Coupling between One-Dimensional Networks Reconciles Conflicting Dynamics in LIP and Reveals Its Recurrent Circuitry

Highlights

- Slow dynamics of LIP local population can appear one-dimensional or high-dimensional

- Model of coupling between local networks reconciles conflicting data

- Reveals LIP's internal, recurrent circuitry underlying surround suppression

- Data show two-dimensional slow dynamics as predicted by model

Authors

Wujie Zhang, Annegret L. Falkner, B. Suresh Krishna, Michael E. Goldberg, Kenneth D. Miller

Correspondence

zhang_wujie@hotmail.com

In Brief

Zhang et al. found that slow LIP network dynamics can appear inconsistent with confinement to a single multi-neuronal dimension, unlike previous observations. Their model reconciles the conflict and reveals the circuitry underlying surround suppression.

CrossMark

CellPress

Coupling between One-Dimensional Networks Reconciles Conflicting Dynamics in LIP and Reveals Its Recurrent Circuitry

Wujie Zhang,[1,2,3,11,12,*] Annegret L. Falkner,[1,3,6,8,9] B. Suresh Krishna,[3,6,8,10] Michael E. Goldberg,[3,5,6,7,8] and Kenneth D. Miller[2,3,4,5]

[1]Doctoral Program in Neurobiology and Behavior
[2]Center for Theoretical Neuroscience
[3]Department of Neuroscience
[4]Swartz Program in Theoretical Neuroscience
[5]Kavli Institute for Brain Science
[6]Mahoney Center for Brain and Behavior
[7]Departments of Neurology, Psychiatry, and Ophthalmology
College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA
[8]New York State Psychiatric Institute, New York, NY 10032, USA
[9]Institute of Neuroscience, New York University School of Medicine, New York, NY 10016, USA
[10]Cognitive Neuroscience Laboratory, German Primate Center — Leibniz Institute for Primate Research, 37077 Goettingen, Germany
[11]Present address: University of California, Berkley, 1951 Oxford Street, Berkeley, CA 94720, USA
[12]Lead Contact
*Correspondence: zhang_wujie@hotmail.com
http://dx.doi.org/10.1016/j.neuron.2016.11.023

SUMMARY

Little is known about the internal circuitry of the primate lateral intraparietal area (LIP). During two versions of a delayed-saccade task, we found radically different network dynamics beneath similar population average firing patterns. When neurons are not influenced by stimuli outside their receptive fields (RFs), dynamics of the high-dimensional LIP network during slowly varying activity lie predominantly in one multi-neuronal dimension, as described previously. However, when activity is suppressed by stimuli outside the RF, slow LIP dynamics markedly deviate from a single dimension. The conflicting results can be reconciled if two LIP local networks, each underlying an RF location and dominated by a single multi-neuronal activity pattern, are suppressively coupled to each other. These results demonstrate the low dimensionality of slow LIP local dynamics, and suggest that LIP local networks encoding the attentional and movement priority of competing visual locations actively suppress one another.

INTRODUCTION

It has become increasingly appreciated that neural functions need to be understood in terms of neuronal populations and the dynamics of the circuits to which they belong (Miller and Wilson, 2008; Shenoy et al., 2013). However, the field of systems neuroscience in nonhuman primates has traditionally been dominated by studies of the properties of single neurons. While we have a wealth of knowledge of single-neuron behaviors in many areas of the primate brain, this knowledge remains largely phenomenological—we know *what* neurons do, but not *how* they do it. Especially on the circuit level, the mechanisms and connectivity underlying neuronal behaviors are often obscure.

Such is the case in the lateral intraparietal area (LIP), where a large body of literature has revealed that the activity of single neurons encodes visual attention and saccadic eye movements, as well as decision making variables, abstract categories, and other cognitive variables (Bisley and Goldberg, 2010; Freedman and Assad, 2011; Gold and Shadlen, 2007; Kable and Glimcher, 2009). However, little is known about the circuitry inside or outside the LIP network that produces such activity, and therefore the role of LIP in many of these functions is controversial. A step in understanding this circuitry was taken by Ganguli et al. (2008), who analyzed LIP network dynamics during two different tasks: a delayed saccade task (Bisley and Goldberg, 2003, 2006) and a random-dot motion discrimination task (Roitman and Shadlen, 2002). They found that the dynamics of the high-dimensional LIP network are dominated by one multi-neuronal dimension on slow timescales, which could be explained by a simple circuit model. This one-dimensionality was key to explaining an unexpected correspondence between LIP single-neuron responses and the timing of attentional shifts (examined in more detail below). More recently, Fitzgerald et al. (2013) found further evidence for one-dimensional dynamics in three experiments in which LIP encoded learned associations between visual stimuli.

Using a delayed-saccade task similar to the task of Bisley and Goldberg (2003, 2006) (hereafter BG), Falkner, Krishna et al., 2010 (hereafter FK) reported "surround suppression" in LIP (see also Louie et al., 2011), i.e., stimuli outside the receptive field

(RF) of a cell suppress the cell's activity. In the FK study, the population-averaged activity over time is very similar to that in the BG study, as expected given the very similar tasks. However, we find that the pattern of activity across neurons changes over time in a very different way in the FK study. In particular, the network dynamics in the FK dataset markedly deviate from the one-dimensional dynamics observed in the BG dataset, calling into question the validity of the one-dimensional LIP model of Ganguli et al. We show that the two sets of conflicting results can be reconciled and well characterized by a more general low-dimensional model: each of two local LIP networks in isolation has its own single dominant dimension, and suppressive coupling between them gives rise to two dominant dimensions. We further show that the FK data directly confirm the two-dimensional dynamics predicted by the model. Our study thus represents a step forward in discovering circuit mechanisms and connectivity from single-neuron recordings, and in understanding mechanisms behind LIP functions.

RESULTS

One-Dimensional Dynamics in LIP

We begin by describing the first of the two conflicting datasets (Bisley and Goldberg, 2003), along with the one-dimensional model (Ganguli et al., 2008) to which it gave rise.

The delayed-saccade task of BG is illustrated in Figure 1A (details in Supplemental Information [SI] section 1, available online). During this task, LIP neurons exhibit a large transient visual response to the onset of a saccade target or distractor in the RF, and sustained delay period activity (delay activity) when a saccade is planned to the RF (Figure 1C). When a distractor is flashed away from the target location during the delay period, attention is transiently attracted away from the target location to the distractor location. At the same time, the average visual response level of LIP neurons whose RFs contain the distractor location (the distractor population) rises above the average delay activity level of neurons whose RFs contain the target location (the target population). As the visual activity of the distractor population decays back to baseline, the locus of attention shifts back to the target location. This shift in attention coincides with the shift in the peak of LIP activity from the distractor population to the target population: when the decaying visual activity of the distractor population drops to a level statistically indistinguishable from the sustained delay activity of the target population (the "crossing time"—when the decaying red trace crosses the blue trace in Figure 1C), neither the target nor the distractor location has attentional advantage, whereas 100–250 ms before or after this crossing time, the distractor or target location, respectively, is the clear locus of attention.

Further analyses of these results (Bisley and Goldberg, 2006) revealed that this correspondence between activity crossing and attentional switching also held at the level of single LIP neurons. The crossing time of a single neuron is defined as the time at which the neuron's decaying response to a distractor, on trials in which a distractor is in its RF (distractor trials), crosses its own level of delay activity on trials in which a target is in its RF (target trials). These single-neuron crossing times are surprisingly invariant across neurons and closely aligned with the monkey's attentional switching time, despite high variability across neurons in their peak visual responses, time constants of visual response decay, and delay period responses.

Ganguli et al. (2008) explained this observation with the proposal that the dynamics of a local network (LN) of LIP neurons are dominated on slow timescales by one multi-neuronal activity pattern (i.e., a pattern, or vector, of relative firing rates across the cells of the network). Throughout this paper, we use the term "local network" (or LN) to mean a network of LIP neurons that share the same RF (explained more fully in the section "Simple model of coupled local networks reconciles the results"). Ganguli et al. proposed that the recurrent connectivity of an LN causes certain multi-neuronal activity patterns to persist longer in the absence of input; given steady input, these slowly decaying patterns also build up to be strongly amplified. If the network has only a single pattern that decays slowly, we refer to it as the network's "slow mode," where "mode" is a term borrowed from physics that describes a characteristic pattern of a system's response. As the visual response to a distractor decays, it becomes dominated by this slow mode after all other patterns decay away. Because the slow mode is more strongly amplified than other patterns, it also dominates steady-state responses, such as delay activity and activity during the initial fixation before target onset (fixation activity). Thus, after the other patterns in the distractor response decay away, the decaying distractor activity and the ongoing delay activity are both dominated by the slow mode, meaning that the pattern of distractor activity across neurons is very nearly a scaled-up version of the delay activity pattern. Then as the distractor activity decays further, it becomes very nearly identical to the delay activity pattern, which happens at the crossing time. Thus, each individual neuron has roughly the same activity in its delay response as in its distractor response at the crossing time, so that all neurons have about the same single-neuron crossing time.

