Supplementary Information

# Orchestrated Excitatory and Inhibitory Learning Rules Lead to the Unsupervised Emergence of Up-states and Balanced Network Dynamics

(Soldado-Magraner, Motanis, Laje & Buonomano 2020)
Author: Rodrigo Laje

December 30, 2020

## 1 Summary of results

### 1.1 *Homeostatic* learning rule

The equations for the Homeostatic learning rule are

$$\begin{aligned}\frac{dW_{EE}}{dt} &= +\alpha g_E\, E(E_{set} - E)\\ \frac{dW_{EI}}{dt} &= -\alpha g_E\, I(E_{set} - E)\\ \frac{dW_{IE}}{dt} &= +\alpha g_I\, E(I_{set} - I)\\ \frac{dW_{II}}{dt} &= -\alpha g_I\, I(I_{set} - I)\end{aligned} \tag{1}$$

and the condition for the Up state to be unstable under this rule is

$$W_{EEup}\, g_E - 1 > \frac{(E_{set} W_{IEup} - \Theta_I) g_E^2}{I_{set} g_I} \tag{2}$$

which is satisfied for biologically backed parameter values. See the step-by-step derivation of this instability condition in Section 2.3.

### 1.2 *Cross-Homeostatic* learning rule

The equations for the Cross-Homeostatic learning rule are

$$\begin{aligned}\frac{dW_{EE}}{dt} &= +\beta(I_{set} - I)\\ \frac{dW_{EI}}{dt} &= -\beta(I_{set} - I)\\ \frac{dW_{IE}}{dt} &= -\beta(E_{set} - E)\\ \frac{dW_{II}}{dt} &= +\beta(E_{set} - E)\end{aligned} \tag{3}$$

This rule has a very simple expression for the stability condition of the Up state when written in terms of $W_{EI}$ and $W_{IE}$:

$$W_{EIup} + W_{IEup} > 0 \tag{4}$$

which is always satisfied since the weights are positive definite. See the step-by-step derivation of this stability condition in Section 2.4.

### 1.3 *Balanced-Homeostatic* learning rule

The equations for the Balanced-Homeostatic learning rule are

$$\begin{aligned}\frac{dW_{EE}}{dt} &= +\alpha_1 g_E E(E_{set} - E)\\ \frac{dW_{EI}}{dt} &= \frac{1}{\tau_0}(W_{EIup} - W_{EI})\\ \frac{dW_{IE}}{dt} &= -\alpha_3 g_I I(I_{set} - I)\\ \frac{dW_{II}}{dt} &= \frac{1}{\tau_0}(W_{IIup} - W_{II})\end{aligned} \tag{5}$$

and the conditions for the Up state to be stable under this rule are

$$\begin{aligned}a_1 + (b_1 - c_1)d &< (b_1' + c_1')e\\ a_2 + (b_2 - c_2)d &< (b_2' + c_2')e\end{aligned} \tag{6}$$

where

$$
\begin{aligned}
a_1 &= (\Theta_E \Theta_I \, g_E g_I + E_{set}^2 \frac{\alpha_3}{\alpha_1}) I_{set} \, g_E g_I \\
b_1 &= I_{set}^2 \Theta_E \, g_E^2 g_I \\
c_1 &= E_{set}^2 I_{set} \, g_E^2 \\
b_1' &= E_{set} I_{set} \Theta_I \, g_E g_I^2 \\
c_1' &= E_{set} I_{set}^2 \, g_I^2 \frac{\alpha_3}{\alpha_1} \\
a_2 &= 2\Theta_E \Theta_I \, g_E^2 g_I^2 \\
b_2 &= 2 I_{set} \Theta_E \, g_E^2 g_I \\
c_2 &= E_{set}^2 \, g_E^2 \\
b_2' &= 2 E_{set} \Theta_I g_E \, g_I^2 \\
c_2' &= E_{set} I_{set} \, g_I^2 \frac{\alpha_3}{\alpha_1} \\
d &= W_{II\,up} \, g_I + 1 \\
e &= W_{EE\,up} \, g_E - 1
\end{aligned}
$$

Conditions Eq. 6 are satisfied for biologically backed parameter values. See the step-by-step derivation of the stability condition in Section 2.5.

### 1.4 ***SynapticScaling*** learning rule

The equations for the SynapticScaling learning rule are

$$
\begin{aligned}
\frac{dW_{EE}}{dt} &= +\gamma g_E (E_{set} - E) W_{EE} \\
\frac{dW_{EI}}{dt} &= -\gamma g_E (E_{set} - E) W_{EI} \\
\frac{dW_{IE}}{dt} &= +\gamma g_I (I_{set} - I) W_{IE} \\
\frac{dW_{II}}{dt} &= -\gamma g_I (I_{set} - I) W_{II}
\end{aligned}
\tag{7}
$$

and the condition for the Up state to be unstable unde this rule is

$$
\begin{aligned}
&(W_{EE\,up} \, g_E - 1)(2 I_{set} W_{II\,up} \, g_I + \Theta_I \, g_I + I_{set}) \\
&> (W_{II\,up} \, g_I + 1)(2 E_{set} W_{EE\,up} \, g_E - \Theta_E \, g_E - E_{set})
\end{aligned}
\tag{8}
$$

This instability condition holds for biologically backed parameter values. See the step-by-step derivation of the stability condition in Section 2.6.

# 2 Detailed calculations

## 2.1 Overview

We analyze the whole neural+synaptic system for every synaptic learning rule considered in this work, and study their stability. In every case, the general prescription is:

1. Take the combined neural+synaptic system and nondimensionalize all variables [1, Sections 1.2 and 1.4][2, Section 3.5], so that the two different time scales are evident (fast neural, slow synaptic).

2. Make a quasi-steady state (QSS) approximation of the neural subsystem [1, 2]. This means we will consider the neural subsystem is fast enough so that it converges "instantaneously" (when compared to the synaptic subsystem) to its corresponding fixed point. For this we will require that the stability conditions of the neural subsystem are satisfied (see below).

3. Find the steady-state solution of the synaptic subsystem, i.e. the Up state fixed point; compute the Jacobian of the synaptic subsystem at the Up state; compute the eigenvalues of the Jacobian [2, 3]. Two out of the four eigenvalues are expected to be zero because the Up state is not an isolated fixed point of the system but a continuous 2D plane in 4D weight space.

4. Address (linear) stability. If both nonzero eigenvalues have negative real part, then the Up state is stable under this learning rule; if at least one of the nonzero eigenvalues has positive real part, then the Up state is unstable [2, 3]. (A note on abuse of notation: we might say indistinctly "the Up state is stable/unstable" and "the learning rule is stable/unstable")

Eigenvalues and stability in the presence of continuous, i.e. non-isolated, attractors have been discussed in the context of neural networks for eye position control [4, 5] (keep in mind that their eigenvalues' critical value is 1 instead of zero because they consider eigenvalues of the connectivity matrix alone, whereas we consider eigenvalues of the full system). As the Up state is a collection of non-isolated fixed points that form a 2D plane, there is no dynamics along the plane, and the linear stability analysis is enough to fully address stability—we do have two zero eigenvalues, but there is no need to compute the center manifold [3] because the other two eigenvalues represent the whole dynamics around the fixed point and have nonzero real part.

In order to apply the tools from Dynamical Systems' theory for flows in a unified way for both the neural and synaptic subsystems, we will switch from a discrete-time description of synaptic weight dynamics (where the change in weight $W$ is represented by $\Delta W$ applied every certain time interval) to a continuous-time description (where the weights are continuously evolving albeit with a long time scale $\tau_0$):

$$\Delta W \to \tau_0 \frac{dW}{dt}$$

In the following we first define the neural subsystem and compute its stability conditions (next subsection). Then we consider every learning rule in detail (following subsections).

## 2.2 Neural dynamics

For the neural+synaptic system in the QSS approximation to be stable under a specific synaptic learning rule, it is necessary that the neural subsystem is stable so it remains in its QSS solution as the weights evolve. In this section we define the neural subsystem and compute its stability conditions.

(SageMath code in the Supplementary Material: `up states - Neural subsystem stability.ipynb`)

### 2.2.1 System's equations and fixed points

We consider a two-subpopulation model with firing-rate units $E$ and $I$ with ReLU activation functions (gain $g_X$, threshold $\Theta_X$, with $X = E, I$). The dynamics for synaptic currents above threshold is given by:

$$\begin{aligned} \frac{dE}{dt} &= \frac{1}{\tau_E}(-E + g_E(W_{EE}E - W_{EI}I - \Theta_E)) \\ \frac{dI}{dt} &= \frac{1}{\tau_I}(-I + g_I(W_{IE}E - W_{II}I - \Theta_I)) \end{aligned} \tag{9}$$

All variables and parameters are definite positive. In this subsection the synaptic weights $W_{XY}$ are fixed.

