PNAS RESEARCH ARTICLE | NEUROSCIENCE OPEN ACCESS

Paradoxical self-sustained dynamics emerge from orchestrated excitatory and inhibitory homeostatic plasticity rules

Saray Soldado-Magraner[a], Michael J. Seay[a], Rodrigo Laje[b,c,1], and Dean V. Buonomano[a,d,1,2]

Edited by Terrence Sejnowski, Salk Institute for Biological Studies, La Jolla, CA; received January 17, 2022; accepted August 30, 2022

Self-sustained neural activity maintained through local recurrent connections is of fundamental importance to cortical function. Converging theoretical and experimental evidence indicates that cortical circuits generating self-sustained dynamics operate in an inhibition-stabilized regime. Theoretical work has established that four sets of weights ($W_{E\leftarrow E}$, $W_{E\leftarrow I}$, $W_{I\leftarrow E}$, and $W_{I\leftarrow I}$) must obey specific relationships to produce inhibition-stabilized dynamics, but it is not known how the brain can appropriately set the values of all four weight classes in an unsupervised manner to be in the inhibition-stabilized regime. We prove that standard homeostatic plasticity rules are generally unable to generate inhibition-stabilized dynamics and that their instability is caused by a signature property of inhibition-stabilized networks: the paradoxical effect. In contrast, we show that a family of "cross-homeostatic" rules overcome the paradoxical effect and robustly lead to the emergence of stable dynamics. This work provides a model of how—beginning from a silent network—self-sustained inhibition-stabilized dynamics can emerge from learning rules governing all four synaptic weight classes in an orchestrated manner.

inhibition-stabilized networks | paradoxical effect | homeostatic plasticity

Self-sustained patterns of neural activity maintained by local recurrent excitation underlie many cortical computations and dynamic regimes, including the persistent activity associated with working memory (1–3), motor control (4, 5), asynchronous states associated with the default cortical dynamic regime (6–10), and up-states (8, 11). Recurrent excitation also has the potential to drive pathological and epileptiform regimes (12, 13). Converging theoretical and experimental evidence indicates that cortical circuits that generate self-sustained dynamics operate in an inhibition-stabilized regime, in which positive feedback is held in check by recurrent inhibition (7, 14–20). There is also evidence that inhibition-stabilized regimes may comprise the default awake cortical dynamic regime (8, 9, 20).

At the computational level inhibition-stabilized networks are often modeled as a simplified circuit composed of interconnected excitatory (*E*) and inhibitory (*I*) neural populations with four classes of synaptic weights: $W_{E\leftarrow E}$, $W_{E\leftarrow I}$, $W_{I\leftarrow E}$, and $W_{I\leftarrow I}$. In the inhibition-stabilized regime, recurrent excitation produces positive feedback, which is held in check by rapid inhibition. The dynamics settles into a stable fixed-point attractor and instantiates an inhibition-stabilized network. A signature of the inhibition-stabilized regime is the presence of the paradoxical effect, in which an *increase* in excitatory drive to inhibitory neurons produces a net *decrease* in the firing rate of those same inhibitory neurons (14–16, 18), a phenomenon that has been observed in the awake resting cortex (Fig. 1*A*) (20). Analytical and numerical studies have demonstrated that in order to support inhibition-stabilized dynamics, the four weight classes must obey certain "balanced" relationships; for example, if excitation is too strong, runaway (or saturated) excitation occurs, whereas if inhibition is too strong the activity falls into a quiescent fixed point (6, 7, 14–16, 19). However, in most computational models the set of four weights is determined analytically or through numerical searches—or in a few cases by allowing one or two weights to be plastic while appropriately hardwiring the others. In contrast, experimental studies both in vitro and in vivo have shown that self-sustained activity emerges autonomously during development (21–26), indicating that synaptic plasticity rules are in place to orchestrate the unsupervised emergence of inhibition-stabilized dynamics. Additionally, because the four weight classes have been observed to undergo synaptic plasticity in experimental studies (27–32), here we ask how inhibition-stabilized dynamics might emerge in a self-organizing manner.

