The Geometry of Plasticity-Induced Sensitization in Isoinhibitory Rate Motifs

Gautam Kumar
gautam.kumar@wustl.edu
ShiNung Ching
shinung@wustl.edu
Electrical and Systems Engineering Washington University in St. Louis,
St. Louis, MO 63130, U.S.A.

A well-known phenomenon in sensory perception is desensitization, wherein behavioral responses to persistent stimuli become attenuated over time. In this letter, our focus is on studying mechanisms through which desensitization may be mediated at the network level and, specifically, how sensitivity changes arise as a function of long-term plasticity. Our principal object of study is a generic isoinhibitory motif: a small excitatory-inhibitory network with recurrent inhibition. Such a motif is of interest due to its overrepresentation in laminar sensory network architectures. Here, we introduce a sensitivity analysis derived from control theory in which we characterize the fixed-energy reachable set of the motif. This set describes the regions of the phase-space that are more easily (in terms of stimulus energy) accessed, thus providing a holistic assessment of sensitivity. We specifically focus on how the geometry of this set changes due to repetitive application of a persistent stimulus. We find that for certain motif dynamics, this geometry contracts along the stimulus orientation while expanding in orthogonal directions. In other words, the motif not only desensitizes to the persistent input, but heightens its responsiveness (sensitizes) to those that are orthogonal. We develop a perturbation analysis that links this sensitization to both plasticity-induced changes in synaptic weights and the intrinsic dynamics of the network, highlighting that the effect is not purely due to weight-dependent disinhibition. Instead, this effect depends on the relative neuronal time constants and the consequent stimulus-induced drift that arises in the motif phase-space. For tightly distributed (but random) parameter ranges, sensitization is quite generic and manifests in larger recurrent E-I networks within which the motif is embedded.

1 Introduction

Desensitization to persistent stimuli is a pervasive aspect of sensory perception (Thompson & Spencer, 1966) and is thought to, in part, mediate the...
detection of novel features in the environment (Rankin et al., 2009; Bradley, Lang, & Cuthbert, 1993; Marsland, Nehmzow, & Shapiro, 2000). Understanding the neural substrate for desensitization is a fundamental question in cognitive neuroscience (Knight, 1984, 1996). At the circuit level, answering this question amounts to a characterization of network sensitivity—how neural circuits respond (over time) to their afferent inputs, mediated through the dynamics of neural plasticity. The goal of this letter is to study how plasticity continually shapes the sensitivity of recurrent excitatory-inhibitory networks, mediating desensitization and high-level effects such as heightened responsiveness to novelty.

At the sensory network level, desensitization (or habituation) has been demonstrated across several sensory modalities (Buchwaldt & Humphrey, 1973; Picton, Hillyard, & Galambos, 1976; Wright et al., 2001; Groves & Thompson, 1970), notably in olfaction (Das et al., 2011; Linster, Menon, Singh, & Wilson, 2009; Sudhakaran et al., 2012). Computational models of such networks have previously revealed basic circuit mechanisms for habituation that involve potentiation of feedback inhibition onto early layers in the network architecture (Lee, Tsunada, & Cohen, 2013; Veale & Scheutz, 2012).

However, the bulk of these efforts center on neuronal responses at the level of individual cells, such as through the characterization of input-output strength-response activation functions. The generalizability of these mechanisms to networks is not as well studied. In a network setting, sensitivity can be measured along any or all directions in the (high-dimensional) space of neuronal activity, that is, the network phase-space. Plasticity can be conceived of as morphing this sensitivity, making neuronal activation in some directions harder and others easier. In other words, plasticity alters the geometry of the network’s responsiveness to extrinsic stimuli.

In this letter, we are principally concerned with elucidating network desensitization that occurs on timescales that are orders of magnitude slower than the natural dynamics of individual neurons. Specifically, we use a combination of computational modeling and systems-level analysis to study how long-term plasticity, formulated via BCM functions, shapes the sensitivity of rate-based networks that are excited by a repetitive stimulus. We focus on recurrent, isoinhibitory motifs, wherein two or more excitatory units are mutually coupled via a single inhibitory unit (see Figure 1A). The study of small motifs is increasingly used in neuronal network analysis, owing to their overrepresentation in larger network topologies (Sporns & Kötter, 2004; Qian, Hintze, & Adami, 2011; Gollo & Breakspear, 2014; Haeusler, Schuch, & Maass, 2009; Whalen, Brennan, Sauer, & Schiff, 2015). The isoinhibitory (or, competitive) motif, in particular, with its inherent recurrent inhibition, is a fundamental building block in the laminar architecture of sensory networks (Turkheimer, Leech, Expert, Lord, & Vernon, 2015; Das et al., 2011). Thus, its analysis can provide a set of mechanistic insights that may be readily generalized to larger E-I network structures.
Figure 1: Characterizing the geometry of sensitivity in isoinhibitory motifs. (A) Three-neuron isoinhibitory motif. (B) Via plasticity, a repetitive conditioned stimulus (e.g., here, applied exclusively to $E_1$) continually shapes the motif sensitivity. (C) We examine the ensuing effect through the geometry of the fixed energy reachable set—the space of inducible trajectories given a stimulus energy. In essence, such a set provides a holistic characterization of how a network tunes itself to an ongoing stimulus. In this schematic, the reachable set contracts in the direction of the conditioned stimulus, while expanding (sensitizing) in the orthogonal direction.

Our study of geometric sensitivity is approached through the lens of control theory and, specifically, reachability analysis. The key object of study is the space of possible (i.e., inducible) output trajectories subject to an input of fixed energy, that is, the so-called reachable set (see Figure 1). Such a set provides a holistic characterization of the network sensitivity with respect to its afferent excitation by describing those regions of the motif phase-space that are most easily accessed. While control-theoretic analyses have been increasingly used to study neural network dynamics (Hennequin, Vogels, & Gerstner, 2014; Whalen et al., 2015; Schiff, 2012; Sritharan & Sarma, 2014), especially in the context of neurostimulation (Ching, Brown, & Kramer, 2012; Ching & Ritt, 2013; Kumar & Ching, 2014; Dasanayake & Li, 2014; Nabi, Mirzadeh, Gibou, & Moehlis, 2013), characterizing how neural plasticity may alter such properties is not well studied. Our overall goal is to characterize how the geometry of the reachable set changes, via the plasticity, as a function of the repetitive (conditioning) stimulus (see Figures 1B and 1C). As we show in our results, it turns out that for certain network dynamics, the geometry of the reachable set changes so as to deemphasize persistent stimuli and enhance (i.e., heighten) responses to orthogonal (novel) stimuli (e.g., depicted schematically in Figure 1C, where $y_1$ and $y_2$ denote activity of $E_1$ and $E_2$, respectively).
Thus, the main contributions of this letter are:

1. A general dynamical systems-based geometric sensitivity analysis in order to quantify directions in the network state-space that are harder and easier to elicit
2. A perturbation analysis framework that links changes in synaptic weights to network sensitivity
3. A specific instantiation of this sensitivity analysis for a generic, overrepresented network motif with long-term plasticity modeled at timescales ranging from $30 \times$ to $400 \times$ the timescales of neurons
4. The demonstration of active sensitization, not specific to the exact timescale of plasticity, wherein the motif becomes more receptive to inputs that are orthogonal to the conditioning stimulus

The remainder of this letter is organized as follows. We describe the rate-based E-I network dynamics, synaptic adaptation rule, and our method of sensitivity analysis in section 2. In section 3, we pursue our geometric characterization of motif sensitivity, highlighting the emergence of orthogonal sensitization. We disassociate the roles of intrinsic motif dynamics and afferent gain in mediating this effect by considering two models for how the afferent input (stimulus) impinges on the motif. Through a simulation study, we demonstrate that analysis of the motif can yield insights that generalize to larger networks. A discussion concludes the letter in section 4.

2 Models and Methods

2.1 Models of Network Dynamics and Neural Plasticity. In this section, we outline the model classes that form the basis of our subsequent development. We used two different firing rate models to represent the neuronal network dynamics and a modified BCM rule to represent the dynamics of synaptic plasticity. We now describe the details of these models.

2.1.1 Firing Rate Model with Additive Synaptic and External Inputs. We consider recurrent, E-I rate networks wherein the dynamics of the $i$th neuron is described by (Dayan & Abbott, 2001):

$$
\tau_i \frac{dr_i}{dt} = -r_i + \frac{c_i}{1 + \exp(-\left(\sum_{j \neq i=1}^{N} (-1)^a w_{ij} r_j + I_{b,i} + u_i(t))\right)}.
$$

(2.1)

Here, $r_i \in [0, 1]$ is the time-varying firing rate of the neuron, $\tau_i$ is the time constant in milliseconds, $c_i$ is the maximum firing rate of the neuron, $w_{ij}$ is the synaptic weight (connectivity strength) of the synapse from neuron $j$ to neuron $i$, $I_{b,i}$ is a background baseline input, $u_i(t)$ is the external input (stimulus) to the neuron $i$, and $N$ is the total number of neurons in the
network. The parameter $\alpha = 1$ if the neuron $j$ is inhibitory and $\alpha = 2$ if the neuron $j$ is excitatory.

2.1.2 Firing Rate Model with Disassociated Input. In equation 2.1, the external input $u_i$ acts on the neuronal dynamics in the same way as other synaptic inputs: through the sigmoidal function. In one of our analyses, we disassociate these inputs (i.e., synaptic and external), by considering the following dynamics of the $i$th neuron:

$$
\frac{dr_i}{dt} = -r_i + \frac{0.5}{1 + \exp(-\sum_{j \neq i=1}^{N}(-1)^{\alpha} w_{ij} r_j))} + \frac{1}{1 + \exp(-(u_i(t) + I_{b,i}))} - \frac{1}{2^2}
$$

(2.2)

It should be noted that $r_i \in [0, 1]$ for $u_i(t) \geq 0$ and $I_{b,i} \geq 0$.

2.1.3 Synaptic Adaptation Rule. To include the effect of long-term plasticity, we consider the following modified Bienenstock-Cooper-Munro (BCM) learning rule (Bienenstock, Cooper, & Murno, 1982) for the activity-dependent dynamical evolution of the synaptic weight $w_{ij}$:

$$
\tau_{ij} \frac{dw_{ij}}{dt} = r_i r_j (r_i - \theta_j) - \epsilon w_{ij},
$$

(2.3a)

$$
\eta \frac{d\theta_i}{dt} = r_i^2 - \theta_i.
$$

(2.3b)

Here, $r_i$ and $r_j$ are firing rates of the postsynaptic and the presynaptic neurons, respectively. The parameter $\theta_i$ is a sliding threshold, and $\tau_{ij}$ and $\eta_i$ are time constants of the synaptic weight dynamics and the sliding threshold, respectively.