**Down state** There's a trivial fixed point (i.e. a steady-state solution $dX/dt = 0$) at $E = I = 0$ when the inputs to both subpopulations are subthreshold. This fixed point is stable: if the currents are below the ReLU threshold, any value $E > 0$ and $I > 0$ will have a negative velocity leading to an exponential decrease towards zero.

**Up state** The other, non-trivial fixed point is the Up state:

$$\begin{aligned} E_{up} &= (W_{EI}\, g_E\, g_I\, \Theta_I - (W_{II}\, g_I + 1)\, g_E\, \Theta_E)/C \\ I_{up} &= ((W_{EE}\, g_E - 1)\, g_I\, \Theta_I - W_{IE}\, g_E\, g_I\, \Theta_E)/C \end{aligned} \tag{10}$$

where

$$C = W_{EI}W_{IE}\, g_E\, g_I - (W_{II}\, g_I + 1)(W_{EE}\, g_E - 1) \tag{11}$$

The activity of the excitatory and inhibitory subpopulations at the Up state, $E_{up}$ and $I_{up}$, depend on all weight values. Only some of the combinations, however, lead to a stable steady state. We compute the stability conditions in the following subsection.

### 2.2.2 Stability of neural fixed point (Up state)

The Jacobian matrix, that is the matrix of first derivatives, gives information regarding the stability of fixed points: if the real parts of its eigenvalues are all negative, then the fixed point is stable.

The Jacobian of the neural system (Eq. 9) is

$$J = \begin{pmatrix} (W_{EE}g_E - 1)/\tau_E & -W_{EI}g_E/\tau_E \\ W_{IE}g_I/\tau_I & -(W_{II}g_I + 1)/\tau_I \end{pmatrix} \tag{12}$$

Its eigenvalues can be expressed as:

$$\lambda_{1,2} = \frac{1}{2}\left(Tr \pm \sqrt{Tr^2 - 4Det}\right) \tag{13}$$

where $Tr$ and $Det$ are the trace and determinant of the matrix, respectively. For eigenvalues either complex or purely real, their real parts are negative (and thus the Up state is stable) when $Det > 0$ and $Tr < 0$, that is:

$$W_{EI}W_{IE}g_E g_I > (W_{EE}g_E - 1)(W_{II}g_I + 1) \tag{14}$$
$$(W_{II}g_I + 1)\tau_E > (W_{EE}g_E - 1)\tau_I \tag{15}$$

Note that the positive determinant condition, Eq. 14, is equivalent to $C > 0$ (Eq. 11).

In the following, we will require that the stability conditions of the neural subsystem, Eqs. 14 and 15, are satisfied.

#### 2.2.3 Paradoxical effect

The paradoxical effect arises when an external depolarization of the inhibitory subpopulation (increase of $I$) produces an actual *decrease* of $I$. In this model, an external depolarization of $I$ can be mimicked by a decrease of its threshold $\Theta_I$, thus there is a paradoxical effect whenever the coefficient of $\Theta_I$ in the numerator of $I_{up}$ is positive. The coefficient is $g_I\,(W_{EE}\,g_E - 1)/C$ and thus there is paradoxical effect if

$$W_{EE}\,g_E - 1 > 0 \tag{16}$$

The paradoxical effect can also be seen in a plot of the Up-state values $E_{up}$ and $I_{up}$ (Eq. 10) as a function of each individual weight. Specifically, from a naive point of view $I_{up}$ should increase when $W_{IE}$ is increased, and decrease when $W_{II}$ is increased; however, it does the opposite in either case (see Figure S3).

## 2.3 Synaptic dynamics: *Homeostatic* learning rule

(SageMath code in the Supplemeentary Material: `up states - Homeostatic stability.ipynb`)

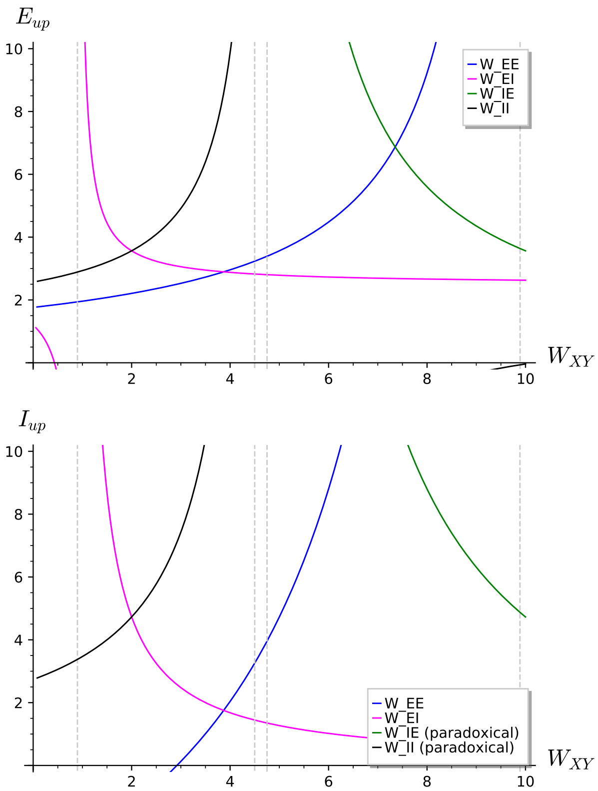


Figure S3: Paradoxical effect in the neural subsystem. The excitatory activity at the Up state, $E_{up}$, behaves as expected when each weight is varied. The inhibitory activity $I_{up}$, however, shows paradoxical behavior when either $W_{IE}$ or $W_{II}$ are varied. Dashed lines are the vertical asymptote of every case.

#### 2.3.1 Definition of the learning rule

In continuous-time dynamics, the Homeostatic learning rule reads:

$$\begin{aligned}
\frac{dW_{EE}}{dt} &= +\alpha g_E E(E_{set} - E)\\
\frac{dW_{EI}}{dt} &= -\alpha g_E I(E_{set} - E)\\
\frac{dW_{IE}}{dt} &= +\alpha g_I E(I_{set} - I)\\
\frac{dW_{II}}{dt} &= -\alpha g_I I(I_{set} - I)
\end{aligned} \tag{17}$$

where $\alpha$ is the learning rate (with appropriate units) setting the time scale of the weight dynamics, and $E_{set}$ and $I_{set}$ are the set points of the excitatory and inhibitory subpopulations, respectively.

The fixed points of the system (i.e. steady states) are determined by setting all derivatives to zero. There is a non-trivial fixed point compatible with the neural subsystem being above threshold: the Up state, that is the set of weight values such that:

$$\begin{aligned} E_{up} &= E_{set} \\ I_{up} &= I_{set} \end{aligned} \tag{18}$$

The values of the weights corresponding to the Up state are given by the (underdetermined) system defined by equating Eqs. 18 and 10. Since it is a two-equation system for a set of four unknown weights, there are two free weights that we choose to be $W_{EEup}$ and $W_{IEup}$. The other two have the following values:

$$\begin{aligned} W_{EIup} &= \frac{(E_{set}W_{EEup} - \Theta_E)g_E - E_{set}}{I_{set}g_E} \\ W_{IIup} &= \frac{(E_{set}W_{IEup} - \Theta_I)g_I - I_{set}}{I_{set}g_I} \end{aligned} \tag{19}$$

This means that the Up-state fixed point is actually a continuous set of non-isolated fixed points forming a 2D plane in 4D weight space. In other words, there is an infinite number of weight values compatible with the Up state (possibly not all stable, though).