One possibility is that standard homeostatic forms of plasticity underlie the emergence of inhibition-stabilized dynamics. Homeostatic plasticity rules generally assume that excitatory weights are regulated in a manner proportional to the difference between some ontogenetically determined setpoint and average neural activity (for both excitatory and inhibitory neurons)—and conversely that inhibitory weights onto excitatory neurons are regulated in the opposite direction (33–38). However, it remains an open question whether homeostatic rules can lead to the self-organized emergence of

Significance

Cortical networks have the remarkable ability to self-assemble into dynamic regimes in which excitatory positive feedback is balanced by recurrent inhibition. This inhibition-stabilized regime is increasingly viewed as the default dynamic regime of the cortex, but how it emerges in an unsupervised manner remains unknown. We prove that classic forms of homeostatic plasticity are unable to drive recurrent networks to an inhibition-stabilized regime due to the well-known paradoxical effect. We next derive a novel family of cross-homeostatic rules that lead to the unsupervised emergence of inhibition-stabilized networks. These rules shed new light on how the brain may reach its default dynamic state and provide a valuable tool to self-assemble artificial neural networks into ideal computational regimes.

Author affiliations: [a]Department of Neurobiology, University of California, Los Angeles, CA 90095; [b]Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Bernal, B1876BXD Argentina; [c]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, C1425FQB Argentina; and [d]Department of Psychology, University of California, Los Angeles, CA 90095

Author contributions: S.S.-M. and D.V.B. designed research; S.S.-M., M.J.S., and R.L. performed research; R.L. contributed new reagents/analytic tools; S.S.-M., M.J.S., and R.L. analyzed data; and S.S.-M. and D.V.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1] R.L. and D.V.B. are joint senior authors on this work.

[2] To whom correspondence may be addressed. Email: dbuono@ucla.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2200621119/-/DCSupplemental.

Published October 17, 2022.

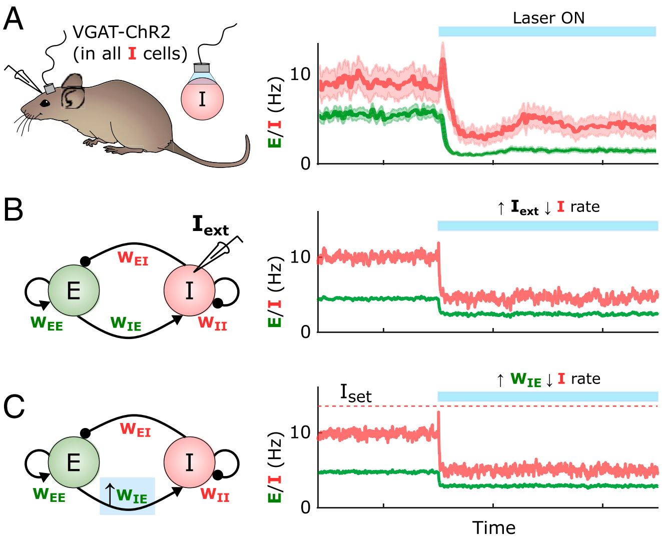


Fig. 1. The paradoxical effect in cortical circuits and its implications for plasticity. (*A*) Average inhibitory (red) and excitatory (green) firing rates in the visual cortex of awake mice in the absence of explicit external sensory stimulation. When inhibitory neurons are optogenetically activated, the firing rates of the inhibitory neurons show a paradoxical decrease in activity during the stimulation, indicative of an inhibition-stabilized network. Adapted from Sanzeni et al. (20). (*B*) Two-population firing rate model of self-sustained cortical activity. The dynamics of the excitatory (green) and inhibitory (red) populations are governed by four synaptic weights, $W_{E\leftarrow E}$, $W_{E\leftarrow I}$, $W_{I\leftarrow E}$, and $W_{I\leftarrow I}$. Similar to the experimental case shown in *A*, the model shows the paradoxical effect when the inhibitory population is excited via an external current $I_{ext} = 7$. Weights were initialized to $W_{EE} = 5$, $W_{EI} = 1.52$, $W_{IE} = 10$, and $W_{II} = 2.25$. (*C*) As in *B*, if an inhibitory population is firing below its homeostatic setpoint, and one were to increase its excitatory weights according to standard homeostatic rules, the increase in excitatory weights would produce a paradoxical and antihomeostatic decrease in inhibitory neuron firing. Weights were initialized to $W_{EE} = 5$, $W_{EI} = 1.52$, $W_{IE} = 10$, and $W_{II} = 2.25$, with a later increase of $W_{IE} = 12$.

