Top-down Control of Inhibition Reshapes Neural Dynamics Giving Rise to a Diversity of Computations

Zhen Chen 1, Krishnan Padmanabhan 2

1. Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627;

2. Department of Neuroscience, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642.

* Krishnan Padmanabhan

Email: Krishnan_Padmanabhan@urmc.rochester.edu

Zhen Chen: ORCID iD: https://orcid.org/0000-0002-5590-3552

Krishnan Padmanabhan: ORCID iD: https://orcid.org/0000-0002-3255-8346

Abstract: 174 words

Significance Statement: 118 words

Introduction: 461 words

Methods: 1081 words

Results: 3669 words

Discussion: 1106 words

Supplemental Materials:
Classification

Major Classification: BIOLOGICAL SCIENCES

Minor Classification: Neuroscience, Applied Mathematics

Keywords

Paste the keywords here. There should be at least three and no more than five.

Author Contributions

K.P. conceived and supervised the project. Z.C. performed all the experiments and analysis. Z.C. and K.P. made the figures and wrote the manuscript.

This PDF file includes:

Main Text

Figures 1 to 5
Abstract

Growing evidence shows that top-down projections from excitatory neurons in higher brain areas selectively synapse onto local inhibitory interneurons in sensory systems. While this connectivity is conserved across sensory modalities, the role of this feedback in shaping the dynamics of local circuits, and the resultant computational benefits it provides remains poorly understood. Using rate models of neuronal firing in a network consisting of excitatory, inhibitory and top-down populations, we found that changes in the weight of feedback to inhibitory neurons generated diverse network dynamics and complex transitions between these dynamics. Additionally, modulation of the weight of top-down feedback supported a number of computations, including both pattern separation and oscillatory synchrony. A bifurcation analysis of the network identified a new mechanism by which gamma oscillations could be generated in a model of neural circuits, which we termed Top-down control of Inhibitory Neuron Gamma (TING). We identified the unique roles that top-down feedback of inhibition plays in shaping network dynamics and computation, and the ways in which these dynamics can be deployed to process sensory inputs.
Significance Statement

The functional role of feedback projections, connecting excitatory neurons in higher brain areas to inhibitory neurons in primary sensory regions, remains a fundamental open question in neuroscience. Growing evidence suggests that this architecture is recapitulated across a diverse array of sensory systems, ranging from vision to olfaction. Using a rate model of top-down feedback onto inhibition, we found that changes in the weight of feedback support both pattern separation and oscillatory synchrony, including a mechanism by which top-down inputs could entrain gamma oscillations within local networks. These dual functions were accomplished via a codimension-2 bifurcation in the dynamical system. Our results highlight a key role for this top-down feedback, gating inhibition to facilitate often diametrically different local computations.
Main Text

Introduction

While inhibitory interneurons constitute a comparatively small fraction (20%) of neurons in the neocortex (Meinecke and Peters 1987), they encompass a diverse array of types, distinct in gene expression, morphology, and connectivity (Kepecs and Fishell 2014). Thought to be largely local in their influence (Yoshimura and Callaway 2005, Fino and Yuste 2011), the interneurons nonetheless play critical roles in shaping the dynamics of neural circuits (Fishell and Kepecs 2019) across all sensory and motor systems (Urban and Sakmann 2002, Okun and Lampl 2008, Atallah and Scanziani 2009, Cardin, Carlén et al. 2009, Poo and Isaacson 2009, Sohal, Zhang et al. 2009) (Silberberg and Markram 2007, Packer and Yuste 2011).

Interestingly, recent anatomical evidence has identified top-down centrifugal afferents from higher cortical areas that specifically target inhibitory interneurons. In visual system of the mouse for example, axons from the cingulate (cG) of the frontal cortex project to GABAergic inhibitory neurons in V1, and activation of cG neurons improved performance on a visual discrimination task (Zhang, Xu et al. 2014). Similarly, in the olfactory system, axons from excitatory neurons in the piriform cortex synapse onto the local inhibitory granule cells in olfactory bulbs, whereby they can modulate the function of the mitral and tufted cells, the principal relays of olfactory information from the bulb to the brain (Shipley and Adamek 1984, Padmanabhan, Osakada et al. 2016, Padmanabhan, Osakada et al. 2019). These top-down higher area neurons receive input from only
excitatory populations in sensory areas, but they exert their influence on the local
circuit dynamics via synapses onto inhibitory populations. Consequently, whereas
the information relayed to these top-down cells comes from the local excitatory
population, they intervene in network dynamics through the inhibitory population.
Although this architecture is recapitulated across multiple sensory systems, the
computational role of this top-down control of inhibition and the dynamic
mechanisms that give rise to such computations remain largely unknown.

To address this question, we built a three-node network model consisting
of an excitatory (E), inhibitory (I) and top-down (T) population, and studied how
firing rate dynamics were influenced by the structure of the top-down connection
weight onto inhibition. Modulating the weight of the top-down connections to local
inhibitory neurons reshaped the dynamics of the local excitatory-inhibitory (E-I)
circuit in a way that enhanced sensory discrimination as well as generated
oscillatory synchrony including entraining gamma oscillations in the local circuit
(Top-down control of Inhibitory Neuron Gamma, (TING)). Finally, the mechanism
underlying the dynamics, as well as the functional roles played by top-down control
of inhibition occurred via a codimension-2 bifurcation in the dynamical system. By
gating the magnitude of top-down feedback, a number of seemingly disparate
computations could be supported by a single circuit, providing an additional
framework for the diversity of inhibitory interneuron function.
**Materials and Methods**

**Network Model:**

The network model was composed of three nodes, a local excitatory (E) and inhibitory (I) nodes which were reciprocally coupled, and a top-down node (T) that received input from the local E node and projected back to the inhibitory I node (Fig.1A). In the model, \( r_i(t), i = 1,2,3 \) represented the firing rates of the three neuron populations respectively, whose dynamics were determined by Wilson-Cowan equations (Wilson and Cowan 1972) as follows:

\[
\begin{align*}
\tau_1 \dot{r}_1 &= -r_1 + S(w_{11}r_1 + w_{12}r_2 + \mu) \\
\tau_2 \dot{r}_2 &= -r_2 + S(w_{21}r_1 + w_{22}r_2 + w_{23}r_3) \\
\tau_3 \dot{r}_3 &= -r_3 + \tanh(w_{31}r_1 + w_{33}r_3)
\end{align*}
\]

(1)

where \( S \) is the sigmoid function:

\[
S(x) = \frac{1}{1+e^{-x}}
\]

(2)

which described the nonlinear relationship between the mean synaptic input and average firing rate (normalized to a range between 0 and 1). The parameter \( \tau_i, i = 1,2,3 \) was the time constant for each population, characterizing how quickly the dynamics of each population evolved. The excitatory neuron population (E) received an external stimulus \( \mu \), that represented the only external input to the system (Fig. S1). The connection weight from population \( j \) to population \( i \) was denoted by \( w_{ij}, i,j = 1,2,3 \), among which \( w_{11}, w_{21}, w_{31}, w_{23}, w_{33} > 0 \) and \( w_{12}, w_{22} < 0 \). The connection weight \( w_{ij}, i,j = 1,2,3 \) represented the average synaptic input...
received by the neuron population i from the population j. Throughout this paper, we set the parameters as follows: \( w_{11} = 8.7, w_{12} = -10, w_{21} = 7.0, w_{22} = -13, w_{31} = 1.5, w_{33} = 0.5 \). As there was no inhibitory synaptic input into the top-down population (T), a hyperbolic tangent function \( \tanh(x) \) was used for \( r_3(t) \) to cover the full range of firing rates \([0,1]\) (Fig.1C).

**Metric definition**

As representations of the network (\( \omega \)-limit sets) could take on different forms, an equilibrium or a limit cycle, we defined two quantitative metrics: \( d^E(\Omega_1, \Omega_2) \) and \( d^S(r_3 |_{\Omega_1}, r_3 |_{\Omega_2}) \) which served to measure the distance between different types of \( \omega \)-limit sets in responses to any given stimulus pair \((\mu_1, \mu_2)\) where \( \mu_2 = \mu_1 + \Delta \).

\( d^E(\Omega_1, \Omega_2) \) denoted the average Euclidean distance between \( \Omega_1 \) and \( \Omega_2 \) in the three-dimensional (3D) phase space of firing rates \((r_1, r_2, r_3)\), and the spectrum distance \( d^S(r_3 |_{\Omega_1}, r_3 |_{\Omega_2}) \) was a sum of the squared differences between both direct components (DC) and alternating components (AC) in the amplitude-frequency domain of the Fourier transforms to the signals \( r_3(t) \) associated with the two stimuli (Fig. S2-S3, Supplementary material). When both \( \Omega_1 \) and \( \Omega_2 \) were equilibria, the spectrum distance \( d^S \) was a projection of the Euclidean distance \( d^E \) onto the \( r_3 \) axis. The spectrum distance was however more sensitive to bifurcations when one of the two \( \omega \)-limit sets \( \Omega_1 \) and \( \Omega_2 \) transitioned into a limit cycle as the frequency of the limit cycle started from non-zero values at the onset of a bifurcation (Fig.2C1), causing discontinuous jumps in the spectrum distance.
However, for the purpose of pattern separation, the two metrics did not give qualitatively different results when assessing the distances due to changes the feedback weight $w_{23}$ (Fig.S4, Supplementary materials). Additionally, we found that the non-monotonic dependence of distance in both $d^E$ and $d^S$ on the feedback weight $(w_{23}^{max})$ persisted, and that an optimal value for any given pair of stimuli could be found (Fig.S5, Supplementary materials).