#### 2.3.2 Nondimensionalization

Next we nondimensionalize all variables in order to have a simpler system and make the QSS approximation in a safe way. We define new (nondimensional) variables $e$, $i$, $\tau$, $w_{EE}$, $w_{EI}$, $w_{IE}$, and $w_{II}$, and their corresponding scaling parameters. We substitute the new variables into the full system (neural+synaptic, Eqs. 9 and 17) and choose the values of the scaling parameters such that all nondimensional variables are of order 1 (see SageMath code in the Supplementary Material). With this, the full system reads:

$$\begin{aligned} \epsilon_E \frac{de}{d\tau} &= -e + Regw_{EE} - \frac{giw_{EI}}{R} - \theta_E \\ \epsilon_I \frac{di}{d\tau} &= -i + \frac{Rew_{IE}}{g} - \frac{iw_{II}}{Rg} - \theta_I \\ \frac{dw_{EE}}{d\tau} &= -e(e-1) \\ \frac{dw_{EI}}{d\tau} &= +i(e-1) \\ \frac{dw_{IE}}{d\tau} &= -e(i-1) \\ \frac{dw_{II}}{d\tau} &= +i(i-1) \end{aligned} \tag{20}$$

where we defined the new parameters

$$\begin{aligned} \epsilon_E &= \tau_E/\tau_0 \\ \epsilon_I &= \tau_I/\tau_0 \\ \tau_0 &= 1/(\alpha g_E g_I E_{set} I_{set}) \\ R &= E_{set}/I_{set} \\ g &= g_I/g_E \\ \theta_E &= (g_E/E_{set})\Theta_E \\ \theta_I &= (g_I/I_{set})\Theta_I \end{aligned}$$

#### 2.3.3 Quasi-steady state approximation

Neural dynamics evolves in a much shorter time scale ($\tau_E$ and $\tau_I$) than synaptic dynamics ($\tau_0$):

$$\begin{aligned} \tau_E \ll \tau_0 &\implies \epsilon_E \ll 1 \\ \tau_I \ll \tau_0 &\implies \epsilon_E \ll 1 \end{aligned}$$

which implies

$$\begin{aligned} \epsilon_E \frac{de}{d\tau} &\sim 0 \\ \epsilon_I \frac{di}{d\tau} &\sim 0 \end{aligned} \tag{21}$$

thus we can safely assume $e$ and $i$ very quickly reach quasi-equilibrium values, i.e. practically instantaneous convergence to quasi-steady state (QSS) values as if the weights were fixed, while the synaptic weights evolve according to their slow dynamics. This allows us to reduce the system's dimensionality from six to four.

In the QSS approximation, the values of the nondimensionalized excitatory and inhibitory activities instantaneously track the slow dynamics of the learning rule. They are determined by applying Eq. 21 to the first two rows of Eq. 20; solving for $e$ and $i$ leads to

$$\begin{aligned} e_{qss} &= (g^2\theta_I w_{EI} - (Rg + w_{II})\theta_E)/c \\ i_{qss} &= (Rg\theta_I(Rgw_{EE} - 1) - R^2\theta_E w_{IE})/c \end{aligned} \tag{22}$$

where

$$c = Rgw_{EI}w_{IE} - (w_{II} + Rg)(Rgw_{EE} - 1)$$

The full system in the QSS approximation reads

$$\begin{aligned} \frac{dw_{EE}}{d\tau} &= -e_{qss}(e_{qss} - 1) \\ \frac{dw_{EI}}{d\tau} &= i_{qss}(e_{qss} - 1) \\ \frac{dw_{IE}}{d\tau} &= -e_{qss}(i_{qss} - 1) \\ \frac{dw_{II}}{d\tau} &= i_{qss}(i_{qss} - 1) \end{aligned} \tag{23}$$

where $e_{qss}$ and $i_{qss}$ are nonlinear functions of the weights as defined by Eq. 22.

Note that the Up state fixed point, defined by making all derivatives equal to zero, can be expressed as

$$\begin{aligned} e_{qss} &= 1 \\ i_{qss} &= 1 \end{aligned} \tag{24}$$

which is the nondimensionalized version of Eq. 18. The weight values compatible with this condition are defined by equating Eqs. 22 and 24:

$$\begin{aligned} w_{EIup} &= Rw_{EEup} - \frac{R(\theta_E + 1)}{g} \\ w_{IIup} &= R^2 w_{IEup} - Rg(\theta_I + 1) \end{aligned} \tag{25}$$

($w_{EEup}$ and $w_{IEup}$ are free). This is the nondimensionalized version of Eq. 19.

### 2.3.4 Instability condition

The program for assessing linear stability of the Up state is as follows: a) compute the Jacobian (the matrix of first derivatives) of Eq. 23 and evaluate it at the Up state; b) compute the eigenvalues of the Jacobian (two of them will be zero because the fixed points form a continuous 2D plane in phase space); c) If the real part of the two nonzero eigenvalues is negative then the Up state is stable; if at least one of the nonzero eigenvalue has positive real part then the Up state is unstable.

**Jacobian matrix** Let the full system in the QSS approximation (Eq. 23) be written as

$$\begin{aligned} \frac{dw_{EE}}{d\tau} &= f_{EE}(e_{qss}, i_{qss}) \\ \frac{dw_{EI}}{d\tau} &= f_{EI}(e_{qss}, i_{qss}) \\ &\text{etc} \ldots \end{aligned}$$

where $e_{qss}$ and $i_{qss}$ are functions of the weights as defined by Eq. 22. By applying the chain rule the elements $J_{ij}$ $(i, j = 1 \ldots 4)$ of the Jacobian matrix can be expressed as

$$\begin{aligned} J_{11} &= \frac{df_{EE}}{dw_{EE}} = \frac{df_{EE}}{de_{qss}}\frac{de_{qss}}{dw_{EE}} + \frac{df_{EE}}{di_{qss}}\frac{di_{qss}}{dw_{EE}} \\ J_{12} &= \frac{df_{EE}}{dw_{EI}} = \frac{df_{EE}}{de_{qss}}\frac{de_{qss}}{dw_{EI}} + \frac{df_{EE}}{di_{qss}}\frac{di_{qss}}{dw_{EI}} \\ J_{13} &= \ldots \\ J_{21} &= \frac{df_{EI}}{dw_{EE}} = \frac{df_{EI}}{de_{qss}}\frac{de_{qss}}{dw_{EE}} + \frac{df_{EI}}{di_{qss}}\frac{di_{qss}}{dw_{EE}} \\ J_{22} &= \ldots \\ &\text{etc} \ldots \end{aligned}$$

In order to have the Jacobian specialized in the Up state, these expressions are to be substituted by Eqs. 22-25.

**Eigenvalues of the Jacobian matrix** The Jacobian matrix has two zero eigenvalues and two nonzero eigenvalues. The nonzero eigenvalues have the form:

$$\lambda_\pm = \frac{\left(A \pm \sqrt{A^2 - DC}\right)(R^2 + 1)}{C} \tag{26}$$

where

$$\begin{aligned} A &= g^2\theta_I + Rgw_{EEup} - Rgw_{IEup} - 1 \\ C &= 2R(Rg^2\theta_I w_{EEup} - R\theta_E w_{IEup} - g\theta_I) \\ D &= 2g/R \end{aligned} \tag{27}$$

**Sign of the eigenvalues** To determine the sign of the real part of Eq. 26, first note that the factors $(R^2+1)$ and $D$ are positive definite. Second, $C$ must be positive because it is related to one of the stability conditions of the neural subsystem (Eq. 14, after substituting back to dimensionalized quantities). Note next that $A^2 - DC$ is less than $A^2$ (since $C$ and $D$ are positive), and thus the square root is either real and less than $|A|$ or imaginary, both cases leading to $\mathrm{Re}(A \pm \sqrt{A^2 - DC}) > 0$ if $A > 0$. The learning rule is then unstable (both eigenvalues have positive real part) if $A > 0$, which in terms of the original parameters and variables reads:

$$W_{EEup}\, g_E - 1 > \frac{(E_{set} W_{IEup} - \Theta_I) g_E^2}{I_{set} g_I} \tag{28}$$

### 2.3.5 Analysis of the instability condition

By using the Up state relationship Eq. 19, this condition can be re-expressed in a more useful form in terms of $W_{IEup}$ and $W_{EIup}$:

$$\begin{aligned} I_{set}^2 W_{EIup}\, g_I + I_{set} \Theta_E\, g_I \\ > E_{set}^2 W_{IEup}\, g_E - E_{set} \Theta_I\, g_E \end{aligned} \tag{29}$$

Note that this relationship can be written as

$$a + b > a' - b'$$

where the left-hand side is a sum of positive definite terms and the right-hand side is a difference of positive definite terms, with

$$\begin{aligned} a &= I_{set}^2 W_{EIup}\, g_I \\ a' &= E_{set}^2 W_{IEup}\, g_E \\ b &= I_{set} \Theta_E\, g_I \\ b' &= E_{set} \Theta_I\, g_E \end{aligned}$$

Note that for a biologically backed set of parameter values the following relations hold:

$$\begin{aligned} I_{set} &> E_{set} \\ g_I &> g_E \\ \Theta_I &> \Theta_E \end{aligned}$$

and thus it is likely that

$$\begin{aligned} a &> a' \\ b &\sim b' \end{aligned}$$

(despite $W_{EIup} \sim 0.1 W_{IEup}$), leading to $a+b > a'-b'$ and thus making the instability condition Eq. 29 to hold.