This one-dimensional model predicts that multi-neuronal activity patterns that change on slow timescales are all highly correlated with one another because all are dominated by the same strongly amplified pattern. These include fixation and delay activity patterns and, to a lesser extent, slowly decaying visual activity patterns and slowly increasing activity patterns during decision-making tasks. On the other hand, during the initial transient visual response, many other activity patterns are activated, so the transient visual activity pattern is not highly correlated with the steady-state activity patterns. Ganguli et al. (2008) confirmed these predictions using the following analysis, which reveals network dynamics from the activity of a population of singly recorded neurons. At any millisecond time point t, we represent the trial-averaged activity of a population of N neurons as an N-dimensional vector, $\vec{r}(t)$, in an N-dimensional multi-neuronal firing rate space; each of the N elements of $\vec{r}(t)$ is the activity of one neuron at time t, averaged over trials. We also compute the N-dimensional fixation activity vector, $\vec{F}$, where each element is the activity of one neuron averaged over the fixation period before target onset and over target trials. Then, at each time point t over the course of the trial, a correlation coefficient is computed between $\vec{F}$ and $\vec{r}(t)$. Figure 1E shows that the correlation to fixation activity is indeed high for
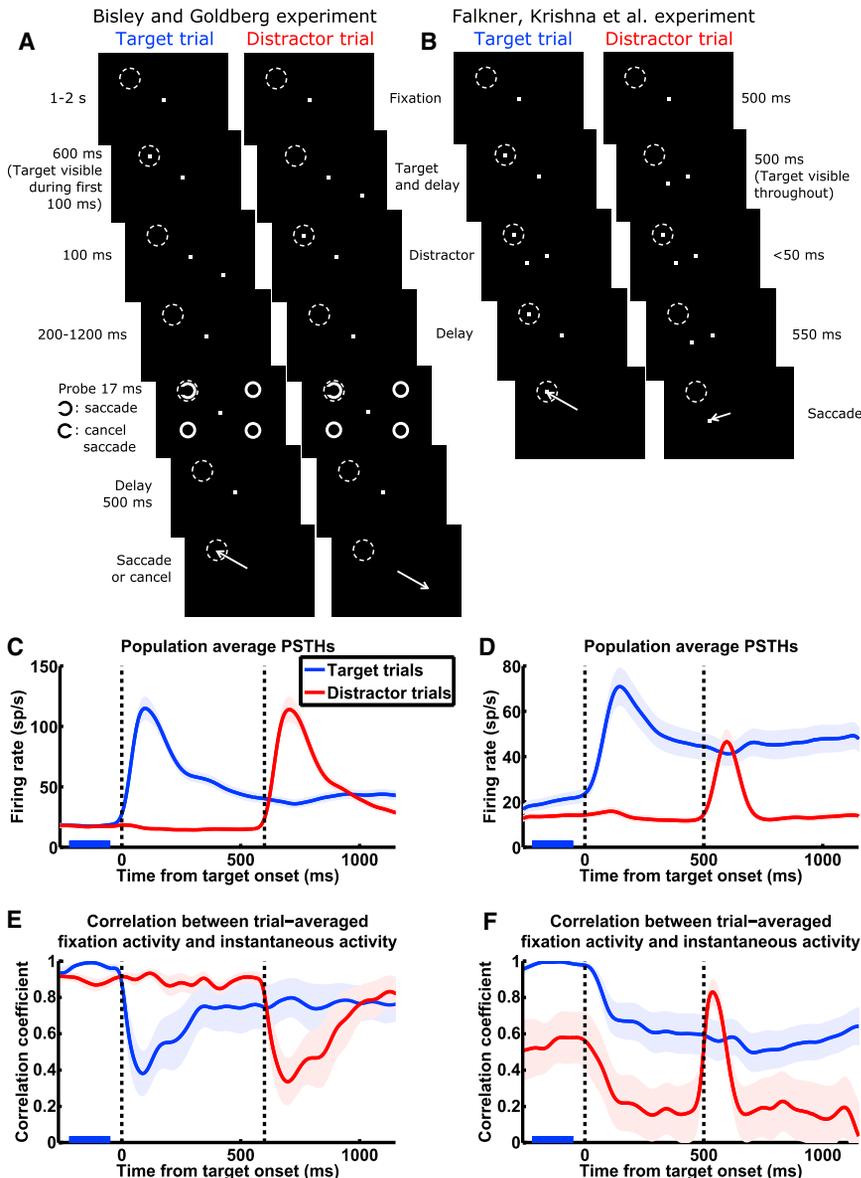
Bisley and Goldberg experiment
Target trial   Distractor trial

Falkner, Krishna et al. experiment
Target trial   Distractor trial

Figure 1. The Conflicting Population Dynamics Observed by Bisley and Goldberg and Falkner, Krishna et al.

(A, C, and E) Bisley and Goldberg (BG).

(B, D, and F) Falkner, Krishna et al. (FK).

(A and B) Task schematics. While the monkey fixates on a central spot, a target appears. The monkey holds fixation until the disappearance of the fixation spot, at which time it makes a saccade to the location of the target (except on "no-go" trials in BG; see below). During the delay between target onset and fixation spot disappearance, a task-irrelevant distractor stimulus is flashed. We call a given trial a target trial or distractor trial when the target or distractor, respectively, is in the RF (dashed circles) of the neuron being recorded. In the BG task, the target and distractor are in opposite visual quadrants and equidistant from the fixation spot; in the FK task, either the target or the distractor is in the RF, and the other stimulus is at the location that elicits maximum surround suppression for the recorded neuron. In the BG task, between 200 and 1,200 ms after the distractor disappears, a probe (a Landolt ring) is flashed at either the target or the distractor location, along with three complete rings elsewhere; a left-facing or right-facing ring instructs the monkey to proceed with the planned saccade ("go" trial) or cancel it and maintain fixation ("no-go" trial), respectively. In (C) and (E), we only include trials in which the probe appeared at least 700 ms after distractor onset. For task details, see SI section 1.

(C and D) Population average peristimulus time histograms (PSTHs) in the BG (C; n = 41 cells) and FK (D; n = 27) studies. Blue/red traces denote trials in which the target/distractor appears in the RF of the neuron being recorded; every neuron was recorded during both target and distractor trials and contributes to both traces. The first and second vertical dashed lines denote the onset of the target and the distractor, respectively. Shading around traces indicates SEM. PSTHs have been smoothed by convolution with a Gaussian kernel (σ = 30 ms; firing rates and correlations appearing to change before stimuli onset in C–F are artifacts of this smoothing).

(E and F) Correlation analysis for the BG (E) and FK (F) datasets. We define a trial-averaged pop-

ulation fixation activity vector $\vec{F}$, each element of which is the activity of one cell on target trials, averaged over trials and over the period from 220 to 50 ms before target onset (marked by blue bars in C–F). At each millisecond time point over the course of the target trial (blue traces) or distractor trial (red traces), the correlation coefficient was computed between the trial-averaged population instantaneous activity vector at that point in time and $\vec{F}$. The BG correlation patterns (E; presented in similar format in Ganguli et al., 2008) exhibit one-dimensional dynamics on slow timescales (high correlations during stable fixation activity and delay activity), while the FK correlation patterns (F) markedly deviate from one dimension (on distractor trials, low correlations during stable activity, and transient increase in correlation during distractor visual response). Vertical dashed lines are as in (C) and (D). Shading around traces indicates SE estimated from 1,000 bootstrap samples.

See Figure S1A for correlations between distractor trial fixation activity and instantaneous activity for the FK data, and Figures S2A–S2D for the FK data plotted separately for different reward conditions.

delay activity or distractor activity after the transient visual response decays away, indicating that fixation, delay, and post-transient distractor activity patterns all lie roughly in a single dimension, corresponding to the dominant activity pattern. The drop in correlation coefficient during the visual response indicates the transient deviation of activity from this one dimension caused by the transient activation of other non-dominant patterns.

Surround Suppression and Violations of One-Dimensional Dynamics

We continue by describing the second of the two conflicting datasets (Falkner, Krishna et al., 2010) and how it exhibits large deviations on both fast and slow timescales from the predictions of the one-dimensional model.

The task of FK (Figure 1B) is very similar to that of BG. For both tasks, we analyze data in each trial during time windows ending

shortly after distractor onset (i.e., before the onset of the probe in the BG task; see Figure 1A), up to which point the two tasks are virtually identical aside from three differences. First, BG used a flashed target while FK presented a target that stayed visible during the delay. This does not result in qualitatively different delay activity levels (compare delay activity between Figures 1C and 1D), consistent with LIP encoding the attentional and saccadic priority of the target location regardless of the visibility of the target. Second, BG randomly interleaved target trials and distractor trials, while FK presented target and distractor trials in blocks. Thus, in the FK experiment, on almost every trial the monkey had an expectation of where the target and distractor would be. This is reflected in higher anticipatory firing on target trials compared to distractor trials during the fixation period before target onset. The third difference is likely to be the key difference that led to different neural responses observed during the two tasks. In the BG task, the target and distractor are in opposite visual quadrants and equidistant from the fixation spot. In the FK task, in contrast, either the target or the distractor is in the RF of the cell being recorded in a given session, and the other stimulus is at the location eliciting maximum surround suppression of the recorded neuron. With this placement of stimuli, a saccade plan to the surround significantly suppressed the visual response to the distractor, while distractor appearance in the surround transiently and weakly, but significantly, suppressed delay activity during saccade planning (Figure 1D; quantified in Falkner, Krishna et al., 2010). Surround suppression was not observed in the BG dataset (quantified in Bisley and Goldberg, 2006), in which the stimulus locations were not selected for suppression. Other than the surround suppression of response amplitudes, the FK dataset displays the same overall pattern of fixation, visual, and delay activity as the BG dataset (compare Figures 1C and 1D).

However, beneath this apparent similarity in population average activity, the network dynamics are radically different; moreover, the FK dynamics appear to clearly violate the predictions of the one-dimensional model. Figure 1F shows the result of the correlation analysis on the FK data. Most strikingly, on distractor trials (red trace), even though the appearance of the target in the surround only minimally affects the mean firing rate of the population, target appearance causes a large, sustained drop in correlation, when the one-dimensional model would predict an unchanging and high level of correlation, as in Figure 1E. This indicates that the activity pattern of the population has changed dramatically while its mean firing rate has remained about the same. Furthermore, the later appearance of the distractor in the RF causes a large, transient rise in correlation that subsequently returns to the steady low level present before distractor onset, when the one-dimensional model would predict the opposite change—a large and transient drop in correlation upon distractor onset, as in Figure 1E. In target trials (blue traces), the difference is more subtle, with target onset evoking a small, sustained drop in correlation, similar to the sustained drop in the BG case, but without the initially larger transient decrease.

Note that in the BG dataset, the two trial types are randomly interleaved; thus, the monkey does not know the trial type during the initial fixation, and fixation activities are the same in the two trial types. In the FK dataset, however, fixation activities are different on the two trial types due to the block design. We chose to use the fixation activity on target trials as opposed to distractor trials to calculate correlations because it reveals salient patterns in the network dynamics. Using distractor trial fixation activity is another angle from which to examine the network dynamics that give less informative results, i.e., correlations do not rise and drop saliently over time (Figure S1A).

Thus, the results of BG and of FK seem incompatible. The robust one-dimensional dynamics observed in the BG data require that the local anatomical connectivity of LIP selectively amplify only one multi-neuronal activity pattern. How can this same anatomical connectivity realize dynamics that deviate so far from the one activity pattern that it so strongly amplifies?

Simple Model of Coupled Local Networks Reconciles the Results

We found the answer in a simple model of the interactions between two coupled LIP LNs. This model replicates the FK findings and yet reduces to the one-dimensional dynamics that characterize the BG findings when the two LNs are not coupled.

We model two LNs in LIP, each composed of excitatory (E) and inhibitory (I) neurons that share an RF, with randomly distributed neuronal time constants (Figures 2A and 2B; see SI section 2.2 for details of the model). Connections between the neurons are sparse, and their strengths are randomly distributed. Within each LN, excitatory connections are, on average, stronger than inhibitory connections. This dominance of excitation is consistent with evidence based on dendritic structure of increased connectivity between excitatory cells in LIP compared to primary sensory cortices (Elston and Rosa, 1997). Such connectivity within an LN, when it's not connected to another LN, amplifies a single pattern, one of increased activity across most cells, more strongly than all other patterns.

