoding, the ability to perturb the system will decrease (Fig. 4D). The next two predictions are described and tested below.

With the FOF and SC each playing fully causal roles in memory maintenance, the model predicts that simultaneous FOF and SC delay period inactivation should produce a substantial bias, much bigger than the minimal effect seen after inactivating one region alone (which was 2.5 % for FOF during last 250 ms in Fig. 2). More specifically, the effect of silencing both regions should be greater than the sum of silencing each region alone (supralinear summation). In contrast, if the FOF and SC have both mostly lost their causal role in memory maintenance by the end of the delay period, we would expect little effect from inactivating both. Other alternatives, even when positing causal roles for FOF and SC, can predict that inactivating both FOF and SC would have an effect no greater than inactivating one of them alone. For example, if the FOF and the SC were both single-region bottlenecks within a feedforward linear chain, inactivating both would break the chain just as much as inactivating one alone, leading to no summation. In other words, supralinear summation is not an inevitable prediction of both regions playing causal roles. By fitting the model to the single-region inactivation data, and then doubling the silencing fraction ($1-h$) (see equation 4), we can use the model to quantitatively predict how much supralinear summation should result from simultaneous FOF and SC inactivation.

In a new group of rats, we unilaterally infected both the FOF and SC (in the same hemisphere, either both left or both right) with the same AAV-eNpHR3.0 as used previously, and implanted a chemically sharpened optical fiber in each of the two regions. Confirming our previous results (Figs. 2 and 3), silencing each region independently during a 1 second window overlapping the cue and memory periods led to a substantial bias (FOF: 18.5%, $p<0.01$, n=3 rats; SC: 20.9%, $p<0.01$, n=4 rats, Table S2), while inactivation only during the delay period produced a smaller bias (FOF: grey bars in Fig. 5A, 6.47%, $p<0.01$, n=3 rats; SC: grey bars in Fig. 5B, 6.46%, $p<0.01$, n=4 rats; here we used a 400 ms inactivation window beginning 100 ms into the delay period). Simultaneous inactivation of both the FOF and SC during the same memory delay period window yielded a statistically significant supralinear increase in the subject's ipsilateral response bias (FOF & SC: grey bar in Fig. 5C, 22.29%, $p<0.01$ dual compared to single site inactivation, $p<0.01$ dual compared to the sum of FOF only and SC only, n=3 rats). We replicated this supralinear summation result in a second group of rats, this time inactivating during the final 250 ms of the memory delay period, silencing either the FOF, SC, or both. Again consistent with our previous results, silencing either region alone resulted in a significant but very small ipsilateral response bias (FOF: grey bars Fig. 5D, 3.5%, $p<0.01$, n=4 rats; SC: grey bars in Fig. 5E, 3.1%, $p=0.017$, n=4 rats). As with the 400-ms inactivation window, silencing both regions simultaneously during the final 250-ms window resulted in a substantial, supralinear increase in bias (FOF & SC: grey bar in Fig. 5F, 9.8%, $p<0.01$ dual compared to single site inactivation, $p=0.018$ dual compared to the sum of FOF only and SC only, n=4 rats).

We fit the model to the FOF inactivation and control data (Fig. 4D,E Fig. 5A,D), leaving out the SC and the dual-region inactivation data. The resulting fits are shown in green in Fig. 5A,D. Following our previous findings (Figs. 2, 3), we predicted that silencing the SC alone would have the same effect as silencing the FOF alone, which indeed was again observed in the data and was consequently well accounted for by the model (Fig. 5B,E, magenta bars). We then used the model to predict the bias that would result from doubling the number of inactivated neurons in the mutual inhibition circuit. The resulting predictions are shown in Fig. 5C,F in magenta, and are very close to the experimental data, although the bias found after 400-ms dual-region is slightly larger than the model prediction. Overall, the model's predictions of results from dual FOF and SC inactivations were consistent with the data.

We now turn to the final prediction of the model. During the cue period, easy trials are predicted to be less perturbable than hard trials because they involve a greater difference in the excitatory input to the network (ϕ is far away from 0.5; see Equation 2), and thus move more rapidly away from the unstable manifold, settling more quickly into the stable attractor. But subsequently, as the network settles into an attractor, information about how it got there is lost, and therefore whether the cue stimulus was easy or hard no longer matters to perturbability. We analyzed the difference in bias between hard and easy trials, both for inactivations in the FOF (Fig. 6A–D) and the SC (Fig. 6E,F). Consistent with the model's prediction for inactivation periods that overlapped the cue presentation, we observed larger biases on hard trials compared with easy trials, and the magnitude of the difference was close to that predicted by the model (colored triangles above each histogram in Fig. 6 indicate the histogram mean; black triangles indicate the model's prediction; we emphasize again that the difference between hard and easy trials was not used to fit the model). We then

performed the same analysis for inactivations during the delay period. We focused on the dual-region inactivations, since these produced a substantial bias that would facilitate observing a difference, if present, between hard and easy trials. In the model, the dual-region delay period inactivations are predicted to produce large average biases comparable to single-region cue period inactivations (compare Fig. 5C, dual-region inactivation of last 400 ms of the delay period, to Fig. 3, single-region inactivation of the cue period) but are still predicted to result in little difference in bias magnitude for hard versus easy trials. This prediction was confirmed by the data (Fig. 6G,H).

Across the conditions analyzed, and over a 9-fold range in predicted hard minus easy bias difference (ranging from only 1.0%, for dual-region inactivation of the last 250 ms to 9.0% for single-region 1 s inactivation of both cue and delay period), we saw a strong correlation between the bias difference predicted by the model and that observed in the data (Fig. 6I, $r=0.96$, $p<0.01$). Although there appeared to be a slight tendency for the model to underpredict the bias difference (data points in Fig. 6I tend to be slightly to the right of the x=y diagonal), the model's prediction lay within the data error bars for all 8 conditions, indicating that, overall, the prediction was quantitatively very good. We saw no correlation between each condition's average bias and its hard minus easy trial bias difference (Fig. 6J, $r=0.10$, $p=0.80$) confirming that the agreement in Fig. 6 between the model prediction and rat data is not merely a consequence of a good model fit to each condition's average bias.

DISCUSSION

Short-term memory has been thought to be primarily subserved by cortical structures. The maintenance of short-term memory for orienting, in particular, has been thought to be subserved by structures including the FEF in primates (Bruce et al., 1985) and its suggested homologue, the FOF, in rodents (Erlich et al., 2011). eNpHR3.0-mediated inhibition, with its associated high temporal resolution, provided an opportunity to probe the dynamics of the system and to causally test this hypothesis. Together with computational modeling and optogenetic inactivation of the SC, the results suggested instead that short-term memory is subserved by a dynamical attractor mechanism that bridges cortical and subcortical regions, with both the FOF and the SC playing equal roles in the dynamics of memory formation and maintenance. Early in the decision process, while the system is still moving towards one of the basins of attraction and firing rates only weakly predict the animal's later motion, the network is susceptible to perturbations that silence any individual region involved. However, later in a trial during short-term memory maintenance or readout, periods by which firing rates more strongly predict the upcoming motion and the system has already settled into a stable attractor, the network is more resistant to perturbations. We fit a simple adaptation of a two-node mutual inhibition dynamical system model of decision-making and short-term memory (Hopfield, 1984; Machens et al., 2005; Wong & Wang, 2006), and tested two of its predictions. First, we demonstrated that simultaneous inactivation of the FOF and SC during the short-term memory delay period led to a predicted supralinear increase in the response bias compared to inactivating either region alone, with the magnitude of the effect close to that predicted by the model. Second, consistent with the model's dynamics, we demonstrated that brief inactivations overlapping the cue presentation led to a greater bias on more difficult trials, but once the network had settled into a stable attractor during the

memory delay period, the effect of trial difficulty on inactivation-induced bias was eliminated.

The attractor model we used here is very simple, with each of the two possible memories represented by a single, static state of neural activities. Nevertheless, we speculate that the key perturbation results and concepts (relatively high perturbability on the unstable manifold dividing the basins of attraction, lower perturbability deeper in the basins) do not depend on the particulars of the attractors at the bottom of the basins, and could thus remain similar even if the attractors were more complex, for example if they were limit cycles representing choice-dependent sequences of neural activity (Harvey et al., 2012).