### 2.3.6 Numerical analysis

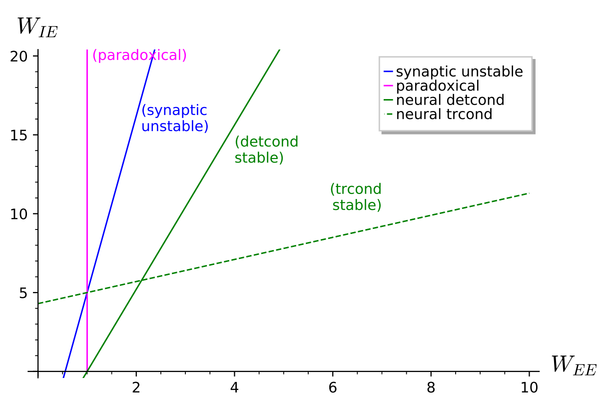


Figure S4: Regions of stability. The Homeostatic learning rule is unstable where the neural subsystem is stable. Synaptic: Eq. 28; Paradoxical: Eq. 16; Neural detcond: Eq. 14; Neural trcond: Eq. 15.

As an illustration of the reasoning above, we plot the instability condition Eq. 28 with parameter values as in Table 1.

| | | | | | |
|---|---|---|---|---|---|
| $I_{set}$ | = | 14 | $E_{set}$ | = | 5 |
| $g_I$ | = | 4 | $g_E$ | = | 1 |
| $\Theta_I$ | = | 25 | $\Theta_E$ | = | 4.8 |

Table 1: Parameter values.

It is clear from Figure S4 that the learning rule is stable in a region that doesn't overlap with the stability region of the neural subsystem.

## 2.4 Synaptic dynamics: *Cross-Homeostatic* learning rule

(SageMath code in the Supplementary Material: `up states - CrossHomeostatic stability.ipynb`)

In this section we follow a path similar to the one in the previous section, so we will skip some details.

### 2.4.1 Definition of the learning rule

In continuous-time dynamics, the CrossHomeostatic learning rule reads:

$$\begin{aligned} \frac{dW_{EE}}{dt} &= +\beta(I_{set} - I) \\ \frac{dW_{EI}}{dt} &= -\beta(I_{set} - I) \\ \frac{dW_{IE}}{dt} &= -\beta(E_{set} - E) \\ \frac{dW_{II}}{dt} &= +\beta(E_{set} - E) \end{aligned} \tag{30}$$

The only fixed point is the Up state and is the set of weight values such that:

$$\begin{aligned} E_{up} &= E_{set} \\ I_{up} &= I_{set} \end{aligned} \tag{31}$$

The values of the weights corresponding to the Up state are given by the (underdetermined) system defined by equating Eq. 31 and Eq. 10. Since it is a two-equation system for a set of four unknown weights, there are two free weights that we choose to be $W_{EEup}$ and $W_{IEup}$. The other two have the following values:

$$\begin{aligned} W_{EIup} &= \frac{(E_{set}W_{EEup} - \Theta_E)g_E - E_{set}}{I_{set}g_E} \\ W_{IIup} &= \frac{(E_{set}W_{IEup} - \Theta_I)g_I - I_{set}}{I_{set}g_I} \end{aligned} \tag{32}$$

This means that, as in Section 2.3, there is an infinite number of weight values compatible with the Up state (possibly not all stable, though).

### 2.4.2 Nondimensionalization

The full system (Eqs. 30 and 9) in nondimensionalized form reads:

$$\begin{aligned} \epsilon_E \frac{de}{d\tau} &= -e + ew_{EE} - \frac{iw_{EI}}{R} - \theta_E \\ \epsilon_I \frac{di}{d\tau} &= -i + Rgew_{IE} - giw_{II} - \theta_I \\ \frac{dw_{EE}}{d\tau} &= -i + 1 \\ \frac{dw_{EI}}{d\tau} &= +i - 1 \\ \frac{dw_{IE}}{d\tau} &= +R(e - 1) \\ \frac{dw_{II}}{d\tau} &= -R(e - 1) \end{aligned} \tag{33}$$

where we defined the new parameters

$$\begin{aligned} \epsilon_E &= \tau_E/\tau_0 \\ \epsilon_I &= \tau_I/\tau_0 \\ \tau_0 &= 1/(\beta g_E I_{set}) \\ R &= E_{set}/I_{set} \\ g &= g_I/g_E \\ \theta_E &= (g_E/E_{set})\Theta_E \\ \theta_I &= (g_I/I_{set})\Theta_I \end{aligned}$$

### 2.4.3 Quasi-steady state approximation

As before, we assume that the neural variables evolve in a much shorter time scale than synaptic variables. In the QSS approximation, the values of the nondimensionalized excitatory and inhibitory activities are

$$\begin{aligned} e_{qss} &= (\theta_I w_{EI} - R\theta_E(gw_{II} + 1))/(Rc) \\ i_{qss} &= (\theta_I(w_{EE} - 1) - Rg\theta_E w_{IE})/c \end{aligned} \tag{34}$$

where

$$c = gw_{EI}w_{IE} - (w_{II} + 1)(w_{EE} - 1)$$

The full system in the QSS approximation reads

$$\begin{aligned}\frac{dw_{EE}}{d\tau} &= -i_{qss} + 1 \\ \frac{dw_{EI}}{d\tau} &= +i_{qss} - 1 \\ \frac{dw_{IE}}{d\tau} &= +R(e_{qss} - 1) \\ \frac{dw_{II}}{d\tau} &= -R(e_{qss} - 1)\end{aligned} \tag{35}$$

where $e_{qss}$ and $i_{qss}$ are nonlinear functions of the weights as defined by Eq. 34.

Note that the Up state fixed point, defined by making all derivatives equal to zero, can be expressed as

$$\begin{aligned}e_{qss} &= 1 \\ i_{qss} &= 1\end{aligned} \tag{36}$$

which is the nondimensionalized version of Eq. 31. The weight values compatible with this condition are defined by equating Eqs. 34 and 36:

$$\begin{aligned}w_{EIup} &= R(w_{EEup} - 1) - R\theta_E \\ w_{IIup} &= Rw_{IEup} - \frac{\theta_I + 1}{g}\end{aligned} \tag{37}$$

($w_{EEup}$ and $w_{IEup}$ are free). This is the nondimensionalized version of Eq. 32.

#### 2.4.4 Stability condition

Now we show that this learning rule is stable. The Jacobian matrix evaluated at the Up state has two zero eigenvalues (as expected) and two nonzero eigenvalues. The nonzero eigenvalues have the form:

$$\lambda_{\pm} = \frac{\left(A \pm \sqrt{A^2 - DC}\right)(R+1)}{C} \tag{38}$$

where

$$\begin{aligned}A &= (R\theta_E - Rw_{EEup} + R - w_{IEup})g \\ C &= 2(\theta_I w_{EEup} - Rg\theta_E w_{IEup} - \theta_I) \\ D &= 2g\end{aligned}$$

To determine the sign of Eq. 38, first note that the factors $(R+1)$ and $D$ are positive definite. Second, $C$ must be positive because it is related to one of the stability conditions of the neural subsystem (Eq. 14, after substituting back to dimensionalized quantities). Note next that $A^2 - DC$ is less than $A^2$ (since $C$ and $D$ are positive), and thus the square root is either real and less than $|A|$ or imaginary, both cases leading to $\mathrm{Re}(A \pm \sqrt{A^2 - DC}) < 0$ if $A < 0$. The learning rule is then stable (both eigenvalues with negative real part) if $A < 0$. The instability condition in terms of the original parameters and variables reads:

$$\begin{aligned}&(E_{set}W_{EEup}\, g_E + I_{set}W_{IEup}\, g_E \\ &\quad - \Theta_E g_E - E_{set})g_I/(I_{set}g_E) > 0\end{aligned} \tag{39}$$

which, after switching $W_{EE} \to W_{EI}$ via Eq. 32, leads to

$$W_{EIup} + W_{IEup} > 0 \tag{40}$$

This condition is always satisfied because the weights are positive definite and thus the rule is stable for any choice of parameter values (as long as the neural subsystem is).

#### 2.4.5 Two-dimensional dynamics

The general program to assess stability of an isolated fixed point involves computing the eigenvalues of the Jacobian at the fixed point and looking at their real parts: a single eigenvalue with positive real part means the fixed point is unstable, while if all eigenvalues have negative real part the fixed point is stable. In case no eigenvalues have positive real part but at least one has zero real part, the linear stability analysis is inconclusive and one must go to higher orders by computing the center manifold.