**Bifurcation analysis**

We denoted the dynamical system of Eqn.(1) as a parameterized form

$$\dot{r} = f(r, \Theta)$$  \hspace{1cm} (3)

where $r \in \mathbb{R}^3$ was the vector of firing rates and $\Theta \in \mathbb{R}^p$ was the vector of parameters. The vector field $f = (f_1, f_2, f_3)^T$ was a smooth function on some open set of $\mathbb{R}^3 \times \mathbb{R}^p$. The dimensionality of $\Theta$ could be up to eight dimensions maximally to include all connection weights $w_{ij}$ and the external stimulus $\mu$. However, since we were only interested in the top-down control, $\Theta$ was restricted to be two dimensions ($p = 2$) including the top-down input $w_{23}$ and the external stimulus $\mu$. 
All the other parameters were fixed as constants determined based on previous experimental work (Whittington, Traub et al. 2000).

As the system Eqn.(3) had an equilibrium at \((r, \Theta) = (r_0, \Theta_0)\), i.e.,

\[
f(r_0, \Theta_0) = 0
\]  

the stability of this equilibrium could be determined from the linearized vector field of Eqn.(3) given by

\[
\dot{\xi} = D_r f(r_d, \Theta_d) \xi, \xi \in \mathbb{R}^3
\]

where \(D_r f \equiv \partial f / \partial r\) was the Jacobian matrix of the vector field \(f\).

If none of the eigenvalues of \(D_r f(r_0, \Theta_0)\) lied on the imaginary axis (i.e., the equilibrium was hyperbolic), the local stability of \((r_0, \Theta_0)\) in the nonlinear system (4) could be determined by the linear system (5). The equilibrium was stable if all eigenvalues of \(D_r f(r_0, \Theta_0)\) had negative real parts. In the case of a hyperbolic equilibrium, varying slightly the parameter \(\Theta\) would not change the stability as taking Eqn.(4) and the invertibility of \(D_r f(r_0, \Theta_0)\), there existed a unique smooth function \(h: \mathbb{R}^p \to \mathbb{R}^3\) such that

\[
h(\Theta_0) = r_0 \text{ and } f(h(\Theta), \Theta) = 0
\]  

for \(\Theta\) sufficiently close to \(\Theta_0\). By continuity of the eigenvalues with respect to parameters, \(D_r f(h(\Theta), \Theta)\) had no eigenvalue on the imaginary axis for \(\Theta\) sufficiently close to \(\Theta_0\). Therefore, the hyperbolicity of the equilibrium persisted.
and its stability type remained unchanged for in close vicinity of $\Theta_0$. By contrast, when some of the eigenvalues of $D_r f(r_0, \Theta_0)$ lied on the imaginary axis, for example, a zero eigenvalue or a pair of purely imaginary eigenvalues, new topologically different dynamical behaviors occurred by a small change in $\Theta$. Equilibria could be created or annihilated, and periodic dynamics could emerge.

The parameterized system (3) thus underwent a bifurcation at $(r_0, \Theta_0)$ if the Jacobian matrix $D_r f(r_0, \Theta_0)$ has an eigenvalue of zero real part. In our model, a saddle-node bifurcation (SN) occurred when $D_r f(r_0, \Theta_0)$ had a single zero eigenvalue (in addition to some nondegenerate conditions), and a Hopf bifurcation (H) occurred when $D_r f(r_0, \Theta_0)$ had a pair of purely imaginary eigenvalues. The bifurcation point was found numerically by XPPAUT or the Matlab toolbox MATCONT.

The number of parameters that must be varied simultaneously to evoke a bifurcation is defined as the codimension of this bifurcation (Guckenheimer and Holmes 2013, Kuznetsov 2013). Considering the infinite-dimensional space $\mathcal{H}$ of all vector fields defined on the $n$-dimensional Euclidean space $\mathbb{R}^n$, a vector field $f_0$ undergoing a bifurcation (Supplemental materials, Fig. S6-S7), for example, a Hopf bifurcation, corresponds to a point in the space $\mathcal{H}$. All nearby vector fields with the same singularity as $f_0$ (i.e., vector fields that are orbitally topologically equivalent to $f_0$) form a submanifold $\mathcal{L}$ of co-dimension $k$, which is an equivalence class of the singular vector field $f_0$. Therefore, within the space $\mathcal{H}$ it requires another submanifold $\tilde{\mathcal{L}}$ of at least $k$ dimensions to intersect transversely with $\mathcal{L}$ at
point $f_0$, such that the singularity of $f_0$ persists under small perturbations of the vector field. The submanifold $\mathcal{L}$ was obtained through a parametrized family of vector fields involving at least $k$ parameters. The least number $k$ is then defined as the codimension of $f_0$. The parametrized vector field $f(r, \Theta)$ in Eqn.(3) can be thought of as one realization of the submanifold $\mathcal{L}$ which passes through the vector field $f_0 \triangleq f(r_0, \Theta_0)$ undergoing a bifurcation with the two parameters corresponding to the top-down input $w_{23}$ and the external stimulus $\mu$. 
Results

Reduced network model generates complex dynamics

To understand the functional role of top-down projections onto inhibitory neurons, we built a three-node network model (Fig.1A, see Methods) that recapitulates a circuit architecture identified both structurally (Padmanabhan, Osakada et al. 2018) and functionally (Boyd, Sturgill et al. 2012, Markopoulos, Rokni et al. 2012) across a number of brain areas. For different stimuli $\mu$, the network exhibited a variety of dynamics (Fig.1B, Fig. S1). For instance, when the stimulus was small, the firing rates $r_i, i = 1,2,3$ had a fast-transient increase followed by damping oscillations that converged to a stationary state (Fig.1B, top). A sufficiently large stimulus $\mu$ elevated the firing rates to near saturation, where they then remained at the upper bound of the nonlinear sigmoid function throughout the duration of the stimulus (Fig.1B, bottom). For small or large stimuli, the network responses converged to a constant firing rate after transient dynamics. By contrast, for medium values of $\mu$, more complex firing rate dynamics emerged, including oscillations (middle in Fig.1B). To visualize the collective behaviors of E, I, and T populations to these different stimuli, we turned to a three dimensional dynamical system representation of the model where the time evolution of the firing rates (i.e., state variables) was as a trajectory (or an orbit) in the phase space $(r_1, r_2, r_3)$ and the tangent vector defining the velocity of each point along a trajectory given by the vector field $\mathbf{f} = (f_1, f_2, f_3)^T$ (see Methods) of Eqn.(1). The firing rates over time in Fig.1B thus corresponded to trajectories in Fig.1D starting from the origin $\mathbf{0}$. 


(where all three populations were silent). For small stimuli, the trajectory made an excursion before spiraling into an equilibrium indicated by the solid dot (Fig.1D, orange). Similarly, when the stimulus $\mu$ was large, the trajectory again settled into an equilibrium, but one that was translated within the phase space to the top-right corner (Fig.1D, black). Finally, medium stimuli, the time-varying oscillation of the firing rates manifested as a periodic orbit (or a limit cycle) in the 3D phase space (Fig.1D, brown). In this space, we defined the steady-state dynamics as the $\omega$-limit set of the system.

**Top-down input reshapes network dynamics and modulates neural oscillations**

To explore how top-down down projections onto the inhibitory neuron population (I) shaped the dynamics of the network, we studied the effects of changes in the connection weight $w_{23}$ on firing rate dynamics. A number of experimental studies have shown that inhibitory neuron activity can be dynamically gated, on short timescales corresponding to changes in behaviors or internal states such as attention or arousal (Saalmann, Pigmarev et al. 2007, Batista-Brito, Zagha et al. 2018), or on longer timescales corresponding to learning and memory (Tomita, Ohbayashi et al. 1999, Cohen, Wilmes et al. 2011). To model these general phenomenological changes, we varied the weight $w_{23}$ (Fig.2A, top) from the top-
down population (T) to the inhibitory population (I) and studied the effect of these changes on the firing rate dynamics of the network.

For a fixed stimulus ($\mu = 1.5$) the dynamics of the firing rates $r_i(t), i = 1,2,3$ were sensitive to different values of $w_{23}$ (Fig.2A, bottom). When the top down weight was small ($w_{23} = 4$), firing rates approached the equilibrium exponentially (Fig.2A bottom and Fig.2B left, black traces). Conversely, when the top-down weight was large ($w_{23} = 10.5$), the firing rate of excitatory cells ($r_1$) increased initially, but was suppressed as inhibition reduced the activity, until the firing rates ultimately settled to an equilibrium (Fig.2A top and Fig.2B left, light magenta traces). When the magnitude of top-down input was changed to an intermediate value ($w_{23} = 5.5$), the same stimulus generated oscillatory activity in the network, with the steady-state dynamics transitioning to a periodic orbit (a limit cycle). Changing the weight of top-down projections onto the local inhibitory population (I) in a single stimulus produced the same diversity of firing rate dynamics that occurred from changes in the stimulus. Furthermore, for a given top-down input weight ($w_{23}$), the effects on the network dynamics for a particular stimulus was unique to that stimulus (Fig.2B right vs. left).