The LIP cortical surface contains rough topological maps of visual space (Blatt et al., 1990; Patel et al., 2010). Neurons sharing an RF, which are more likely to be located close to each other on the cortical surface, make up an LN in our model. We model the connections of I cells to be restricted to the LN to which they belong, for inhibitory interneurons generally only make short-range projections, whereas E cells can potentially make long-range projections to the other LN. Since no significant interaction between RFs was observed in the BG dataset (quantified in Bisley and Goldberg, 2006), we infer that for these RFs, the corresponding LNs are not directly connected (Figure 2A). In contrast, by maximizing surround suppression, FK selected for RFs that did interact. Since the interaction observed was predominantly suppressive, it's likely that the excitatory connections from each LN are stronger to the I cells than to the E cells of the other LN. For simplicity, we model the across-network connections as being from the E cells of each LN to the other LN's I cells only, with sparse and random connectivity (Figure 2B). Our results do not change if we include weaker across-network E-to-E connections (data not shown).

We use a standard linear firing rate model (SI section 2.2; Dayan and Abbott, 2005) to simulate the trial-averaged activity in the experiments. We do not explicitly simulate single trials, for we have no knowledge of the single-trial population dynamics
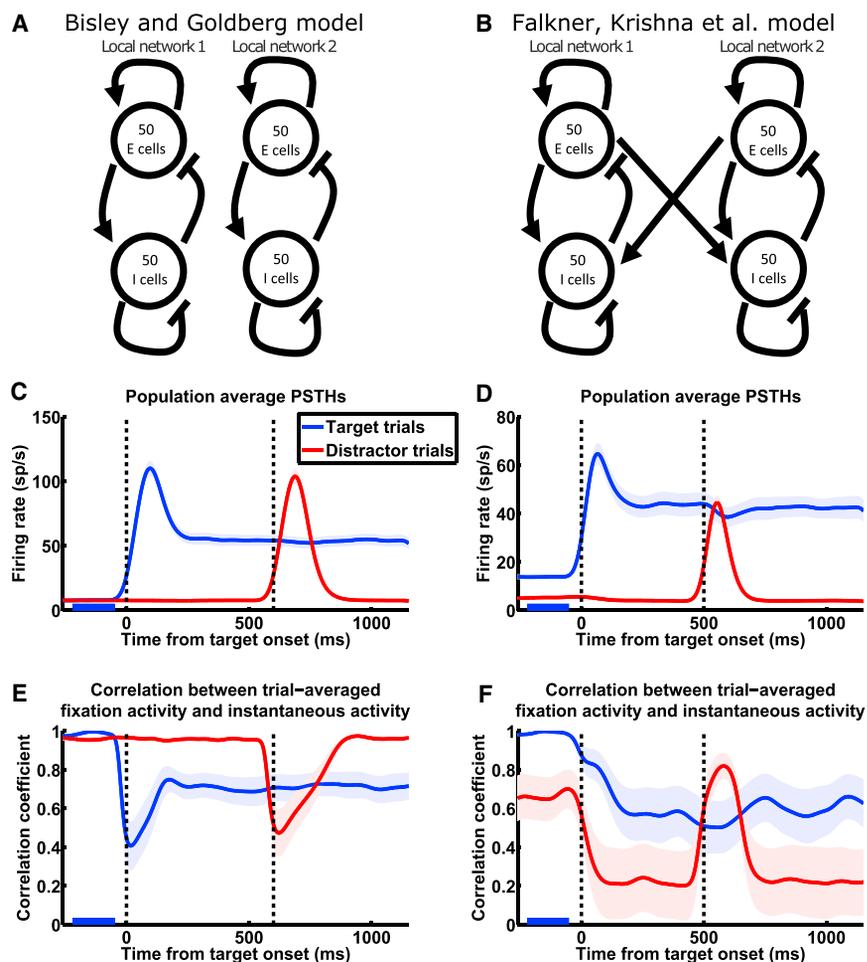
A Bisley and Goldberg model

Local network 1   Local network 2

B Falkner, Krishna et al. model

Local network 1   Local network 2

C   Population average PSTHs

D   Population average PSTHs

E   Correlation between trial−averaged
fixation activity and instantaneous activity

F   Correlation between trial−averaged
fixation activity and instantaneous activity

Figure 2. Model Reproduces the Response and Network Dynamics of Bisley and Goldberg and Falkner, Krishna et al.

(A and B) Schematics of the model network connectivity for the BG (A) and the FK (B) scenarios. In both cases, we model two recurrently connected local LIP networks (together referred to as a global network) corresponding to two RF locations, with each LN consisting of E and I cells. Connectivity within each LN is such that each LN by itself amplifies a single multi-neuronal activity pattern much more strongly than other patterns (see Results for details). The FK model network (B) differs from the BG model network (A) in the addition of coupling between the LNs that mediates interactions between them.

(C and D) Model reproduces LIP activity observed by BG (C; n = 41 cells; in all simulation results except where noted, each neuron was "recorded" from a different simulated global network) and FK (D; n = 27). Population average PSTHs with same conventions as Figures 1C and 1D. PSTHs have been smoothed by convolution with a Gaussian kernel (σ = 30 ms; firing rates and correlations appearing to change before stimuli onset in C–F are artifacts of this smoothing).

(E and F) Model reproduces LIP network dynamics observed by BG (E) and FK (F). Correlation analysis with same conventions as Figures 1E and 1F. See Figure S1B for correlations between distractor trial fixation activity and instantaneous activity for the FK simulation, and Figures S2E–S2H for separate simulations of the different reward conditions of the FK experiment.

during the tasks. The experiments involve a variety of sensory, motor, and cognitive processes that likely give rise to a variety of external inputs to LIP during a trial, which we model as the following four types. (1) Fixation input, which is spontaneous firing from the external input sources when there is no stimulus in or saccade plan to the RF, such as during the fixation period. (2) Visual input, which is bottom-up input to an LN when a visual stimulus is in the RF, which is strong upon stimulus onset and becomes weak as the stimulus is sustained. Visual input arrives from areas that could include V2, V3, V3A, V4, middle temporal area (MT), and inferotemporal cortex (Baizer et al., 1991; Blatt et al., 1990; Lewis and Van Essen, 2000). (3) Delay input, which is persistent top-down input to an LN when a saccade is being planned to the RF, arriving from frontal areas such as the frontal eye field (FEF) or dorsolateral prefrontal cortex (dlPFC; Blatt et al., 1990; Stanton et al., 1995; in SI section 3, we discuss other possible mechanisms underlying delay activity and their implications for our model). (4) Expectation input, which is top-down input to one LN during the fixation period before target onset, when the animal is in a block of trials during which the target always appears in the RF of that LN (as in the blocked experiment of FK). Expectation input likely also arrives from frontal areas such as FEF or dlPFC (Coe et al., 2002; Roesch and

Olson, 2003). The total external input to the neurons at any time is the sum of one or more of these four types of input. For each of the four types of input, input to each cell is independently drawn from a uniform distribution, with ranges of the distributions chosen to fit experimentally observed neural responses. Thus, it is important that the inputs from different sources are uncorrelated. In addition, the external input contains weak, temporally correlated noise that is independent for different neurons, simply to produce small firing rate fluctuations similar to those seen in the experiments.

In the experiments, different neurons are recorded from different LIP locations and have different RF positions. As in Ganguli et al. (2008), we interpret this to mean that these neurons are situated in different LNs, which share the same set of connectivity, neuronal, and input statistics. To model this, we run the simulation multiple times, each time with a different random instantiation of network connectivity, neuronal time constants, and input patterns, and "record" from a single randomly chosen cell during each simulation. Each simulation includes target and distractor trials for the recorded cell. Figures 2C and 2D show the population peristimulus time histograms (PSTHs) from such simulations of the BG (Figure 2C) and FK (Figure 2D) experiments, which reproduce the experimentally observed firing patterns,

including the observed absence or presence of surround interactions. More significantly, our model reproduces the apparently conflicting network dynamics of the two experiments, as revealed from the correlation analysis: the BG model shows one-dimensional dynamics on slow timescales (Figure 2E), and the FK model shows the same higher-dimensional dynamics as experimentally observed (Figure 2F).

If we compute correlations of instantaneous activity to distractor trial fixation activity, rather than to target trial fixation activity, the model also qualitatively reproduces the experimental results (Figure S1B). Furthermore, modeling higher reward levels as resulting in higher levels of delay input (Leon and Shadlen, 1999; Kennerley and Wallis, 2009), we reproduce the results found when the data of FK, which consist of trials with large or small reward, are analyzed separately by reward level (Figure S2). Because the activity and correlation patterns are qualitatively similar across reward levels in the data (Figures S2A–S2D), in all other simulations we simply modeled the average reward level.

In addition to the correlation analysis, another way to examine network dynamics is to calculate, for each instantaneous activity vector, the norms of its component parallel to $\vec{F}$ and its component orthogonal to $\vec{F}$. This reveals a similar picture for the FK data to the correlation analysis, which our model reproduces (Figure S3).

Conceptual Picture: Coupling of Local Slow Modes Explains LIP Dynamics

We fully analyze the behaviors of the model in the next sections, but first, in this section, we presage those results by presenting a simplified conceptual understanding.

Each LN has its own single dominant activity pattern (its slow mode), and therefore each on its own would follow one-dimensional dynamics. The circuitry that creates surround suppression causes these two patterns to suppress one another, and this mutual suppression in turn qualitatively explains the FK correlation patterns, as follows.

Suppose we are recording in one of the LNs, call it LN1, and let the other LN be LN2. $\vec{F}$, the fixation activity of LN1 on target trials, is dominated by LN1's slow mode because it is driven by both fixation input and expectation input. At any given time, the correlation of LN1's instantaneous activity with $\vec{F}$ is high or low according to whether or not that instantaneous activity is dominated by the slow mode. Now consider LN1 on distractor trials. During the initial fixation period, LN1 receives fixation input, but not expectation input. Thus, its slow mode is activated less than on target trials; in addition, its slow mode is suppressed by the more activated slow mode of LN2, which is receiving both fixation and expectation inputs. As a result, the relative contribution of activity patterns other than the slow mode to LN1's activity is larger than on target trials, resulting in reduced correlation between distractor trial fixation activity and $\vec{F}$. After the target appears in LN2's RF, LN1's slow mode continues to be driven only by fixation input; in addition, it is strongly suppressed by the slow mode of LN2, which is strongly driven by both visual stimulation and the subsequent top-down delay input. This greatly reduces the correlation. Finally, when the distractor appears, strong visual stimulation transiently drives up LN1's

slow mode as well as other patterns. In BG, this would lower the correlation—LN1 was dominated by the slow mode before distractor onset, and is now less so. In FK, LN1's slow mode was strongly suppressed before distractor onset, and is now driven up to dominate LN1's activity, resulting in the transient rise in correlation.