One of the principal simplifications in our model is that it treats each side of the brain as if it held only contralateral-preferring neurons. However, in rodents, each side of the brain contains an almost equal mixture of neurons with contra- and ipsi- firing rate preferences (Fig. 2A, 56 versus 44 % contra versus ipsi preference in the FOF, Erlich et al., 2011; see also Felsen & Mainen, 2008; Li et al., 2015). Unilateral inactivations nevertheless cause a strong ipsilateral bias (Felsen & Mainen, 2008; Gage et al., 2010; Erlich et al., 2011; Guo et al., 2014; Li et al., 2015). Recent data from the mouse ALM during a directional licking task show that the contralateral bias is much greater in pyramidal tract-projecting neurons than in intratelencephalic-projecting neurons, and that this can explain the inactivation bias (Li et al., 2015). Future elaborations of our model should include mixed ipsi- and contra-preferring neurons, and distinguish between neuronal classes with different projection patterns.

The timing of the effects we found contrast with those recently described by Svoboda and colleagues (Guo et al., 2014; Li et al., 2015). Using a memory-guided directional licking task and optogenetic inactivation, they demonstrated that mouse anterior lateral motor cortex (ALM) is required during the memory delay period but not during the stimulus cue period. This temporal trend is opposite to the one we found for the FOF and SC in our orienting task. Directional licking may engage circuits and mechanisms different to those engaged by eye and body orientation. For example, it is not yet known whether directional licking is distributed across cortex and the SC the way orienting is (Fig. 3), nor is it yet known whether ALM is preferentially required for memory-guided responses compared to sensory-guided responses, as FOF and SC are for orienting responses (Fig. 1). Furthermore, ALM might be more directly tied to motor control (Komiyama et al., 2010) than the FOF and SC, which are not required during the orienting motor acts themselves (Figs. 2, 3, and S4). Finally, if the subjects form their decision during the stimulus period, a lack of requirement for the ALM during that period would suggest that, unlike the FOF and the SC, the ALM is not involved in the decision itself. Behavioral tasks that allow a more precise assessment of the moment in which the decision commitment occurs will be instrumental in resolving some of these questions.

One such task is the “Poisson Clicks” task, for which behavioral evidence indicates that rats and humans gradually accumulate sensory evidence (a process that requires using short-term memory), over many hundreds of milliseconds, and form their decision commitment at the end of the evidence accumulation period (Brunton et al., 2013). Thus, in contrast to the Memory-Guided Orienting task used here, in which the decision can be formed early during

the nose fixation period, in the Poisson Clicks task the decision is not formed until the end of sensory evidence delivery, which coincided with the end of the nose fixation period. Hanks and Kopec et al., 2015 found a corresponding contrast in the timing of effects of unilateral FOF silencing. As described above, in our Memory-Guided Orienting task, FOF silencing has its greatest effect early in the nose fixation period (Fig. 2). In the Poisson Clicks task, the same perturbation affects the rat's choice only at the end of the nose fixation period (see Fig. 4 in Hanks and Kopec, et al. 2015). Thus, in both tasks unilateral optogenetic inhibition of the FOF is most able to perturb the subjects choice when it occurs at the time of decision commitment, (presumably) early during the auditory cue in the Memory-Guided Orienting task, and at the end of the accumulation of evidence period in the Poisson Clicks task.

Unilateral inhibition of either the FOF or SC is suggested by our model to be equivalent to silencing approximately one third of the output of either node, with 95% confidence error bars encompassing a range of one half to one fifth (Fig. S6A). Future studies will be necessary to identify the location(s) of the remainder of this potentially distributed network, with previous work suggesting that posterior parietal (Harvey et al., 2012) and/or prefrontal (Chafee & Goldman-Rakic, 1998; Wimmer et al., 2014) cortices may play a role. It has also been demonstrated that rats generate small postural changes during the memory delay period (Erlich et al., 2011), suggesting that other sensory or motor regions could participate in maintaining the memory of the upcoming orienting movement.

Our results with single-region versus dual-region FOF and SC inactivation at the end of the memory delay period (Fig. 5) are reminiscent of pioneering work in the visuomotor system of primates. Schiller and colleagues demonstrated that after recovering from permanent lesions of either the FEF or SC alone, there was a modest to no behavioral impairment in sensory-guided saccades, but permanently lesioning both structures led to persistent deficits (Schiller et al., 1979). Later work, using memory-guided saccades and acute (tens-of-minutes to hours-long) pharmacological inactivations, demonstrated that single-region silencing of either the SC alone (Hikosaka & Wurtz, 1985) or the FEF alone (Sommer & Tehovnik, 1997; Dias & Segraves, 1999), or in rats, the FOF alone (Erlich et al., 2011) was sufficient to substantially impair memory-guided saccades or orienting, respectively, even while sensory-guided saccades or orienting were left relatively intact. Now, using optogenetic inactivation with tens of milliseconds resolution and again focusing on memory-guided orienting, we return to a distributed circuit concept, by showing that towards the end of the memory maintenance phase, substantial behavioral effects require silencing of both the FOF and SC. Our model assumes that the FOF and SC play equivalent roles in this network. Future experiments may elucidate functional differences between the two regions.

Here we used a simple binary choice memory-guided orienting task, appropriate for rodents. Unraveling the full spatial topology of SC and frontal cortical interactions during short-term memory is likely to require experiments in the primate visual system (Cavanaugh et al., 2012), which can take advantage of fine-scale topographic maps in both the SC and FEF. Disambiguating short-term memory for sensory information versus motor decisions would require additional modification to the task, as has been demonstrated for primates (Gold & Shadlen, 2003; Horwitz et al., 2004) and rats (Duan et al., 2015). Overall, the work

presented here is an example of how dynamical network models, combined with high temporal resolution optogenetic perturbations, can provide insight into the sub-second dynamics of cognitive function.

METHODS

Rat Housing and General Training

Male Long-Evans rats were pair housed in Technoplast cages on a reversed dark/light cycle. Rats had free access to food but had restricted water access limited to 1 hour per day (starting 30 minutes following the end of training) and what they could earn during training, as approved by Princeton University IACUC. Training environment was as previously described (Erlich et al., 2011).

Training the Memory-Guided Orienting Task

Training was fully automated utilizing the Bcontrol System (http://brodylab.princeton.edu/bcontrol) and similar to that previously described (Erlich et al., 2011) except for the following differences: the fixation period was fixed at 1 s, the cue and memory delay periods where equivalent in duration at 500 ms, and only memory-guided trials were presented to the rats. Training required on average 16 weeks to achieve full performance.

If a rat failed to maintain fixation a 1s white noise sound was played to signal the violation, and the next trial was then initiated. Such fixation violation trials (~24%) were excluded from all analysis. Reaction time, defined as the time from the “go” cue (center port light off), to the time of pulling out from the center nose port, was on average 170 ms +/− 263 ms (standard deviation). Fixation violations result in a negative reaction time and are therefore excluded from this measure. Movement time is defined as the duration from when the rat withdraws from the center port until it enters either of the side nose ports (Figure S4).

Training the Sensory-Guided Orienting Task

The sensory-guided orienting task was similar to the memory-guided orienting task except that the auditory cue began at the end of fixation rather that at the beginning. For the stereo balance version of this task (Table S1 and S2) the 4 auditory cues were trains of clicks presented at 50 Hz with the following left/right stereo balances: 100–0% and 67–33% paired with reward on the left, and 33–67% and 0–100% paired with reward on the right. Training required on average 12 weeks to reach full performance for either version.

Fiber Optic Chemical Sharpening

Construction was as previously described (Hanks and Kopec et al., 2015). Briefly, a standard off the shelf 50/125 μm FC-FC duplex fiber cable (FiberCables.com) was stripped of all but the innermost plastic jacket (clear). 2 mm was submerged in concentrated hydrofluoric acid (48%) topped with mineral oil for 85 minutes, then water for 5 minutes (submerging 5 mm), and acetone for 2 minutes. The plastic jacket was then cut with a razor and removed with tweezers to reveal a 1 mm sharp etched fiber tip. Enough plastic was removed, to ensure that only the glass fiber optic was inserted into the brain.