In regimes where specific weights of top-down input generated sustained oscillatory activity for a given stimulus $\mu$, we determined the frequency and amplitude of these oscillations (Fig.2B, left $w_{23} = 5.5$, right $w_{23} = 10.5$) as changes in both have been tied to circuit function and behavior (Buzsaki and Draguhn 2004, Kay, Beshel et al. 2009). For a given stimulus, oscillations emerged between two
values of $w_{23}$, with the frequency of the oscillation varying monotonically (Fig.2C1).

By contrast, while the amplitudes of the oscillations started from zero at the two critical values of $w_{23}$, they reached a maximum in between (Fig.2D1). Thus, by changing the top-down input ($w_{23}$) the oscillations generated covered a wide range of frequencies (Fig.2C2, 0 – 45 Hz) and amplitudes (Fig.2D2, 0 – 1 A.U.), including frequencies in the alpha, beta and gamma bands (Fig. S2).

**Top-down input contributes to pattern separation**

As changing the top-down weight onto inhibitory neurons could generate complex activity patterns we next asked what computations could be performed by this control. For example, both behavioral and neurophysiological measures show that as the representations of two stimuli by neuronal circuits become different, distinguishing between them becomes easier (Leutgeb, Leutgeb et al. 2007, Yassa and Stark 2011). Examples include distinguishing two similar odors that differ slightly in a physiochemical feature (Friedrich and Laurent 2001), or distinguishing between two similarly oriented bars of light (Long, Jiang et al. 2015). In the olfactory bulb, local inhibitory granule cells appear critical for discrimination tasks (Abraham, Egger et al. 2010, Nunez-Parra, Maurer et al. 2013); in cortical circuits, local parvalbumin positive inhibitory neurons play an analogous critical role in discrimination (Korotkova, Fuchs et al. 2010, Lovett-Barron and Losonczy 2014).

Both olfactory bulb granule cells and cortical parvalbumin interneurons receive feedback inputs from higher processing areas, suggesting that control of inhibition,
via top-down centrifugal projections, may be one way that stimulus discrimination is implemented by the circuit.

To test this hypothesis, we presented our network with a pair of stimuli, denoted by $\mu_1$ and $\mu_2$ (corresponding to stimuli arranged along a one-dimensional axis) and studied how top-down control of inhibition altered the firing rate representations of the two stimuli (Fig.3A). For a set of stimuli $\mu_i, i = 1,2$, the steady-state representation of the network activity was the $\omega$-limit set $\Omega_i, i = 1,2$ for the associated stimulus with the distance between the two stimuli $\mu_1$ and $\mu_2 \Delta \mu$ in the stimulus space, and the resultant distance in the firing rate phase space between the two $\omega$-limit sets ($\Omega_1$ and $\Omega_2$) as a metric $d$ (see Methods). The smaller the $\Delta \mu$, the more similar the two stimuli were. Thus, for any given pair of stimuli separated by $\Delta \mu$, we hypothesized that changes in the weight of feedback onto the inhibitory neuron population ($w_{23}$) could increase the value of $d$. Such increases in distances would result in the representations of the two stimuli $\mu_1$ and $\mu_2$ becoming more distinct (Fig.3A).

In a representative example where $\mu_1 = 1.0$ and $\mu_2 = 1.1$, when the top-down input was low ($w_{23} = 3$), the representations of the two stimuli were close (Fig.3B-3C left, $d = 0.01$). As $w_{23}$ was increased, the adjacent equilibria were pushed apart ($w_{23}=3.8$, Fig.3B-3C, middle, $d = 0.5$), separating the patterns of activity that corresponded to the two stimuli. Interestingly, as $w_{23}$ was increased further ($w_{23} = 10$), the representations of the two stimuli became close to one another again (Fig.3B-3C, right, $d=0.02$). Consequently, the distance $d$ between
the two equilibria ($\Omega_1$ and $\Omega_2$) was a non-monotonic function of $w_{23}$ (Fig.3D, Fig. S3). For any given pair of stimuli an optimal value of $w_{23}^{max}$ could be found which maximized the distance between the two resultant representations $\Omega_1$ and $\Omega_2$. A landscape of the optimal $w_{23}^{max}$ across all pairs of stimuli ($\mu_1, \mu_2$) was shown in Fig.3E (Supplemental materials, Fig. S4). Changing the amount of top-down projections onto the inhibitory neuron population (I) could be used to facilitate stimulus separation dynamically (Fig. S5). To understand why top-down inputs achieved this optimal $w_{23}^{max}$, we examined the effect that varying the top-down projection weight had on the stationary responses of both local excitatory and inhibitory populations ($r_1$ and $r_2$). For a given stimulus, an increase in $w_{23}$ led to a monotonic decrease in $r_1$, suggesting persistent suppression onto the local E population. By contrast, the response of the local I population $r_2$ was elevated first with increasing $w_{23}$ until reaching the maximum $r_2^{max}$, after which $r_2$ dropped significantly (Fig.3F). Across different levels of stimuli $\mu$, the shape of the firing rate $r_2$ as a function of $w_{23}$ remained the same but shifted vertically (Fig.3F). To determine if these differences in the firing rate of inhibitory neurons ($r_2$) were related to the optimal values of top-down inputs, for any given pair of stimuli ($\mu_1, \mu_2$), we plotted the $w_{23}^{max}$ (abscissa) obtained from the landscape in Fig.3E versus the midpoint $w_{23}^{mid}$ (ordinate) of the connecting line between the $r_2^{max}$ of the same stimulus pair (inset of Fig.3G). The response $r_2$ to one stimulus on the left of the midpoint dropped significantly, while the response to a similar stimulus on the right still had a high inhibitory firing rate. The optimal $w_{23}^{max}$ for separating the
responses of two similar stimuli was correlated to $w_{23}^\mu$ ($R^2 = 0.98$), the value at which inhibitory neuron activity from one stimulus was suppressed while activity from the other similar stimulus remained persistently high (Fig.3G). Consequently, stimulus separation arose from the differential sensitivity of inhibitory neurons to the balance between top-down feedback and recurrent inhibition, i.e., when an imbalance occurred between the top-down feedback and the recurrent inhibition for one stimulus while that balance was preserved for the second stimulus.

**Top-down input contributes to oscillation synchrony**

In addition to equilibrium attractors in the firing rate phase space across different stimuli, our model was also capable of generating sustained oscillations of firing rates, due to presence of limit cycles. Stimulus-evoked oscillations were modulated by the top-down inputs (Fig.2) and covered a wide range of frequency and amplitude, suggesting the possibility that these oscillatory responses to different stimuli could be synchronized by tuning $w_{23}$. To explore this question, we first examined the oscillations in the firing rate generated by two different stimuli $\mu_1$ and $\mu_2$. At a given value of top-down input ($w_{23} = 12.0$), one stimulus ($\mu_1 = 1.15$) generated oscillations ($f_1 = 33.5$Hz, Fig.4A1, before) in the firing rates of the top-down (T) population that were different in both frequency and amplitude from the oscillations ($f_2 = 37.5$Hz, Fig.4A1, before) in response to a second stimulus ($\mu_2 = 0.6$). However, a change in the weight of the top-down input ($w_{23} = 8$), resulted in the frequencies of the firing rate oscillations becoming more similar for the same two stimuli (Fig.4A1, after, $f_1 = 30.4$Hz, $f_2 = 30.8$Hz). This increase
the firing rate synchrony was also apparent when visualized in the 3D phase space (Fig.4A2). To quantify the synchrony between the two oscillations in response to $\mu_1$ and $\mu_2$, before and after changes in top-down input, we calculated the spectrum distance $d^S$ (see Supplementary material), a measure of the total difference in both frequency and amplitude between the two limit cycles (Fig.4B1). Changing the top-down input resulted in synchronized activity in the network for two different stimuli (Fig.4B2). While the effect was greatest when stimuli were similar, we found examples where oscillations from two different stimuli that were initially as far as 20 Hz apart could nonetheless become synchronous by changing the weight of feedback $w_{23}$ onto inhibitory neuron populations. As with stimulus discrimination, a systematic relationship emerged corresponding to the optimal top-down input $w_{23}^{\text{min}}$ across combinations of stimuli ($\mu_1$ vs. $\mu_2$) that minimized the distance $d^S$ (Fig.4C).

Although we thus far treated stimulus discrimination and response synchrony separately, neural circuits perform both operations simultaneously, bringing the network representation of one stimulus closer to another, while simultaneously pushing the representation of that stimulus farther from a third. We therefore tested if a single change in the top-down input $w_{23}$ accomplish both of these operations; minimize the distance between the responses to one pair of stimuli ($\mu_1$ vs. $\mu_2$) while also maximize the response distance to another other pair of stimuli ($\mu_1$ vs. $\mu_3$, Fig.4D). To do this, we identified values of $w_{23}$ that were optimal for both synchrony between the oscillations generated by stimulus $\mu_1$ and
\( \mu_2 \) (Fig. 4C) and also produced a maximum separation between the representations of stimulus \( \mu_1 \) and \( \mu_3 \). A 3D scatter plot of \( \mu_1 - \mu_2 - \mu_3 \) (Fig. 4D), showed all examples corresponding to two pairs of stimuli: \(( \mu_1, \mu_2 )\) whose oscillatory responses to stimuli were synchronous for a given \( w_{23} \) and two stimuli \(( \mu_1, \mu_3 )\) whose responses were also maximally separable at the same \( w_{23} \). The values of top-down input \( w_{23} \) for each point that fulfilled these diametrically distinct functions were coded by color and size (Fig. 4E). Importantly, we found that the top-down input scaled with the stimuli, such that when \( \mu_1, \mu_2 \) and \( \mu_3 \) were small, the top-down input that achieved both synchrony and separation was also small, but as the three stimuli increased in magnitude, the top-down input needed to synchronize one pair and separate the other pair also increased, suggesting that the optimal weights for the top-down projections onto inhibitory neurons that performed both pattern separation and oscillation synchrony corresponded to the magnitude of the stimuli. Finally, we found a strong correlation between the values of stimulus differences: \(| \mu_1 - \mu_2 | \) and \(| \mu_1 - \mu_3 | \) (Fig. 4F) at which an \( w_{23} \) weight was optimal for stimulus separation and oscillatory synchrony.