However, the learning rules we study have nonisolated fixed points: the 2D planes defined by e.g. Eq. 32 where two weights are free. In this case, two out of the four eigenvalues are expected to be zero, meaning there is no dynamics along the 2D plane and thus there is no need to compute the center manifold.

In order to show that this is the case and build confidence in our results, here we perform a simple linear transformation of our variables and arrive without any further approximation at a system where two of the variables have exact zero dynamics. See

SageMath code in the Supplementary Material: `up states - CrossHomeostatic stability.ipynb`.

As a first step we switch from the nondimensionalized weights $w_{EE}, w_{EI}, w_{IE}, w_{II}$ to new variables $w_1, w_2, w_3, w_4$ that are zero at the fixed point:

$$\begin{aligned} w_1 &= w_{EE} - w_{EEup} \\ w_2 &= w_{EI} - w_{EIup} \\ w_3 &= w_{IE} - w_{IEup} \\ w_4 &= w_{II} - w_{IIup} \end{aligned} \tag{41}$$

where $w_{EEup}$ and $w_{IEup}$ are set to arbitrary values and $w_{EIup}$ and $w_{IIup}$ are defined by Eq. 37 (all satisfying the stability conditions of the neural subsystem, Eqs. 14 and 15). In the new variables, the learning rule Eq. 35 reads:

$$\begin{aligned} \frac{dw_1}{d\tau} &= f_1(w_1, w_2, w_3, w_4) \\ \frac{dw_2}{d\tau} &= f_2(w_1, w_2, w_3, w_4) \\ \frac{dw_3}{d\tau} &= f_3(w_1, w_2, w_3, w_4) \\ \frac{dw_4}{d\tau} &= f_4(w_1, w_2, w_3, w_4) \end{aligned} \tag{42}$$

where

$$\begin{aligned} f_1 = -i_{qss}(&w_1 + w_{EEup}, w_2 + w_{EIup}, \\ &w_3 + w_{IEup}, w_4 + w_{IIup}) + 1 \\ f_2 = \ldots&\text{etc.} \end{aligned}$$

according to the transformation Eq. 41.

The second step is a textbook diagonalization of the system. The Jacobian matrix of Eq. 42 has the following eigenvalues:

$$(0, 0, \lambda_3, \lambda_4)$$

where $\lambda_{3,4} < 0$ are the eigenvalues shown in Eq. 38. With this we compute the corresponding eigenvectors $v_1, v_2, v_3, v_4$ (the double eigenvalue 0 has a geometric multiplicity of 2 so there is no need to compute generalized eigenvectors and Jordan form) and arrange them in columns to form the diagonalizing matrix $\mathbf{T}$. The transformation of the variables $(w_1, w_2, w_3, w_4) \to (w, x, y, z)$ is such that, in matrix notation,

$$(w_1, w_2, w_3, w_4) = \mathbf{T}(w, x, y, z) \tag{43}$$

The transformation of the whole vector field, in matrix notation with $\mathbf{F} = (f_1, f_2, f_3, f_4)$, is

$$\frac{d(w, x, y, z)}{d\tau} = \mathbf{T}^{-1}\mathbf{F}(\mathbf{T}(w, x, y, z)) \tag{44}$$

which unfolded is

$$\begin{aligned} \frac{dw}{d\tau} &= 0 \\ \frac{dx}{d\tau} &= 0 \\ \frac{dy}{d\tau} &= \lambda_3 y + g_3(w, x, y, z) \\ \frac{dz}{d\tau} &= \lambda_4 z + g_4(w, x, y, z) \end{aligned} \tag{45}$$

showing that two directions have dynamics that is identically zero. Note that this result didn't involve any further approximation other than the QSS approximation of the neural subsystem.

## 2.5 Synaptic dynamics: *Balanced-Homeostatic* learning rule

(SageMath code in the Supplementary Material: `up states - Balanced stability.ipynb`)

In this section we follow a similar path as in previous sections, so we will skip some details.

### 2.5.1 Definition of the learning rule

In continuous-time dynamics, the Balanced-Homeostatic learning rule reads:

$$\begin{aligned} \frac{dW_{EE}}{dt} &= +\alpha_1 g_E E(E_{set} - E) \\ \frac{dW_{EI}}{dt} &= \frac{1}{\tau_0}(f(W_{EE}) - W_{EI}) \\ \frac{dW_{IE}}{dt} &= -\alpha_3 g_I I(I_{set} - I) \\ \frac{dW_{II}}{dt} &= \frac{1}{\tau_0}(g(W_{IE}) - W_{II}) \end{aligned} \tag{46}$$

where

$$f(W_{EE}) = \frac{(E_{set}W_{EE} - \Theta_E)g_E - E_{set}}{I_{set}g_E}$$
$$g(W_{IE}) = \frac{(E_{set}W_{IE} - \Theta_I)g_I - I_{set}}{I_{set}g_I} \tag{47}$$

are the Up state fixed point values of the weights (see below). The only fixed point compatible with the neural subsystem being suprathreshold is the Up state

$$\begin{aligned} E_{up} &= E_{set} \\ I_{up} &= I_{set} \\ W_{EIup} &= f(W_{EEup}) \\ W_{IIup} &= g(W_{IEup}) \end{aligned} \tag{48}$$

where $W_{EEup}$ and $W_{IEup}$ are free (not all values will lead to stable Up states, though).

### 2.5.2 Nondimensionalization

The full system (Eqs. 46 and 9) in nondimensionalized form reads:

$$\begin{aligned} \epsilon_E \frac{de}{d\tau} &= -e + Rgew_{EE} - iw_{EI} - \theta_E \\ \epsilon_I \frac{di}{d\tau} &= -i + \frac{ew_{IE}}{\alpha g} - iw_{II} - \theta_I \\ \frac{dw_{EE}}{d\tau} &= -(e-1)e \\ \frac{dw_{EI}}{d\tau} &= -w_{EI} + Rgw_{EE} - \theta_E - 1 \\ \frac{dw_{IE}}{d\tau} &= +(i-1)i \\ \frac{dw_{II}}{d\tau} &= -w_{II} + \frac{w_{IE}}{\alpha g} - \theta_I - 1 \end{aligned} \tag{49}$$

where we defined the new parameters

$$\begin{aligned} \epsilon_E &= \tau_E/\tau_0 \\ \epsilon_I &= \tau_I/\tau_0 \\ \tau_0 &= 1/(\alpha_1 g_E g_I E_{set} I_{set}) \\ R &= E_{set}/I_{set} \\ \alpha &= \alpha_1/\alpha_3 \\ g &= g_I/g_E \\ \theta_E &= (g_E/E_{set})\Theta_E \\ \theta_I &= (g_I/I_{set})\Theta_I \end{aligned}$$

### 2.5.3 Quasi-steady state approximation

As before, we assume that the neural variables evolve in a much shorter time scale than synaptic variables. In the QSS approximation, the values of the nondimensionalized excitatory and inhibitory activities are

$$\begin{aligned} e_{qss} &= (\theta_I w_{EI} - \theta_E(w_{II}+1))\alpha g/c \\ i_{qss} &= (\alpha g \theta_I (Rgw_{EE} - 1) - \theta_E w_{IE})/c \end{aligned} \tag{50}$$

where

$$c = (w_{II}+1)\alpha g(Rgw_{EE} - 1) - w_{EI}w_{IE}$$

The full system in the QSS approximation reads

$$\begin{aligned} \frac{dw_{EE}}{d\tau} &= -(e_{qss}-1)e_{qss} \\ \frac{dw_{EI}}{d\tau} &= -w_{EI} + Rgw_{EE} - \theta_E - 1 \\ \frac{dw_{IE}}{d\tau} &= +(i_{qss}-1)i_{qss} \\ \frac{dw_{II}}{d\tau} &= -w_{II} + \frac{w_{IE}}{\alpha g} - \theta_I - 1 \end{aligned} \tag{51}$$

where $e_{qss}$ and $i_{qss}$ are nonlinear functions of the weights as defined by Eq. 50.

Note that the Up state fixed point, defined by making all derivatives equal to zero, are the weight values that satisfy

$$\begin{aligned} e_{qss} &= 1 \\ i_{qss} &= 1 \end{aligned} \tag{52}$$

and

$$\begin{aligned} w_{EIup} &= Rgw_{EEup} - \theta_E - 1 \\ w_{IIup} &= \frac{w_{IEup}}{\alpha g} - \theta_I - 1 \end{aligned} \tag{53}$$

which are the nondimensionalized versions of Eqs. 48 and 47, respectively. Weights $w_{EEup}$ and $w_{IEup}$ are free, so the Up state is a 2D plane of non-isolated fixed points.