Virus Injection and Fiber Implantation

All surgical procedures were as previously described (Hanks and Kopec et al., 2015) and performed in accordance with and approved by Princeton University IACUC. Coordinates relative to bregma (FOF: 2 mm anterior, 1.3 mm lateral; SC: 6.8 mm posterior, 2 mm lateral). 2 μl of AAV virus (AAV5-CaMKIIα-eYFP-eNpHR3.0 or AAV2/9-hEF1a-eYFP-ChR2) was lightly died with fast green powder and front loaded into a glass pipette mounted to a Nanoject (Drummond Scientific) prefilled with mineral oil. The pipette tip was manually cut to ~30 μm diameter. At the targeted coordinates, two injections of 9.2 nl were made every 100 μm in depth starting 200 μm below brain surface for FOF and 3.5mm below brain surface for SC for 1.5 mm. Four additional injection tracts were completed, one 500 μm anterior, posterior, medial, and lateral from the central tract. Each injection was followed by a 10 s pause, with 1 minute following the final injection in a tract before the pipette was removed. A total of 1.5 μl of virus was injected over a 30 minute period. The fiber tip was lowered down the central injection tract to a depth of 1 mm for FOF and 4.2 mm for SC. For LED based implants, a 3 mm diameter blue 473 nm LED (Superbrightleds.com) was positioned 500 μm above the brain surface within the craneotomy prefilled with wet kwik-sil (World Precision Instruments). After 1 week recovery the rat was returned to water restriction and resumed training. eNpHR3.0 expression was allowed to develop for 6–8 weeks before behavioral testing began.

eNpHR3.0 Activation

The animal's implant was connected to a 1m patch cable connected to a single fiber rotary joint (Princetel) mounted in the ceiling of the sound attenuation chamber. This was connected to a 200 mW 532 nm laser (OEM Laser Systems) operating at 25 mW which was triggered with a 5V TTL controlled by BControl. Laser illumination occurred on a random 25% of trials.

Computing Bias and Confidence Intervals

The fraction of trials where the animal responded ipsilateral to their implant location was first computed for each of the 4 trial types. Each trial type was selected with equal probability, however random noise resulted in an unequal number of trials for each of the four trial types. The mean of these four values was then taken as the observed "go ipsi" rate (this corrected for any small difference in the number of trials of each type collected). The bias was taken as the difference between the observed "go ipsi" rate for inactivation trials and control trials. A positive value represented an increase in ipsilateral responses on laser illumination trials.

The observed "go ipsi" rate is a noisy estimate of the underlying true"go ipsi" rate (which would be the observed rate given an infinite number of trials). The probability of any given rate being the true underlying "go ipsi" rate is given by:

$$p(r|o,n)=\left(\frac{(n+1)!}{o!\cdot(n-o)!}\right)\cdot r^{o}\cdot(1-r)^{(n-o)} \quad \text{(Equation 5)}$$

where n is the total number of trials, o is the observed number of ipsi responses, and r is the rate to be tested. The distribution of possible true "go ipsi" rates was computed independently for laser and no laser trials. The distribution of possible true underlying biases was taken as the difference of these two distributions, that for the true underlying "go ipsi" rate for laser and no laser trials. The 95% confidence interval on the bias was then the range that encompassed 95% of the distribution of possible biases centered on the observed bias. The p-value for an ipsilateral bias is taken as the fraction of the distribution that is less than zero, i.e. what is the probability of the data being explained by a contralateral bias.

To compute the confidence intervals where only the sessions or rats are considered true independent observations, not the individual trials as explained above, we first computed the distribution of possible bias values for each group of trials (by rat or session) as described above. For example, if 10 training sessions contributed to a particular experiment, 10 distributions of possible biases were computed, one for each session, independently. We then took the mean of N samples from these distributions with replacement (where N is the number of groups). Given our example, we would combine these 10 distributions into one and take the mean of 10 samples drawn with replacement. Repeating this mean of 10 samples would give a distribution of possible means. Confidence intervals were then taken on this distribution of means and the p-values were computes as described above. Table S2 contains the N and p-value for each experiment given the assumption that either trials, sessions, or rats are true independent observations.

Histology

The rat was fully anesthetized with 400 µl ketamine (100 mg/ml) and 200 µl xylazine (100 mg/ml) IP, followed by transcardial perfusion of 100 ml saline (0.9% NaCl, 0.3× PBS pH 7.0, 50 µl heparin 10,000 USP units/ml), and finally transcardial perfusion of 250 ml 10% formalin neutral buffered solution (Sigma HT501128). The brain was removed and post fixed in 10% formalin solution for a minimum of 2 days. 100 µm sections were prepared on a Leica VT1200S vibratome, mounted on Superfrost Plus glass slides (Fisher) with Fluoromount-G (SouthernBiotech) mounting solution and glass cover slips. Images were acquired on a Nikon Eclipse Ti fluorescence microscope under 4× magnification. Brain region targeting confirmed with AtlasFitter software (http://brodylab.princeton.edu/wiki/index.php/AtlasFitter) (Kopec et al., 2011).

Model

Equations governing the model are discussed in the main text. We used ϕ values of 0, 0.333, 0.667, and 1 to represent the four cue stimuli.

The model contains 6 free parameters: M, I, σ, B, E_{cue}, and n. These were fit using a gradient ascent (MatLab's fmincon function) to maximize the likelihood of the model producing the biases measured from data collected from rats with implants in FOF. The model was fit to the following 12 data points: fraction "go ipsi" on each of the 4 trial types for control trials; and eNpHR3.0 induced bias for 1 s (fixation period), 500 ms (cue period and memory period), 400 ms (memory period), and 250 ms (cue early, cue late, memory early, and memory late periods) laser durations. The model from Figure 4D was fit to the

overall net bias for all four trial types taken together (easy ipsi, hard ipsi, easy contra, and hard contra) which was calculated from the aggregate rat data. Here τ = 100ms was used though equivalent fits were obtained with τ = 1s. The best fit parameter values used in Fig. 4, 5, and 6 were M = 41.27, I = 28.91, σ = 1.42, B = 3.95, E_{cue} = 10.26, and n = 3.03 with 95% confidence range M = 14.25–68.30, I = 11.17–46.65, σ = 1.10–1.74, B = 0–14.65, E_{cue} = 6.78–13.77, and n = 2.15–5.10. Confidence intervals were determined by first computing the likelihood for a series of parameter values spaced in a grid around the best fit value (6 dimensions, one for each parameter). The likelihood profile for a single parameter was then calculated by summing the likelihoods across each of the remaining dimensions. The resulting likelihood profiles for each parameter were individually fit with a Gaussian function and the range encompassing 95% of the likelihood centered on the mean was calculated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. Teran and L.K. Osorio for all rat handling, K. Miller, T. Hanks, A. Akrami, I. Witten, and M. Yartsev for comments on the manuscript, A. Kepecs and B. Burbach for help designing the fiber optic implant, A. Zador for supplying the ChR2 virus, M. Taylor for help with microscopy, and K. Zalocusky for help with the AAV eNpHR3.0 virus.