The conceptual picture just given is simplified in that it describes each LN as having only one dimension of activity that is strongly amplified. In reality, a second strongly amplified dimension is created in each LN by the coupling between the two LNs. When LN2 is the more strongly driven LN, it strongly drives the I cells of LN1. In addition to suppressing LN1's slow mode, this amplifies a pattern of differential firing between the E and I cells in LN1, making the slow mode an even less dominant part of LN1's activity. This will become clear with the detailed analysis below.

Detailed Analysis: Two-Dimensional Dynamics Result from the Coupling of Local Slow Modes

We now take a closer look at the mechanisms of the model. We modeled the BG scenario with two unconnected LNs, each with a slow mode. The model simply behaves like two copies of the one-dimensional model of Ganguli et al., reproducing one-dimensional dynamics and the absence of surround interaction.

The only difference in network architecture in our model of the FK scenario is the presence of connections between the two LNs. Thus, the dominant activity patterns of the two LNs influence each other and are no longer independent. To understand the activity patterns of the global network consisting of the two coupled LNs, we examine the global connectivity matrix, which describes all connections, both within and between the LNs. This connectivity between neurons determines how strongly each neuron excites or inhibits other neurons. The connectivity can equivalently be described as connections between sets of activity patterns, determining how strongly activity in one pattern excites or inhibits activity in itself and in other patterns. For a network composed of separate excitatory and inhibitory neurons, it is often informative to analyze its connectivity as the connections between its Schur activity patterns (Murphy and Miller, 2009; Goldman, 2009; described in more detail in SI section 4). These are an ordered set of orthogonal activity patterns whose connections with each other are as simple as possible for a set of orthogonal patterns: each Schur pattern has a self-connection, and in addition, there is a set of purely feedforward connections between the patterns. We choose to order the patterns by their self-connection strength. Then, activity in pattern 1 (the pattern with the strongest self-excitation) can only influence itself by its self-connection; activity in pattern 2 can excite or inhibit activity in pattern 1, in addition to influencing itself; pattern 3 can excite or inhibit pattern 1, pattern 2, and itself, etc. Thus, given similar external inputs to the patterns, the dominance of any pattern in the network's dynamics can be predicted by the strengths of its self-connection and the feedforward connections it receives. The activity of the network at any moment can be uniquely decomposed as a weighted sum of all Schur patterns. The dominant patterns will generally have weights with the largest absolute values (the weights can be positive or negative).
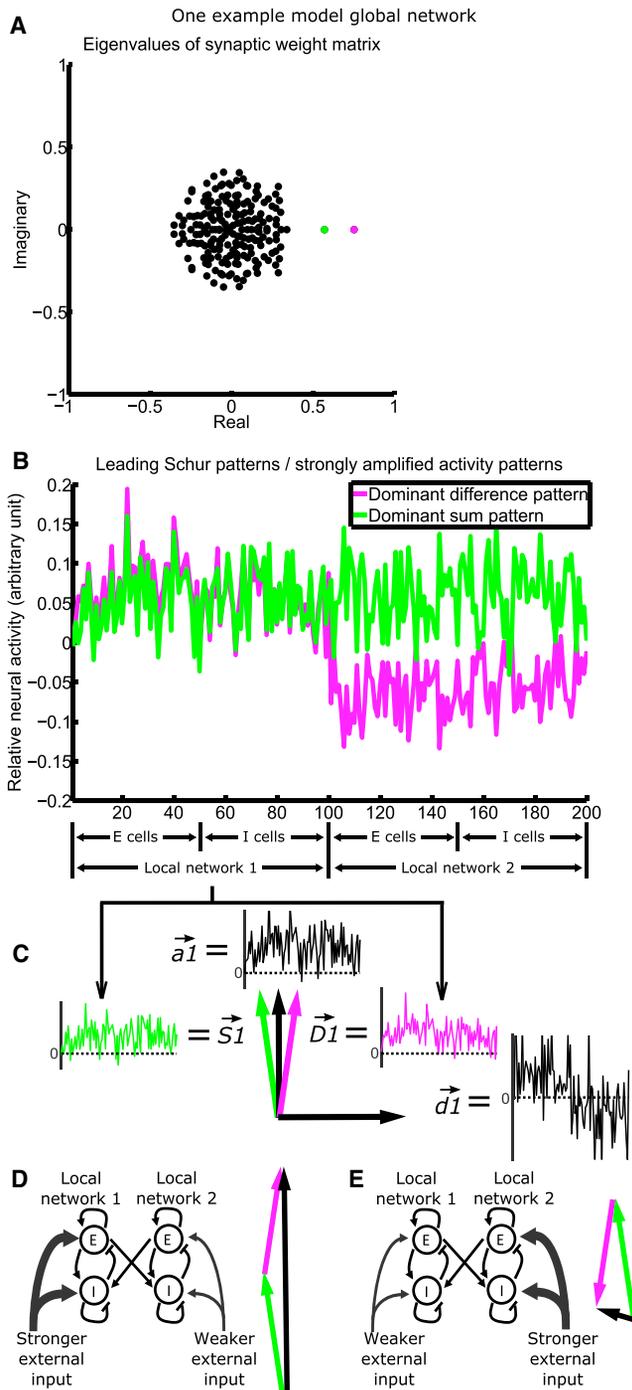
A

One example model global network
Eigenvalues of synaptic weight matrix

B

Leading Schur patterns / strongly amplified activity patterns

Dominant difference pattern
Dominant sum pattern

E cells   I cells   E cells   I cells

Local network 1   Local network 2

C

$\vec{a1} =$

$= \vec{S1}$ $\vec{D1} =$

$\vec{d1} = 0$

D

Local
network 1   Local
network 2

Stronger
external
input   Weaker
external
input

E

Local
network 1   Local
network 2

Weaker
external
input   Stronger
external
input

A set of numbers called eigenvalues can be calculated from the connectivity matrix; each eigenvalue is associated with a Schur pattern, and the real part of the eigenvalue corresponds to the strength of that pattern's self-connection. Plotting the eigenvalues of the global connectivity matrix from one representative simulation, we see that two eigenvalues have real parts more positive than the rest, indicating that there are two strongly self-excitatory activity patterns (Figure 3A; SI section 5). Analysis of the feedforward connections between patterns (SI section 4; Figure S4) shows that (1) these two patterns are nearly independent, with only a very weak feedforward connection from one to the other (Figure S4G); (2) these two patterns receive approximately orthogonal sets of feedforward inputs from the less self-excitatory patterns (Figures S4B, S4C, and S4F); and (3) the less self-excitatory patterns form much stronger feedforward connections to the two strongly self-excitatory patterns than to each other (Figure S4G). Thus, given similar external input to all patterns, these two patterns will dominate the network's activity.

To understand the structure of these two potentially dominant activity patterns, in Figure 3B we plot the relative activation of different neurons in these two Schur patterns. We have arbitrarily chosen the overall sign of each pattern in Figure 3B such that both have mostly positive elements in LN1, and we have arbitrarily set the amplitude (i.e., the vector norm) of each pattern to 1. We note two key points about these two global patterns, which together show that they represent the coupled activation of the two local slow modes. (1) The two patterns represent two different forms of coupled activation of the two LNs: one is a "sum pattern," representing roughly equal activation of the two LNs; the other is a "difference pattern," representing differential activation of the two LNs, i.e., this pattern increases the activity of one LN and decreases the activity of the other. (2) The portions of the sum and difference patterns within a given LN are very similar to each other (e.g., in Figure 3B, compare the two patterns restricted to neurons 1–100; for this comparison, the overall sign of activation within an LN is arbitrary; also see Figure S5A), as well as to the slow mode of that LN if it were not connected with the other LN (Figure S5A), which reflects the connectivity within that LN.

The sum and difference patterns, in addition to being strongly amplified by recurrent connectivity, also typically receive stronger external input than the other patterns. The intuition for this is the following: The external input is non-negative, since we

Figure 3. Recurrent Connectivity Strongly Amplifies Two Activity Patterns

(A) The eigenvalue spectrum of the connectivity matrix (matrix plotted in Figure S4D) for a model network composed of two interconnected LNs of 100 neurons each. Each eigenvalue is associated with a Schur vector, representing a pattern of relative activation across neurons (see Results for details). The more positive the real part of an eigenvalue, the more strongly the network amplifies the corresponding Schur activity pattern. Two patterns (magenta and green) are more strongly amplified than others and are plotted in (B).

(B) Relative activation across neurons in the dominant difference pattern (differential activation of the two LNs; magenta) and the dominant sum pattern

(equal activation of the two LNs; green), or, equivalently, the two leading Schur vectors of the connectivity matrix. The difference/sum pattern is driven by the difference/sum of the mean inputs to the two LNs. Note the similarity of the two patterns across cells of the same LN.

(C) The LN1 portion of the sum $(\vec{S1})$ and difference $(\vec{D1})$ patterns can be represented as vectors in the two-dimensional space they define. We can take the axes of the 2D space to be $\vec{a1}$, a vector proportional to the average of $\vec{S1}$ and $\vec{D1}$, and $\vec{d1}$, a vector proportional to their difference.

(D and E) When LN1 receives stronger (D)/weaker (E) mean external input than LN2, $\vec{S1}$ is activated positively, and $\vec{D1}$ is activated positively (D)/negatively (E). Thus, the $\vec{a1}$ components of $\vec{S1}$ and $\vec{D1}$ add (D)/cancel (E), while the $\vec{d1}$ components of $\vec{S1}$ and $\vec{D1}$ cancel (D)/add (E). The actual activity vectors (black) thus point in very different directions in (D) and (E).

See Figure S4 for analysis of the connectivity and the Schur patterns, and Figure S5A for comparisons of the directions of dominant activity patterns.

assume it's carried by purely excitatory projection neurons. The purely excitatory input to an LN most strongly activates patterns that represent concerted activation of cells in the LN, and the sum and difference patterns are the only such patterns (Figure S6). We now present the math behind this intuition.