Our key modification to the standard BCM rule is the inclusion of a decay term $\epsilon w_{ij}$, $\epsilon > 0$ in equation 2.3a, similar to formulations in Zenke, Hennequin, and Gerstner (2013). For a small $\epsilon$, this term keeps the weights positive and ensures that the systems 2.1 to 2.3, and 2.2 to 2.3, possess a fixed point.

2.2 Sensitivity via Fixed-Energy Finite-Time Reachability Analysis. In order to assess the sensitivity of our isoinhibitory rate motifs, we perform reachability analysis on a local linearization obtained at a prescribed time. We briefly describe the key enabling theory here.

Consider a linear dynamical system,
\[
\frac{dy(t)}{dt} = Ay(t) + Bu(t).
\] (2.4)

Here, \(y(t) \in \mathbb{R}^{N \times 1}\) is the state of the system 2.4, \(u(t) \in \mathbb{R}^{m \times 1}\) is the afferent input vector, and \(A \in \mathbb{R}^{N \times N}\) and \(B \in \mathbb{R}^{N \times m}\) are the state transition matrix and the input gain matrix, respectively.

It is well known that in order to drive system 2.4 from \(y(0)\) at \(t = 0\) to \(y(t_f)\) at \(t = t_f\), the afferent (driving) input \(u(t)\) must have (at least) a minimum energy given by (Kirk, 2012)

\[
\varepsilon(t_f) = (y(t_f) - e^{A t_f} y(0))^T (W_c(t_f))^{-1} (y(t_f) - e^{A t_f} y(0)),
\] (2.5)

where \(W_c(t_f)\), a symmetric positive-definitive matrix, is known as the controllability Gramian:

\[
W_c(t_f) = \int_0^{t_f} e^{A(t_f-t)} BB^T e^{A^T(t_f-t)} dt.
\] (2.6)

Here, \((\cdot)^T\) and \((\cdot)^{-1}\) represent the transpose and the inverse of the underlying argument, respectively. Assuming \(y_0 = 0\), the ellipsoid

\[
\Xi(t_f) = \{ y(t_f) | \varepsilon(t_f) \leq 1 \},
\] (2.7)

centered at the origin, characterizes the unit-energy reachable set—the set of points that could be reached by \(y(t)\) for inputs \(u(t)\) having at most unit energy (i.e., \(\int_0^{t_f} ||u(t)||_2^2 dt \leq 1\)). The geometry of the reachable set thus characterizes the overall sensitivity (measured relative to a given external input subject to a fixed stimulus energy) of system 2.4 by describing the regions of the phase-space that are more easily (in terms of stimulus energy) traversed. In this letter, we study this geometry for the isoinhibitory motif by computing the energies (via equation 2.5) to drive the rate trajectories along cardinal directions—parallel to the \(E_i\) and \(I_i\) axes in the motif phase space. It is worth emphasizing here that there is a unique minimum energy corresponding to each cardinal direction, that is, directions that are parallel \(E_i\) and \(I_i\) axes. We also study the overall sensitivity in terms of the total volume of the ellipsoid, equation 2.7, given by

\[
V(\Xi(t_f)) = \frac{\pi^{N/2}}{\Gamma(N/2 + 1)} \prod_{i=1}^{N} \frac{1}{\sqrt{\lambda_i}}.
\] (2.8)

Here, \(\Gamma(\cdot)\) is the gamma function, and \(\lambda_i\) denote the eigenvalues of \(W_c^{-1}(t_f)\).
2.3 Characterizing Motif Sensitization to a Repetitive Stimulus. We specifically focus on how a repetitive stimulus shapes the sensitivity of the motif (via the plasticity rule, equation 2.3). To do so, we first obtain the fixed point of equations 2.1 to 2.3 (or 2.2 to 2.3) in the absence of external inputs (i.e., $u_i(t) = 0$) in order to obtain the equilibrium synaptic weights as well as the steady-state (baseline) neuronal firing rates.

With a slight abuse of nomenclature, we define a conditioning stimulus $u_a = (a_1, a_2, \ldots, a_N)$, where

$$a_i = 0, \forall i \notin A. \quad (2.9)$$

$A \subset \{1, \ldots, N\}$ denotes the set of conditioned neurons. As per our motif, only excitatory neurons can receive an external input. An orthogonal stimulus $u_o$ has the property that $u_a \cdot u_o = 0$, that is, it impinges on the set of unconditioned neurons, denoted $U \equiv \{1, \ldots, N\} \setminus A$.

We will study the evolution of the motif sensitivity as a function of repetition of the (conditioning) stimulus, which is delivered according to

$$u_i(t) = \begin{cases} a_i & \text{for } t \in (0, T_1] \\ 0 & \text{for } t \in (T_1, T_2]. \end{cases} \quad (2.10)$$

Thus, in each successive repetition, the stimulus is presented for the time duration $(T_1)$, then turned off for the duration $T_2 - T_1$ (see Figure 1B), which is sufficient to allow the firing rates of the motif to return to near steady state (see also Figure 5A).

To characterize the motif sensitivity, we freeze the synaptic weights $w_{i,j}$ in equations 2.1 and 2.2 at the end of each repetition and locally linearize these equations with respect to the state $r_i(t)$ and the input $u_i(t)$ for $i = 1, 2, \ldots, N$ (see the appendix). The linearized network thus takes the form of equation 2.4. We then used equations 2.5 and 2.6 to compute the minimum energy required to drive the linearized equation system, (see the appendix) a distance $\Delta r = 0.01$, for $t_f = 0.3$ ms, in each of the motif cardinal directions, parallel to the $E_i$ and $I_i$ axes. The sensitivity profile is computed as the relative change in minimum energies as a function of $k$, the number of repetitions of the conditioned stimulus:

$$\Delta \varepsilon_{i,k}(\%) = \frac{\varepsilon_{i,k} - \varepsilon_{i,0}}{\varepsilon_{i,0}} \times 100. \quad (2.11)$$

$\Delta \varepsilon_{i,k}(\%) > 0$ and $\Delta \varepsilon_{i,k}(\%) < 0$ imply desensitization (DS) and sensitization (S), respectively, of the motif along the direction $i \in \{E_1, E_2, I\}$. In particular, we can characterize the sensitivity profiles along the conditioned and orthogonal input directions.
3 Results

We now present our main findings. We begin by developing an analytical framework for assessing network sensitivity, which consists of both the reachability analysis presented above and a novel perturbation analysis to link synaptic weights to the reachable set. These results form the basis of subsequent numerical studies that characterize the reachable geometry of the network in the presence of plasticity for the three-neuron isoinhibitory motif. We highlight dynamic sensitization in unconditioned directions. Finally, we generalize our findings by simulating large motifs of E-I networks.

3.1 Approximate Sensitivity Analysis via Perturbation Method. We consider the three-neuron isoinhibitory motif shown in Figure 1A. The rate dynamics of neurons $E_1$ (excitatory), $E_2$ (excitatory), and $I$ (inhibitory) are given by (see equation 2.1)

$$\tau_1 \frac{dr_1}{dt} = -r_1 + \frac{1}{1 + \exp(w_{13}r_3 - u_1(t))}, \quad (3.1a)$$

$$\tau_2 \frac{dr_2}{dt} = -r_2 + \frac{1}{1 + \exp(w_{23}r_3 - u_2(t))}, \quad (3.1b)$$

$$\tau_3 \frac{dr_3}{dt} = -r_3 + \frac{1}{1 + \exp(-w_{31}r_1 - w_{32}r_2)}. \quad (3.1c)$$

Here $r_1$, $r_2$, and $r_3$ are the firing rates of neurons $E_1$, $E_2$, and $I$, respectively. $\tau_1$, $\tau_2$, and $\tau_3$ are the time constant of neurons $E_1$, $E_2$, and $I$, respectively. $w_{13}$, $w_{23}$, $w_{31}$, and $w_{32}$ are synaptic weights from $I$ to $E_1$, $I$ to $E_2$, $E_1$ to $I$, and $E_2$ to $I$, respectively. $u_1(t)$ and $u_2(t)$ are the external stimulus to neurons $E_1$ and $E_2$. We here assumed that $\tilde{I}_{b,i} = 0$ and $c_i = 1$ for $i = 1, 2, 3$ in equation 2.1.

For given synaptic weights $w_{13}$, $w_{23}$, $w_{31}$, and $w_{32}$, the equilibrium firing rates of $E_1$, $E_2$, and $I$ at $u_1 = u_2 = 0$ can be obtained from equation 3.1 as

$$r_{1,s} = \frac{1}{1 + \exp(w_{13}r_{3,s})}, \quad (3.2a)$$

$$r_{2,s} = \frac{1}{1 + \exp(w_{23}r_{3,s})}, \quad (3.2b)$$

$$r_{3,s} = \frac{1}{1 + \exp(-w_{31}r_{1,s} - w_{32}r_{2,s})}. \quad (3.2c)$$

Here $r_{1,s}$, $r_{2,s}$, and $r_{3,s}$ are equilibrium (steady-state) rates of neurons $E_1$, $E_2$, and $I$, respectively.
3.1.1 Effect of Synaptic Weights on Equilibrium Firing Rates. We investigate how the equilibrium firing rates (i.e., $r_{1,s} \rightarrow r_{1,s} + \Delta r_{1}$, $r_{2,s} \rightarrow r_{2,s} + \Delta r_{2}$, $r_{3,s} \rightarrow r_{3,s} + \Delta r_{3}$) change in response to an infinitesimal perturbation in synaptic weights in equation 3.2 (i.e., $w_{13} \rightarrow w_{13} + \Delta w_{13}$, $w_{23} \rightarrow w_{23} + \Delta w_{23}$, $w_{31} \rightarrow w_{31} + \Delta w_{31}$ and $w_{32} \rightarrow w_{32} + \Delta w_{32}$). For this, we performed perturbation analysis and computed $\Delta r_{i}, i = 1, 2, 3$. Final expressions are provided below (see details of the derivations in the appendix):

$$
\Delta r_{1} = - \frac{(r_{3,s} + w_{23}w_{32}r_{3,s}\beta\gamma\delta)\Delta w_{13} + (w_{32}w_{13}r_{3,s}\beta\gamma\delta)\Delta w_{23}}{1 + w_{13}w_{31}\beta\delta + w_{23}w_{32}\gamma\delta},
$$

$$
\Delta r_{2} = \frac{(w_{23}w_{31}r_{3,s}\beta\gamma\delta)\Delta w_{13} - (r_{3,s} + w_{31}w_{13}r_{3,s}\beta\gamma\delta)\Delta w_{23}}{1 + w_{13}w_{31}\beta\delta + w_{23}w_{32}\gamma\delta},
$$

$$
\Delta r_{3} = - \frac{(w_{31}r_{3,s}\beta\delta)\Delta w_{13} - (w_{32}r_{3,s}\gamma\delta)\Delta w_{23} + (r_{1,s})\Delta w_{31} + (r_{2,s})\Delta w_{32}}{1 + w_{13}w_{31}\beta\delta + w_{23}w_{32}\gamma\delta}.
$$

Here $\beta = r_{1,s}(1 - r_{1,s})$, $\gamma = r_{2,s}(1 - r_{2,s})$, and $\delta = r_{3,s}(1 - r_{3,s})$. As an example, consider $\Delta w_{13} > 0$, $\Delta w_{23} = \Delta w_{31} = \Delta w_{32} = 0$. It is easy to see from equation 3.3 that $\Delta r_{1} < 0$, $\Delta r_{2} > 0$, and $\Delta r_{3} < 0$: as the inhibition to $E_{1}$ increases, the equilibrium firing rate of $E_{1}$ and $I$ decreases, while the equilibrium firing rate of $E_{2}$ increases.