**Bifurcation mechanism for top-down control of inhibition**

Finally, to understand mathematically how such operations emerged as a result of changes in the top-down weight to inhibition, we studied the structure of the transitions in the dynamics of firing rates in the network (Fig. 2A). These transitions were associated with qualitative or topological changes in the \( \omega \)-limit sets of the system, indicative of the occurrence of bifurcations in the system. To explore this
further, we first examined the how the $\omega$-limit sets of the system changed with different top-down inputs. An equilibrium corresponding to constant firing rates in the network (Fig.5A, left and right) arose from the intersection of three nullclines (Fig.5A, yellow thick lines), each characterizing the geometric shape on which the firing rate derivatives of two nodes equaled to zero (Fig.S6, Supplementary materials). Global phase structures for two representative values of top-down inputs (Fig.5A, left: $\omega_{23} = 15$, right: $\omega_{23} = 4$) illustrated how these equilibria varied within the firing rate phase space, where the purple arrows showed the velocities (directions and magnitudes) of the flows along which trajectories followed. Using this information, for a given combination of stimulus $\mu$ and top-down input $w_{23}$, we calculated a family of $\omega$-limit sets on which dynamics settled from any set of initial values in the firing rates of the excitatory (E), inhibitory (I), and top-down (T) populations. In these two examples, both equilibria were stable and attractive, with all nearby trajectories (black thin lines) moving towards them. This was, however, not true for all values of $w_{23}$. At some critical values of $w_{23}$, the equilibrium lost stability, and a small-amplitude limit cycle branched from that unstable equilibrium, resulting in the oscillations observed in the dynamics (Fig.5A, middle). This transition signified a Hopf bifurcation of the system (see Methods and Supplementary materials, Fig. S6), which arose when the top-down input $w_{23}$ was tuned within a specific regime. Across combinations of external stimuli $\mu$ and top-down input $w_{23}$, we obtained a smooth manifold in the phase space (Fig.5B, left), composed of stable and unstable equilibria and divided into several regions. Sustained oscillations corresponded to the red region-LC (LC: limit cycle) where
each equilibrium (unstable) was paired with exactly one limit cycle that were born simultaneously via a Hopf bifurcation (the purple empty square vs. the dot-dashed curve). Constant firing rates corresponded to the grey region-EE (EE: exponential equilibrium) and green region-SE (SE: spiraling equilibrium), where the equilibria were stable, approached either exponentially (region-EE) or via damping-oscillations (region-SE). Finally, the two blue regions on the manifold were bounded by another type of bifurcation near Bogdanov–Takens (BT) and an example global phase structure (Fig.S7). The equilibrium manifold thus defined the entire family of network representations (i.e., the of $\omega$-limit sets) for all possible combinations of stimuli $\mu$ and top-down input $w_{23}$.

Within the space of $\mu$ and $w_{23}$, we identified a transition boundary $\Gamma_H$ (black solid curves, Fig.5B, middle) corresponding to the separatrix enclosing the region-LC on the equilibrium manifold. $\Gamma_H$ specified the parameter pairs ($\mu, w_{23}$) at which a Hopf bifurcation occurred, thereby dividing the parameter space into regimes where different dynamics were present (same color coded as Fig.5B1, left). For a given pair of stimuli (for example, $\mu_1 = 2.0, \mu_2 = 3.0$), changing $w_{23}$ corresponded to shifting the horizontal dashed line vertically (three representatives were shown in Fig.5B1, middle), thereby shifting the junctions with the two stimuli (vertical solid lines, Fig.5B1, middle) across different regimes in the parameter space. Correspondingly, in the firing rate phase space (Fig.5B, left), changes in $w_{23}$ for one stimulus moved the equilibrium through different regions of the manifold: EE-LC-SE (the part in region-LC was plotted in a dashed curve), while for another
stimulus, a parallel curve on the manifold could also be traced by changing $w_{23}$. When the two junctions in Fig.5B1 (middle) were on opposite sides of the transition boundary $\Gamma_H$, with one equilibrium in region-LC and the other in region-SE (Fig.5B1, left), the two $\omega$-limit sets (i.e., network representations) became topologically different from each other; the former a limit cycle, and the latter an equilibrium point. For this specific combination of stimulus pairs and top-down input, the distance between the network representations of the two stimuli became maximal (Fig.5B1, right) such that the optimal $w_{23}^{max}$ for pattern separation was then achieved when the $\omega$-limit sets were on different sides of transition boundary $\Gamma_H$.

Furthermore, when the two junctions were both inside the transition boundary $\Gamma_H$ (Fig.5B2, middle) for a pair of stimuli, this time, when the top-down input was altered for both, two limit cycles emerged (one for each stimulus, for example, $\mu_1 = 3.45, \mu_2 = 3.95$, Fig.5B2). As both the frequencies and amplitudes of the oscillations were modulated by $w_{23}$ (Fig.2C and 2D) and synchrony between the network responses for the two stimuli were achieved when the top-down input was tuned to place both junctions within this space. As changes in top-down input moved the junctions for pairs of stimuli within the parameter space, a shared mechanism supported both stimulus separation and oscillation synchrony,
depending on the relative positions of the junctions with respect to the transition boundary.

We next wished to determine if the transition boundary $\Gamma_H$ identified via analysis of the dynamical system corresponded to the $w_{23}^{\text{max}}$ matrix found in Fig.3Ei, a measure of the computational optimization performed by the network. To do this, we considered a set of $\mu_1$, i.e., $\{\mu_1 | \mu_1 \in [0,4]\}$, and a set of $\Delta \mu$, wherein each value was an array of stimuli $\mu_1 + \Delta \mu$ that defined a unique set of stimulus pairs $\{(\mu_1, \mu_1 + \Delta \mu) | \mu_1 \in [0,4]\}$. For a given $\Delta \mu > 0$ (distinguishing the pair $(\mu_1, \mu_1 + \Delta \mu)$ was the same task as distinguishing $(\mu_1 + \Delta \mu, \mu_1)$ in terms of a pattern separation), for instance, $\Delta \mu = 0.1$ or $\Delta \mu = 0.5$ (Fig.5C, left and middle), the set of stimulus pairs had a unique transition boundary $\Gamma_{\Delta}$ (Fig.5C, right). The section of the $w_{23}^{\text{max}}$ matrix in Fig.3Ei for the set of stimulus pairs $\{(\mu_1, \mu_1 + \Delta \mu) | \mu_1 \in [0,4], \Delta \mu \text{ is given}\}$ was correlated to the transition boundary $\Gamma_{\Delta}$ of the same value $\Delta \mu$ (Fig.5C). For small $\Delta \mu = 0.1$, the slice of the $w_{23}^{\text{max}}$ followed closely with the transition boundary $\Gamma_{\Delta}$ (Fig.5C, left). As the stimulus difference increased (Fig.5C, middle), $w_{23}^{\text{max}}$ deviated from the boundary $\Gamma_{\Delta}$. Larger $\Delta \mu$’s increased the bifurcation lag between two stimuli such that the one that bifurcated first had more parameter space to develop before the bifurcation of the other stimulus. This suggested that as the stimulus discrimination became harder, i.e., as $\Delta \mu$ decreased, the range of top-down input $w_{23}$ to separate two stimuli shrank significantly around a close vicinity of the transition boundary. Consequently, accurate and subtle adjustments of top-down input around the transition boundary
were then required to separate similar stimuli from each other. Taken together, we identified a bridge linking the mechanisms that gave rise to the dynamics of this neural circuit with the computations that were performed by the circuit.
Discussion

Using a three-node model, which includes top-down centrifugal projections onto inhibitory neurons, we identified a network capable of complex dynamic behaviors, ranging from an equilibrium attractor to stable oscillations across a range of frequencies and amplitudes. By changing the weight of top-down projections onto inhibitory neurons, the network could either facilitate pattern separation across two similar stimuli, or synchronize the oscillatory activity produced by two different stimuli. A bifurcation analysis of the dynamical system revealed that both mechanisms emerged from the transition boundary of Hopf bifurcations which branched from co-dimensional two bifurcation points (i.e., the Bogdanov-Takens bifurcation). Our results provide both a mathematical framework for how top-down control of inhibition confers unique properties to the dynamics of networks, and a link as to how these dynamics support the computations the neural circuits are often performing.

An important point is to consider how these changes in top-down projections may be instantiated biologically? The mechanisms governing such changes are dependent on the timescale of weight changes.