References

Bruce CJ, Goldberg ME. Primate frontal eye fields. i. single neurons discharging before saccades. J Neurophysiol. 1985; 53:603–35. [PubMed: 3981231]

Bruce CJ, Goldberg ME, Bushnell MC, Stanton GB. Primate frontal eye fields. ii. physiological and anatomical correlates of electrically evoked eye movements. J Neurophysiol. 1985; 54:714–34. [PubMed: 4045546]

Cavanaugh J, Monosov IE, McAlonan K, Berman R, Smith MK, Cao V, Wang KH, Boyden ES, Wurtz RH. Optogenetic inactivation modifies monkey visuomotor behavior. Neuron. 2012; 76:901–7. [PubMed: 23217739]

Chafee MV, Goldman-Rakic PS. Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. J Neurophysiol. 1998; 79:2919–40. [PubMed: 9636098]

Dias EC, Segraves MA. Muscimol-induced inactivation of monkey frontal eye field: effects on visually and memory-guided saccades. J Neurophysiol. 1999; 81:2191–214. [PubMed: 10322059]

DiCarlo JJ, Maunsell JHR. Using neuronal latency to determine sensory-motor processing pathways in reaction time tasks. J Neurophysiol. 2005; 93:2974–86. [PubMed: 15548629]

Duan CA, Erlich JC, Brody CD. Requirement of prefrontal and midbrain regions for rapid executive control of behavior in the rat. Neuron. 2015; 86:1491–503. [PubMed: 26087166]

Erlich JC, Bialek M, Brody CD. A cortical substrate for memory-guided orienting in the rat. Neuron. 2011; 72:330–43. [PubMed: 22017991]

Felsen G, Mainen ZF. Neural substrates of sensory-guided locomotor decisions in the rat superior colliculus. Neuron. 2008; 60:137–48. [PubMed: 18940594]

Felsen G, Mainen ZF. Midbrain contributions to sensorimotor decision making. J Neurophysiol. 2012; 108:135–47. [PubMed: 22496524]

Funahashi S, Bruce CJ, Goldman-Rakic PS. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol. 1989; 61:331–49. [PubMed: 2918358]

Fuster, J. The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe. 3rd. Philadelphia: Lippincott-Raven; 1997.

Gage GJ, Stoetzner CR, Wiltschko AB, Berke JD. Selective activation of striatal fast-spiking interneurons during choice execution. Neuron. 2010; 67:466–79. [PubMed: 20696383]

Gold J, Shadlen M. The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. Journal of Neuroscience. 2003; 23:632–51. [PubMed: 12533623]

Goldman-Rakic PS. Regional and cellular fractionation of working memory. Proc Natl Acad Sci U S A. 1996; 93:13473–80. [PubMed: 8942959]

Gradinaru V, Zhang F, Ramakrishnan C, Mattis J, Prakash R, Diester I, Goshen I, Thompson KR, Deisseroth K. Molecular and cellular approaches for diversifying and extending optogenetics. Cell. 2010; 141:154–65. [PubMed: 20303157]

Guo ZV, Li N, Huber D, Ophir E, Gutnisky D, Ting JT, Feng G, Svoboda K. Flow of cortical activity underlying a tactile decision in mice. Neuron. 2014; 81:179–94. [PubMed: 24361077]

Hanks T, Kopec CD, Brunton B, Duan C, Erlich J, Brody C. Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature. 2015; 520:220–3. [PubMed: 25600270]

Harvey CD, Coen P, Tank DW. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature. 2012; 484:62–8. [PubMed: 22419153]

Hikosaka O, Wurtz RH. Modification of saccadic eye movements by gaba-related substances. i. effect of muscimol and bicuculline in monkey superior colliculus. J Neurophysiol. 1985; 53:266–91. [PubMed: 2983037]

Hopfield JJ. Neurons with graded response have collective computational properties like those of two-state neurons. Proc Natl Acad Sci U S A. 1984; 81:3088–92. [PubMed: 6587342]

Horwitz G, Batista A, Newsome W. Representation of an abstract perceptual decision in macaque superior colliculus. Journal of Neurophysiology. 2004; 91:2281–96. [PubMed: 14711971]

Komatsu H, Suzuki H. Projections from the functional subdivisions of the frontal eye field to the superior colliculus in the monkey. Brain Res. 1985; 327:324–7. [PubMed: 2985177]

Komiyama T, Sato TR, O'Connor DH, Zhang Y, Huber D, Hooks BM, Gabitto M, Svoboda K. Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. Nature. 2010; 464:1182–6. [PubMed: 20376005]

Kopec CD, Bowers AC, Pai S, Brody CD. Semi-automated atlas-based analysis of brain histological sections. J Neurosci Methods. 2011; 196:12–9. [PubMed: 21194546]

Lambelet P, Sayah A, Pfeffer M, Philipona C, Marquis-Weible F. Chemically etched fiber tips for near-field optical microscopy: a process for smoother tips. Appl Opt. 1998; 37:7289–92. [PubMed: 18301560]

Li N, Chen TW, Guo ZV, Gerfen CR, Svoboda K. A motor cortex circuit for motor planning and movement. Nature. 2015; 519:51–6. [PubMed: 25731172]

Machens CK, Romo R, Brody CD. Flexible control of mutual inhibition: a neural model of two-interval discrimination. Science. 2005; 307:1121–4. [PubMed: 15718474]

Munoz DP, Wurtz RH. Saccade-related activity in monkey superior colliculus. i. characteristics of burst and buildup cells. J Neurophysiol. 1995a; 73:2313–33. [PubMed: 7666141]

Munoz DP, Wurtz RH. Saccade-related activity in monkey superior colliculus. ii. spread of activity during saccades. J Neurophysiol. 1995b; 73:2334–48. [PubMed: 7666142]

Nummela SU, Krauzlis RJ. Inactivation of primate superior colliculus biases target choice for smooth pursuit, saccades, and button press responses. J Neurophysiol. 2010; 104:1538–48. [PubMed: 20660420]

Port NL, Wurtz RH. Target selection and saccade generation in monkey superior colliculus. Exp Brain Res. 2009; 192:465–77. [PubMed: 19030853]

Reep RL, Corwin JV, Hashimoto A, Watson RT. Efferent connections of the rostral portion of medial agranular cortex in rats. Brain Res Bull. 1987; 19:203–21. [PubMed: 2822206]

Riehle A, Requin J. The predictive value for performance speed of preparatory changes in neuronal activity of the monkey motor and premotor cortex. Behav Brain Res. 1993; 53:35–49. [PubMed: 8466666]

Robinson DA, Fuchs AF. Eye movements evoked by stimulation of frontal eye fields. J Neurophysiol. 1969; 32:637–48. [PubMed: 4980022]

Schall JD, Thompson KG. Neural selection and control of visually guided eye movements. Annu Rev Neurosci. 1999; 22:241–59. [PubMed: 10202539]

Schiller PH, True SD, Conway JL. Effects of frontal eye field and superior colliculus ablations on eye movements. Science. 1979; 206:590–2. [PubMed: 115091]

Sommer MA, Tehovnik EJ. Reversible inactivation of macaque frontal eye field. Exp Brain Res. 1997; 116:229–49. [PubMed: 9348123]

Sommer MA, Wurtz RH. What the brain stem tells the frontal cortex. i. oculomotor signals sent from superior colliculus to frontal eye field via mediodorsal thalamus. J Neurophysiol. 2004; 91:1381–402. [PubMed: 14573558]

Sommer MA, Wurtz RH. Brain circuits for the internal monitoring of movements. Annu Rev Neurosci. 2008; 31:317–38. [PubMed: 18558858]

Stanton GB, Goldberg ME, Bruce CJ. Frontal eye field efferents in the macaque monkey: Ii. topography of terminal fields in midbrain and pons. J Comp Neurol. 1988; 271:493–506. [PubMed: 2454971]

Stubblefield E, Costabile J, Felsen G. Optogenetic investigation of the role of the superior colliculus in orienting movements. Behavioural Brain Research. 2013

Tye KM, Prakash R, Kim S-YY, Fenno LE, Grosenick L, Zarabi H, Thompson KR, Gradinaru V, Ramakrishnan C, Deisseroth K. Amygdala circuitry mediating reversible and bidirectional control of anxiety. Nature. 2011; 471:358–62. [PubMed: 21389985]

Wang X-JJ. Probabilistic decision making by slow reverberation in cortical circuits. Neuron. 2002; 36:955–68. [PubMed: 12467598]

Wang X-JJ. Decision making in recurrent neuronal circuits. Neuron. 2008; 60:215–34. [PubMed: 18957215]

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

Wong K-FF, Wang X-JJ. A recurrent network mechanism of time integration in perceptual decisions. J Neurosci. 2006; 26:1314–28. [PubMed: 16436619]

Wurtz RH, Albano JE. Visual-motor function of the primate superior colliculus. Annu Rev Neurosci. 1980; 3:189–226. [PubMed: 6774653]