3.1.2 Effect of Synaptic Weights on Controllability Gramian. We linearized equation 3.1 at $r_{i} = r_{i,s}$ and $u_{i}(t) = u_{2}(t) = 0$, which takes the form of equation 2.4 with

$$
A_{i,i} = - \frac{1}{\tau_{i}}, \quad i = 1, 2, 3,
$$

$$
A_{1,2} = A_{2,1} = 0,
$$

$$
A_{1,3} = \frac{- w_{13}\exp(w_{13}r_{3,s})}{\tau_{1}(1 + \exp(w_{13}r_{3,s}))^2} = \frac{- w_{13}r_{1,s}(1 - r_{1,s})}{\tau_{1}},
$$


\[ A_{2,3} = \frac{-w_{23} \exp(w_{23}r_{3,s})}{\tau_1 (1 + \exp(w_{23}r_{3,s}))^2} \]

\[ = \frac{-w_{23}r_{2,s}(1 - r_{2,s})}{\tau_2}, \quad (3.4e) \]

\[ A_{3,1} = \frac{w_{31} \exp(-w_{31}r_{1,s} - w_{32}r_{2,s})}{\tau_3 (1 + \exp(-w_{31}r_{1,s} - w_{32}r_{2,s}))^2} \]

\[ = \frac{w_{31}r_{3,s}(1 - r_{3,s})}{\tau_3}, \quad (3.4g) \]

\[ A_{3,2} = \frac{w_{32} \exp(-w_{31}r_{1,s} - w_{32}r_{2,s})}{\tau_3 (1 + \exp(-w_{31}r_{1,s} - w_{32}r_{2,s}))^2} \]

\[ = \frac{w_{32}r_{3,s}(1 - r_{3,s})}{\tau_3}, \quad (3.4i) \]

\[ B_{1,2} = B_{2,1} = B_{3,1} = B_{3,2} = 0, \quad (3.4k) \]

\[ B_{1,1} = \frac{\exp(w_{13}r_{3,s})}{\tau_1 (1 + \exp(w_{13}r_{3,s}))^2}, \quad (3.4l) \]

\[ = \frac{r_{1,s}(1 - r_{1,s})}{\tau_1}, \quad (3.4m) \]

\[ B_{2,2} = \frac{\exp(w_{23}r_{3,s})}{\tau_1 (1 + \exp(w_{23}r_{3,s}))^2}, \quad (3.4n) \]

\[ = \frac{r_{2,s}(1 - r_{2,s})}{\tau_2}, \quad (3.4o) \]

with \( y = (r_1 - r_{1,s}, r_2 - r_{2,s}, r_3 - r_{3,s})^T \). \( A_{i,j} \) is the \((i, j)\)th element of a \(3 \times 3\) matrix, and \( B_{i,j} \) is the \((i, j)\)th element of a \(3 \times 2\) matrix. Again by performing perturbation analysis, we computed the changes in \( A_{i,j} \) and \( B_{i,j} \) as

\[ \Delta A_{1,3} = -\frac{(r_{1,s}(1 - r_{1,s})) \Delta w_{13}}{\tau_1} - \frac{(w_{13}(1 - 2r_{1,s})) \Delta r_{1}}{\tau_1}, \quad (3.5a) \]

\[ \Delta A_{2,3} = -\frac{(r_{2,s}(1 - r_{2,s})) \Delta w_{23}}{\tau_2} - \frac{(w_{23}(1 - 2r_{2,s})) \Delta r_{2}}{\tau_2}, \quad (3.5b) \]

\[ \Delta A_{3,1} = \frac{(r_{3,s}(1 - r_{3,s})) \Delta w_{31}}{\tau_3} + \frac{(w_{31}(1 - 2r_{3,s})) \Delta r_{3}}{\tau_3}, \quad (3.5c) \]

\[ \Delta A_{3,2} = \frac{(r_{3,s}(1 - r_{3,s})) \Delta w_{32}}{\tau_3} + \frac{(w_{32}(1 - 2r_{3,s})) \Delta r_{3}}{\tau_3}, \quad (3.5d) \]
Plasticity-Induced Sensitization in Rate Motifs

\[ \Delta B_{1,1} = \frac{(1 - 2r_{1,s}) \Delta r_1}{\tau_1}, \]  
\[ \Delta B_{2,2} = \frac{(1 - 2r_{2,s}) \Delta r_2}{\tau_2}. \]  

As an example, we again consider \( \Delta w_{13} > 0, \Delta w_{23} = \Delta w_{31} = \Delta w_{32} = 0, \) \( r_{1,s} < 0.5, r_{2,s} < 0.5, \) and \( r_{3,s} > 0.5. \) Using equation 3.2 and 3.5, it is easy to verify that \( \Delta A_{2,3} < 0, \Delta A_{3,1} > 0, \Delta A_{3,2} > 0, \Delta B_{1,1} < 0, \) and \( \Delta B_{2,2} > 0. \) Since \( A_{2,3} < 0, A_{3,1} > 0, \) and \( A_{3,2} > 0, \) both the net inhibition to \( E_2 \) and the net excitation to \( I \) in the linearized network increase with the increase in the inhibition to \( E_1 \) in the nonlinear network. It should be noted here that the sign of \( \Delta A_{1,3} \) (i.e., change in net inhibition to \( E_1 \)) will depend on the balance between the positive term associated with \( \Delta r_1 \) and the negative term associated with \( \Delta w_{13}. \)

Next we performed perturbation analysis of the Lyapunov equation, equation 2.6, to compute deviation in the finite time controllability Gramian with respect to infinitesimal changes in matrices \( A \) and \( B \) that is, \( \Delta A \) and \( \Delta B. \) By substituting \( \Delta W_c(t_f) + \Delta \Delta W_c(t_f), A + \Delta A, \) and \( B + \Delta B \) in equation 2.6 for \( \Delta W_c(t_f), A, \) and \( B, \) respectively, and ignoring terms higher than first-order changes, we obtained

\[ A(\Delta W_c(t_f)) + (\Delta W_c(t_f))A^T + \Delta Q = 0, \]  

where

\[ \Delta Q = \int_0^{t_f} e^{A(t_f-t)}(\Delta A)BB^Te^{A^T(t_f-t)}dt + \int_0^{t_f} e^{A(t_f-t)}BB^Te^{A^T(t_f-t)}(\Delta A)^Tdt \]

\[ + \int_0^{t_f} e^{A(t_f-t)}(B(\Delta B)^T + (\Delta B)B^T)e^{A^T(t_f-t)}dt. \]  

Equation 3.6 is a Lyapunov equation with the unperturbed matrix \( A, \) which can be solved numerically for infinitesimal changes in synaptic weights. It should be noted that \( W_c(t_f) \) and \( \Delta W_c(t_f) \) are functions of both the synaptic weights and the time constant of neurons.

Although not explicitly used here, we note that it is possible to compute changes in synaptic weights that affect the eigenvalues and eigenvectors of \( W_c(t_f) \) via a perturbation analysis on

\[ W_c(t_f)x_i = \lambda_i x_i. \]  

Here \( x_i \) is the \( i \)th eigenvector of \( W_c(t_f) \) corresponding to the \( i \)th eigenvalue \( \lambda_i. \) By replacing \( W_c(t_f), \lambda_i \) and \( x_i \) with \( W_c(t_f) + \Delta W_c(t_f), \lambda_i + \Delta \lambda_i \) and \( x_i + \Delta x_i, \) we obtained

\[ W_c(t_f) + \Delta W_c(t_f) = (\lambda_i + \Delta \lambda_i) (x_i + \Delta x_i). \]
\[ \Delta x_i = x_i^T (\Delta W_c(t_f)) x_i, \quad (3.9a) \]

\[ \Delta \lambda_i = \frac{\sum_{j \neq i} x_j^T (\Delta W_c(t_f)) x_j}{\lambda_i - \lambda_j} x_i, \quad (3.9b) \]

It should be noted that equations 3.6 and 3.9 are generalizable for \( n > 3 \)-dimensional linear systems (see the appendix for derivations).

Together, equations 3.2 to 3.9 provide an approximate method to compute changes in eigenvalues and eigenvectors of the controllability Gramian \( W_c(t_f) \) with respect to infinitesimal changes in synaptic weights.

3.1.3 Effect of Inhibition on the Motif’s Directional Sensitivity. We study how the increase in inhibition to \( E_1 \) affects the geometry of the motif’s sensitivity to afferent stimuli. It is clear from equations 3.2 to 3.6 that synaptic weights and time constants of neurons have independent effects on the controllability Gramian \( W_c(t_f) \).

Our first step is to study how a perturbation in synaptic weights manifests in terms of the linearized system. We consider \( \Delta w_{13} > 0, \Delta w_{23} = \Delta w_{31} = \Delta w_{32} = 0 \) in equations 3.2 to 3.7 with \( r_{1,s} < 0.5, r_{2,s} < 0.5, \) and \( r_{3,s} > 0.5 \). That is, we formulate a perturbation in the synaptic weight from \( I \rightarrow E_1 \) only. As a function of this perturbation, the equilibrium firing rate of \( E_1 \) (i.e., \( r_{1,s} \)) and \( I \) (i.e., \( r_{3,s} \)) decreases while the equilibrium firing rate of \( E_2 \) (i.e., \( r_{2,s} \)) increases. This effect is shown in Figure 2 for the example parameterization \( \tau_1 = \tau_2 = 25.855 \text{ms}, \tau_3 = 11.7223 \text{ms}, w_{23} = 6, w_{31} = w_{32} = 30, w_{13} \in [6, 10], \) and \( t_f = 0.3 \text{ms} \). Also, the net inhibition to \( E_2 \) (i.e., \( A_{2,3} \)) and the net excitation to \( I \) (i.e., \( A_{3,1} \) and \( A_{3,2} \)) increase while the net inhibition to \( E_1 \) (i.e., \( A_{1,3} \)) decreases. Moreover, the external input gain (i.e., elements of \( B \) matrix) to \( E_1 \) decreases while the input gain to \( E_2 \) increases.