First, evidence suggests that behavioral state can strongly modulate the local dynamics of a circuit, including the firing rates of inhibitory neurons (Pakan, Lowe et al. 2016, Hattori, Kuchibhotla et al. 2017). For example, in primate v4, inhibitory neurons show strong attentional modulation (Fries, Reynolds et al. 2001, Mitchell, Sundberg et al. 2007). Similarly, synchronization of parvalbumin (PV) inhibitory neurons and elevated gamma has been tied to attention (Kim, Ährlund-
Richter et al. 2016). Additionally, it has been suggested that neuromodulatory signals may gate the effect of attention (Thiele and Bellgrove 2018). Changes in the amount of top-down feedback that occur on short timescales such as during attention, may thus be implement by neuromodulators that affect either the release of neurotransmitter, or gate the responses at the post-synaptic target of glutamate release. Either would change the weight of top-down inputs onto inhibition dynamically. Both dopamine (Gao and Goldman-Rakic 2003, Tritsch and Sabatini 2012) and serotonin (Athilingam, Ben-Shalom et al. 2017) alter excitatory post-synaptic potential (EPSP) amplitudes in inhibitory interneurons, and such alterations would result in changes in the weight of synapses from top-down projections to local inhibitory interneurons. In this way, either pattern separation, or oscillatory synchrony could happen on relatively short timescales corresponding to epochs of attention. Here the modulation of top-down weight would be flexible and dynamic.

Second, top-down weights could change on longer timescales as a result of synaptic plasticity (Isaacson and Scanziani 2011, Kullmann, Moreau et al. 2012). While initial reports identified plasticity at glutamatergic synapses on inhibitory interneurons (Buzsaki and Eidelberg 1982, Tomasulo and Steward 1996), the rules governing this plasticity appear to be more complex than the rules for glutamatergic plasticity on excitatory neuron synapses. Nonetheless, classic plasticity induction paradigms, such as pairing theta burst stimulation with post-synaptic depolarization (Perez, Morin et al. 2001, Lamsa, Irvine et al. 2007), result in changes in the weight of glutamatergic synapses on inhibitory interneurons. In
this example, top-down changes in the weight would correspond to long time-
scales, operating to codify stable changes in synapses that render representations
permanently more distinct or similar.

While the biological mechanisms by which the weights of top-down
synapses change onto inhibitory neuronal populations may be diverse depending
on the timescale of that change, they can give rise to functionally equivalent
changes to the dynamics of neural circuits that support an array of computations.
For instance, changes in the top-down weight to inhibitory neurons would render
two stimuli more distinct at the level of firing rates in the population, a process
referred to as pattern separation (Cayco-Gajic and Silver 2019) or decorrelation
(Friedrich and Laurent 2001). Our model predicts that pattern separation may
arise as a feature of the non-monotonic change in firing when the weight of
feedback is gradually altered. The amount of top-down projections onto inhibitory
neurons thus sets a gate, allowing some stimuli to cross the threshold of inhibition,
while others do not; the consequence of which is increasing their distance in
representational space.

In parallel, changing top-down weights onto inhibitory neurons could
increase the oscillatory synchrony in the network for two stimuli that were initially
asynchronous. A number of experimental and theoretical studies have explored
the privileged role that inhibitory interneurons play in generating gamma
oscillations (Whittington, Traub et al. 1995, Hasenstaub, Shu et al. 2005, Cardin,
Carlén et al. 2009, Sohal, Zhang et al. 2009, Tiesinga and Sejnowski 2009).
Among these, the two most common models are those where gamma arises from
reciprocal coupling between pyramidal cells and inhibitory interneurons (PING), and recurrent connections among inhibitory interneurons (ING) (Whittington, Traub et al. 2000, Tiesinga and Sejnowski 2009). In both, the generation of oscillatory activity arises as a result of motifs of local connectivity. In our work, we identified another motif by which gamma oscillations arise – **Top-down control of Inhibitory Neuron Gamma (TING).** Local excitatory cells broadcast activity patterns to a top-down population, that then synapses onto a local inhibitory interneuron population. As the system crosses the transition boundary defined by a combination of stimulus and top-down weight, gamma oscillations emerge. In addition to identifying how gamma oscillations can emerge in a circuit with top-down connections made exclusively onto local inhibitory interneuron populations, we also demonstrate how synchronous oscillatory activity within local circuits can be changed by changing the weight of top-down projections onto inhibition.

Finally, our model links two key areas of computational neuroscience, the dynamical systems framework for evaluating how neuronal responses evolve over time, and principles of neural coding, the aim of which is to understand what can be done computationally with the temporal evolution of neural dynamics. The capacity of the top-down feedback to modulate not only the dynamics of the local network, but to reshape its own input information via inhibition reveals an essential feature of neural circuits. Using bifurcation analysis, we found that the system was governed by a codimension-2 bifurcation, with distinct domains within a manifold of the system defined by the stimulus and the weight of feedback. These domains corresponded to transitions in the dynamics of the system. Changes in the top-
down inputs onto inhibition operated by moving a transition boundary for dynamics across different external stimuli. When two stimuli were on opposites sides of this transition boundary, their dynamics operated under two different regimes, and their representations were pushed further apart. By contrast, when the stimuli were both on the same side of the transition boundary, within regimes corresponding to similar dynamics, their activity became more synchronous; effectively binding those stimuli together. Changes in top-down weight were therefore changes in the location of the transition boundary that could either marshal the representations of two stimuli together or push them apart.

In conclusion, we identified a model that links the dynamics of neural systems with the computations they are hypothesized to perform, and may be used as a generalized framework to study the diverse effects of feedback onto inhibitory populations.
Acknowledgments

This study was supported by funding from the National Institutes of Health (NIH) and the National Science Foundation (NSF). KP was funded by NIH R01 MH113924, NSF CAREER 1749772, the Cystinosis Research Foundation, and the Kilian J. and Caroline F. Schmitt Foundation.
References

Abraham, N. M., V. Egger, D. R. Shimshek, R. Renden, I. Fukunaga, R. Sprengel, P. H. Seeburg, M. Klugmann, T. W. Margrie, A. T. Schaefer and T. Kuner (2010). "Synaptic inhibition in the olfactory bulb accelerates odor discrimination in mice." Neuron 65(3): 399-411.

Atallah, B. V. and M. Scanziani (2009). "Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition." Neuron 62(4): 566-577.

Athilingam, J. C., R. Ben-Shalom, C. M. Keeshen, V. S. Sohal and K. J. Bender (2017). "Serotonin enhances excitability and gamma frequency temporal integration in mouse prefrontal fast-spiking interneurons." Elife 6: e31991.

Batista-Brito, R., E. Zagha, J. M. Ratliff and M. Vinck (2018). "Modulation of cortical circuits by top-down processing and arousal state in health and disease." Current opinion in neurobiology 52: 172-181.

Boyd, A. M., J. F. Sturgill, C. Poo and J. S. Isaacson (2012). "Cortical feedback control of olfactory bulb circuits." Neuron 76(6): 1161-1174.

Buzsáki, G. (1984). "Feed-forward inhibition in the hippocampal formation." Progress in neurobiology 22(2): 131-153.

Buzsaki, G. and A. Draguhn (2004). "Neuronal oscillations in cortical networks." Science 304(5679): 1926-1929.
Buzsaki, G. and E. Eidelberg (1982). "Direct afferent excitation and long-term potentiation of hippocampal interneurons." Journal of Neurophysiology 48(3): 597-607.

Cardin, J. A., M. Carlén, K. Meletis, U. Knoblich, F. Zhang, K. Deisseroth, L.-H. Tsai and C. I. Moore (2009). "Driving fast-spiking cells induces gamma rhythm and controls sensory responses." Nature 459(7247): 663-667.

Cayco-Gajic, N. A. and R. A. Silver (2019). "Re-evaluating circuit mechanisms underlying pattern separation." Neuron 101(4): 584-602.

Cohen, M. R. and J. H. Maunsell (2009). "Attention improves performance primarily by reducing interneuronal correlations." Nat Neurosci 12(12): 1594-1600.

Cohen, M. X., K. A. Wilmes and I. van de Vijver (2011). "Cortical electrophysiological network dynamics of feedback learning." Trends in cognitive sciences 15(12): 558-566.

Fino, E. and R. Yuste (2011). "Dense inhibitory connectivity in neocortex." Neuron 69(6): 1188-1203.

Fishell, G. and A. Kepecs (2019). "Interneuron types as attractors and controllers." Annual review of neuroscience 43.

Friedrich, R. W. and G. Laurent (2001). "Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity." Science 291(5505): 889-894.
Fries, P., J. H. Reynolds, A. E. Rorie and R. Desimone (2001). "Modulation of oscillatory neuronal synchronization by selective visual attention." *Science* **291**(5508): 1560-1563.

Gao, W.-J. and P. S. Goldman-Rakic (2003). "Selective modulation of excitatory and inhibitory microcircuits by dopamine." *Proceedings of the national academy of sciences* **100**(5): 2836-2841.

Guckenheimer, J. and P. Holmes (2013). *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Springer Science & Business Media.

Hasenstaub, A., Y. Shu, B. Haider, U. Kraushaar, A. Duque and D. A. McCormick (2005). "Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks." *Neuron* **47**(3): 423-435.

Hattori, R., K. V. Kuchibhotla, R. C. Froemke and T. Komiyama (2017). "Functions and dysfunctions of neocortical inhibitory neuron subtypes." *Nat Neurosci* **20**(9): 1199-1208.

Isaacson, J. S. and M. Scanziani (2011). "How inhibition shapes cortical activity." *Neuron* **72**(2): 231-243.

Kay, L. M., J. Beshel, J. Brea, C. Martin, D. Rojas-Libano and N. Kopell (2009). "Olfactory oscillations: the what, how and what for." *Trends Neurosci* **32**(4): 207-214.