Zenon A, Krauzlis RJ. Attention deficits without cortical neuronal deficits. Nature. 2012; 489:434–7. [PubMed: 22972195]

HIGHLIGHTS

- Optogenetics probes precisely when FOF and SC are needed for memory-guided orienting
- Behavioral effect of silencing FOF or SC decreases monotonically during each trial
- Attractor model reconciles decreasing perturbability with increasing neural encoding
- Key attractor model predictions are confirmed

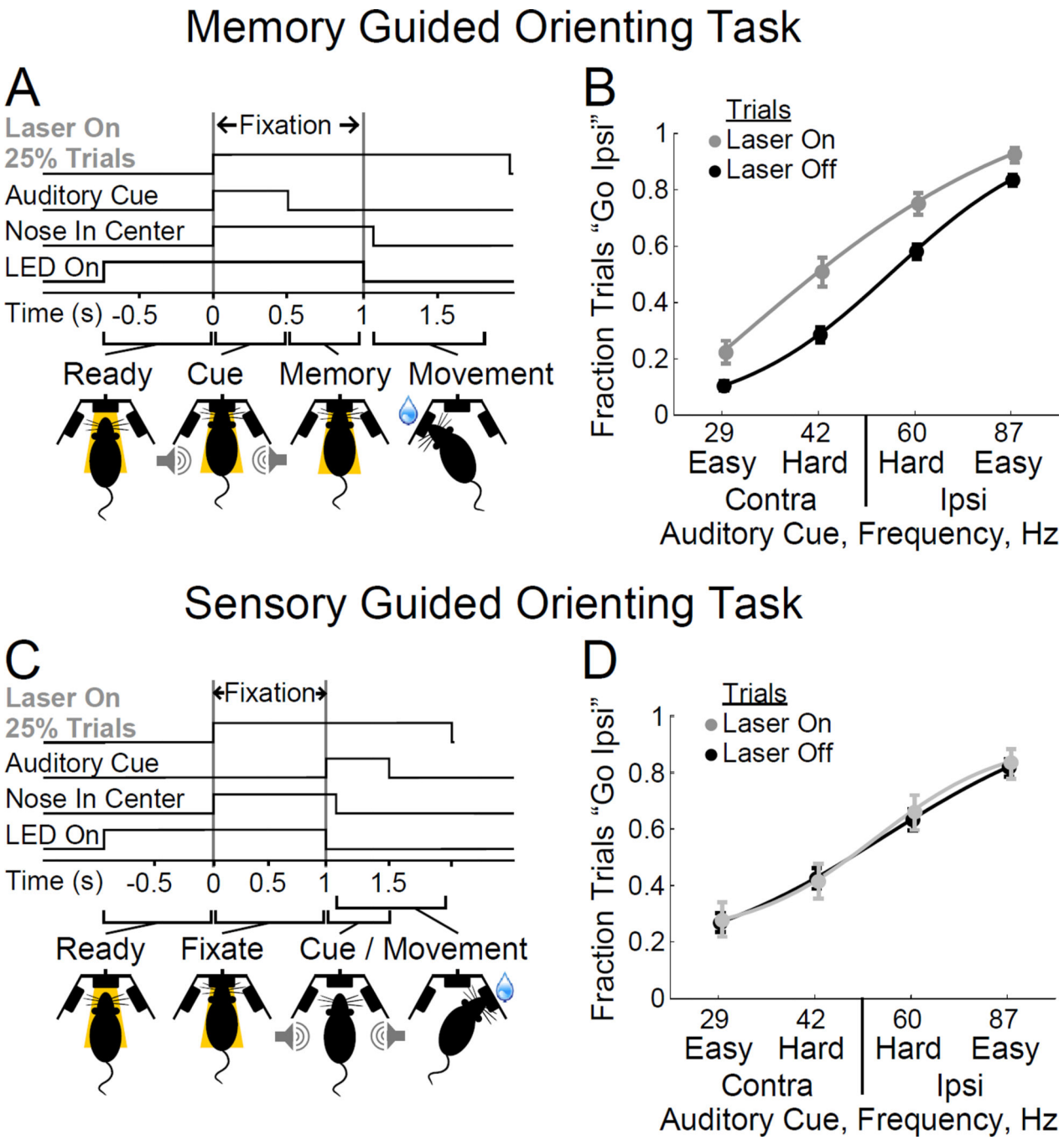

Figure 1.

Transient inactivation of the FOF impairs a memory-guided orienting task but has no effect on a sensory-guided orienting task. (A) Sequence of events in each trial of the memory-guided orienting task. Following the onset of an LED in the center nose port, trained rats placed their nose into the port, and were required to "fixate" their nose there until the LED was turned off (1 sec). During the first 500 ms of the fixation period, a regular train of auditory clicks was played. A frequency f > 50 clicks/sec indicated that a water reward was available at the right nose port, while f < 50 clicks/sec indicated that the reward was on the

left. A trial was easy(hard) if |f-50| was large(small). After a further silent 500 ms short-term memory delay period, the center LED was turned off as the “go” cue signaling the end of fixation. The rat was then rewarded if it made a nose poke into the correct side port (Erlich et al., 2011). On 25% of randomly chosen trials, a laser light in the eNpHR3.0-expressing side of the FOF was turned on (“Laser On”). (B) Performance as a function of click frequency for control trials (Laser Off, black) and for intermingled trials with eNpHR3.0-mediated inactivation of one side of the FOF (Laser ON, grey). “Ipsi” refers to the same side as the eNpHR3.0-expressing side of the brain. Cue frequencies are shown for subjects expressing eNpHR3.0 in the right FOF. For those expressing in the left, the order is reversed. n=4 rats, 14 sessions. (C) In a sensory-guided orienting task, the auditory cue is not presented until after the nose fixation period ends, and subjects can respond to the cue immediately, without a short-term memory delay period. The sensory cue, a 500 ms regular train of auditory clicks, is the same as for the Memory-Guided Orienting Task. (D) In the sensory-guided orienting task of panel (C), performance on control trials (Laser Off, black) was identical to FOF inactivation trials (Laser On, grey), regardless of trial difficulty or orienting direction. n=3 rats, 9 sessions. Error bars 95% confidence.