For a given global network, consider the vector of external inputs $\vec{I}$, each of whose elements is the input to one neuron of the network. Let's decompose it as $\vec{I} = \vec{I}_{mean} + \vec{I}_{res}$, where the LN1 elements of $\vec{I}_{mean}$ all equal the mean input to LN1, which we call I_1, and similarly the LN2 elements all equal the mean input to LN2, I_2. $\vec{I}_{res}$ contains the residuals, which sum to zero over each LN. Similarly, we can decompose a Schur pattern $\vec{P}$ of the given global network as $\vec{P} = \vec{P}_{mean} + \vec{P}_{res}$, where the elements of $\vec{P}_{mean}$ are P_1 and P_2, the LN means of $\vec{P}$. Given the orthonormality of the Schur patterns, the external input to $\vec{P}$ is $\vec{I} \cdot \vec{P}$. Over each LN, the residuals sum to zero while the mean vectors are constant, so the dot product of any residual vector with any mean vector is 0. Thus, $\vec{I} \cdot \vec{P} = \vec{I}_{mean} \cdot \vec{P}_{mean} + \vec{I}_{res} \cdot \vec{P}_{res}$. The first term $\vec{I}_{mean} \cdot \vec{P}_{mean} = N(I_1 P_1 + I_2 P_2)$, where N is the number of neurons in an LN. The second term is a dot product of random vectors drawn independently for each global network and input pattern. By the central limit theorem, for large N, $\vec{I}_{res} \cdot \vec{P}_{res}$ across different random instantiations of networks and inputs approaches a Gaussian distribution with mean zero and SD $\sqrt{N}\sqrt{2}\sigma_I\sigma_P$, where σ_I and σ_P are the SDs across the elements of $\vec{I}_{res}$ and $\vec{P}_{res}$, respectively. Thus, the typical order of magnitude of $\vec{I}_{res} \cdot \vec{P}_{res}$ in any given global network will be $\sqrt{N}\sqrt{2}\sigma_I\sigma_P$. To compare the magnitude of $N(I_1 P_1 + I_2 P_2)$ and $\sqrt{N}\sqrt{2}\sigma_I\sigma_P$, we note that N is much greater than $\sqrt{N}$, and I_1 and I_2 are larger than or comparable to σ_I (since external inputs are positive). For the sum and difference patterns, the absolute values of P_1 and P_2 are comparable to σ_P (Figure S6). Thus, $N(I_1 P_1 + I_2 P_2) >> \sqrt{N}\sqrt{2}\sigma_I\sigma_P$ for these two patterns, so their inputs are approximately $N(I_1 P_1 + I_2 P_2)$. For the other patterns, P_1 and P_2 are close to zero, much smaller than σ_P (specifically, for our model with N = 100, P_1, $P_2 < \sigma_P / \sqrt{N}$; Figure S6) and much smaller than the LN means of the sum and difference patterns (Figure S6). Thus, their mean-driven inputs are small, and their input is dominated by the relatively small input from the residual terms—intuitively, they represent random activations of cells and are weakly driven by the random fluctuations across cells of inputs about their mean. In our model, the range of the uniform distribution from which visual inputs for different cells are drawn is larger than that for delay inputs, which is larger than those for fixation and expectation inputs (SI section 2.2). A larger input variance (i.e., σ_I) means larger random fluctuations of inputs across cells; thus, patterns other than the sum and difference patterns are activated more strongly by visual and delay inputs than by fixation and expectation inputs. This is consistent with the large variance of activity across cells during visual and delay responses (Figures 1C and 1D), and the strong activation of weak patterns by the visual input (Ganguli et al., 2008).

When inputs are randomly redrawn for simulations of different global networks, the means of the inputs will be consistent across simulations. The large means of the sum and difference patterns within each LN will also be consistent across simula-

tions (Figure S6) because these are primarily determined by the statistically consistent strength of the overall excess of excitation in each LN's connectivity. The means of the other patterns will be consistently small (Figure S6) because they are determined by random factors in the connectivity. Thus, across different global networks, the same analysis will apply and the sum and difference patterns will consistently receive strong input, while the other patterns will receive weaker inputs that vary from simulation to simulation.

We note that for a small proportion of random instantiations of connectivity matrices, a pair of complex patterns (which are complex conjugates in the eigenvector basis) take the place of the single real sum pattern described above. We show in SI section 6 and Figure S5 that in these cases, the complex pattern pair behaves effectively like the single real sum pattern.

We can use our understanding of the two dominant global activity patterns to understand the activity within a single LN, which we take to be LN1. We will call the LN1 portions of the sum and difference patterns $\vec{S1}$ and $\vec{D1}$, respectively, and take them to be normalized to unit vector length. Because these two patterns are not exactly equal to one another, they define a two-dimensional space of strongly amplified activity patterns in LN1. A convenient orthogonal pair of vectors to serve as a basis for this space is a vector $\vec{a1}$ proportional to the average of $\vec{S1}$ and $\vec{D1}$, and a vector $\vec{d1}$ proportional to their difference (again, both normalized to unit vector length; Figure 3C). $\vec{a1}$ represents concerted firing of the cells in LN1, while $\vec{d1}$ represents differential firing of the E and I cells in LN1. $\vec{a1}$ is almost precisely the slow mode of LN1 if it were isolated, while $\vec{d1}$ is very nearly orthogonal to that slow mode (Figure S5A). From the analysis above, the activation of $\vec{S1}$ and $\vec{D1}$ is largely determined by the mean inputs to the two LNs, as illustrated in Figures 3D and 3E.

In Figure S7 and SI section 7, we show that because there are two strongly amplified activity patterns, single neurons in both the FK data and model no longer have a common "crossing time," as observed in BG by Ganguli et al. (2008), and we discuss possible consequences of this for attentional switching.

Detailed Analysis: Two-Dimensional Dynamics Explain Correlation Patterns

We are now in a position to understand the behavior of correlations between fixation and instantaneous activities in the FK model. Here we consider a population of neurons simultaneously recorded from a single LN, part of a single global network (Figure 4). In SI section 8, we explain why the conclusions we reach remain valid for a population in which each neuron is recorded from a different global network (the case of our main simulations in Figure 2 and likely of the experiments in Figure 1).

First, we see in simulations of a single global network that, indeed, the two dominant activity patterns, $\vec{S1}$ and $\vec{D1}$, largely explain the population-averaged activity of LN1 (Figures 4A and 4E; the results and analysis are identical for LN2). Moreover, we can see the contributions of $\vec{S1}$ and $\vec{D1}$ activity to the correlation patterns by breaking up the correlations into two
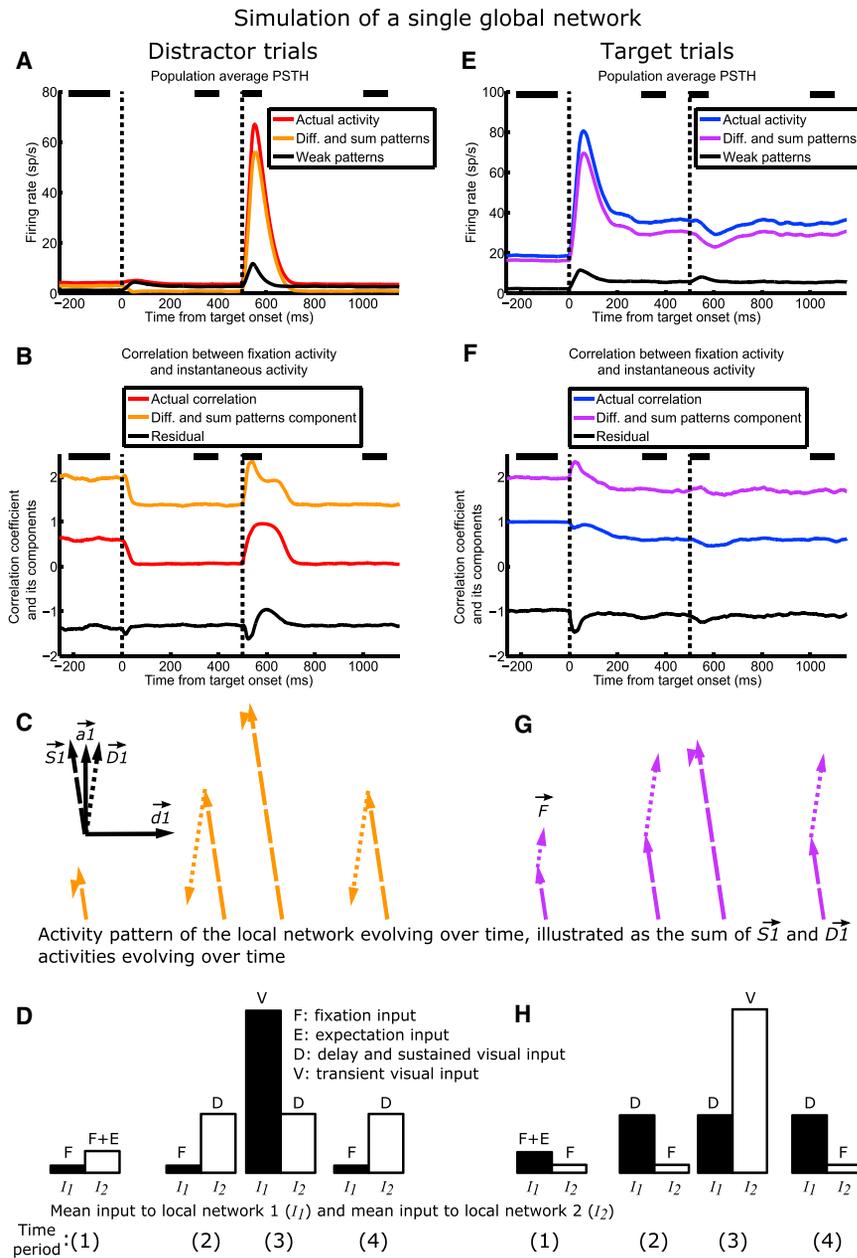
Simulation of a single global network

A Distractor trials

Population average PSTH

E Target trials

Population average PSTH

B

Correlation between fixation activity and instantaneous activity

F

Correlation between fixation activity and instantaneous activity

C

Activity pattern of the local network evolving over time, illustrated as the sum of $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activities evolving over time

G

D

F: fixation input
E: expectation input
D: delay and sustained visual input
V: transient visual input

Mean input to local network 1 (I_1) and mean input to local network 2 (I_2)

Time period: (1)   (2)   (3)   (4)

H

Time period: (1)   (2)   (3)   (4)

Figure 4. Two Multi-neuronal Activity Patterns Explain LIP Dynamics

One global network composed of LNs 1 and 2 is simulated, and the dynamics in LN1 on distractor trials (A–D) and target trials (E–H) are analyzed.

(A and E) $\overrightarrow{S1}$ and $\overrightarrow{D1}$ patterns dominate activity. Population average activity (red/blue), its component in the space of $\overrightarrow{S1}$ and $\overrightarrow{D1}$ (orange/purple), and its component in the space of all other patterns (black) on distractor/target (A/E) trials (n = 100 cells). In (A), the orange and black traces add up to the red trace, and in (E), the purple and black traces add up to the blue trace.