inhibition-stabilized networks. Here we use computational models and analytical methods to explore families of homeostatic plasticity rules that operate in parallel in all four synapse classes and lead to inhibition-stabilized dynamics. We show that when driving the network toward inhibition-stabilized regimes, standard forms of homeostatic plasticity are stable only in a narrow region of parameter space. Here we prove that this instability arises from the paradoxical effect. Indeed, it can be seen that like increasing external input to an inhibitory neuron, increasing the excitatory weights onto an inhibitory neuron firing below its setpoint produces a paradoxical (and antihomeostatic) decrease in inhibitory activity (Fig. 1 *B* and *C*). While inhibition-stabilized regimes can operate in the presence or absence of external input, here we focus primarily on "fully self-sustained" activity in the absence of any external tonic input; however, we show that our results also apply when external inputs are present.

We conclude that homeostatic manipulations in the inhibitory population lead to paradoxical outcomes, making the rules unstable in this context. Therefore, homeostatic plasticity rules that aim to bring the network to the relevant dynamic regime of the cerebral cortex must work in paradoxical conditions. We developed a family of homeostatic plasticity rules that include "cross-homeostatic" influences and lead to the unsupervised emergence of fully self-sustained dynamics in the inhibition-stabilized regime in a robust manner. These rules are consistent with experimental data and generate explicit predictions regarding the effects of manipulations of excitatory and inhibitory neurons on synaptic plasticity.

Results

Standard Homeostatic Plasticity Rules Cannot Account for the Emergence of Stable Self-Sustained Activity. We first ask, when starting from a network in a silent regime similar to cortical circuits early in development, whether standard homeostatic rules can drive networks to stable self-sustained dynamics in the absence of any external input. Based on experimental studies we assume that both excitatory and inhibitory neurons exhibit ontogenetically programmed firing rate setpoints (38–42) and ask whether homeostatic plasticity of excitatory and inhibitory weights can drive neurons to these setpoints. Homeostatic plasticity rules are traditionally defined by changes in synaptic weights that are proportional to an "error term" defined by the difference between the setpoint and the neurons' average activity levels (34–37, 39, 43, 44), for example, $\Delta W_{E\leftarrow E} \propto E_{set} - E_{avg}$, where any departure of the excitatory activity E_{avg} from the setpoint E_{set} would lead to a compensatory correction in the value of the weight $W_{E\leftarrow E}$.

We first examined whether stable self-sustained dynamics can emerge in the standard two-population model (Fig. 1*B*) (19) through homeostatic mechanisms. We initialized the four weights ($W_{E\leftarrow E}$, $W_{E\leftarrow I}$, $W_{I\leftarrow E}$, and $W_{I\leftarrow I}$) of the model to small values and applied a standard family of homeostatic plasticity rules to all four weight classes (Fig. 2*A*). It is well established that PV^{+}-inhibitory neurons have higher firing rates than pyramidal neurons during periods of self-sustained activity (45, 46); thus, based on previous data using intracellular recordings, we set the setpoints for the *E* and *I* populations to 5 and 14 Hz, respectively (46). We first asked whether the family of four standard homeostatic plasticity rules can lead to a stable self-sustained dynamic regime in response to a brief external input. Since in the absence of external input (or intrinsic spontaneous activity) networks capable of self-sustained activity have a trivial stable silent (down-state) regime, at the beginning of each trial we administered a brief external input to engage the network (low levels of noise were used to avoid fluctuation-induced transitions). Although the rules are homeostatic in nature (e.g., if *I* is below I_{set}, an increase in $W_{I\leftarrow E}$ and a decrease in $W_{I\leftarrow I}$ would be induced), in the example shown in Fig. 2 *B* and *C* the network failed to converge to a stable self-sustained regime (Fig. 2 *B* and *C*). Initially (trial 1) an external input to the excitatory population does not engage recurrent activity because $W_{E\leftarrow E}$ is too weak. By trial 200 the weights have evolved and the brief external input triggers self-sustained activity, but activities *E* and *I* do not match the corresponding setpoints; the network is in a nonbiologically observed regime in which $E > I$, so the weights keep evolving. By trial 600 $E = E_{set}$ but $I < I_{set}$, and rather than converging to I_{set}, the network returns to a regime without self-sustained activity by trial 1,000. At that point both setpoint error terms have increased, leading to continued weight changes (Fig. 2*C*). Results across 100 simulations with different weight initializations (*SI Appendix*, *Supplementary Methods*) further indicate that the standard homeostatic rules are ineffective at driving *E* and *I* toward their respective setpoints and generating stable self-sustained dynamics (Fig. 2*D*).