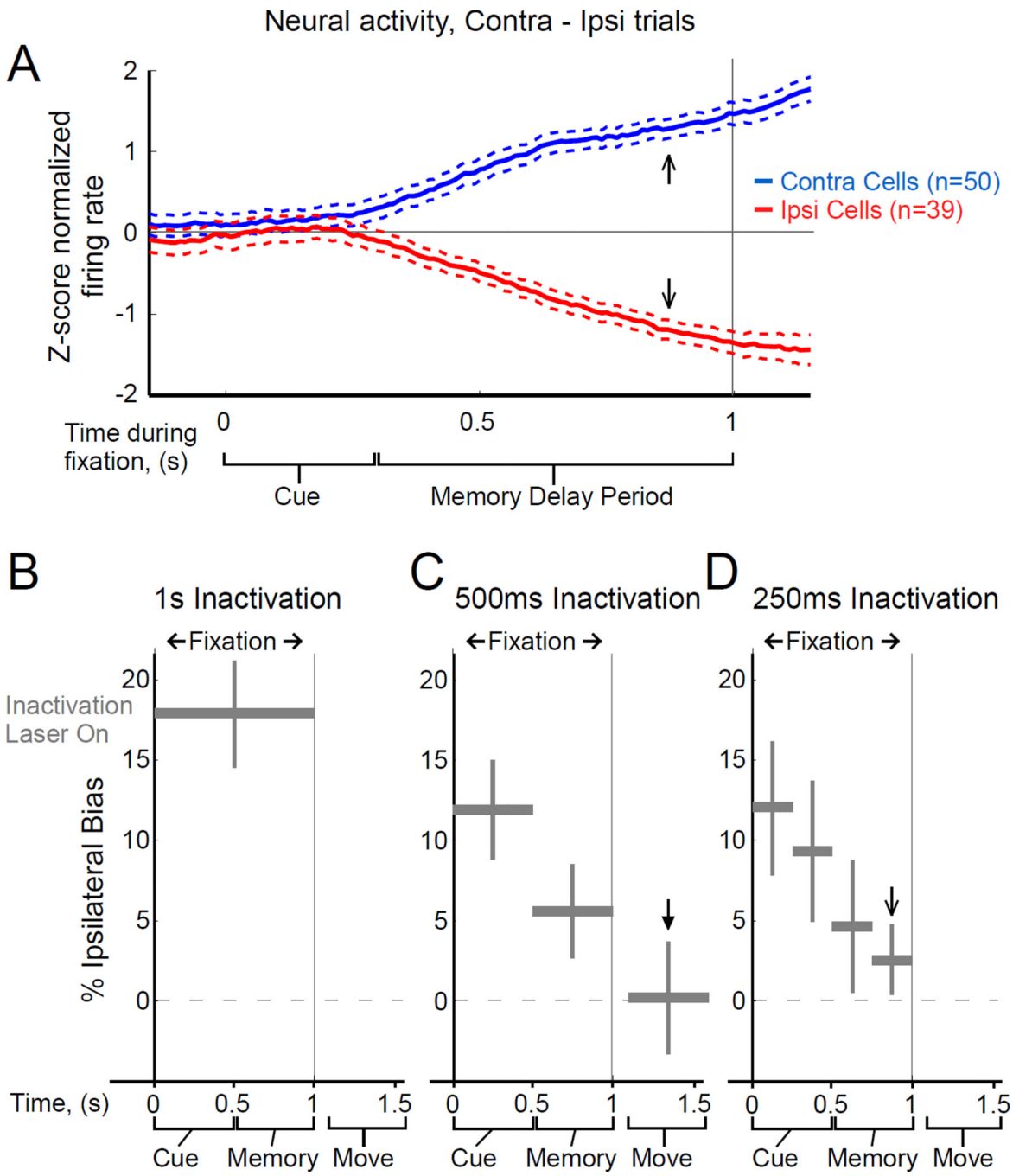

Figure 2.

FOF inactivation effects have a temporal profile opposite to FOF neural encoding. (A) Neural encoding of the direction of the upcoming motor response grows monotonically in FOF during cue presentation and the short-term memory delay period. The thick lines shows normalized firing rate in trials resulting in an orienting motion contralateral to the recorded neurons, minus normalized firing rate in trials resulting in an ipsilateral orienting motion (blue, average of n=50 contralateral-preferring neurons, red, 39 ipsi-preferring neurons). Thin dashed lines are +/− s.e.m. Data is reanalyzed from (Erlich et al., 2011) and is from

correct trials only. (B), (C), (D), Average bias towards the implanted FOF side caused by inactivation of different time windows. The vertical position of the grey bars indicates the measured bias, while the bar's horizontal extent indicates the time period over which the laser was turned on. Twenty-five percent of trials were chosen randomly to be inactivation trials; one inactivation time window was used in each of the inactivation trials. (B) Inactivation of the entire fixation period. n=7 rats, 33 sessions. (C) Inactivation of 500 ms-long windows. n=7 rats, 33 sessions. (D) Inactivation of 250 ms-long windows. n=3 rats, 29 sessions for the three time windows between 0 and 0.75 s, n=7 rats, 49 session for the final time window. Error bars indicate 95% confidence intervals centered on the mean for the combined data (See Methods for a description of how confidence intervals are computed, and Table S1 for individual rat biases for each experiment). Arrows in (A), (C), and (D) indicate time periods referred to in the text.

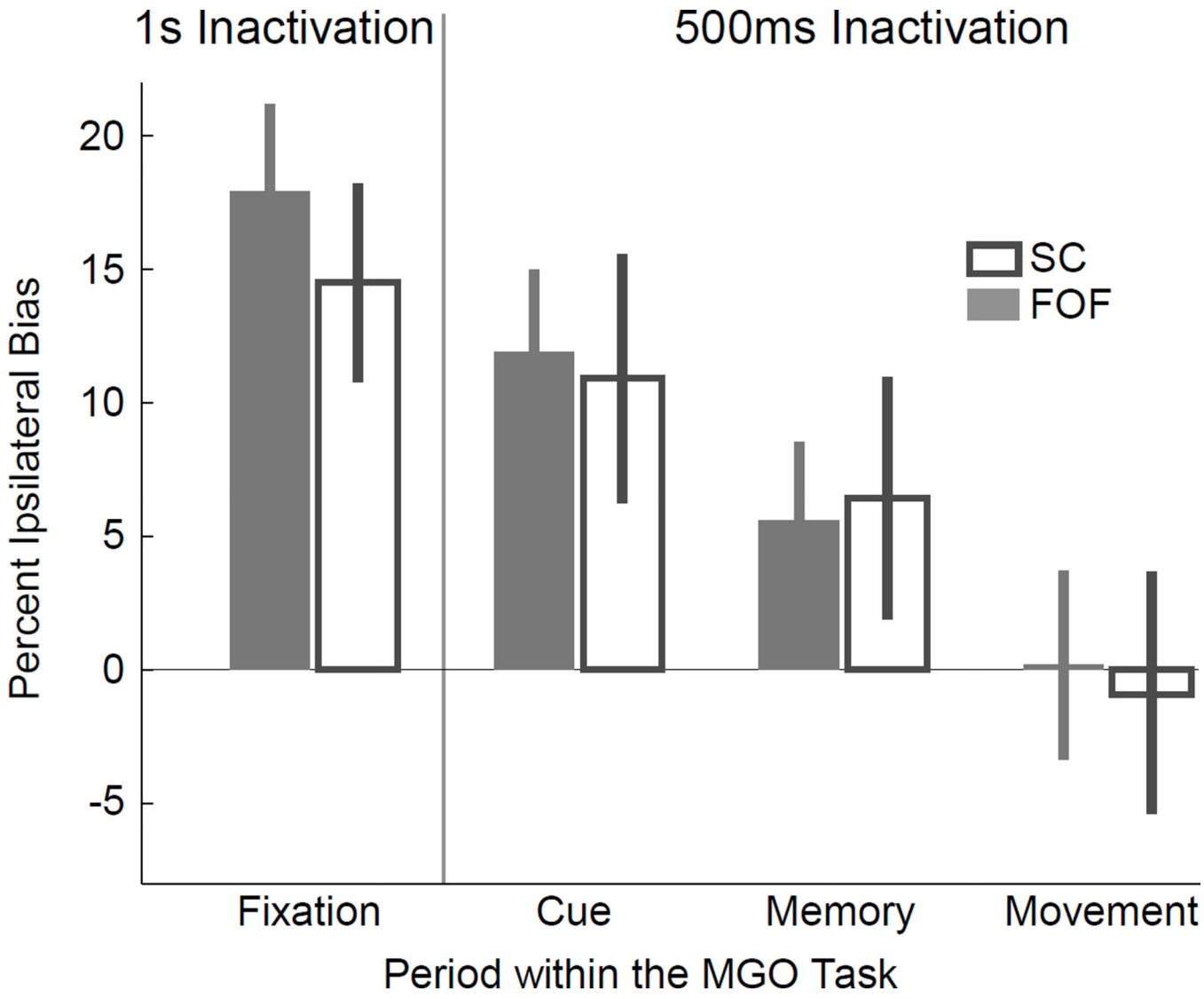

Figure 3.

Inactivation of the SC produces the same behavioral effects as inactivation of the FOF. Open grey bars show the effect of unilateral inactivation of the superior colliculus during performance of the memory-guided orienting task. The same time windows as in Fig. 2B ("1s Inactivation") and Fig. 2C ("500 ms Inactivation") were used. n=5 rats, 10 sessions for 1 s Halo experiment. n=3 rats, 16 sessions for 500 ms Halo experiment. As benchmark comparison data, the FOF inactivation effects from Figs. 2B and 2C are shown again here in the solid grey bars. Error bars indicate 95% confidence intervals centered on the mean for the combined data (See Methods).