### 2.5.4 Stability condition

Now we show the conditions for which this learning rule is stable. The Jacobian matrix of Eq. 51 evaluated at the Up state has two zero eigenvalues (as expected) and two nonzero eigenvalues. The nonzero eigenvalues have the form:

$$\lambda_{\pm} = \frac{A \pm \sqrt{A^2 + FC}}{C} \tag{54}$$

where

$$C = 2(R\alpha g^2 \theta_I w_{EEup} - \theta_E w_{IEup} - \alpha g \theta_I)$$

and $A$ and $F$ are long expressions that can be found in the corresponding Jupyter notebook.

To determine the sign of Eq. 54, first note that $C$ must be positive because it is related to one of the stability conditions of the neural subsystem (Eq. 14, after substituting back to dimensionalized quantities). Next note that both eigenvalues have negative real part if $F < 0$ and $A < 0$: if $F < 0$ then the argument of the square root is less than $A^2$, and then the square root itself is either real and less than $|A|$, or imaginary. In any case, if in addition $A < 0$ then both eigenvalues have negative real part. The stability conditions are then $F < 0$ and $A < 0$, which can be written in terms of the dimensionalized parameters and variables as

$$\begin{aligned} a_1 + (b_1 - c_1)d &< (b_1' + c_1')e \\ a_2 + (b_2 - c_2)d &< (b_2' + c_2')e \end{aligned} \tag{55}$$

where

$$\begin{aligned}
a_1 &= (\Theta_E \Theta_I \, g_E g_I + E_{set}^2 \frac{\alpha_3}{\alpha_1}) I_{set} \, g_E g_I \\
b_1 &= I_{set}^2 \Theta_E \, g_E^2 g_I \\
c_1 &= E_{set}^2 I_{set} \, g_E^2 \\
b_1' &= E_{set} I_{set} \Theta_I \, g_E g_I^2 \\
c_1' &= E_{set} I_{set}^2 \, g_I^2 \frac{\alpha_3}{\alpha_1} \\
a_2 &= 2\Theta_E \Theta_I \, g_E^2 g_I^2 \\
b_2 &= 2 I_{set} \Theta_E \, g_E^2 g_I \\
c_2 &= E_{set}^2 \, g_E^2 \\
b_2' &= 2 E_{set} \Theta_I g_E \, g_I^2 \\
c_2' &= E_{set} I_{set} \, g_I^2 \frac{\alpha_3}{\alpha_1} \\
d &= W_{IIup} \, g_I + 1 \\
e &= W_{EEup} \, g_E - 1
\end{aligned}$$

### 2.5.5 Numerical analysis

It is hard to analytically determine whether the stability conditions Eq. 55 are satisfied or not in the general case. However, here we show that for the biologically backed set of parameter values in Table 1 (with the addition of $\alpha_1 = 0.001$ and $\alpha_3 = 0.0001$) the conditions are satisfied and the learning rule is thus stable.

In Figure S5 we plot the stability conditions of this rule. It is clear that the stability region of the neural subsystem lies well within the stability region of the learning rule, making the full system stable.

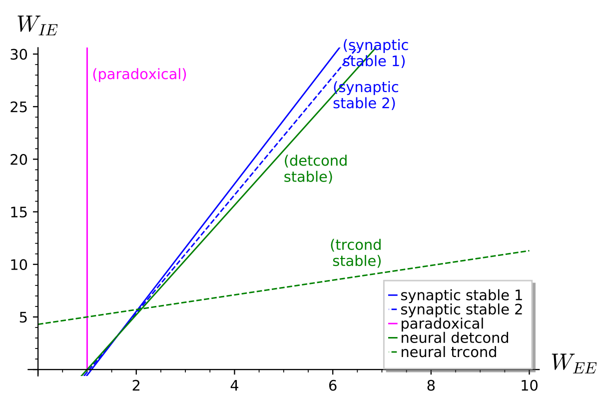


Figure S5: Regions of stability. The neural subsystem is stable in a region where the Balanced-Homeostatic rule is also stable. Synaptic 1 and 2: Eq. 55; Paradoxical: Eq. 16; Neural detcond: Eq. 14; Neural trcond: Eq. 15.

## 2.6 Synaptic dynamics: *Synaptic-Scaling* learning rule

(SageMath code in the Supplementary Material: `up states - SynapticScaling stability.ipynb`)

In this section we follow a similar path as in previous sections, so we will skip some details.

### 2.6.1 Definition of the learning rule

In continuous-time dynamics, the SynapticScaling learning rule [6] reads:

$$\begin{aligned}
\frac{dW_{EE}}{dt} &= +\gamma g_E(E_{set} - E)W_{EE} \\
\frac{dW_{EI}}{dt} &= -\gamma g_E(E_{set} - E)W_{EI} \\
\frac{dW_{IE}}{dt} &= +\gamma g_I(I_{set} - I)W_{IE} \\
\frac{dW_{II}}{dt} &= -\gamma g_I(I_{set} - I)W_{II}
\end{aligned} \tag{56}$$

The only fixed point compatible with the neural subsystem being above threshold is the Up state and is the set of weight values such that:

$$\begin{aligned}
E_{up} &= E_{set} \\
I_{up} &= I_{set}
\end{aligned} \tag{57}$$

The values of the weights corresponding to the Up state are given by the (underdetermined) system defined by equating Eq. 57 and Eq. 10. Since it is a two-equation system for a set of four unknown weights, there are two free weights that we choose to be $W_{EEup}$ and $W_{IEup}$. The other two have the following values:

$$\begin{aligned}
W_{EIup} &= \frac{(E_{set}W_{EEup} - \Theta_E)g_E - E_{set}}{I_{set}g_E} \\
W_{IIup} &= \frac{(E_{set}W_{IEup} - \Theta_I)g_I - I_{set}}{I_{set}g_I}
\end{aligned} \tag{58}$$

This means that there is an infinite number of weight values compatible with the Up state (possibly not all stable, though).

### 2.6.2 Nondimensionalization

The full system (Eqs. 56 and 9) in nondimensionalized form reads:

$$\begin{aligned}
\epsilon_E \frac{de}{d\tau} &= -e + ew_{EE} - iw_{EI} - \theta_E \\
\epsilon_I \frac{di}{d\tau} &= -i + ew_{IE} - iw_{II} - \theta_I \\
\frac{dw_{EE}}{d\tau} &= -(e-1)w_{EE} \\
\frac{dw_{EI}}{d\tau} &= +(e-1)w_{EI} \\
\frac{dw_{IE}}{d\tau} &= -(i-1)w_{IE}/R \\
\frac{dw_{II}}{d\tau} &= +(i-1)w_{II}/R
\end{aligned} \tag{59}$$

where we defined the new parameters

$$\begin{aligned}
\epsilon_E &= \tau_E/\tau_0 \\
\epsilon_I &= \tau_I/\tau_0 \\
\tau_0 &= 1/(\gamma E_{set}) \\
R &= E_{set}/I_{set} \\
g &= g_I/g_E \\
\theta_E &= (g_E/E_{set})\Theta_E \\
\theta_I &= (g_I/I_{set})\Theta_I
\end{aligned}$$

### 2.6.3 Quasi-steady state approximation

As before, we assume that the neural variables evolve in a much shorter time scale than synaptic variables. In the QSS approximation, the values of the nondimensionalized excitatory and inhibitory activities are

$$\begin{aligned}
e_{qss} &= (\theta_I w_{EI} - R\theta_E(w_{II}+1))/c \\
i_{qss} &= (\theta_I(w_{EE}-1) - \theta_E w_{IE})/c
\end{aligned} \tag{60}$$

where

$$c = w_{EI}w_{IE} - (w_{II}+1)(w_{EE}-1)$$

The full system in the QSS approximation reads

$$\begin{aligned}\frac{dw_{EE}}{d\tau} &= -(e_{qss}-1)w_{EE} \\ \frac{dw_{EI}}{d\tau} &= +(e_{qss}-1)w_{EI} \\ \frac{dw_{IE}}{d\tau} &= -(i_{qss}-1)w_{IE}/R \\ \frac{dw_{II}}{d\tau} &= +(i_{qss}-1)w_{II}/R\end{aligned} \tag{61}$$

where $e_{qss}$ and $i_{qss}$ are nonlinear functions of the weights as defined by Eq. 60.