(B and F) Actual correlation (red/blue), $Corr_{sum,diff}$ (orange/purple, the component of correlation due to the $\overrightarrow{S1}$ and $\overrightarrow{D1}$ patterns alone), and $Corr_{residual}$ (black, the residual component) on distractor/target (B/F) trials. In (B), the orange and black traces add up to the red trace, and in (F), the purple and black traces add up to the blue trace. $Corr_{sum,diff}$ mirrors the salient ups and downs in the actual correlation, while $Corr_{residual}$ largely does not change with time—thus, the changes in actual correlation over a trial are largely due to the $\overrightarrow{S1}$ and $\overrightarrow{D1}$ patterns. See Results for how the correlation was broken down into two components. Note that only the actual correlation, but not $Corr_{sum,diff}$ or $Corr_{residual}$, is restricted to lie within −1 and 1.

(C) Top left inset: the two-dimensional space spanned by the two 100-dimensional dominant activity patterns of LN1, $\overrightarrow{S1}$ (dashed vector) and $\overrightarrow{D1}$ (dotted vector), with $\overrightarrow{a1}$ and $\overrightarrow{d1}$ as axes.

(C and G) The evolution of $\overrightarrow{S1}$ (dashed vectors) and $\overrightarrow{D1}$ (dotted vectors) activity during distractor (C; orange vectors) and target (G; purple vectors) trials, as a result of the evolving inputs illustrated in (D) and (H). For each trial type, activities in the $\overrightarrow{S1}$ and $\overrightarrow{D1}$ directions are each averaged over each of four time periods (spanned by black bars in A, B, E, and F), and are illustrated in their two-dimensional space, with the angle between $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activities and their relative lengths accurately rendered. In this 2D space, at a given time, the activity pattern across the cells of LN1 is the vector sum of $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activities at that time. Thus, $\overrightarrow{F}$, the vector of target trial fixation activities, is the vector sum of the $\overrightarrow{S1}$ and $\overrightarrow{D1}$ vectors at time (1) in (G). The angle between $\overrightarrow{F}$ and the vector sum of $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activities at a given time period generally determines the actual correlation at that time: the larger the angle, the lower the correlation,

and vice versa (see Figure S8 for the precise relationship between the vectors and correlation). For example, the angle between the vector sum during the delay on distractor trials (vector sum of the $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activities at time (2) in C) and $\overrightarrow{F}$ is large, so the correlation during that time period is low; the vector sum following distractor onset on distractor trials (vector sum of $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activity at time (3) in C) points in similar directions as $\overrightarrow{F}$, so the correlation during that time period is high.

(D and H) The relative inputs to LNs 1 and 2 during the four time periods on distractor (D) and target (H) trials. Black and white bars denote the mean input to LN1 (I_1) and mean input to LN2 (I_2), respectively. As illustrated in Figures 3D and 3E, during each time period, the sum of (difference between) the black and white bars largely determines the magnitude and direction of $\overrightarrow{S1}$ ($\overrightarrow{D1}$) activity, which is plotted directly above the bars in (C) and (G) (see SI section 9 for explanations of why the inputs here do not perfectly predict the activations in C and G).

---

components, the component due to activity in the $\overrightarrow{S1}$ and $\overrightarrow{D1}$ patterns alone and the residual component, as follows. At any given time point, the correlation between instantaneous activity and fixation activity is $\hat{r} \cdot \hat{F}/(|\hat{r}||\hat{F}|)$, where $\hat{r}$ and $\hat{F}$ are the vectors of mean-subtracted instantaneous activities and mean-sub-

tracted fixation activities, respectively (each element of $\hat{r}$ is the instantaneous activity of one neuron minus the population mean instantaneous activity, and similarly for $\hat{F}$), and $|\cdot|$ denotes vector norm. We break $\hat{r}$ into components $\hat{r}_{sum}$, $\hat{r}_{diff}$, and $\hat{r}_{weak}$, the mean-subtracted instantaneous activity in the $\overrightarrow{S1}$ pattern,

the $\overrightarrow{D1}$ pattern, and all other patterns, respectively, and do likewise for $\widehat{F}$:

$$\frac{\widehat{r} \cdot \widehat{F}}{|\widehat{r}||\widehat{F}|} = \frac{(\widehat{r}_{sum} + \widehat{r}_{diff} + \widehat{r}_{weak}) \cdot (\widehat{F}_{sum} + \widehat{F}_{diff} + \widehat{F}_{weak})}{|\widehat{r}||\widehat{F}|}$$

$$= \frac{(\widehat{r}_{sum} + \widehat{r}_{diff}) \cdot (\widehat{F}_{sum} + \widehat{F}_{diff})}{|\widehat{r}||\widehat{F}|}$$

$$+ \frac{(\widehat{r}_{sum} + \widehat{r}_{diff}) \cdot \widehat{F}_{weak} + \widehat{r}_{weak} \cdot (\widehat{F}_{sum} + \widehat{F}_{diff}) + \widehat{r}_{weak} \cdot \widehat{F}_{weak}}{|\widehat{r}||\widehat{F}|}$$

$$= Corr_{sum,diff} + Corr_{residual}.$$

The two terms $Corr_{sum,diff}$ and $Corr_{residual}$ that sum to the actual correlation are plotted in Figures 4B and 4F—we see that $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activities largely explain the qualitative changes in correlations over time.

Thus, the actual activity pattern across cells of the LN, $\overrightarrow{r}$, can be approximated as the vector sum of $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activities, which determines the correlation patterns. Figures 4C, 4G, and S8 illustrate $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activities evolving over four time periods during a trial, as well as how their dynamics explain the correlation patterns. In SI section 10, we discuss why correlations at the peaks of visual responses are higher in FK than in BG. In Figures S9 and S10 and SI section 11, we show how the model dynamics change smoothly from the BG to the FK case for varying strengths of surround suppression.

Now we turn to examine how the dynamics of $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activities are determined by their inputs. We have shown above that the external input to the sum or difference pattern is approximately $N(I_1 P_1 + I_2 P_2)$. We note that the absolute values of P_1 and P_2 for the sum and difference patterns are all about equal (Figure 3B), which we can call m. That is, for the sum pattern, $P_1 \approx P_2 \approx m$, while for the difference pattern, $P_1 \approx m$ and $P_2 \approx -m$. Thus, the inputs to the sum and difference patterns are approximately $Nm(I_1+I_2)$ and $Nm(I_1-I_2)$, respectively. When the network is in a steady state, these inputs are amplified by the connectivity: $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activations are approximately given by $1/\sqrt{2}\ Nm(I_1+I_2)/(1-\lambda_S)$ and $1/\sqrt{2}\ Nm(I_1-I_2)/(1-\lambda_D)$ respectively, where λ_S and λ_D are the eigenvalues of the sum and difference patterns, respectively (for this approximation, we ignore the feedforward input from the other patterns, which is weak relative to the mean-driven input; the $1/\sqrt{2}$ factor arises because $\overrightarrow{S1}$ is only the LN1 half of the sum pattern, and similarly for $\overrightarrow{D1}$). As N, m, λ_S, and λ_D are all fixed properties of the network, $\overrightarrow{S1}$ and $\overrightarrow{D1}$ activities just depend on the dynamics of the mean inputs I_1 and I_2, being simply proportional to I_1+I_2 and I_1-I_2, respectively (Figures 4C, 4D, 4G, and 4H).

Direct Evidence for Two-Dimensional Dynamics in the FK Dataset

Since we propose that the BG data are predominantly one-dimensional and the FK data two-dimensional, we used principal component analysis (PCA) to directly examine the dimensionality of the two datasets. We focus on distractor trials because our correlation analysis revealed that they show the most salient dynamical differences between BG and FK. Furthermore, our

model predicts that the activity on FK distractor trials has both large $\overrightarrow{a1}$ and $\overrightarrow{d1}$ activations, and thus is likely to reveal two dynamical dimensions, whereas the activity on FK target trials is more strongly one-dimensional, dominated by the single $\overrightarrow{a1}$ direction (Figures 4C and 4G). We excluded the transient visual responses to the distractor, as they involve activation of weak patterns, and performed PCA on the remaining slowly varying activity patterns of distractor trials. The results indeed confirm the one-dimensionality of the BG data and two-dimensionality of the FK data (Figures 5A and 5B).

Given the 2D space spanned by the top two principal components (PCs) identified from the FK data, we ask further, do activity patterns in this dominant 2D space actually behave as our model predicts? To answer this question, we first estimate the activity directions in the data that correspond to those in our model. We take the direction with the maximum mean firing rate within the 2D space of the two PCs as the putative $\overrightarrow{a1}$ direction, since $\overrightarrow{a1}$ is a direction representing concerted firing of neurons in an LN, and take the direction orthogonal to the putative $\overrightarrow{a1}$ as the putative $\overrightarrow{d1}$ (Figure 5C). In Figures 5D–5G, we plot the activities over time in the $\overrightarrow{a1}$ and $\overrightarrow{d1}$ directions in the data and model (in SI section 12, we discuss differences between data and model). The activities in the putative $\overrightarrow{a1}$ and $\overrightarrow{d1}$ directions in the data match those predicted by the model, providing direct evidence that our proposed two-dimensional dynamics underlie the FK data.

In SI section 13, we discuss the dynamics and dimensionality of E and I subpopulations.

Two-Dimensional Dynamics Suggest a Recurrent Origin for LIP Surround Suppression

Surround suppression is observed in multiple cortical areas (reviewed in Rubin et al., 2015) and has been extensively studied as a model for understanding cortical computations and circuit mechanisms (e.g., in V1; Ozeki et al., 2009; Rubin et al., 2015; and see review by Nurminen and Angelucci, 2014). When considering surround suppression in a given cortical area, a key mechanistic question is the following: to what extent is it inherited from surround suppression in other areas, i.e., resulting from a withdrawal of input from those areas, and to what extent is it due to reciprocal, suppressive coupling within the given area?

Of areas that directly or indirectly project to LIP (Blatt et al., 1990; Clower et al., 2001), surround suppression has been observed in MT (Hunter and Born, 2011; Tsui and Pack, 2011), V4 (Desimone and Schein, 1987; Sundberg et al., 2009), superior colliculus (Dorris et al., 2007), FEF (Schall and Hanes, 1993; Cavanaugh et al., 2012), and dlPFC (Suzuki and Gottlieb, 2013). Thus, it is possible that LIP surround suppression is inherited from one or more of these areas. However, according to our model, the observed pattern of correlation between fixation and instantaneous activity arises from the coupling of local LIP networks. We argue that the experimentally observed correlation pattern is a signature indicating that surround suppression of external input alone cannot account for LIP surround suppression.