It is important to note that while the perturbation is only to the inhibition to \( E_1 \), the effect of this appears in all connections in the linearized motif due to nonlinearity in the dynamics and recurrent connections.

Next, we link the linearized system to the controllability Gramian. For this, we approximate the matrix exponential \( \exp(A(t-t_f)) \) in the controllability Gramian, equation 2.6, by \( I + A(t-t_f) \), that is, using the first two terms of the Taylor series expansion. With this, we can compute the Gramian and derive \( \varepsilon(E_1), \varepsilon(E_2), \) and \( \varepsilon(I) \) analytically:

\[
\varepsilon(E_1) = \frac{4A_{3,1}^2 B_{1,1}^2 (t_f A_{2,2}^2 + 3t_f A_{2,2} + 3) + 3A_{3,2}^2 B_{2,2}^2}{t_f B_{1,1}^2 (A_{3,1}^2 B_{1,1}^2 (t_f A_{2,2}^2 + 3t_f A_{2,2} + 3) + A_{3,2}^2 B_{2,2}^2 (t_f A_{1,1}^2 + 3t_f A_{1,1} + 3))},
\]

(3.10a)
Plasticity-Induced Sensitization in Rate Motifs

Figure 2: Changes in firing rates and connectivity matrices of the linearized network as the inhibition to $E_1$ increases.

$$
\varepsilon_{(E_2)}(\tau_2) = \frac{4A_{3,2}^2B_{2,2}^2(t_2^2A_{1,1}^2 + 3t_2fA_{1,1} + 3) + 3A_{3,1}^2B_{1,1}^2}{t_2B_{2,2}^2(A_{3,1}^2B_{1,1}^2(t_2^2A_{2,2}^2 + 3t_2fA_{2,2} + 3) + A_{3,2}^2B_{2,2}^2(t_2^2A_{1,1}^2 + 3t_2fA_{1,1} + 3))},
$$
(3.10b)

$$
\varepsilon_{(E_2)}(\tau_1) = \frac{4(t_2^2A_{1,1}^2 + 3t_2fA_{1,1} + 3)(t_2^2A_{2,2}^2 + 3t_2fA_{2,2} + 3)}{t_2^2(A_{3,1}^2B_{1,1}^2(t_2^2A_{2,2}^2 + 3t_2fA_{2,2} + 3) + A_{3,2}^2B_{2,2}^2(t_2^2A_{1,1}^2 + 3t_2fA_{1,1} + 3))},
$$
(3.10c)

Together, equations 3.4 and 3.10 provide an analytical framework to link changes in synaptic weights to motif sensitivity. As an example, we consider the same set of parameters used to generate Figure 2, with the addition of allowing $\tau_2$ to vary. We find that the motif always desensitizes ($\Delta\varepsilon_{(E_2)}(\%) > 0$) in the $E_1$ direction irrespective of the ratio of $\tau_2$ and $\tau_1$, as shown in Figure 3A, while we find four qualitatively different sensitivity regimes along the $E_2$ direction: desensitization (DS) ($\Delta\varepsilon_{(E_2)}(\%) > 0$), sensitization (S) ($\Delta\varepsilon_{(E_2)}(\%) < 0$), desensitization followed by sensitization...
The motif sensitivity in the \( f \) direction, we find three qualitatively different sensitivity regimes: desensitization (DS, \( \tau_{2} < \tau_{1} \)), sensitization (S, \( \tau_{2} > \tau_{1} \)), and desensitization followed by sensitization (DS-S, \( \tau_{2} > \tau_{1} > \tau_{1} \)).

Figure 3: Effect of increased inhibition to \( E_{1} \) on the directional sensitivity of the motif of the three neurons. (A) The motif always desensitizes (i.e., \( \Delta \varepsilon_{(E_{1})}^{\text{DS}}(\% > 0) \) in the \( E_{1} \) direction. (B, C) The motif sensitivity in the \( E_{2} \) and \( f \) directions depend on the ratio of the time constants of excitatory neurons (i.e., \( \frac{\tau_{2}}{\tau_{1}} \)).

(DS-S), and sensitization followed by desensitization (S-DS) corresponding to \( \frac{\tau_{2}}{\tau_{1}} = 2, \frac{\tau_{2}}{\tau_{1}} = 0.5, \frac{\tau_{2}}{\tau_{1}} = 1 \), and \( \frac{\tau_{2}}{\tau_{1}} = 4 \), respectively, as shown in Figure 3B. In the \( f \) direction, we find three qualitatively different sensitivity regimes: desensitization (DS, \( \frac{\tau_{2}}{\tau_{1}} = 2, 4 \)), sensitization (S, \( \frac{\tau_{2}}{\tau_{1}} = 0.5 \)), and desensitization followed by sensitization (DS-S, \( \frac{\tau_{2}}{\tau_{1}} = 1 \)), as shown in Figure 3C.

To understand this dependence on time constants, we further simplify equation 3.10 for the case when \( w_{31} = w_{32} \). The resultant expressions are

\[
\varepsilon_{(E_{1})} = \left( \frac{1}{t_{f} B_{1,1}^{2}} \right) \left( 1 \right) \left( \frac{4 \left( \frac{t_{f}^{2}}{\tau_{1}^{2}} - 3 \frac{t_{f}^{2}}{\tau_{2}^{2}} + 3 \right) \left( \frac{r_{3,1}^{-1}(1-r_{3,1})}{r_{3,1}^{-1}(1-r_{3,2})} \right)^{2} \left( \frac{\tau_{2}}{\tau_{1}} \right)^{2} \right) \right),
\]

(3.11a)

\[
\varepsilon_{(E_{2})} = \left( \frac{1}{t_{f} B_{2,2}^{2}} \right) \left( 1 \right) \left( \frac{4 \left( \frac{t_{f}^{2}}{\tau_{1}^{2}} - 3 \frac{t_{f}^{2}}{\tau_{2}^{2}} + 3 \right) \left( \frac{r_{3,1}^{-1}(1-r_{3,1})}{r_{3,1}^{-1}(1-r_{3,2})} \right)^{2} \left( \frac{\tau_{2}}{\tau_{1}} \right)^{2} \right) \right),
\]

(3.11b)

\[
\varepsilon_{(f)} = \left( \frac{1}{t_{f} B_{2,2}^{2} A_{3,1}} \right) \left( 1 \right) \left( \frac{4 \left( \frac{t_{f}^{2}}{\tau_{1}^{2}} - 3 \frac{t_{f}^{2}}{\tau_{2}^{2}} + 3 \right) \left( \frac{t_{f}^{2}}{\tau_{2}^{2}} - 3 \frac{t_{f}^{2}}{\tau_{1}^{2}} + 3 \right) \left( \frac{r_{3,2}^{-1}(1-r_{3,2})}{r_{3,2}^{-1}(1-r_{3,1})} \right)^{2} \left( \frac{\tau_{1}}{\tau_{2}} \right)^{2} \right) \right).
\]

(3.11c)

Let us assume \( \tau_{1} \gg t_{f} \) and \( \tau_{2} \gg t_{f} \); the time constants of excitatory neurons are much larger than the timescale of the gramian. In this case,
\[
\frac{t_f^2}{t_0^2} - 3 \frac{t_f}{t_1} + 3 \approx 3 \text{ and } \frac{t_f^2}{t_2^2} - 3 \frac{t_f}{t_2} + 3 \approx 3.
\]
With this, we analyze equation 3.11 when \( \frac{t_f}{t_0} \ll 1 \) or \( \frac{t_f}{t_2} \gg 1 \). In either case, it is straightforward to see from equation 3.11a that \( \varepsilon(E_1) \approx \frac{1}{I_f B_{21}^0} \). Then, since \( B_{11} \) decreases (see Figure 2), the motif desensitizes along the \( E_1 \) direction (i.e., \( \Delta \varepsilon(E_1) > 0 \)). Similarly from equation 3.11b, we obtain \( \varepsilon(E_2) \approx \frac{1}{I_f B_{22}^0} \). Since \( B_{22} \) increases (see Figure 2), \( \varepsilon(E_2) \) decreases (i.e., \( \Delta \varepsilon(E_2) < 0 \)); the motif sensitizes along the \( E_2 \) direction. Finally we analyze the sensitivity in the \( I \) direction. For \( \frac{t_f}{t_1} \ll 1 \), we obtain \( \varepsilon(I) \approx \frac{12}{I_f B_{12}^0 A_{31}^1} \) from equation 3.11c. Since both \( B_{22} \) and \( A_{31} \) increase, \( \varepsilon(I) \) decreases (i.e., \( \Delta \varepsilon(I) < 0 \)), and thus the motif sensitizes along the \( I \) direction. For \( \frac{t_f}{t_1} \gg 1 \), we obtain \( \varepsilon(I) \approx \frac{12}{I_f B_{12}^0 A_{31}^1} \). Since \( B_{11} \) decreases and \( A_{31} \) increases, the increase or the decrease in \( \varepsilon(I) \) with an increase in \( w_{13} \) will depend on the trade-off between \( B_{11} \) and \( A_{31} \). For intermediate values of \( \frac{t_f}{t_1} \), it is difficult to make any substantive conclusions on the motif sensitivity by analyzing equation 3.11.

Thus, a key take-away from this analysis is that knowing how the synaptic weights change as a function of time is not sufficient to fully characterize the motif sensitivity.

### 3.1.4 Effect of Excitation on the Motif’s Directional Sensitivity

We repeat the above analysis for an increase in excitation to \( I \) from \( E_1 \) and study how this increase in excitation affects the geometry of the motif’s sensitivity to afferent stimuli. For this, we consider \( \Delta w_{31} > 0, \Delta w_{13} = \Delta w_{23} = \Delta w_{31} = 0 \) in equation 3.2 to 3.7 with \( r_{1,5} < 0.5, r_{2,5} < 0.5, \) and \( r_{3,5} > 0.5 \). As a result, the equilibrium firing rate of \( E_1 \) and \( E_2 \) decreases, while the equilibrium firing rate of \( I \) increases (clear from equation 3.4). Also, the net inhibition to \( E_1 \) and \( E_2 \) and the net excitation to \( I \) decrease.