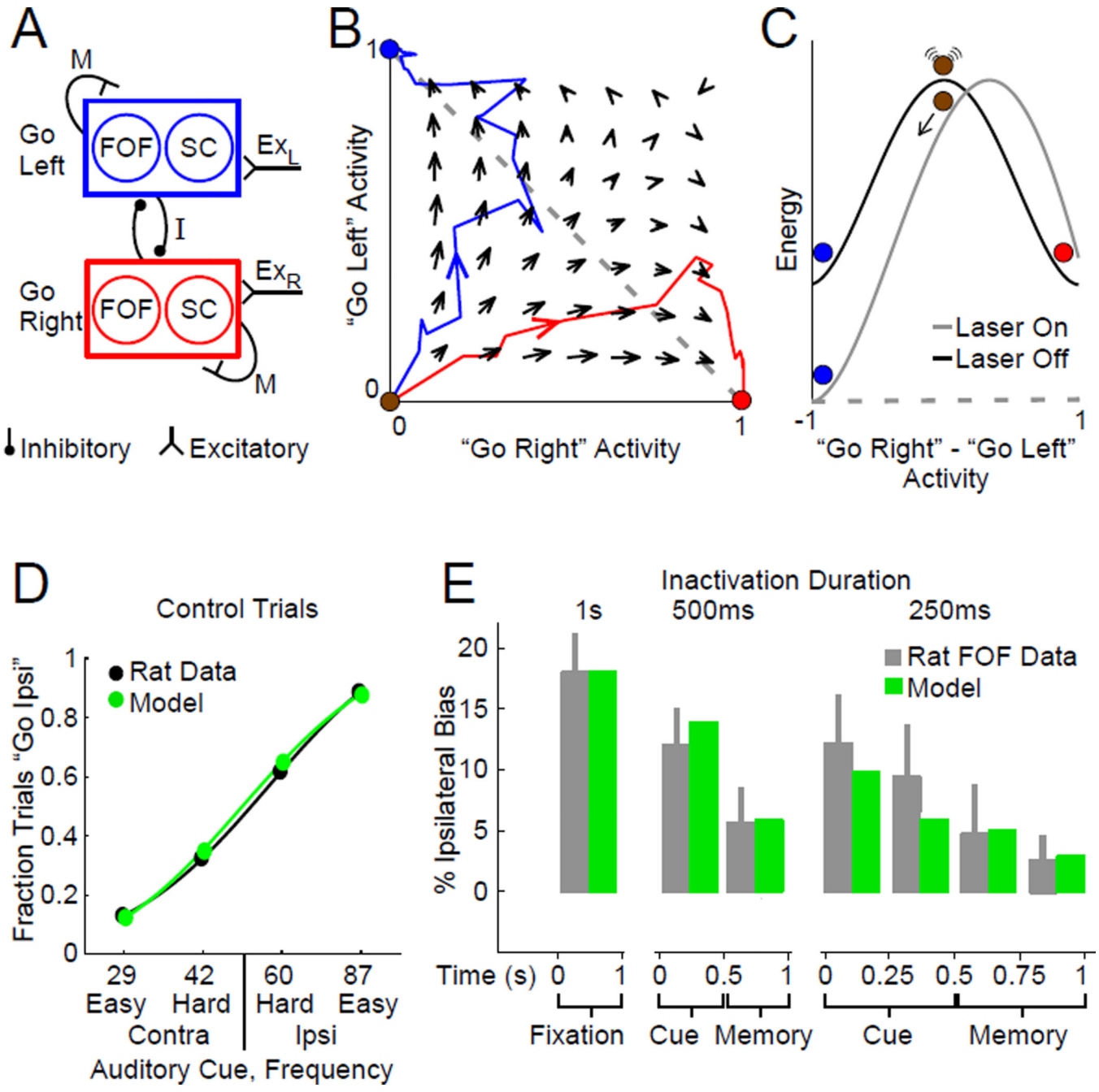

Figure 4.

A simple dynamical mechanism bridging cortical and subcortical regions reproduces the experimental data. (A) Circuit architecture in the model. There is self-excitation M between all "Go Left" and between all "Go Right" neurons, and mutual inhibition I between the two groups. One third of the neurons in each group are in the FOF, and another third are in the SC (n = 3.03 Equ. 4). A noise parameter σ controls the amount of random activity. Sensory inputs (Ex_R and Ex_L) bias activity towards one or another group, and the mutual inhibition leads to dynamics where one of the two groups will end each trial as the active "winner". (B) Phase space diagram for neural activity in the model. Arrows indicate the direction of flow

at each point in phase space. Red and blue lines show trajectories for two identically prepared trials, with equal and unbiased click frequency f = 50 clicks/sec. The only difference between the two example trials is different instantiations of noise. (C) Energy landscape along the “Go Right” - “Go Left” line (dashed grey line in (B)). At the beginning of each trial, the system is at the top of an energy hill and is susceptible to perturbations. Inactivation of one third of the neurons in a group (“Laser On”) has a strong effect of which attractor the system will flow towards. Once the system has reached an attractor (red or blue points), the same perturbation is less able to change attractor. (D) Comparison of rat and model performance on control trials. (E) Comparison of rat and model performance during inactivation trials.

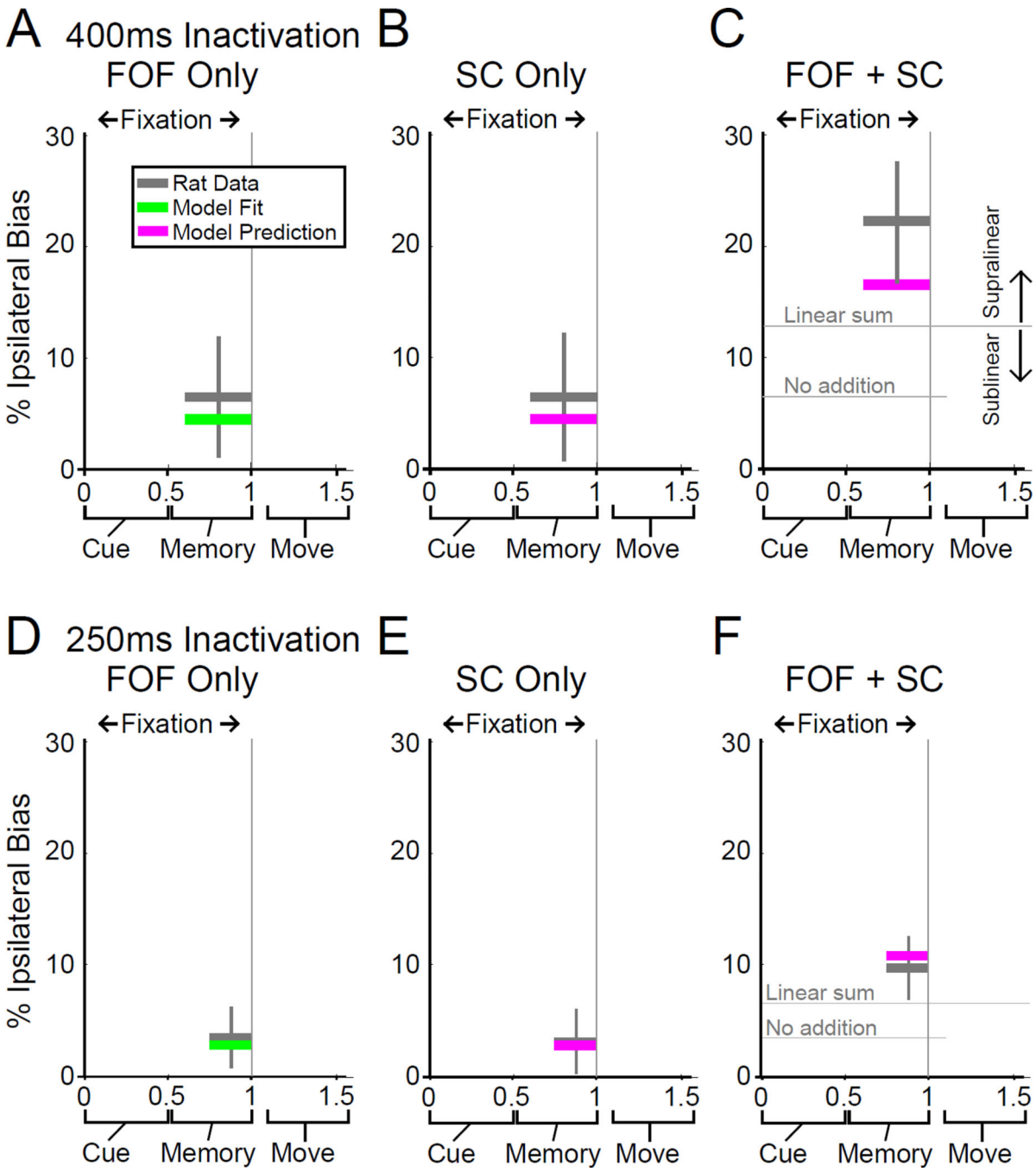

Figure 5.