Kepecs, A. and G. Fishell (2014). "Interneuron cell types are fit to function." *Nature* **505**(7483): 318-326.

Kim, H., S. Ährlund-Richter, X. Wang, K. Deisseroth and M. Carlén (2016). "Prefrontal parvalbumin neurons in control of attention." *Cell* **164**(1-2): 208-218.
Korotkova, T., E. C. Fuchs, A. Ponomarenko, J. von Engelhardt and H. Monyer (2010). "NMDA receptor ablation on parvalbumin-positive interneurons impairs hippocampal synchrony, spatial representations, and working memory." Neuron 68(3): 557-569.

Kullmann, D. M., A. W. Moreau, Y. Bakiri and E. Nicholson (2012). "Plasticity of inhibition." Neuron 75(6): 951-962.

Kuznetsov, Y. A. (2013). Elements of applied bifurcation theory, Springer Science & Business Media.

Lamsa, K., E. E. Irvine, K. P. Giese and D. M. Kullmann (2007). "NMDA receptor-dependent long-term potentiation in mouse hippocampal interneurons shows a unique dependence on Ca2+/calmodulin-dependent kinases." The Journal of physiology 584(3): 885-894.

Leutgeb, J. K., S. Leutgeb, M.-B. Moser and E. I. Moser (2007). "Pattern separation in the dentate gyrus and CA3 of the hippocampus." science 315(5814): 961-966.

Long, M., W. Jiang, D. Liu and H. Yao (2015). "Contrast-dependent orientation discrimination in the mouse." Scientific reports 5(1): 1-14.

Lovett-Barron, M. and A. Losonczy (2014). "Behavioral consequences of GABAergic neuronal diversity." Current opinion in neurobiology 26: 27-33.

Markopoulos, F., D. Rokni, D. H. Gire and V. N. Murthy (2012). "Functional properties of cortical feedback projections to the olfactory bulb." Neuron 76(6): 1175-1188.
Meinecke, D. L. and A. Peters (1987). "GABA immunoreactive neurons in rat visual cortex." Journal of Comparative Neurology 261(3): 388-404.

Mitchell, J. F., K. A. Sundberg and J. H. Reynolds (2007). "Differential attention-dependent response modulation across cell classes in macaque visual area V4." Neuron 55(1): 131-141.

Nunez-Parra, A., R. K. Maurer, K. Krahe, R. S. Smith and R. C. Araneda (2013). "Disruption of centrifugal inhibition to olfactory bulb granule cells impairs olfactory discrimination." Proceedings of the National Academy of Sciences 110(36): 14777-14782.

Okun, M. and I. Lampl (2008). "Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities." Nature neuroscience 11(5): 535.

Packer, A. M. and R. Yuste (2011). "Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition?" Journal of Neuroscience 31(37): 13260-13271.

Padmanabhan, K., F. Osakada, A. Tarabrina, E. Kizer, E. M. Callaway, F. H. Gage and T. J. Sejnowski (2016). "Diverse Representations of Olfactory Information in Centrifugal Feedback Projections." J Neurosci 36(28): 7535-7545.

Padmanabhan, K., F. Osakada, A. Tarabrina, E. Kizer, E. M. Callaway, F. H. Gage and T. J. Sejnowski (2018). "Centrifugal inputs to the main olfactory bulb revealed through whole brain circuit-mapping." Frontiers in neuroanatomy 12.

Padmanabhan, K., F. Osakada, A. Tarabrina, E. Kizer, E. M. Callaway, F. H. Gage and T. J. Sejnowski (2019). "Centrifugal inputs to the main olfactory bulb
revealed through whole brain circuit-mapping." Frontiers in neuroanatomy 12: 115.

Pakan, J. M., S. C. Lowe, E. Dylda, S. W. Keemink, S. P. Currie, C. A. Coutts and N. L. Rochefort (2016). "Behavioral-state modulation of inhibition is context-dependent and cell type specific in mouse visual cortex." Elife 5: e14985.

Perez, Y., F. Morin and J.-C. Lacaille (2001). "A hebbian form of long-term potentiation dependent on mGluR1a in hippocampal inhibitory interneurons." Proceedings of the National Academy of Sciences 98(16): 9401-9406.

Poo, C. and J. S. Isaacson (2009). "Odor representations in olfactory cortex:"sparse" coding, global inhibition, and oscillations." Neuron 62(6): 850-861.

Ruff, D. A. and M. R. Cohen (2019). "Simultaneous multi-area recordings suggest that attention improves performance by reshaping stimulus representations." Nat Neurosci 22(10): 1669-1676.

Saalmann, Y. B., I. N. Pigarev and T. R. Vidyasagar (2007). "Neural mechanisms of visual attention: how top-down feedback highlights relevant locations." Science 316(5831): 1612-1615.

Shipley, M. T. and G. D. Adamek (1984). "The connections of the mouse olfactory bulb: a study using orthograde and retrograde transport of wheat germ agglutinin conjugated to horseradish peroxidase." Brain research bulletin 12(6): 669-688.

Silberberg, G. and H. Markram (2007). "Disynaptic inhibition between neocortical pyramidal cells mediated by Martinotti cells." Neuron 53(5): 735-746.
Sohal, V. S., F. Zhang, O. Yizhar and K. Deisseroth (2009). "Parvalbumin neurons and gamma rhythms enhance cortical circuit performance." Nature 459(7247): 698-702.

Thiele, A. and M. A. Bellgrove (2018). "Neuromodulation of attention." Neuron 97(4): 769-785.

Tiesinga, P. and T. J. Sejnowski (2009). "Cortical enlightenment: are attentional gamma oscillations driven by ING or PING?" Neuron 63(6): 727-732.

Tomasulo, R. A. and O. Steward (1996). "Homosynaptic and heterosynaptic changes in driving of dentate gyrus interneurons after brief tetanic stimulation in vivo." Hippocampus 6(1): 62-71.

Tomita, H., M. Ohbayashi, K. Nakahara, I. Hasegawa and Y. Miyashita (1999). "Top-down signal from prefrontal cortex in executive control of memory retrieval." Nature 401(6754): 699-703.

Tritsch, N. X. and B. L. Sabatini (2012). "Dopaminergic modulation of synaptic transmission in cortex and striatum." Neuron 76(1): 33-50.

Urban, N. N. and B. Sakmann (2002). "Reciprocal intraglomerular excitation and intra-and interglomerular lateral inhibition between mouse olfactory bulb mitral cells." The Journal of physiology 542(2): 355-367.

Whittington, M. A., R. Traub, N. Kopell, B. Ermentrout and E. Buhl (2000). "Inhibition-based rhythms: experimental and mathematical observations on network dynamics." International journal of psychophysiology 38(3): 315-336.
Whittington, M. A., R. D. Traub and J. G. Jefferys (1995). "Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation." *Nature* **373**(6515): 612-615.

Wilson, H. R. and J. D. Cowan (1972). "Excitatory and inhibitory interactions in localized populations of model neurons." *Biophys J* **12**(1): 1-24.

Yassa, M. A. and C. E. Stark (2011). "Pattern separation in the hippocampus." *Trends in neurosciences* **34**(10): 515-525.

Yoshimura, Y. and E. M. Callaway (2005). "Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity." *Nature neuroscience* **8**(11): 1552-1559.