This can be demonstrated by simulating the scenario of the null hypothesis—LIP surround suppression being inherited from external inputs. In this version of the model, the two LNs
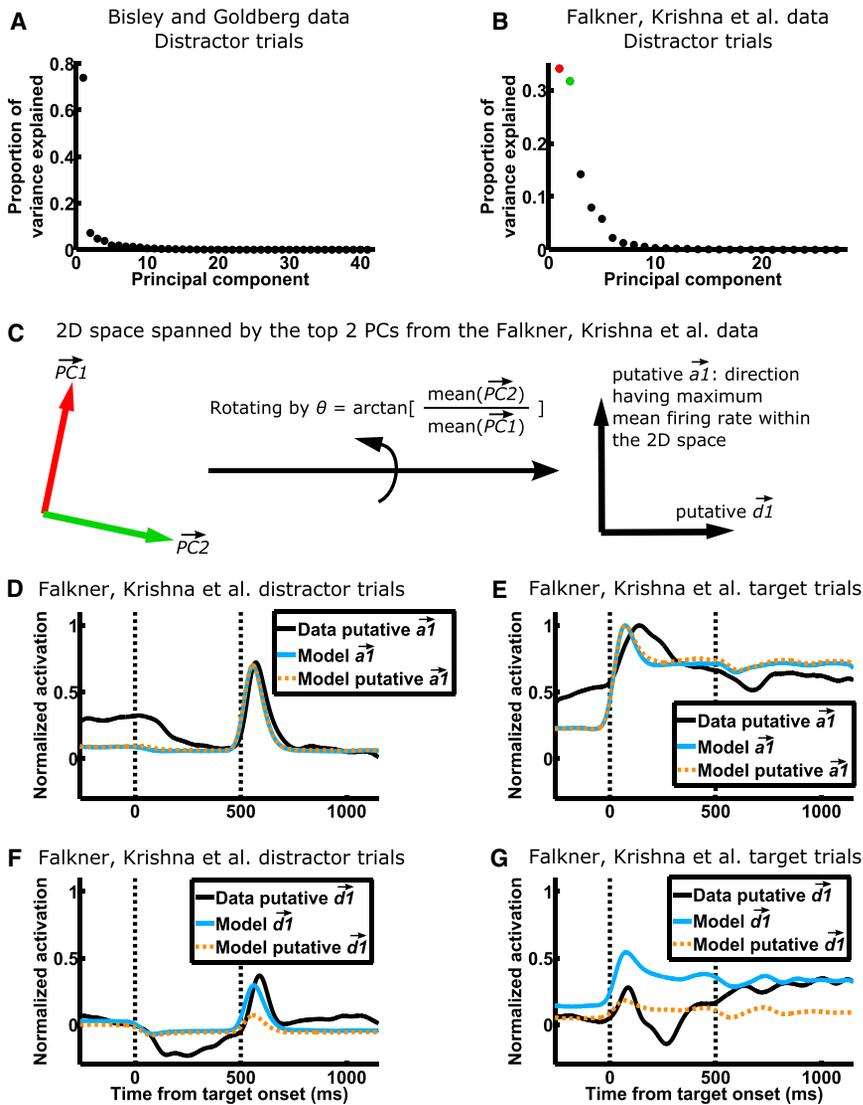
A Bisley and Goldberg data
Distractor trials

B Falkner, Krishna et al. data
Distractor trials

Proportion of variance explained (A: 0.8, 0.6, 0.4, 0.2, 0)
Principal component (0, 10, 20, 30, 40)

Proportion of variance explained (B: 0.3, 0.2, 0.1, 0)
Principal component (0, 10, 20)

C 2D space spanned by the top 2 PCs from the Falkner, Krishna et al. data

$\vec{PC1}$

Rotating by $\theta = \arctan[\frac{\mathrm{mean}(\vec{PC2})}{\mathrm{mean}(\vec{PC1})}]$

$\vec{PC2}$

putative $\vec{a1}$: direction having maximum mean firing rate within the 2D space

putative $\vec{d1}$

D Falkner, Krishna et al. distractor trials

Normalized activation (1, 0.5, 0)
— Data putative $\vec{a1}$
— Model $\vec{a1}$
⋯ Model putative $\vec{a1}$

E Falkner, Krishna et al. target trials

Normalized activation (1, 0.5, 0)
— Data putative $\vec{a1}$
— Model $\vec{a1}$
⋯ Model putative $\vec{a1}$

F Falkner, Krishna et al. distractor trials

Normalized activation (1, 0.5, 0)
— Data putative $\vec{d1}$
— Model $\vec{d1}$
⋯ Model putative $\vec{d1}$
Time from target onset (ms) (0, 500, 1000)

G Falkner, Krishna et al. target trials

Normalized activation (1, 0.5, 0)
— Data putative $\vec{d1}$
— Model $\vec{d1}$
⋯ Model putative $\vec{d1}$
Time from target onset (ms) (0, 500, 1000)

Figure 5. Direct Evidence for Two-Dimensional Dynamics in the Falkner, Krishna et al. Data

(A and B) PCA, where the variables are neurons and the observations are the instantaneous activity vectors during distractor trials, for the BG (A) and FK (B) datasets. Activity vectors during the transient visual responses to the distractor (600–1,100 ms after target onset for BG; 450–750 ms after target onset for FK) were not included for this analysis because they involve activation of weak patterns. The majority of the variance is explained by one PC in BG, while a comparable proportion is explained by two PCs in FK, consistent with one-dimensional dynamics in BG and two-dimensional dynamics in FK.

(C) We hypothesize that the 2D space spanned by the top two PCs (colored as in B) in the FK data is the 2D space of $\vec{S1}$ and $\vec{D1}$ (Figures 3C, 4C, and 4G). We further hypothesize that the direction with the maximum mean firing rate within the 2D space of the two PCs is the putative $\vec{a1}$ direction, since $\vec{a1}$ represents concerted firing of neurons in an LN. We can thus find the putative $\vec{a1}$ and $\vec{d1}$ of the FK data by rotating the two PCs by an angle of arctan[mean ($\vec{PC1}$) /mean ($\vec{PC2}$)], where mean(·) denotes mean over the elements of a vector.

(D–G) The activation of $\vec{a1}$ (D and E) and $\vec{d1}$ (F and G) on FK distractor trials (D and F) and target trials (E and G). In D and E, the data putative $\vec{a1}$ was derived as in (C). To determine activation in the model, one cell was "recorded" from each of multiple simulated global networks to form the model population. To determine the model $\vec{a1}$, suppose the ith cell of the model population is the jth cell from LN1 of the ith global network. Then the ith element of the model $\vec{a1}$ is the jth element of the actual $\vec{a1}$ of the ith global network. The model putative $\vec{a1}$ was derived as in (C), but from the model population. The $\vec{d1}$ directions are determined similarly. Each set of activations (e.g., the four activation traces of data putative $\vec{a1}$ and $\vec{d1}$ on target and distractor trials comprise a set) is normalized by its peak $\vec{a1}$ activation on target trials—thus, (D)–(G) share the same scale. Vertical dashed lines denote the onsets of the target and distractor.

---

are uncoupled. Whenever a stimulus appears or a saccade is planned, the LN with the corresponding RF is activated by visual or delay input; at the same time, the external input to the other LN is reduced, modeling surround suppression inherent in one or more input sources (see SI section 2.2 for model details). Figure 6A shows the population average PSTHs from a simulation of the FK experiment using this model. On the surface, if we examine only the firing rates, this model of surround suppression reproduces the experimental data. However, if we examine the underlying network dynamics using the correlation analysis (Figure 6B), we find that this model cannot reproduce the experimentally observed correlation pattern. Specifically, the dynamics of each LN here are dominated by its slow mode, as in BG (Figures 1E and 2E).

There are alternative mechanisms that conceivably could account for the FK correlations. We consider these in SI section 14 but conclude, for multiple reasons, that the most likely and parsimonious interpretation is that surround suppression in LIP arises primarily from direct suppressive coupling between LIP LNs.

DISCUSSION

By demonstrating that recurrent interactions between LIP LNs are likely to drive LIP surround suppression, our study suggests the active involvement of LIP in attentional and saccadic selection. LIP is part of a fronto-parietal-collicular network that mediates attentional guidance and eye movements, and the attentional and saccadic priorities associated with different locations (a "priority map") are encoded in the activity of neurons in this network with the corresponding RF locations (Bisley and Goldberg, 2010). It has long been theorized that different
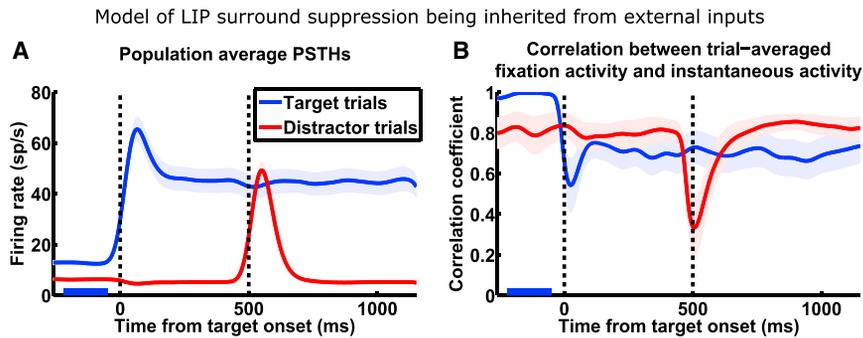
Model of LIP surround suppression being inherited from external inputs

A **Population average PSTHs**

B Correlation between trial−averaged
fixation activity and instantaneous activity

Figure 6. Model of Inherited Surround Suppression Cannot Reproduce Observed Network Dynamics

The model: two LIP LNs are uncoupled; whenever one LN receives visual or delay input, the external input to the other LN is reduced.

(A) Population average PSTHs (n = 27 cells; same conventions as Figure 1D) show that this model reproduces activity observed by FK (Figure 1D) during surround interactions.

(B) This model fails to reproduce network dynamics observed by FK (Figure 1F) during surround interactions. Correlation analysis with same conventions as Figure 1F.

locations on this priority map mutually suppress each other to facilitate attentional and saccadic selection, to allow persistent focus by resisting distraction, and to allow the planning and execution of sequential saccades (Itti and Koch, 2001; Constantinidis and Wang, 2004; Xing and Andersen, 2000). However, the neural substrates and mechanisms of these processes are not clear. Our results suggest that LIP directly participates in these processes and shapes the priority map, instead of merely reflecting computations achieved in other areas.

Ganguli et al. (2008) showed that the slow dynamics of a local LIP network (i.e., neurons sharing the same RF) are 1D, but we can now see that this result is restricted to the case in which a single LN is activated without activation of other LNs to which it is coupled. Nonetheless, we can simply understand the more complex, 2D slow dynamics we have found in the FK data as resulting from coupling between networks that, in isolation, each have 1D slow dynamics. This lays a basis for understanding higher-dimensional local dynamics induced by interactions between larger numbers of simultaneously activated LNs. For example, during visual search, it takes longer to find a target when there are more distractors, and, correspondingly, LIP activity is lower when there are more distractors (Balan et al., 2008). This likely results from increased surround suppression by larger numbers of activated LNs, which would yield correlations corresponding to a higher-dimensional dominant activity space (e.g., number of dimensions equal to the number of mutually interacting networks). Our study thus provides a basis for analyzing activity dynamics when multiple stimuli evoke interaction of multiple local LIP networks, as occurs in natural visual environments.