Note that the Up state fixed point, defined by making all derivatives equal to zero, can be expressed as

$$\begin{aligned}e_{qss} &= 1 \\ i_{qss} &= 1\end{aligned} \tag{62}$$

which is the nondimensionalized version of Eq. 57. The weight values compatible with this condition are defined by equating Eqs. 60 and 62:

$$\begin{aligned}w_{EI\,up} &= w_{EE\,up} - 1 - \theta_E \\ w_{II\,up} &= w_{IE\,up} - 1 - \theta_I\end{aligned} \tag{63}$$

($w_{EE\,up}$ and $w_{IE\,up}$ are free). This is the nondimensionalized version of Eq. 58.

### 2.6.4 Instability condition

Now we show that this learning rule is unstable for biologically backed parameter values. The Jacobian matrix evaluated at the Up state has two zero eigenvalues (as expected) and two nonzero eigenvalues. The nonzero eigenvalues have the form:

$$\lambda_{\pm} = \frac{A \pm \sqrt{A^2 - DC}}{C} \tag{64}$$

where

$$\begin{aligned}C &= 2R(\theta_I w_{EE\,up} - \theta_E w_{IE\,up} - \theta_I) \\ D &= 2(\theta_E - 2w_{EE\,up} + 1)(\theta_I - 2w_{IE\,up} + 1)\end{aligned}$$

and $A$ is a long expression that can be found in the corresponding Jupyter notebook.

To determine the sign of Eq. 64, first note that $C$ must be positive because it is related to one of the stability conditions of the neural subsystem (Eq. 14, after substituting back to dimensionalized quantities). Next note that $D$, after switching $w_{EE} \to w_{EI}$ and $w_{IE} \to w_{II}$ by means of Eq. 63, reads

$$D = 2(\theta_E + 2w_{EI\,up} + 1)(\theta_I + 2w_{II\,up} + 1)$$

which is positive definite

Note next that $A^2 - DC$ is less than $A^2$ (since $C$ and $D$ are positive), and thus the square root is either real and less than $|A|$ or imaginary, both cases leading to $\mathrm{Re}(A \pm \sqrt{A^2 - DC}) > 0$ if $A > 0$. The learning rule is then unstable (both eigenvalues with positive real part) if $A > 0$. This condition can be written as

$$(W_{EE\,up}\, g_E - 1)a > (W_{II\,up}\, g_I + 1)a' \tag{65}$$

where

$$\begin{aligned}a &= 2I_{set} W_{II\,up}\, g_I + \Theta_I\, g_I + I_{set} \\ a' &= 2E_{set} W_{EE\,up}\, g_E - \Theta_E\, g_E - E_{set}\end{aligned}$$

### 2.6.5 Analysis of the instability condition

In a biologically backed set of parameter values the following is true:

$$\begin{aligned}I_{set} &> E_{set} \\ g_I &> g_E \\ \Theta_I &> \Theta_E\end{aligned}$$

Keeping this in mind, and taking into account that $a$ is the sum of positive terms while $a'$ is the difference of positive terms, we can safely say that $a > a'$ (despite $W_{II\,up} \leq W_{EE\,up}$), and thus the instability condition Eq. 65 is satisfied despite that one of the stability conditions of the neural subsystem requires that $(W_{EE\,up}\, g_E - 1)\tau_I > (W_{II\,up}\, g_I + 1)\tau_E$, see Eq. 15). The SynapticScaling rule is then unstable.

### 2.6.6 Numerical analysis

As an illustration of the reasoning above, we express the instability condition Eq. 65 in terms of the free weights $W_{EE}$ and $W_{IE}$ by means of Eq. 58, and plot it with parameter values as in Table 1.

The instability condition is a homographic function (i.e. a hyperbola) with instability regions in its first and third quadrants. It is clear from Figure S6 that the neural subsystem is stable in a region that is entirely within the instability region of the synaptic subsystem.

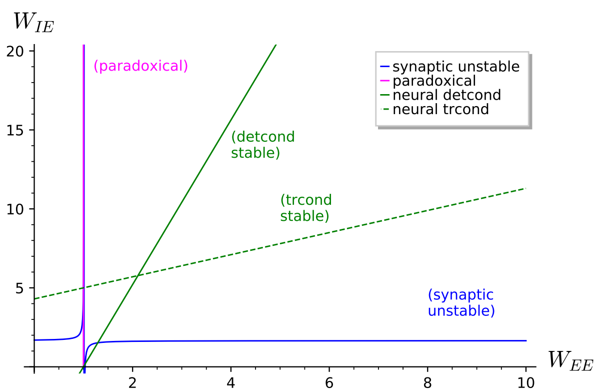


Figure S6: Regions of stability. The SynapticScaling learning rule is unstable where the neural subsystem is stable. Synaptic: Eq. 65; Paradoxical: Eq. 16; Neural detcond: Eq. 14; Neural trcond: Eq. 15.

# 3 Learning rule from loss function

(SageMath code in the Supplementary Material: `up states - Loss function.ipynb`)

Here we show how to compute the learning rule starting from a loss function. Then we make an approximation by considering the weight values are small, and take that as input to interpret several learning rules that don't come from a loss function.

## 3.1 General prescription

We consider the full neural+synaptic system in the QSS approximation (see e.g. Section 2.3). In this approximation the neural subsystem is represented by the quasi-steady-state values

$$\begin{aligned} E &= E_{up}(W_{EE}, W_{EI}, W_{IE}, W_{II}) \\ I &= I_{up}(W_{EE}, W_{EI}, W_{IE}, W_{II}) \end{aligned} \tag{66}$$

where the functions $E_{up}$ and $I_{up}$ are defined by Eq. 10 (see [7] for a related discussion on quasi-steady state, synaptic plasticity, and gradient descent).

The synaptic subsystem, that is the learning rule, will be obtained as a result of considering a specific loss function, and the general prescription to compute the learning rule from a loss function $L$ is the following:

1. Consider a loss function depending on $E$ and $I$ (which in turn depend on all weights):

$$L = L(E, I)$$

Conditions to be satisfied by the loss function are, for instance, to be smooth enough (i.e. continuous and differentiable) and to have a minimum when the activities $E$ and $I$ are at the set points $E_{set}$ and $I_{set}$ (i.e. homeostatic plasticity).

2. The dynamics of the weights is such that it follows a gradient descent on the loss function towards its minimum. In vector notation :

$$\Delta \mathbf{W} = -\alpha \nabla L \tag{67}$$

with learning rate $\alpha$. The unfolded learning rules, that is the equations that govern the weights' dynamics, are then

$$\begin{aligned} \Delta W_{EE} &= -\alpha \frac{\partial L}{\partial W_{EE}} \\ \Delta W_{EI} &= -\alpha \frac{\partial L}{\partial W_{EI}} \\ \Delta W_{IE} &= -\alpha \frac{\partial L}{\partial W_{IE}} \\ \Delta W_{II} &= -\alpha \frac{\partial L}{\partial W_{II}} \end{aligned} \tag{68}$$

3. The partial derivatives of the loss function in Eq. 68 are:

$$\begin{aligned} \frac{\partial L}{\partial W_{EE}} &= \frac{\partial L}{\partial E}\frac{\partial E}{\partial W_{EE}} + \frac{\partial L}{\partial I}\frac{\partial I}{\partial W_{EE}} \\ \frac{\partial L}{\partial W_{EI}} &= \frac{\partial L}{\partial E}\frac{\partial E}{\partial W_{EI}} + \frac{\partial L}{\partial I}\frac{\partial I}{\partial W_{EI}} \\ \frac{\partial L}{\partial W_{IE}} &= \frac{\partial L}{\partial E}\frac{\partial E}{\partial W_{IE}} + \frac{\partial L}{\partial I}\frac{\partial I}{\partial W_{IE}} \\ \frac{\partial L}{\partial W_{II}} &= \frac{\partial L}{\partial E}\frac{\partial E}{\partial W_{II}} + \frac{\partial L}{\partial I}\frac{\partial I}{\partial W_{II}} \end{aligned} \tag{69}$$

or, in vector notation:

$$\nabla L = \frac{\partial L}{\partial E}\nabla E + \frac{\partial L}{\partial I}\nabla I \tag{70}$$

Here we use the chain rule for the derivatives because it gives us much more compact expressions at the end.

4. The partial derivatives in the gradients $\nabla E = \left(\frac{\partial E}{\partial W_{EE}}, \dots\right)$ and $\nabla I = \left(\frac{\partial I}{\partial W_{EE}}, \dots\right)$ etc. are to be taken from the quasi-steady-state values of $E$ and $I$, Eq. 66. We will, however, compute the partial derivatives from the implicit expressions given by setting $dE/dt = dI/dt = 0$ in Eq. 9 without solving for $E$ and $I$.