Using equations 3.4 and 3.10, we compute \( \varepsilon(E_1), \varepsilon(E_2), \) and \( \varepsilon(I) \) for the same set of parameters we used in the previous section, except \( w_{13} = w_{23} = 6, w_{32} = 30, \) and \( w_{31} \in [30, 35] \). We find that the motif always desensitizes in the \( E_1 \) direction irrespective of the ratio of \( t_2 \) and \( t_1 \), as shown in Figure 4A. In the \( E_2 \) direction, we find two qualitatively different sensitivity regimes: desensitization (DS) and sensitization (S) corresponding to \( \frac{t_2}{t_1} = 0.5, 4 \), and \( \frac{t_2}{t_1} = 1, 2 \), respectively, as shown in Figure 4B. In the \( I \) direction, again we find two qualitatively different sensitivity regimes: desensitization (DS, \( \frac{t_f}{t_1} = 0.5 \)) and sensitization (S, \( \frac{t_f}{t_1} = 1, 2, 4 \)) as shown in Figure 4C.

We follow a similar argument as developed above to understand how sensitivity depends on neuronal time constants. Again, we assume \( t_1 \gg t_f \) and \( t_2 \gg t_f \) such that \( \frac{t_f^2}{t_1^2} - 3 \frac{t_f}{t_1} + 3 \approx 3 \) and \( \frac{t_f^2}{t_2^2} - 3 \frac{t_f}{t_2} + 3 \approx 3 \). With this, we
analyze equation 3.10 for the case when $\frac{\tau_2}{\tau_1} \ll 1$ or $\frac{\tau_2}{\tau_1} \gg 1$. It is not difficult to see from equation 3.10a that $\epsilon(E_1) \approx \frac{1}{\tau_1 B_{1,1}}$. Since $B_{1,1}$ decreases, the motif desensitizes along $E_1$ direction. Similarly, from equation 3.10, we obtain $\epsilon(E_2) \approx \frac{1}{\tau_2 B_{2,2}}$. Again since $B_{2,2}$ decreases, $\epsilon(E_2)$ increases. Thus, the motif desensitizes along the $E_2$ direction. Finally we analyze the sensitivity in the $I$ direction. For $\frac{\tau_2}{\tau_1} \ll 1$, we obtain $\epsilon(I) \approx \frac{12}{\tau_1 B_{2,2} A_{3,2}}$ from equation 3.10c. Since both $B_{2,2}$ and $A_{3,2}$ decrease, $\epsilon(I)$ increase, and thus the motif desensitizes along the $I$ direction. For $\frac{\tau_2}{\tau_1} \gg 1$, we obtain $\epsilon(I) \approx \frac{12}{\tau_2 B_{1,1} A_{3,1}}$. Since $B_{1,1}$ decreases and $A_{3,1}$ increases, the increase or the decrease in $\epsilon(I)$ with increase in $w_{31}$ will depend on the trade-off between $B_{1,1}$ and $A_{3,1}$. Again, for intermediate values of $\frac{\tau_2}{\tau_1}$, it is difficult to make any definitive conclusions on the motif sensitivity by analyzing equation 3.10.

### 3.2 Plasticity Induces Sensitization to Novel Stimuli

The above analysis demonstrates that a motif may exhibit geometric sensitization that depends nontrivially on differential effects of synaptic weights and intrinsic time constants. The analysis predicts that neural plasticity, through its effects on synaptic weights, will cause a motif to desensitize in conditioned directions, while sensitizing in unconditioned directions, depending to some degree on the inherent timescales of neurons. To verify this and complete this study, we proceed to numerically evaluate motif sensitivity for networks endowed with long-term plasticity.

#### 3.2.1 Change in the Reachable Set Geometry Mediates Directional Sensitization

Figure 5 shows an example of how the fixed-energy reachable set of the three-neuron motif is shaped by ongoing long-term plasticity with respect to the repetitive conditioned stimulus (see the appendix for parameters). We note that in most of our simulations, we choose a timescale of plasticity...
Figure 5: Plasticity shapes the geometry of sensitivity in a threee-neuron isoinhibitory motif. (A) Each stimulus repetition elicits a firing rate trajectory, which adapts as a function of the number of repetitions, \( k \). (B) The local reachable set (an ellipsoid), depicted as a collection of inhibitory isoclines, computed at the time of the \( k \)th stimulus onset. Over repetitions, the reachable set contracts in the conditioned (i.e., \( E_1 \) direction) but also expands in the direction of the unconditioned neuron \( E_2 \).

that is 30 times the intrinsic timescales of neurons. This is much shorter than some of the common timescales used to model long-term plasticity, which are on the scale of minutes and hours (Zenke et al., 2013). However, our choice of timescale provides a computationally tractable way to explore the
different types of qualitative sensitization trajectories that may arise due to plasticity. Later in the discussion, we expand on this modeling choice and show that the findings are preserved for longer timescales of plasticity.

In Figure 5, the reachable set is depicted in terms of inhibitory isoclines: contours in the statespace associated with constant levels of inhibition. At $k = 0$ (conditioned stimulus not yet applied), the reachable set is biased in the direction of $E_1$: the major axis of the reachable ellipsoid is parallel to the $E_1$-axis. The conditioned stimulus is then applied to $E_1$ repetitively and persistently (as described in section 2). As $k$ (i.e., the number of stimulus repetitions) increases, we observe a change in the geometry of the reachable set: it contracts along the $E_1$-axis while expanding along the $E_2$-axis. In other words, the motif desensitizes to the conditioned stimulus but sensitizes to an orthogonal one (i.e., a stimulus that would impinge on $E_2$).

3.2.2 Sensitization/Desensitization Depends on Relative Timescales. In order to elucidate the underpinnings of orthogonal sensitization, we consider a simulation study of 100 different motif parameterizations. Figure 6A depicts the change in the reachable set geometry for each realization, quantified as the relative change in the minimum stimulus energy, $\Delta \varepsilon_i(i)$ for $i \in \{E_1, E_2, I\}$, required to elicit a response in each of the cardinal directions (i.e., along the $E_1$, $E_2$, and $I$ axes). Here, without loss of generality, the conditioned stimulus is delivered to $E_1$ only. The baseline current was set as $I_{b,i} = 0$.

Perhaps intuitively, the motif always desensitizes in the direction of $E_1$ — the conditioned, persistent stimulus (see Figure 6A). In contrast, the change in sensitivity along the $E_2$ and $I$ axes (not shown in Figure 6A) depict a more complex temporal dependency, with either monotone or biphasic sensitization and desensitization profiles with respect to $k$.

To disentangle the profiles shown in Figure 6A, we cluster them into qualitatively similar groups. As shown in Figure 6B, the reachable set geometry along the $E_2$ axis separates into four regimes: desensitization (DS), sensitization (S), desensitization followed by sensitization (DS-S), and sensitization followed by desensitization (S-DS). Further, the clustering analysis indicates that these regimes are directly associated with the ratio of the time constants of $E_1$ and $E_2$ (i.e., $\frac{\tau_2}{\tau_1}$), as depicted in Figure 6B.

Figure 6C shows the profiles for the synaptic weights over repetitions. As shown, these weight profiles do not correlate in an obvious fashion to the sensitivity profile clusters in Figure 6B. Notably, while we see overall potentiation in the recurrent weight from $E_1$ to $I$ and back to $E_1$ (i.e., an effective increase in recurrent inhibition), we do not observe any significant attenuation in recurrent inhibition onto $E_2$ (see also section 4).

Figure 6D illustrates the overall sensitivity profile of the motif—the change in volume of the reachable ellipsoid. We observe a consistent reduction in volume, indicating that, overall, the motif tends to desensitize
Figure 6: Motif sensitivity is not overtly related to weight trajectories. (A) Sensitivity profiles for 100 random parameterizations, quantified in terms of the minimum stimulus energy required to elicit a response from each excitatory neuron, $\Delta \varepsilon_{(i)}$ for $i \in \{E_1, E_2\}$, expressed relative to $k = 0$. The motif always desensitizes along the $E_1$ direction, while its responses in the $E_2$ direction assume a more varied character. (B) Clustering analysis over the sensitivity profiles for the unconditioned neuron, $\Delta \varepsilon_{(E_2)}$. These responses can be separated into four qualitative responses: (S) sensitization, (DS) desensitization, (S-DS) sensitization followed by desensitization, and (DS-S) desensitization followed by sensitization. Recall that a negative $\Delta \varepsilon_{(E_2)}$ implies that over repetitions, it would require less energy to elicit a response in the unconditioned neuron $E_2$. The four response types are directly associated with the time constants of the excitatory neurons. (C) Changes in synaptic weights over repetitions $k$. There is no systematic relationship between the sensitivity profile clusters and the weight profiles. (D) The overall sensitization profile, defined as the change in the volume of the reachable set. On average, the motif desensitizes.

to input stimuli. Thus, the orthogonal sensitization observed in Figure 6B implies that the reachable ellipsoid becomes eccentric, with the major axis aligned transverse to the conditioned stimulus.

Figure 7 shows that the reachable set geometry along the $I$-axis separates into three regimes: desensitization (DS), sensitization (S), and desensitization followed by sensitization (DS-S). Further, the clustering analysis indicates that these regimes are also directly associated with the ratio of the time constants of $E_1$ and $E_2$ (i.e., $\frac{\tau_2}{\tau_1}$).
Figure 7: Sensitivity profiles for 100 random parameterizations, quantified in terms of the minimum stimulus energy required to elicit a response from the inhibitory neuron, $\Delta \varepsilon_{I}$, expressed relative to $k = 0$. The motif responses in the $I$ direction assume a more varied character. (Top panel) Clustering analysis over the sensitivity profiles for the inhibitory neuron, $\Delta \varepsilon_{I}$. These responses can be separated into three qualitative responses: (S) sensitization, (DS) desensitization, and (DS-S) desensitization followed by sensitization. The three response types are directly associated with the time constants of the excitatory neurons.

To separate out the contribution of $E \rightarrow I$ plasticity from $I \rightarrow E$ plasticity in Figures 6 and 7, we performed simulations for the case where only $I \rightarrow E_1$ and $I \rightarrow E_2$ synaptic connections are plastic. We used the same simulation parameters, including the initial conditions, as before. Figure 8 shows the change in the reachable set geometry for each realization, quantified as the relative change in the minimum stimulus energy, $\Delta \varepsilon_i$ for $i \in \{E_1, E_2, I\}$, required to elicit a response in each of the cardinal directions (i.e., along the $E_1$, $E_2$, and $I$ axes). As we notice in Figure 8C, the motif mostly desensitizes in the $E_2$ and $I$ directions compared to Figures 6 and 7.