Zhang, S., M. Xu, T. Kamigaki, J. P. Hoang Do, W. C. Chang, S. Jenvay, K. Miyamichi, L. Luo and Y. Dan (2014). "Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing." *Science* **345**(6197): 660-665.
Figure 1. Network model of top-down feedback to inhibitory neurons exhibiting complex dynamics. (A). A schematic diagram illustrating the topology of the network model. Each node denotes a neuron population and the connection weights are defined by $w_{ij}$ (excitatory: arrows; inhibitory: circles). (B). Network responses (color-coded firing rates match population nodes in (A)) to different stimulus magnitudes (left: $\mu_{\text{low}} = 0.1$, middle: $\mu_{\text{med}} = 1.5$, right: $\mu_{\text{high}} = 3.0$). (C). Distribution of all possible steady states $\tau_3$ shows the diversity of responses the network can generate. (D). Trajectories of the firing rate responses plotted in (B) visualized in the phase space spanned by $(r_1, r_2, r_3)$. The color bar indicates stimulus for three representative levels as in (B) ($\mu_{\text{low}} = 0.1$, $\mu_{\text{med}} = 1.5$, $\mu_{\text{high}} = 3.0$).
Figure 2. Network dynamics are controlled by weight $w_{23}$ of top-down inputs to inhibitory neurons. (A). Changing top-down input $w_{23}$ (A1, thick arrow) reshapes network activity to the same stimulus (color bar indicates values of $w_{23}$); A2: firing rates at three representative values of $w_{23}$ while the stimulus is held constant ($\mu = 1.5$). (B). Trajectories in the phase space for the same $w_{23}$ as in (A); left: oscillations occur at $w_{23} = 5.5$ for $\mu = 1.5$ (same as (A)); right: oscillations occur at $w_{23} = 10.5$ for a different stimulus $\mu = 0.6$, illustrating that the dynamics are diverse across different combinations of stimuli and top-down input. (C). Modulation of oscillation frequency by top-down input (for $\mu = 1.5$). C1:
dependence of frequency on top-down input \( w_{23} \). Inset: time series of \( r_3(t) \) for two example values of \( w_{23} \) (squares). C2: a distribution of oscillation frequencies that can be generated by the network for all possible combinations of stimulus and top-down weight. (D). Similar to (C) but for oscillation amplitudes.
Figure 3. Pattern separation via top-down control of inhibition. (A). A schematic diagram illustrating the separation of response patterns to a pair of stimuli ($\mu_1$ and $\mu_2$) by changing top-down input $w_{23}$ (color bar indicates values of $w_{23}$). (B). Time series of $r_3(t)$ in response to $\mu_1 = 1.0$ and $\mu_2 = 1.1$ at three representative values of $w_{23}$ shows that $r_3(t)$ are close at some top-down weights (left: $w_{23} = 3$; right: $w_{23} = 10$), but pushed apart at other top-down weights (middle: $w_{23} = 3.8$). (C). Response trajectories and network representations of two stimuli in the phase space from which the distance $d$ between representations was calculated. From left to right the top-down input $w_{23}$ correspond to those in (B) for the same stimulus.
pair \((\mu_1 = 1.0 \text{ and } \mu_2 = 1.1)\). (D). Nonmonotonic dependence of the distance \(d\) on
top-down weight \(w_{23}\) (same parameters as in (B)), with maximum achieved at
\(w_{23} = 3.8\). The squares are color coded as in (B) and (C). (E). The matrix \(w_{23}^{\max}\)
maximizes \(d\) between network representations in response to all combinations of
stimuli \((\mu_1 \text{ and } \mu_2)\) presented to the network. Inset: the same matrix of \(w_{23}^{\max}\)
organized by one stimulus \(\mu_1\) vs. stimulus difference \(\Delta \mu\). (F). Dependence of the
stationary firing rate of inhibitory population \(r_2\) on feedback \(w_{23}\) at different levels
of stimulus (indicated by color bar). (G). The correlation between \(w_{23}^{\max}\) obtained
from (E) for all pairs of stimuli \(\mu_1 \text{ and } \mu_2\) and the \(w_{23}^{\text{mid}}\) corresponding to the mid-
point of the two inhibitory firing rate maxima associated with the same pair of stimuli
(upper left inset) reveals that top-down input optimizes pattern separation by gating
the balance between excitation-inhibition and recurrent-inhibition. The grey line
denotes the utility line. Lower right inset shows the correlation coefficient and the
slope of the linear regression.
**Figure 4.** Oscillation synchrony via top-down control. (A). Oscillatory responses to two example stimuli $\mu_1 = 1.15$ and $\mu_2 = 0.6$ become synchronized after changing the top-down input $w_{23}$. A1. Time series of $r_3(t)$ before and after changing the top-down input. A2. Limit cycles corresponding to the oscillatory responses in A1 are plotted in phase space. (B). Changing $w_{23}$ can make both the frequency and amplitude of two oscillations similar to one another. B1. Frequency and amplitude components of the two oscillations shown in (A). B2. Changing top-down input reduces the frequency differences of responses to two distinct stimuli. (C). The matrix $w_{23}^{min}$ which minimizes the distance $d$ between oscillatory responses to all combinations of stimuli $\mu_1$ and $\mu_2$. Inset: the same matrix of $w_{23}^{min}$ organized by...
one stimulus $\mu_1$ vs. stimulus difference $\Delta \mu$. (D). Schematic diagram illustrating that the same value of $w_{23}$ which minimizes the distance between oscillations responding to stimuli $\mu_1$ and $\mu_2$ can maximize the distance between responses to stimuli $\mu_1$ and $\mu_3$. (E). Scatter plot in $\mu_1 - \mu_2 - \mu_3$ space where each sphere size and color corresponds a top-down weight $w_{23}$. (F). Correlation between the differences of stimuli whose response distances are simultaneously minimized and maximized. Inset: correlation coefficient and the slope of linear regression.
Figure 5. Bifurcation mechanism of top-down control to support both stimulus separation and oscillation synchrony. (A). Global phase structures showing the...
vector field (purple arrows), nullclines (yellow thick curves) and several representative trajectories (black thin curves) for the same stimulus $\mu = 1.5$ and three different values of $w_{23}$. Left, $w_{23} = 15$; middle, $w_{23} = 6$; right, $w_{23} = 4$. (B).

Illustration of the bifurcation mechanism and the transition boundary of dynamics in phase space and parameter space. B1, pattern separation. Left: the equilibrium manifold in the firing rate phase space divided by the separatrix from multiple codimension-2 bifurcation points: BT (Bogdanov–Takens bifurcation) into several regions. Region-LC: limit cycle. Region-SE: spiraling equilibria. Region-EE: exponential equilibria. Blue regions: regions where multiple equilibria coexisted.. For two example stimuli $\mu_1 = 2.0$ and $\mu_2 = 3.0$ given in the middle of B2, two paths of equilibria were induced on the equilibrium manifold and traversed across different regions as changing top-down input $w_{23}$. Middle. Different regions on the equilibrium manifold corresponded to different regimes in the parameter space of $\mu$ and $w_{23}$ in the same color scheme (parameters for blue regions in B1 were largely beyond the range in B2 thus not shown). The transition boundary $\Gamma_H$ specified the pair $(\mu, w_{23})$ at which the network underwent a Hopf bifurcation and corresponded to the separatrix enclosing the region-LC in B1. Two given stimuli were denoted by two vertical lines and three example values of $w_{23}$ corresponded to three horizontal dashed lines, giving rise to a pair of junctions for each. These junctions were also plotted as squares in the left of B1 denoting the corresponding $\omega$-limit sets in the same color (solid square: stable equilibrium; empty square: unstable equilibrium). Right: the distance between the $\omega$-limit sets to represent the two given stimuli. The maximal distance was
achieved when the two junctions were on opposite sides of $\Gamma_H$. B2: same as B1

but for oscillation synchrony occurring when the two junctions were both inside $\Gamma_H$. (C). Comparisons between the translated transition boundary $\Gamma_\Delta$ (dashed curve) depending on $\Delta \mu$ (left: $\Delta \mu = 0.1$, middle: $\Delta \mu = 0.5$) and the sliced section of the $w_{23}^{\text{max}}$ at the same $\Delta \mu$ (solid curve). Right: a series of translated $\Gamma_\Delta$ for three representative values of $\Delta \mu$. 

50
Top-down Control of Inhibition Reshapes Neural Dynamics Giving Rise to a Diversity of Computations

Zhen Chen 1, Krishnan Padmanabhan 2

1. Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627;

2. Department of Neuroscience, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642.

* Krishnan Padmanabhan

Email: Krishnan_Padmanabhan@urmc.rochester.edu

Supplementary materials
Stimulus of intermediate strength evoked oscillations of firing rates

Our network model generated diverse dynamics in response to different stimuli $\mu$. As the input increased from low to high values, the steady states of the dynamics transitioned from equilibria to limit cycles and then back to equilibria again (Fig. 1). We plotted firing rates of the top-down population (T) after transient dynamics as a function of input magnitude (Fig S1). An intermediate interval existed where $r_3$ occupied a continuum of firing rates for one stimulus, indicating the emergence of oscillations. However, when the stimulus strength was beyond that interval, either too low or too high, the firing rate $r_3$ achieved a constant value for the stimulus.

Feedback modulation of neural oscillations across different feedforward levels

For an input of intermediate value ($\mu = 1.5$) changing the connection weight $w_{23}$ corresponding to the top-down projection to inhibitory neurons modulated the oscillations over a range of frequencies as well as amplitudes (Fig. 2). The control of both the frequency and amplitude via changes in $w_{23}$ occurred across an array of weights ($w_{31}$) associated the feedforward drive from the local excitatory population (E) to the feedback population (Fig.S2). Interestingly, the range of $w_{23}$ within which neural oscillations emerged became narrower with increasing $w_{31}$, suggesting that the feedforward drive established the dynamic range within which changes in feedback ($w_{23}$) adjusted both the amplitude and frequency of the oscillations that could be generated by the network.
Definition of period of a limit cycle

From the perspective of dynamical system, a limit cycle in the phase space corresponds to the oscillation of firing rates in the temporal space. Since the time constants in our model had units of millisecond, the frequency of oscillations was defined as $1000/T$, with $T$ denoting the period of limit cycles, which was defined as follows: if $r_i(t), i = 1, 2, 3$ denote the firing rate of neural population in the model, then a limit cycle satisfies the periodicity:

$$r_i(t) = r_i(t + T), i = 1, 2, 3$$

for some $T > 0$ and all $t \in \mathbb{R}$. The minimal $T$ for which the above equality holds is the period of the limit cycle.

Definition of Euclidean distance $d^E$ and spectrum distance $d^S$

Here we defined two types of distances, the Euclidean distance $d^E$ and the spectrum distance $d^S$, to measure the separation between the $\omega$-limit sets for different stimuli. The Euclidean distance was defined as follows: supposing that $\Omega_1$ and $\Omega_2$ are two $\omega$-limit sets composed of $N_1$ and $N_2$ discrete points in three-dimensional phase space $(r_1, r_2, r_3)$ respectively, denoted as $\{\alpha_1, \alpha_2, ..., \alpha_{N_1}\}$ and $\{\beta_1, \beta_2, ..., \beta_{N_2}\}$ where $\alpha_i$ and $\beta_j$ are three-dimensional vectors, then

$$d^E = \langle d(\alpha_i, \beta_j) \rangle, i = 1, ..., N_1, j = 1, ..., N_2$$

Where $\langle \cdot \rangle$ denotes the average and $d(x, y) = \sqrt{(x - y)^T(x - y)}$ which is the standard distance between any two points in the three-dimensional Euclidean space. Note that in the case of equilibria, the number of discrete points in the $\omega$-limit set was one: $N = 1$, and
in the case of a limit cycle, we set $N = T/dt$, where $T$ was the period and $dt$ was time bin
for numerical integration.