Simultaneously silencing both the FOF and SC during the short-term memory delay period induces a greater choice bias than silencing either region alone, Average bias towards the implanted side caused by inactivation during different time windows. The vertical position of the grey bars indicate the measured bias, while the bar's horizontal extent indicates the time period over which the laser was turned on. Error bars represent 95% confidence centered on the mean for the combined data. Green bars represent the bias from the model for data it was fit to. Magenta bars represent the bias predicted by the model for data it was

not fit to. (A) Inactivation of the FOF during a 400 ms-long window in the memory delay period. Here we delayed the inactivation in the memory period by 100 ms to account for any potential response latency between the sensory cue and neural activity in the FOF (Hanks and Kopec et al., 2015) ensuring that the inactivation only occurred during the memory maintenance phase of the task. n = 3 rats, 11 sessions. (B) Same as (A) but for inactivation of the SC. n = 4 rats, 8 sessions. (C) Same as (A) but for simultaneous inactivation of both the FOF and SC. n = 3 rats, 12 session. Induced bias was greater than either the FOF alone ($p<0.01$), the SC alone ($p<0.01$), or the sum of the FOF and SC independently ($p<0.01$). (D–F) Same as (A–C) but with a new group of rats and inactivation during the last 250 ms of the memory delay period. (D) Inactivation of the FOF. n = 4 rats, 20 sessions. (E) Inactivation of the SC. n = 4 rats, 20 sessions. (F) Inactivation of the FOF and SC. n = 4 rats, 17 sessions. Induced bias was greater than either the FOF alone ($p<0.01$), the SC alone ($p<0.01$), or the sum of the FOF and SC independently ($p=0.018$). “No addition” line represents the bias from FOF or SC alone, whichever was greater.

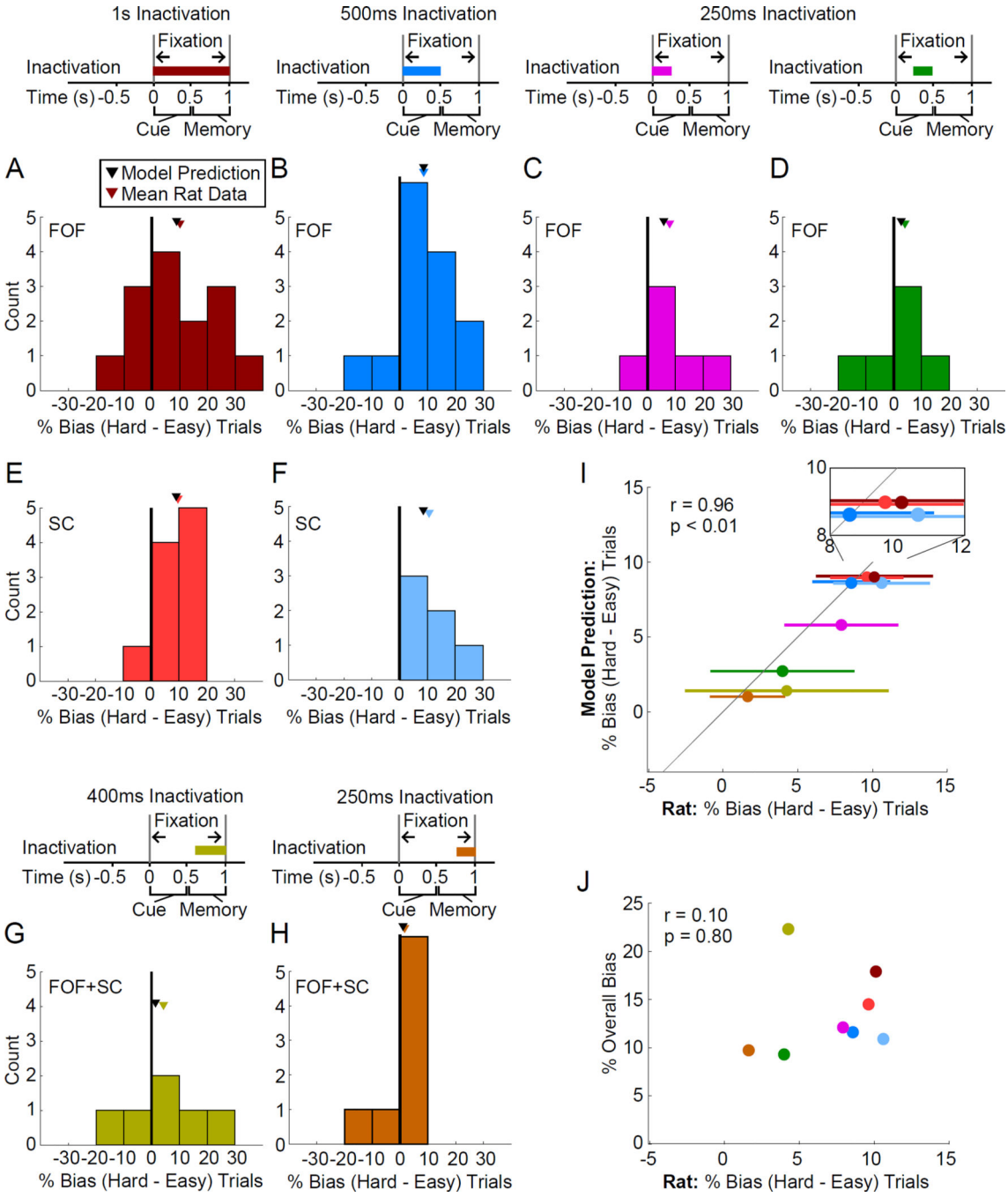

Figure 6.

Consistent with an attractor model, hard trials are perturbed to a greater extent than easy trials for transient inactivation overlapping decision commitment during cue presentation. (A–H) Histograms of difference in percent response bias between hard and easy trials for different inactivation periods. Differences computed separately for ipsiversive and contraversive trials. Therefore each subject contributes two data points to each plot. Mean across rats indicated with a colored triangle. Model prediction indicated with a black triangle. (A) 1 s inactivation of the FOF during the fixation period, n=7 rats, 33 sessions. (B)

500 ms inactivation of the FOF during cue presentation, n=7 rats, 33 sessions. (C) 250 ms inactivation of the FOF during the first half of the cue presentation, n=3 rats, 29 sessions. (D) 250 ms inactivation of the FOF during the second half of the cue presentation, n=3 rats, 29 sessions. (E) 1 s inactivation of the SC during the fixation period, n=5 rats, 10 sessions. (F) 500 ms inactivation of the SC during cue presentation, n=3 rats, 16 sessions. (G) 400 ms simultaneous inactivation of the FOF and SC during the memory delay period, n=3 rats, 12 sessions. (H) 250 ms simultaneous inactivation of the FOF and SC during the memory delay period, n=4 rats, 17 sessions. (I) Mean difference in response bias between hard and easy trials across rats for the different inactivation periods shown in (A–H) plotted against the model predictions for the same inactivation periods. Error bars represent s.e.m. across subjects. The model accurately predicts the mean rat data (correlation $r=0.96$, $p<0.01$; probability that rat and model data are not drawn from separate distributions $p=0.41$) Inset shows enlarged view of the region containing 1 s fixation and 500 ms cue inactivation period data. (J) Mean difference in response bias between hard and easy trials across rats for the different inactivation periods shown in (A–H) plotted against the overall ipsilateral response bias. No significant correlation was observed.