We also performed simulations for the case where only the $E_1 \rightarrow I$ and $E_2 \rightarrow I$ synaptic connections are plastic. Figure 9 shows the change in the reachable set geometry for each realization, quantified as the relative change in the minimum stimulus energy, $\Delta \varepsilon_i$ for $i \in \{E_1, E_2, I\}$, required to elicit a
Figure 8: Motif sensitivity profiles when only the $I \rightarrow E_1$ and $I \rightarrow E_2$ connections are plastic. (A) Sensitivity profiles for 100 random parameterizations, quantified in terms of the minimum stimulus energy required to elicit a response from each excitatory neuron, $\Delta \varepsilon_i$ for $i \in \{E_1, E_2, I\}$, expressed relative to $k = 0$. The motif always desensitizes along the $E_1$ direction, while its responses in the $E_2$ direction assume a more varied character. (B) Changes in synaptic weights over repetitions $k$. (C) Clustering analysis over the sensitivity profiles for $E_2$ neuron, $\Delta \varepsilon_{(E_2)}$. These responses can be separated into four qualitative responses: (S) sensitization, (DS) desensitization, (S-DS) sensitization followed by desensitization, and (DS-S) desensitization followed by sensitization. (D) Clustering analysis over the sensitivity profiles for $I$ neuron, $\Delta \varepsilon_{(I)}$. These responses can be separated into three qualitative responses: S, DS, DS-S.

3.2.3 Sensitization Is Qualitatively Invariant to Baseline Stimulus. Above, we controlled the baseline stimulus to be the same across realizations, so that all realizations exhibited the same initial synaptic weights and the steady-state (baseline) firing rates of neurons (at $k = 0$). We repeated our simulation study, allowing the baseline stimulus to vary along with all others. As suggested by our clustering analysis, we sampled $(\tau_1, \tau_2)$ pairs along a uniform grid and generated 100 random realizations at each grid point (see Figure 10A). Again, the conditioned (persistent) stimulus was applied to $E_1$ without loss of generality.
Figure 9: Motif sensitivity profiles when only $E_1 \rightarrow I$ and $E_2 \rightarrow I$ connections are plastic. (A) Sensitivity profiles for 100 random parameterizations, quantified in terms of the minimum stimulus energy required to elicit a response from each excitatory neuron, $\Delta \epsilon_{(i)}$ for $i \in \{E_1, E_2, I\}$, expressed relative to $k = 0$. The motif always desensitizes along the $E_1$ direction. (B) The motif always sensitizes along the $E_2$ direction. (C) The motif always sensitizes along the $I$ direction. (D) Changes in synaptic weights over repetitions $k$.

We observe from Figure 10 that randomization over baseline currents has little qualitative effect on the motif sensitivity profiles (see Figure 10B). The motif always desensitizes to the conditioned stimulus (i.e., in the $E_1$ direction) and exhibits tuning in the orthogonal ($E_2$) direction that depends on the $\tau_2/\tau_1$ ratio (see Figure 10B). In particular, we observe that desensitization (DS) or transient sensitization followed by desensitization (S-DS) occur when $\tau_2/\tau_1 > 1$, while sensitization (S) occurs for $\tau_2/\tau_1 < 1$. Importantly, $\tau_2/\tau_1 \approx 1$ is associated with brief transient desensitization followed by prolonged sensitization (DS-S).

3.3 Sensitization Is Mediated by Adaptation in Internal Dynamics and Not by Afferent Gain. In equation A.6, both the state matrix $A(W, r_s)$ and the afferent input gain matrix $B(W, r_s)$ depend on the synaptic weight matrix $W$. Thus, both $A(W, r_s)$ and $B(W, r_s)$ are continuously modulated as
a function of stimulus repetitions (see also section 3.1). Since the afferent input affects the motif through \( B(W, r_s) \), a small change in \( B(W, r_s) \) can significantly affect the sensitivity of the network via equations 2.5 and 2.6. Thus, it is not clear whether the major contribution to the motif sensitivity profiles is from \( A(W, r_s) \) or \( B(W, r_s) \). To disambiguate this, we characterize the motif sensitivity in the case where the synaptic and the external inputs are disassociated, as in equations 2.2 and 2.3. In this case, the input gain matrix \( B \) in equation A.12 is independent of the synaptic weight matrix \( W \); consequently, changes in the reachable set geometry over time are mediated strictly by the intrinsic dynamics.

Figure 11A shows the sensitivity profiles \( \Delta \varepsilon_{(E_1)} \) and \( \Delta \varepsilon_{(E_2)} \) for 100 random parameterizations of the disassociated input motif, again assuming the conditioned stimulus acts on \( E_1 \). In contrast to the four profiles exhibited before, here we observe that the motif always sensitizes in the direction of \( E_2 \), orthogonal to the conditioned stimulus. As before, the motif exhibits desensitization in the \( E_1 \) direction. Thus, for the disassociated input motif, plasticity mediates generic sensitization in the direction of the unconditioned neuron \( E_2 \) strictly due to changes in the intrinsic dynamics. Figure 11B illustrates the overall sensitivity profile of the motif, where we observe that the reachable ellipsoid expands over repetitions (in contrast to Figure 6D).

Figure 11C depicts the corresponding evolution in synaptic weights for each realization of the disassociated input model. Again, we observe that the net recurrent inhibition to the conditioned neuron clearly potentiates over repetitions. However, again, little effect, and certainly no reduction, is seen for the recurrent inhibition to the orthogonal neuron (i.e., \( E_2 \)), despite
Figure 11: Generic orthogonal sensitization for motif with disassociated input model. (A) The sensitivity profiles $\Delta \varepsilon_{(E_1)}$ and $\Delta \varepsilon_{(E_2)}$ for 100 random realizations, indicating desensitization to the conditioned stimulus, and sensitization in the direction of the unconditioned neuron $E_2$. (B) The overall sensitivity profile, characterized as the change in volume of the reachable set, $\Delta V_k$. (C) Changes in synaptic weights over repetitions $k$.

the motif sensitizing in the $E_2$ direction (i.e., $\Delta \varepsilon_{(E_2)}$), as shown in Figure 11A. Thus, the mechanism for orthogonal sensitization here is not readily apparent from simple examination of the synaptic weight profiles.

### 3.4 Motif Sensitization Manifests in Larger E-I Networks.

In the previous sections, we focused on studying the generic three-neuron iso-inhibitory motif 1A as a building block for larger E-I networks. To demonstrate how our results may generalize to larger networks, we consider a 10-neuron network, consisting of 8 excitatory and 2 inhibitory neurons (see Figure 12A), with no E-E connectivity and no external inputs to inhibitory neurons. The dynamics of the neurons as well as the long-term synaptic plasticity in the network are given by equations 2.1 and 2.3.

We assume that the time constants of the excitatory neurons are chosen randomly to within 10% of each other (see the appendix). From our analysis of the 3-neuron motif, we thus expect that any repetitive conditioned stimulus would generically lead to a long-term sensitization in directions of unconditioned neurons. To verify this prediction, we simulate 100 random realizations of the 10-neuron network and compute the sensitivity profiles along each of the excitatory cardinal directions. For each realization, the conditioned stimulus impinges on three of the eight excitatory neurons.

Figure 12B shows the mean and range (minimum to maximum) of the sensitivity profiles in the direction of the conditioned $(\Delta \varepsilon_{(E_1)})$, unconditioned $(\Delta \varepsilon_{(E_2)})$, and inhibitory $(\Delta \varepsilon_{(E_3)})$ neurons. Consistent with our prediction, we observe that the network desensitizes in the conditioned direction, while it sensitizes along the directions of all of the unconditioned neurons.
Figure 12: Plasticity-induced directional sensitivity in larger E-I networks. (A) E-I recurrent structure with eight excitatory and two inhibitory neurons and no E-E connectivity. (B) Sensitivity profiles $\Delta \varepsilon^{(E_A)}(\%)$ and $\Delta \varepsilon^{(E_U)}(\%)$ for the conditioned and unconditioned neurons, respectively. The shaded region indicates the entire range (minimum to maximum) of profiles over 100 random realizations. The overall sensitivity profile $\Delta V_k$ and the sensitivity profile for the inhibitory neurons $\Delta \varepsilon^{(E_I)}(\%)$ are also shown.

4 Discussion

4.1 Plasticity Enables the Shaping of Motif Sensitivity. Understanding the advantages of neural plasticity for information processing and computation is in general a wide-ranging neuroscience question. Many putative network properties and computational functions have been characterized, including (1) information-theoretic properties such as information transmission (Toyoizumi, Pfister, Aihara, & Gerstner, 2005), neural coding and computation (Toutounji & Pipa, 2014), self-organized criticality (Meisel & Gross, 2009; Stepp, Plenz, & Srinivasa, 2015), and memory storage capacity (Van Rossum, Shippi, & Barrett, 2012); (2) dynamical systems properties such as stability (Zenke et al., 2013; Litwin-Kumar & Doiron, 2014; Duarte & Morrison, 2014), bifurcation (Ren, Kolwankar, Samal, & Jost, 2012), robustness (Toutounji & Pipa, 2014; Loewenstein, 2008), and synchronization (Talathi, Hwang, & Ditto, 2008); (3) functional properties such as decision making (Katahira, Okanoya, & Okada, 2010) and classification (Srinivasa & Cho, 2014; Duarte & Morrison, 2014).

In this letter, we have studied how synaptic plasticity shapes the sensitivity of isoinhibitory motifs to afferent excitation, toward enabling heightened sensitivity to novel inputs. A reachability (or controllability) analysis was performed to characterize how the geometry of the motif’s reachable space changes due to a repetitive, conditioned stimulus. We found that the space of reachable trajectories (neuronal activation patterns) contracts in the direction associated with the conditioned stimulus, while simultaneously...
expanding in orthogonal directions. In other words, the network desensi-
tizes to the conditioned stimulus, while at the same time sensitizes to po-
tential competing or distracting stimuli, assuming that the latter does not
substantially overlap the former. Thus, our key finding is a set of general
network dynamics that mediate directionally specific sensitivity profiles
and, in particular, heightened sensitization to novel stimuli.

4.2 Motif Dynamics, Not Simply the Synaptic Weights, Underlie Or-
thogonal Sensitization. We found that sensitization to orthogonal stimuli
was mediated by a rotation and contraction in the geometry of the reach-
able set away from the persistent, conditioned stimulus. This effect depends
critically on the relative time constants of the network, as well as their re-
lationship to the frequency of the stimulus repetition. Owing to the decay
term within our plasticity rule, if the time interval between stimuli were
increased, the network would return to its steady state and any effects of
long-term adaptation would, of course, be diminished.