The Euclidean distance $d^E$ worked well in measuring the distance between two
 equilibrium in the phase space, but when one $\omega$-limit set was a limit cycle, the averaging
operation in the definition made it only a coarse and lagged estimate of the separation for
 equilibrium vs. limit cycle and limit cycle vs. limit cycle. In particular, the activity of the top-
down population (T) should be decodable with respect to the stimulus information,
something that was problematic for when using $d^E$. Therefore, we defined the spectrum
distance $d^S$ to address the question of distances between representations that were
sensitive to those representations being oscillations. To calculate the spectrum distance
$d^S$ between two $\omega$-limit sets $\Omega_1$ and $\Omega_2$, only the sequence of their $r_3$ component was
decomposed by Fourier transform which converted the firing rate signal in the temporal
domain into a representation in the frequency domain.

The single-sided amplitude spectrum for the Fourier transform of the firing rate signal
$r_3(t)$, was used to obtain peaks around frequency values. For an equilibrium
corresponding to constant firing rate $r_3 = A$, there existed only one peak around zero
frequency with its amplitude proportional to $A$, since the Fourier transform of a constant
function is a delta-function. We referred this component as the direct component (DC) of
the signal. For the case of a limit cycle, in addition to one peak around zero frequency,
there existed another peak around frequency $1000/T$ where $T$ denoted the period of the
limit cycle. We referred this additional peak of a limit cycle as the alternating component
(AC). Thus, the spectrum distance $d^S$ was a sum of the differences between both direct
components (DC) of two limit sets and alternating components (AC) of two limit sets, which was formalized as follows: supposing that $D_1$ and $D_2$ were the amplitudes of the peaks at zero frequency for two $\omega$-limit sets $\Omega_1$ and $\Omega_2$, and $a_i = (f_i, A_i)$, $i = 1, 2$ denoted the corresponding alternating components of $\Omega_i$, where $f_i$ was the non-zero frequency and $A_i$ was the amplitude of the peak around $f_i$, then we have

$$d^S = |D_1 - D_2| + d(a_1, a_2)$$

Note that we set $a_i = (0, 0)$ if the $\omega$-limit sets $\Omega_i$ was an equilibrium. Thus, when the two limit sets were both equilibria, $d^S$ only contained the first term measuring the difference between the direct components. In this case, $d^S$ was only a linear projection (up to a constant factor) of the Euclidean distance $d^E$.

**Landscape of distance $d^E$ and $d^S$ over different input pairs**

The network model generated an $\omega$-limit set in response to each input, which was taken as the network representation of the external stimulus. Using the distance defined above, we measured the separation between two network representations given any pair of stimuli: $\mu_1$ and $\mu_2 = \mu_1 + \Delta \mu$. Although the absolute value of $d^E$ between a given stimulus pair was different from the value of $d^S$ between the same pair, the overall trends of the landscapes over all combinations of stimuli given by $d^E$ and $d^S$ were found to be similar (Fig.S4). When the feedback was small ($w_{23} = 1.5$), the landscapes of $d^E$ and $d^S$ varied smoothly across all stimulus combinations. This was because the only form of $\omega$-limit sets in this case was an equilibrium. However, a relatively larger feedback weight ($w_{23} = 4$) unmasked complicated waved structures confined within the range of stimuli ($\mu_1 < 1$). Also, the magnitude of both $d^E$ and $d^S$ were increased by an order, suggesting the
representations reflected larger distances, that would improve the discriminability between two inputs. $d^E$ and $d^S$ consequently quantified different attributes of how one limit set was separated from the other one. The Euclidean distance $d^E$ characterized their relative configurations in the three-dimensional phase space, while the spectrum distance $d^S$ accounted for the oscillatory extent by decomposing the one-dimensional signal into DC and AC. Therefore, $d^S$ was more sensitive to the transitions of the forms of limit sets, as well as the difference of frequencies between two limit cycles. The $w_{23}^{max}$ matrices given by the two distances were compared (Fig.S5) and we found that for most input pairs the differences $\Delta w_{23}^{max}$ remained small. The one given by the Euclidean distance was shown in Fig.3.

Global phase structures - nullplanes and nullclines

For some fixed external stimulus ($\mu = 1.5$), changing the feedback weight $w_{23}$ would change the global phase structures, which in turn gave rise to changes of network dynamics. In our three-dimensional model, three “nullplanes” could be obtained by equaling the right hand side (RHS) of each component of Eqn.(1) to zero, on which the velocity vectors only had two non-zero components (Fig.S6). Furthermore, three “nullclines” were pairwise intersections of nullplanes, on which the velocity vectors only had one non-zero component. Therefore, the equilibrium of the system was given by the intersection of all nullclines, where the velocity was a zero vector. Changing feedback weight $w_{23}$ tilted one nullplane (blue) determined by the second equation of Eqn.(1), thus moving the positions of the intersection point, i.e., the equilibrium. Around $w_{23} = 6$
(middle), the equilibrium was unstable and there existed a limit cycle surrounding it; a Hopf bifurcation occurred during this process.

**Conditions for Hopf bifurcation**

First, starting from $w_{23} = 15$, the linearized Jacobian matrix of the vector field of Eqn.(1) at the equilibrium had a pair of complex-conjugate eigenvalues with negative real parts (the third eigenvalue is negative real number. In this case nearly all trajectories in the phase space spiraled into this equilibrium referred to as a *focus-node* so that the time sequences exhibit a transient phase of damping oscillations (Fig.2A). Then slightly altering the parameter $w_{23}$ caused small variations of the equilibrium position as well as the associated eigenvalues, until some critical point of $w_{23}$ when this pair of complex-conjugate eigenvalues crossed the imaginary axis in the complex plane. As the real parts of the complex pairs became zero, the equilibrium lost stability and a small-amplitude limit cycle branched from that unstable equilibrium, i.e., a Hopf bifurcation.

**Bistable regions in the manifold of equilibria**

There were two blue regions (referred to as BS) on the manifold of equilibrium in Fig.5B, left, which were bounded by saddle-node bifurcations near BT. By choosing the representative parameters as $\mu = -1, w_{23} = 0.2$, we found that the nullclines intersected at three points (Fig.S7A, yellow thick lines), two of which were stable and the other one was a saddle point (with two positive eigenvalues and one negative eigenvalue). Therefore, in this case two stable equilibria coexisted in the phase space and the entire phase space was the divided into two basins of attraction (BOA) by the stable manifold.
of saddle point and the nullplane (Fig.S7B) corresponding to the second equation of Eqn.(1). Trajectories initiating from within one BOA (Fig.S7A, black thin lines) approached the equilibrium belonging to it.
Fig.S1. Transitions of dynamics elicited by varying stimulus strength. External stimulus of intermediate strength (approximately $\mu \in [1.4, 2.5]$) evoked sustained oscillations of firing rates. Individual points indicated constant firing rate and continuous vertical lines indicated oscillations.
Fig. S2. Dependence of oscillation frequency and amplitude on the feedforward and feedback weight (fixing $\mu = 1.5$). Oscillations occurred within an interval of feedback weight $w_{23}$ whose size shrank as the feedforward weight $w_{31}$ increased. (A). Heatmap of frequencies. (B). Heatmap of amplitudes.
Fig.S3. Schematic illustration of direct component (DC) and alternating component (AC) extracted from the time series of $r_3(t)$ by Fourier transform. Top: a representative waveform of oscillatory firing rate signal of $r_3(t)$. Bottom: the components of the signal above. The direct component (DC) at zero frequency was proportional to the average value of the periodic signal. The alternating component (AC) measured the maximum deviation of firing rate from DC, which gave the amplitude of AC, and how frequent that deviation happened, which gave the frequency of AC.
Fig. S4. The evolution of distance landscapes between two $\omega$-limit sets generated by different pairs of stimuli ($\mu_1$ and $\mu_2 = \mu_1 + \Delta \mu$) as the feedback weight $w_{23}$ varied. (A-C).

Landscapes given by the Euclidean distance for three example values of $w_{23}$. (D-F).

Landscapes given by the spectrum distance for the same values of $w_{23}$ as in (A)-(C).
Fig. S5. No qualitative difference between results given by the Euclidean distance and spectrum distance. (A). The non-monotonic dependence of both $d^E$ and $d^S$ on the feedback weight $w_{23}$ for the same pair of input stimuli ($\mu_1 = 1.0, \mu_2 = 1.1$). Both metrics gave similar results of $w_{23}^{\text{max}}$. (B). Matrix of the difference in $w_{23}^{\text{max}}$ given by $d^E$ (Fig.3) and $d^S$. The color bar indicated the absolute values of the differences.
Fig. S6. Global phase structures at different example values of $w_{23}$. Three nullplanes were denoted as transparent surfaces whose intersections gave rise to nullclines. Changing the feedback weight tilted the blue nullplane progressively to change the position of the equilibrium as well as its stability.
Fig. S7. Global phase structures of bistability. (A). Nullclines (yellow thick curves) intersected at three equilibria, two of which were stable and attracting nearby trajectories (black thin curves). The middle equilibrium was a saddle point. Purple arrows indicated the velocities of flow. (B). Nullplanes giving rise to the nullclines in (A).