## 3.2 Detailed calculation

**Loss function** We choose a very general loss function that depends homeostatically on both $E$ and $I$ activities:

$$L(E, I) = \frac{1}{2}(E_{set} - E)^2 + \frac{1}{2}(I_{set} - I)^2 \tag{71}$$

This loss function is an elliptic paraboloid in $(E, I)$ space with a global minimum at $(E_{set}, I_{set})$ so a gradient descend learning rule as above should converge to that minimum (see Liapunov function and gradient systems: [3, Section 1.1B][8, Sections 9.3 and 9.4][2, Section 7.2]. Keep in mind, however, that $L$ has a different shape when expressed as a function of the weights, and that $E$ and $I$ are not necessarily monotonic functions of the weights, so the conditions for the set point of $L$ to be stable or a global minimum or even unique are not necessarily satisfied.

**Partial derivatives of $L$** The partial derivatives of $L$ with respect to $E$ and $I$ are simply

$$\begin{aligned}\frac{\partial L}{\partial E} &= -(E_{set} - E)\\ \frac{\partial L}{\partial I} &= -(I_{set} - I)\end{aligned} \tag{72}$$

**Partial derivatives of $E$ and $I$** We compute the partial derivatives $\partial X/\partial W_{XY}$ $(X, Y = E, I)$ by first functions defined by equating the neural subsystem (Eq. 9) to zero:

$$\begin{aligned}E &= g_E(W_{EE}E - W_{EI}I - \Theta_E))\\ I &= g_I(W_{IE}E - W_{II}I - \Theta_I))\end{aligned} \tag{73}$$

then differentiating the implicit functions:

$$\begin{aligned}
\frac{\partial E}{\partial W_{EE}} &= g_E(E + W_{EE}\frac{\partial E}{\partial W_{EE}}) - g_E W_{EI}\frac{\partial I}{\partial W_{EE}}\\
\frac{\partial E}{\partial W_{EI}} &= g_E W_{EE}\frac{\partial E}{\partial W_{EI}} - g_E(I + W_{EI}\frac{\partial I}{\partial W_{EI}})\\
\frac{\partial E}{\partial W_{IE}} &= g_E W_{EE}\frac{\partial E}{\partial W_{IE}} - g_E W_{EI}\frac{\partial I}{\partial W_{IE}}\\
\frac{\partial E}{\partial W_{II}} &= g_E W_{EE}\frac{\partial E}{\partial W_{II}} - g_E W_{EI}\frac{\partial I}{\partial W_{II}}\\
\frac{\partial I}{\partial W_{EE}} &= g_I W_{IE}\frac{\partial E}{\partial W_{EE}} - g_I W_{II}\frac{\partial I}{\partial W_{EE}}\\
\frac{\partial I}{\partial W_{EI}} &= g_I W_{IE}\frac{\partial E}{\partial W_{EI}} - g_I W_{II}\frac{\partial I}{\partial W_{EI}}\\
\frac{\partial I}{\partial W_{IE}} &= g_I(E + W_{IE}\frac{\partial E}{\partial W_{IE}}) - g_I W_{II}\frac{\partial I}{\partial W_{IE}}\\
\frac{\partial I}{\partial W_{II}} &= g_I W_{IE}\frac{\partial E}{\partial W_{II}} - g_I(I + W_{II}\frac{\partial I}{\partial W_{II}})
\end{aligned} \tag{74}$$

and then solving for the derivatives:

$$\begin{aligned}
\frac{\partial E}{\partial W_{EE}} &= -(EW_{II}\, g_E\, g_I + Eg_E)/C\\
\frac{\partial E}{\partial W_{EI}} &= (IW_{II}\, g_E\, g_I + Ig_E)/C\\
\frac{\partial E}{\partial W_{IE}} &= EW_{EI}\, g_E\, g_I/C\\
\frac{\partial E}{\partial W_{II}} &= -IW_{EI}\, g_E\, g_I/C\\
\frac{\partial I}{\partial W_{EE}} &= -EW_{IE}\, g_E\, g_I/C\\
\frac{\partial I}{\partial W_{EI}} &= IW_{IE}\, g_E\, g_I/C\\
\frac{\partial I}{\partial W_{IE}} &= (EW_{EE}\, g_E - E)g_I/C\\
\frac{\partial I}{\partial W_{II}} &= -(IW_{EE}\, g_E - I)g_I/C
\end{aligned} \tag{75}$$

where

$$C = W_{EI}W_{IE}\, g_E\, g_I - (W_{II}\, g_I + 1)(W_{EE}\, g_E - 1)$$

**Exact learning rules** Putting everything together, the learning rules Eq. 68 are:

$$\begin{aligned}
\Delta W_{EE} =& -\frac{\alpha}{C}((I_{set} - I)EW_{IE}\, g_e\, g_I \\
&\quad + (E_{set} - E)E(W_{II}\, g_I + 1)g_E) \\
\Delta W_{EI} =& +\frac{\alpha}{C}((I_{set} - I)IW_{IE}\, g_e\, g_I \\
&\quad + (E_{set} - E)I(W_{II}\, g_I + 1)g_E) \\
\Delta W_{IE} =& +\frac{\alpha}{C}((E_{set} - E)EW_{EI}\, g_e\, g_I \\
&\quad + (I_{set} - I)E(W_{EE}\, g_E - 1)g_I) \\
\Delta W_{II} =& -\frac{\alpha}{C}((E_{set} - E)IW_{EI}\, g_e\, g_I \\
&\quad + (I_{set} - I)I(W_{EE}\, g_E - 1)g_I)
\end{aligned} \tag{76}$$

Note that these are very complicated, nonlinear expressions because both $E$ and $I$ depend on all weights via Eq. 73. Also the denominator $C$ depends on all weights (see previous paragraph).

**Small weights approximation** We want simpler expressions for the learning rules. Note that the exact expressions above all have a homeostatic factor (either $E - E_{set}$ or $I - I_{set}$) and a presynaptic factor ($E$ in $\Delta W_{EE}$ and $\Delta W_{IE}$ and $I$ in $\Delta W_{EI}$ and $\Delta W_{II}$). Despite their complicated dependence on the weights, both factors have simple interpretations so we want to keep them as they are while expanding the rest of the expressions (explicit dependence on the weights including $C$) as a first-order Taylor series around zero. Although this is not a textbook Taylor expansion of the full expressions, it is very informative because the results can be much easily interpreted (for a similar approach see [7]:

$$\begin{aligned}
\Delta W_{EE} =& +\alpha((E_{set} - E)E\, g_E \\
&\quad + (E_{set} - E)EW_{EE}\, g_E^2 \\
&\quad + (I_{set} - I)EW_{IE}\, g_E\, g_I) \\
\Delta W_{EI} =& -\alpha((E_{set} - E)I\, g_E \\
&\quad + (E_{set} - E)IW_{EE}\, g_E^2 \\
&\quad + (I_{set} - I)IW_{IE}\, g_E\, g_I) \\
\Delta W_{IE} =& +\alpha((I_{set} - I)E\, g_I \\
&\quad - (I_{set} - I)EW_{II}\, g_I^2 \\
&\quad - (E_{set} - E)EW_{EI}\, g_E\, g_I) \\
\Delta W_{II} =& -\alpha((I_{set} - I)I\, g_I \\
&\quad - (I_{set} - I)IW_{II}\, g_I^2 \\
&\quad - (E_{set} - E)IW_{EI}\, g_E\, g_I)
\end{aligned} \tag{77}$$

Note that the first terms of these expressions (corresponding to the zeroth order in the approximation) are exactly the standard Homeostatic learning rules, Eq. 17. Also note that the homeostatic factors in the third terms have the sign corresponding to the Cross-Homeostatic rules, Eq. 30.

## References


1. Keener, J. P. & Sneyd, J. *Mathematical physiology* (Springer, 1998).
2. Strogatz, S. H. *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering* (CRC press, 2018).
3. Wiggins, S. *Introduction to applied nonlinear dynamical systems and applications* (Springer-Verlag, 1996).
4. Seung, H. S. How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences* **93,** 13339–13344 (1996).
5. Seung, H. S. Continuous attractors and oculomotor control. *Neural Networks* **11,** 1253–1258 (1998).

6. Van Rossum, M. C., Bi, G. Q. & Turrigiano, G. G. Stable Hebbian learning from spike timing-dependent plasticity. *J Neurosci* **20,** 8812–8821 (2000).

7. Mackwood, O., Naumann, L. B. & Sprekeler, H. Learning excitatory-inhibitory neuronal assemblies in recurrent networks. *bioRxiv.* `https://doi.org/10.1101/2020.03.30.016352` (2020).

8. Hirsch, M. W. & Smale, S. *Differential equations, dynamical systems, and linear algebra* (Academic press, 1974).