Our results have implications for the mechanisms underlying certain types of perceptual decision making, where saccadic decisions are made based on noisy sensory evidence (Roitman and Shadlen, 2002; Gold and Shadlen, 2007). This type of decision making has been posited to involve two neuronal pools that integrate opposing sensory evidence, each either accumulating evidence independently and racing toward a decision (Mazurek et al., 2003) or competing with the other by mutual inhibition (Wong and Wang, 2006; Usher and McClelland, 2001). Recent results show that these neuronal pools are not exclusively in LIP (Katz et al., 2016), but if they are distributed across multiple areas that include LIP, which of the two classes of models applies to each instance of decision making would depend on whether the two neuronal pools are recurrently coupled. When

they are not coupled (like the neuronal pools studied by BG), the independent accumulator model would apply, and when they are coupled, the mechanisms described here would contribute to the competition that leads to decision making. In the future, it would be interesting to study decision making separately for cases in which the two alternative choices engage LIP LNs that do or do not suppress one another (as BG and FK have done in studying attentional switching), examine the neural correlates on the population level, and compare them with behavior.

The interactions between LNs we describe would cause priority assignments to be in part determined in relative terms, as has been observed in certain forms of value-based decision making. When different saccade targets are associated with different magnitudes of reward or reward probabilities, some LIP neurons encode the expected value of different saccades (Platt and Glimcher, 1999; Dorris and Glimcher, 2004). Importantly, these value representations in LIP are relative, such that the response to one saccade target depends on its value relative to those of other possible saccade targets; this relative value encoding is well described by the phenomenological model of divisive normalization (Louie et al., 2011; Carandini and Heeger, 2011). Surround suppression, computed within LIP in ways similar to those described here, provides a circuit mechanism for divisive normalization of value representations (Louie et al., 2014; LoFaro et al., 2014; Rubin et al., 2015).

Regardless of the cognitive context in which LIP function has been investigated, research has often focused on single-neuron activity or the average activity of LIP populations. Our work adds to other recent work (e.g., see review by Cunningham and Yu, 2014) in suggesting that there is much information in the activity patterns across neurons, which change as a function of external stimuli and internal goals such as saccade plans. As we have seen, even when the mean activity of a population changes only subtly, the pattern of activity across neurons can change drastically (e.g., when a target appears in the suppressive surround of a local LIP population). Thus, beyond the information carried by single neurons or their average activity, downstream areas could potentially read out information from the activity pattern across LIP neurons, although whether or how downstream areas do this remains to be examined. This is potentially important in the natural context, where LIP LNs must interact to process a multitude of changing visual stimuli and internal goals to guide behavior.

REFERENCES

Baizer, J.S., Ungerleider, L.G., and Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. J. Neurosci. *11*, 168–190.

Balan, P.F., Oristaglio, J., Schneider, D.M., and Gottlieb, J. (2008). Neuronal correlates of the set-size effect in monkey lateral intraparietal area. PLoS Biol. *6*, e158.

Bisley, J.W., and Goldberg, M.E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. Science *299*, 81–86.

Bisley, J.W., and Goldberg, M.E. (2006). Neural correlates of attention and distractibility in the lateral intraparietal area. J. Neurophysiol. *95*, 1696–1717.

Bisley, J.W., and Goldberg, M.E. (2010). Attention, intention, and priority in the parietal lobe. Annu. Rev. Neurosci. *33*, 1–21.

Blatt, G.J., Andersen, R.A., and Stoner, G.R. (1990). Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. J. Comp. Neurol. *299*, 421–445.

Carandini, M., and Heeger, D.J. (2011). Normalization as a canonical neural computation. Nat. Rev. Neurosci. *13*, 51–62.

Cavanaugh, J., Joiner, W.M., and Wurtz, R.H. (2012). Suppressive surrounds of receptive fields in monkey frontal eye field. J. Neurosci. *32*, 12284–12293.

Clower, D.M., West, R.A., Lynch, J.C., and Strick, P.L. (2001). The inferior parietal lobule is the target of output from the superior colliculus, hippocampus, and cerebellum. J. Neurosci. *21*, 6283–6291.

Coe, B., Tomihara, K., Matsuzawa, M., and Hikosaka, O. (2002). Visual and anticipatory bias in three cortical eye fields of the monkey during an adaptive decision-making task. J. Neurosci. *22*, 5081–5090.

Constantinidis, C., and Wang, X.-J. (2004). A neural circuit basis for spatial working memory. Neuroscientist *10*, 553–565.

Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. *17*, 1500–1509.

Dayan, P., and Abbott, L.F. (2005). Theoretical Neuroscience (The MIT Press).

Desimone, R., and Schein, S.J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. J. Neurophysiol. *57*, 835–868.

Dorris, M.C., and Glimcher, P.W. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron *44*, 365–378.

Dorris, M.C., Olivier, E., and Munoz, D.P. (2007). Competitive integration of visual and preparatory signals in the superior colliculus during saccadic programming. J. Neurosci. *27*, 5053–5062.

Elston, G.N., and Rosa, M.G. (1997). The occipitoparietal pathway of the macaque monkey: comparison of pyramidal cell morphology in layer III of functionally related cortical visual areas. Cereb. Cortex *7*, 432–452.

Falkner, A.L., Krishna, B.S., and Goldberg, M.E. (2010). Surround suppression sharpens the priority map in the lateral intraparietal area. J. Neurosci. *30*, 12787–12797.

Fitzgerald, J.K., Freedman, D.J., Fanini, A., Bennur, S., Gold, J.I., and Assad, J.A. (2013). Biased associative representations in parietal cortex. Neuron *77*, 180–191.

Freedman, D.J., and Assad, J.A. (2011). A proposed common neural mechanism for categorization and perceptual decisions. Nat. Neurosci. *14*, 143–146.

Ganguli, S., Bisley, J.W., Roitman, J.D., Shadlen, M.N., Goldberg, M.E., and Miller, K.D. (2008). One-dimensional dynamics of attention and decision making in LIP. Neuron *58*, 15–25.

Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. Annu. Rev. Neurosci. *30*, 535–574.

Goldman, M.S. (2009). Memory without feedback in a neural network. Neuron *61*, 621–634.

Hunter, J.N., and Born, R.T. (2011). Stimulus-dependent modulation of suppressive influences in MT. J. Neurosci. *31*, 678–686.

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. Nat. Rev. Neurosci. *2*, 194–203.

Kable, J.W., and Glimcher, P.W. (2009). The neurobiology of decision: consensus and controversy. Neuron *63*, 733–745.

Katz, L.N., Yates, J.L., Pillow, J.W., and Huk, A.C. (2016). Dissociated functional significance of decision-related activity in the primate dorsal stream. Nature *535*, 285–288.

Kennerley, S.W., and Wallis, J.D. (2009). Reward-dependent modulation of working memory in lateral prefrontal cortex. J. Neurosci. *29*, 3259–3270.

Leon, M.I., and Shadlen, M.N. (1999). Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. Neuron *24*, 415–425.

Lewis, J.W., and Van Essen, D.C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. J. Comp. Neurol. *428*, 112–137.

LoFaro, T., Louie, K., Webb, R., and Glimcher, P.W. (2014). The temporal dynamics of cortical normalization models of decision-making. Lett. Biomath. *1*, 209–220.

Louie, K., Grattan, L.E., and Glimcher, P.W. (2011). Reward value-based gain control: divisive normalization in parietal cortex. J. Neurosci. *31*, 10627–10639.

Louie, K., LoFaro, T., Webb, R., and Glimcher, P.W. (2014). Dynamic divisive normalization predicts time-varying value coding in decision-related circuits. J. Neurosci. *34*, 16046–16057.

Mazurek, M.E., Roitman, J.D., Ditterich, J., and Shadlen, M.N. (2003). A role for neural integrators in perceptual decision making. Cereb. Cortex *13*, 1257–1269.

Miller, E.K., and Wilson, M.A. (2008). All my circuits: using multiple electrodes to understand functioning neural networks. Neuron *60*, 483–488.

Murphy, B.K., and Miller, K.D. (2009). Balanced amplification: a new mechanism of selective amplification of neural activity patterns. Neuron *61*, 635–648.

Nurminen, L., and Angelucci, A. (2014). Multiple components of surround modulation in primary visual cortex: multiple neural circuits with multiple functions? Vision Res. *104*, 47–56.

Ozeki, H., Finn, I.M., Schaffer, E.S., Miller, K.D., and Ferster, D. (2009). Inhibitory stabilization of the cortical network underlies visual surround suppression. Neuron *62*, 578–592.

Patel, G.H., Shulman, G.L., Baker, J.T., Akbudak, E., Snyder, A.Z., Snyder, L.H., and Corbetta, M. (2010). Topographic organization of macaque area LIP. Proc. Natl. Acad. Sci. USA *107*, 4728–4733.

Platt, M.L., and Glimcher, P.W. (1999). Neural correlates of decision variables in parietal cortex. Nature *400*, 233–238.

Roesch, M.R., and Olson, C.R. (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. J. Neurophysiol. *90*, 1766–1789.

Roitman, J.D., and Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J. Neurosci. *22*, 9475–9489.

Rubin, D.B., Van Hooser, S.D., and Miller, K.D. (2015). The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. Neuron *85*, 402–417.

Schall, J.D., and Hanes, D.P. (1993). Neural basis of saccade target selection in frontal eye field during visual search. Nature *366*, 467–469.

Shenoy, K.V., Sahani, M., and Churchland, M.M. (2013). Cortical control of arm movements: a dynamical systems perspective. Annu. Rev. Neurosci. *36*, 337–359.

Stanton, G.B., Bruce, C.J., and Goldberg, M.E. (1995). Topography of projections to posterior cortical areas from the macaque frontal eye fields. J. Comp. Neurol. *353*, 291–305.

Sundberg, K.A., Mitchell, J.F., and Reynolds, J.H. (2009). Spatial attention modulates center-surround interactions in macaque visual area v4. Neuron *61*, 952–963.

Suzuki, M., and Gottlieb, J. (2013). Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. Nat. Neurosci. *16*, 98–104.

Tsui, J.M.G., and Pack, C.C. (2011). Contrast sensitivity of MT receptive field centers and surrounds. J. Neurophysiol. *106*, 1888–1900.

Usher, M., and McClelland, J.L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. Psychol. Rev. *108*, 550–592.

Wong, K.-F., and Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. J. Neurosci. *26*, 1314–1328.

Xing, J., and Andersen, R.A. (2000). Memory activity of LIP neurons for sequential eye movements simulated with neural networks. J. Neurophysiol. *84*, 651–665.