The mechanism that underlies desensitization (or habituation) to the
conditioned stimulus is in large part a substantial increase in the recurrent
inhibitory gain back onto the conditioned neuron (see the panel $I \rightarrow E_1$ in
Figure 6C). Prima facie, the mechanism that underlies the motif sensitization
would, conversely, involve a reduction in the recurrent inhibition to the
orthogonal neuron. However, as we observed in Figures 6 and 11, the net
recurrent inhibitory gain to the orthogonal neuron does not substantially
decrease (see the panel $I \rightarrow E_2$ in (see Figure 6C). This is particularly notable
in the case of the disassociated input model, where the weights from the
inhibitory neuron onto the orthogonal neuron potentiate (see Figure 11).
In other words, in terms of the synaptic weights, the orthogonal neuron
is actually increasingly inhibited, even though the motif sensitizes in that
direction. Thus, the sensitivity profiles obtained through our analysis are
not just a consequence of adaptation in the synaptic weights, but also the
progressive change in the local dynamics caused by the stimulus-induced
shift in the motif phase space.

4.3 Effects of Long-Term Plasticity. Our choice of plasticity timescale
is faster than the typically indicated timescales of long-term plasticity. Spec-
ifying plasticity at the scale of 30 times that of neurons allowed us to elu-
cidate the types of sensitization profiles that may arise, with the premise
that the nature of these profiles would be qualitatively unchanged at longer
timescales. To verify this, we also simulated our same motif setup with
plasticity at 400 times slower than the intrinsic neuronal dynamics. Figure
13 shows the resultant sensitization profiles, where we find no change in
the character of the results. Based on these findings, including our analyt-
ical framework, we predict the same types of phenomena to occur at even
longer timescales of plasticity, although we did not pursue this here.
Figure 13: Effect of long-term plasticity on the motif sensitivity (compare with Figure 6). Here the timescale of synaptic plasticity is approximately 400 times slower than the timescale of neuronal dynamics. (A) Sensitivity profiles for 100 random parameterizations, quantified in terms of the minimum stimulus energy required to elicit a response from each excitatory neuron, $\Delta \epsilon_i$ for $i \in \{E_1, E_2, I\}$, expressed relative to $k = 0$. The motif always desensitizes along the $E_1$ direction. (B) Clustering analysis over the sensitivity profiles for $E_2$ neuron, $\Delta \epsilon_{(E_2)}$. These responses can be separated into four qualitative responses: (S) sensitization, (DS) desensitization, (S-DS) sensitization followed by desensitization, and (DS-S) desensitization followed by sensitization. The four response types are directly associated with the time constants of the excitatory neurons. (C) Clustering analysis over the sensitivity profiles for $I$ neuron, $\Delta \epsilon_{(I)}$. These responses can be separated into three qualitative responses: S, DS, and DS-S. The three response types are again directly associated with the time constants of the excitatory neurons. (D) Changes in synaptic weights over repetitions $k$.

4.4 Generic Novelty Detection in E-I Networks. In examining the sensitivity profiles of the three-neuron isoinhibitory motif, we noted that when the time constants of the excitatory neurons are equal, the motif produces orthogonal sensitization without loss of generality. That is, excitation of either neuron will cause sensitization of the other. The phenomenon is robust to all other parameters and, indeed, the inclusion of E-E connections, as shown in Figure 14.
Figure 14: Motif sensitivity is robust to the inclusion of E-E connectivity (compare with Figure 6). (A) Sensitivity profiles for 100 random parameterizations, quantified in terms of the minimum stimulus energy required to elicit a response from each excitatory neuron, $\Delta \varepsilon_{(i)}$ for $i \in \{E_1, E_2, I\}$, expressed relative to $k = 0$. The motif always desensitizes along the $E_1$ direction. (B) Clustering analysis over the sensitivity profiles for $E_2$ neuron, $\Delta \varepsilon_{(E_2)}$. These responses can be separated into four qualitative responses: (S) sensitization, (DS) desensitization, (S-DS) sensitization followed by desensitization, and (DS-S) desensitization followed by sensitization. The four response types are directly associated with the time constants of the excitatory neurons. (C) Clustering analysis over the sensitivity profiles for $I$ neuron, $\Delta \varepsilon_{(I)}$. These responses can be separated into three qualitative responses: S, DS, DS-S. The three response types are again directly associated with the time constants of the excitatory neurons. (D) Changes in synaptic weights over repetitions $k$.

If we extrapolate this to larger E-I network structures, we would expect to find a generic property of novelty detection, where long-term plasticity enables a network to favor stimuli that are different or novel from ongoing activity. We showed such an extrapolation in our simulation study of the larger 10-neuron network. Thus, our results provide a general mechanism for novelty enhancement that is innate to recurrent E-I structures. As mentioned in section 1, such structures are pervasive in early sensory networks (Raman, Joseph, Tang, & Stopfer, 2010; Das et al., 2011).
Beyond the direct neurobiological context, our results indicate a set of dynamical systems-based principles for the construction of systems for novelty detection (Pimentel, Clifton, Clifton, & Tarassenko, 2014). Such potential has been demonstrated for engineered sensor arrays based on olfactory network architectures (Das et al., 2011; Raman, Gutierrez-Galvez, & Gutierrez-Osuna, 2006).

4.5 A Geometric Sensitivity Analysis for Neuronal Networks. In characterizing the motif sensitivity, we pursue an approach derived from control theoretic analysis to characterize the motif’s reachable set. The reachable set provides a holistic (albeit local) characterization of the motif sensitivity in all directions of its phase space. While we focused on just the changes along the cardinal directions, more complex assessments of changes to the reachable geometry are possible. Of course, our approach represents only one of many possible ways to characterize sensitivity. Other methods, such as those involving perturbation analysis, have been used to make similar characterizations in networks of spiking neurons (Bell & Parra, 2005; Sengupta, Gurumoorthy, & Banerjee, 2015).

4.6 Assumptions and Limitations. Throughout this work, we have assumed that both excitatory and inhibitory synaptic connections are plastic, as has been characterized in visual and auditory cortices (Vogels et al., 2013; D’Amour & Froemke, 2015). However, the presence of plasticity in excitatory and inhibitory synapses remains an active area of study. Indeed, quite different functional forms and mechanisms have been suggested for excitatory plasticity (STDP) and inhibitory plasticity (i-STDP) (Caporale & Dan, 2008). Thus, a simplifying assumption in our work is the use of the BCM rule to model the plasticity in both types of synapses in the absence of an equivalent rate model for inhibitory plasticity.

Our analysis methods, discussed above, are based on linearization of the nonlinear model and thus provide a local characterization. In order to perform the reachability analysis, we made a tacit assumption that the linearized system is controllable, that is, the controllability Gramian is full rank. In higher dimensions, even if a system is controllable, numerical analysis can become difficult due to poor conditioning of the Gramian. In this case, one may perform similar analysis on subspaces of the overall system.

We performed the reachability analysis for $t_f = 0.3$ ms because we were interested in sensitization over short time horizons. Recall that our analysis is based on a perturbation approach (see section 3.1) that considers small excursions about the local linearization. Thus, a short horizon provides a characterization of sensitivity in the infinitesimal sense, as opposed to considering small excursions over long horizons. Nevertheless, in principle, one could perform the reachability over any desired horizon.
Our focus in this letter is on the specific motif and its manifestation in larger E-I networks. We also focus on a particular form of repetitive stimulus, a common paradigm for studying adaptation in sensory systems (Grill-Spector, Henson, & Martin, 2006; Kaliukhovich & Vogels, 2012; Brunet et al., 2014). Clearly, many other motif structures and stimulus patterns including stochasticity could be considered, which may enrich the characterizations made in this letter. We consider such investigations as topics for future work.

Appendix: Additional Details on Analysis and Derivations

A.1 Linearized Dynamics of Equation 2.1. At the end of each stimulus repetition, we linearized equation 2.1 with respect to \( r_i(t) \) and \( u_i(t) \) to obtain the network dynamics in the form of equation 2.4. By defining \( f_i(r_1, \cdots, r_N, I_{b,i}, u_i) = \frac{c_i}{1+\exp(-\sum_{j\neq i}^{N}(-1)^{\nu} w_{ij} r_j + I_{b,i} + u_i(t)))} \), equation 2.1 can be rewritten as

\[
\frac{dr_i}{dt} = -\frac{r_i}{\tau_i} + \frac{1}{\tau_i} f_i(r_1, \cdots, r_N, I_{b,i}, u_i). \tag{A.1}
\]

By representing \( r(t) = (r_1(t), \cdots, r_N(t))^T \) and assuming that \( r(t) = r_k \) at the end of the stimulus repetition \( k \), linearization of equation A.1 results in

\[
\frac{dr(t)}{dt} = g(r_k, I_b, 0, \tau) + A(W, r_k, \tau, I_b)(r(t) - r_k) + B(W, r_k, \tau, I_b)u(t). \tag{A.2}
\]

Here, \( \tau \) is the diagonal matrix with the \( i \)th diagonal entry \( \tau_i \), the time constant of neuron \( i \), and \( W \) is the weight matrix. \( g(r_k, I_b, 0, \tau) \) is a vector where the \( i \)th component of the vector is given by

\[
g_i(r_k, I_b, 0, \tau) = -\frac{r_{i,s}}{\tau_i} + \frac{1}{\tau_i} f_i(r_{1,k}, \cdots, r_{N,k}, I_{b,i}, 0), \tag{A.3}
\]

where \( r_{j,k} \) is the \( j \)th component of the vector \( r_k \). \( A(W, r_b, \tau, I_b) \) is the state matrix where the \((i, j)\)th component of the matrix is given by

\[
A_{i,j}(W, r_k, \tau, I_b) = \begin{cases} 
(-1)^\nu c_i w_{ij} \exp(-\sum_{l\neq j=1}^{N}(-1)^\nu w_{il} r_{l,k} + I_{b,i})) & \text{if } i \neq j \\
\tau_i (1 + \exp(-\sum_{l\neq i}^{N}(-1)^\nu w_{il} r_{l,k} + I_{b,i})^2) & \text{if } i = j.
\end{cases} \tag{A.4}
\]
Figure 15: The change in $g_i(r_k, I_b, 0, \tau), i \in \{1, 2, 3\}$. See equation A.3 as a function of the repeated stimulus $k$ for the simulation results presented in Figures 6 and 7.

$B(W, r_k, \tau, I_b)$ is the input matrix where the $(i, j)$th component of the matrix is given by

$$B_{i,j}(W, r_k, \tau, I_b) = \begin{cases} 0 & \text{if } i \neq j \\ \exp\left(-\left(\sum_{l \neq i}^N (-1)^a w_{il} r_{l,k} + I_{b,i}\right)\right) / \tau_i(1 + \exp\left(-\left(\sum_{l \neq i}^N (-1)^a w_{il} r_{l,k} + I_{b,i}\right)\right))^2 & \text{if } i = j. \end{cases}$$  

(A.5)

Since $g(r_k, I_b, 0, \tau)$ is close to 0 (see Figure 15), we ignore the contribution of this term in equation A.2 and rewrite it by substituting $y(t) = r(t) - r_k$ in it, which results in

$$\frac{dy(t)}{dt} = A(W, r_k, \tau, I_b) y(t) + B(W, r_k, \tau, I_b) u(t),$$  

(A.6)

with $y(0) = 0$. Analysis now proceed with respect to the state variable $y$.

**A.2 Linearized Dynamics of 2.2.** At the end of each stimulus repetition, we linearized equation 2.2 with respect to $r_i(t)$ and $u_i(t)$ to obtain the network dynamics in the form of equation 2.4. By defining

$$f_i(r_1, \cdots, r_N, I_{b,i}, u_i) = \frac{0.5}{1 + \exp\left(-\left(\sum_{j \neq i}^N (-1)^a w_{ij} r_j\right)\right)} + \frac{1}{1 + \exp\left(-\left(u_i(t) + I_{b,i}\right)\right)} - \frac{1}{2},$$

equation 2.2 can be rewritten as

$$\frac{dr_i}{dt} = -\frac{r_i}{\tau_i} + \frac{1}{\tau_i} f_i(r_1, \cdots, r_N, I_{b,i}, u_i).$$  

(A.7)
By representing \( r(t) = (r_1(t), \cdots, r_N(t))^T \) and assuming that \( r(t) = r_k \) at the end of the repetitive stimulus \( k \), linearization of equation A.7 results in

\[
\frac{dr(t)}{dt} = g(r_k, I_b, 0, \tau) + A(W, r_k, \tau)(r(t) - r_k) + B(\tau, I_b)u(t). \tag{A.8}
\]

Here, \( \tau \) is the diagonal matrix with the \( i \)th diagonal entry \( \tau_i \), the time constant of neuron \( i \), and \( W \) is the weight matrix. \( g(r_k, I_b, 0, \tau) \) is a vector where the \( i \)th component of the vector is given by

\[
g_i(r_k, I_b, 0, \tau) = -\frac{r_{i,k}}{\tau_i} + 1 \quad \text{if } i \neq j
\]

\[
g_i(r_k, I_b, 0, \tau) = \frac{1}{\tau_i} \quad \text{if } i = j. \tag{A.9}
\]

where \( r_{j,s} \) is the \( j \)th component of the vector \( r_k \). \( A(W, r_k, \tau) \) is the state matrix where the \((i, j)\)th component of the matrix is given by

\[
A_{i,j}(W, r_k, \tau) = \begin{cases} 
(\frac{1}{\tau_i} - 1) \sum_{l=1}^N (-1)^a \exp \left( \frac{-\sum_{l=1}^N (-1)^a w_{il} r_{l,k}}{\tau_i} \right) 
& \text{if } i \neq j \\
\frac{1}{\tau_i} & \text{if } i = j.
\end{cases} \tag{A.10}
\]

\( B(\tau, I_b) \) is the input matrix where the \((i, j)\)th component of the matrix is given by

\[
B_{i,j}(\tau, I_b) = \begin{cases} 
0 & \text{if } i \neq j \\
\frac{\exp(-I_{b,i})}{\tau_i(1 + \exp(-I_{b,i}))^2} & \text{if } i = j.
\end{cases} \tag{A.11}
\]

Again we ignore the contribution of \( g(r_k, I_b, 0, \tau) \) in equation A.8 and rewrite it by substituting \( y(t) = r(t) - r_k \) in it, which results in

\[
\frac{dy(t)}{dt} = A(W, r_s)y(t) + B(\tau, I_b)u(t). \tag{A.12}
\]

with \( y(0) = 0 \). Analysis proceeds with respect to the state variable \( y \).

A.3 Derivation of Equation 3.3. By replacing \( r_{1,s} \) by \( r_{1,s} + \Delta r_1 \), \( r_{3,s} \) by \( r_{3,s} + \Delta r_3 \) and \( w_{13} \) by \( w_{13} + \Delta w_{13} \) in equation 3.2a and solving for \( \Delta r_1 \) using that equation, we obtained.
Plasticity-Induced Sensitization in Rate Motifs

\[
\Delta r_1 = -\frac{(r_{3,s}\Delta w_{13} + w_{13}\Delta r_3) \exp(w_{13}r_{3,s})}{(1 + \exp(w_{13}r_{3,s}))^2}.
\]  

(A.13)

Here we ignored the higher-order term \(\Delta w_{13}\Delta r_{3,s}\) and approximated \(\exp(w_{13}r_{3,s} + (r_{3,s}\Delta w_{13} + w_{13}\Delta r_3))\) by \((1 + (r_{3,s}\Delta w_{13} + w_{13}\Delta r_3))\exp(w_{13}r_{3,s})\). We also assumed that \((r_{3,s}\Delta w_{13} + w_{13}\Delta r_3) \ll 1\). Similarly, we obtained

\[
\Delta r_2 = -\frac{(r_{3,s}\Delta w_{23} + w_{23}\Delta r_3) \exp(w_{23}r_{3,s})}{(1 + \exp(w_{23}r_{3,s}))^2}
\]  

(A.14)

using equation 3.2b and

\[
\Delta r_3 = \frac{(r_{1,s}\Delta w_{31} + w_{31}\Delta r_1 + r_{2,s}\Delta w_{32} + w_{32}\Delta r_2) \exp(-w_{31}r_{1,s} - w_{32}r_{2,s})}{(1 + \exp(-w_{31}r_{1,s} - w_{32}r_{2,s}))^2}
\]  

(A.15)

using equation 3.2c. By solving equations A.13 to A.15, we obtained expressions in equation 3.3.

**A.4 Derivation of Equation 3.9.** By replacing \(W_c(t_f)\) by \(W_c(t_f) + \Delta W_c(t_f)\), \(x_i\) by \(x_i + \Delta x_i\) and \(\lambda_i\) by \(\lambda_i + \Delta \lambda_i\) in equation 3.8 and neglecting higher-order terms, we obtained

\[
W_c(t_f)\Delta x_i + \Delta W_c(t_f)x_i = \lambda_i \Delta x_i + \Delta \lambda_i x_i.
\]  

(A.16)

Since the unperturbed eigenvectors \(x_i\) are orthonormal, we used \(x_i\) as a basis to construct

\[
\Delta x_i = \sum_{j=1}^{N} \epsilon_{ij} x_j.
\]  

(A.17)

Here \(\epsilon_{ij}\) are small constants. By substituting equation A.17 in A.16 and using 3.8, we obtained

\[
\sum_{j=1}^{N} \epsilon_{ij} \lambda_j x_j + \Delta W_c(t_f)x_i = \lambda_i \sum_{j=1}^{N} \epsilon_{ij} x_j + \Delta \lambda_i x_i.
\]  

(A.18)

By multiplying both sides of equation A.18 with \(x_i^T\) from the left and using the fact that \(x_i\) are orthonormal, we obtained equation 3.9a. To compute \(\epsilon_{ij}\)
in equation A.17, we again substituted equation A.17 in A.16 and multiplied the resultant equation with $x_k^T$, $k \neq i$ from left, which resulted in

$$x_k^T W_c(t_f) \sum_{j=1}^{N} \varepsilon_{ij} x_j + x_k^T \Delta W_c(t_f) x_i = x_k^T \lambda_i \sum_{j=1}^{N} \varepsilon_{ij} x_j + x_k^T \Delta \lambda_i x_i. \tag{A.19}$$

Using equations 3.9a and 3.8 and the fact that $x_i'$ are orthonormal, we obtained

$$\varepsilon_{ij} = \frac{x_j^T \Delta W_c(t_f) x_i}{\lambda_i - \lambda_j} \quad \text{if} \ j \neq i. \tag{A.20}$$

To compute $\varepsilon_{ii}$, we used the fact that the new eigenvectors are also orthonormal:

$$(x_i + \Delta x_i)^T (x_i + \Delta x_i) = 1. \tag{A.21}$$

By using equation A.17 and neglecting higher-order changes, we obtained $\varepsilon_{ii} = 0$. This completes the derivation of equation 3.9b.

**A.5 Network and Simulation Parameters.** In all simulations (except Figures 2, 3, 4, and 13) the time constants $\tau_i$, $\tau_{i,j}$ and $\eta_i$ were chosen from uniform distributions $\mathcal{U}(10 \text{ ms}, 60 \text{ ms})$, $\mathcal{U}(1000 \text{ ms}, 1500 \text{ ms})$ and $\mathcal{U}(100 \text{ ms}, 150 \text{ ms})$ respectively. The maximum firing rate of neuron $i$, that is, $c_i$ was set to 1 for all neurons in the motif and $\epsilon = 10^{-4}$. The stimulus repetition parameters $T_1$ and $T_2$ in equation 2.10 were set to 500 ms and 1500 ms, respectively. The amplitude of the stimulus, $a_i$ in equation 2.10, was chosen from the uniform distribution $\mathcal{U}(2, 6)$. Except for the results shown in Figure 10 where the baseline input $I_{b,i}$ for neurons 1, 2, and 3 was chosen from uniform distributions $\mathcal{U}(0.05, 0.25)$, $\mathcal{U}(0.1, 0.6)$, and $\mathcal{U}(0.01, 0.06)$, respectively, $I_{b,i} = 0$. For Figure 13, we used the same parameters as above except that $\tau_{i,j}$ is chosen from $\mathcal{U}(20,000 \text{ ms}, 25,000 \text{ ms})$, $\eta_i$ is chosen from $\mathcal{U}(1000 \text{ ms}, 1500 \text{ ms})$, and $T_2 = 5000 \text{ ms}. t_f = 0.3 \text{ ms}$. In Figure 5B, we computed the ellipsoid using the infinite-time Gramian,

$$AW_{c,s} + W_{c,s} A^T + BB^T = 0. \tag{A.22}$$

All of our simulation codes are available on our server and can be downloaded at braindynamics.engineering.wustl.edu/code/. Also, simulated data for generating figures are available on request.
Acknowledgments

This work was partially supported by AFOSR 15RT0189 and NSF ECCS 1509342. S. C. holds a Career Award at the Scientific Interface from the Burroughs-Wellcome Fund. We thank Baranidharan Raman for helpful discussions in the development of this research. We also thank the reviewers for providing constructive suggestions that helped us to improve our letter.

References


Received August 3, 2015; accepted April 29, 2016.