To gain insights into why a family of homeostatic plasticity rules that might intuitively converge fails to do so, we can consider the case in which a network is initialized to a set of weights that already match E_{set} and I_{set} (Fig. 2*E*). Although the neural subsystem alone is stable at this condition (trial 1), small fluctuations in *E* and *I* cause the homeostatic rules to drive the weight values and the average activity of the network away

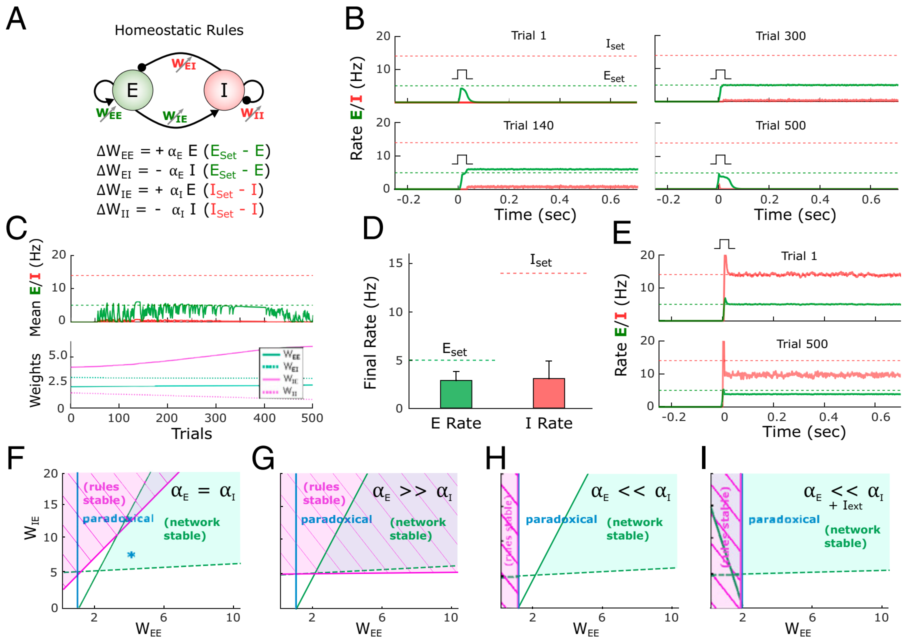


Fig. 2. Standard homeostatic rules are stable only in a narrow parameter regime. (*A*) Schematic (*Top*) of the population rate model in which the four weights are governed by a family of homeostatic plasticity rules (*Bottom*). (*B*) Example simulation of the network over the course of simulated development. Each plot shows the firing rate of the excitatory and inhibitory populations over the course of a trial in response to a brief external input (I_{ext} = 7, I_{dur} = 10 ms). Note that the pulse is applied on every trial at t = 0. E_{set} = 5 and I_{set} = 14 represent the target homeostatic setpoints. Weights were initialized to W_{EE} = 2.1, W_{EI} = 3, W_{IE} = 4, and W_{II} = 2. The learning rate was set to $\alpha_E = \alpha_I = 1e^{-4}$. Note that while the network supports self-sustained activity in trial 200, the firing rates do not converge to their setpoints, and by trial 500 the self-sustained dynamics are no longer observed. (*C*) Average rate across trials (Top) for the excitatory and inhibitory populations for the data shown in *B*. Weight dynamics (*Bottom*) produced by the homeostatic rules across trials for the data shown in *B*. (*D*) Average final rate for 100 independent simulations with different weight initializations. Data represent mean ± SEM. (*E*) Simulation of a network starting with the weights sets that generate self-sustained activity at the target setpoints (E_{set} = 5 and I_{set} = 14 Hz; trial 1, *Top*). After 500 trials the network has diverged from its setpoints, indicating the synaptic plasticity rules are unstable. Weights were initialized to W_{EE} = 5, W_{EI} = 1.09, W_{IE} = 10, and W_{II} = 1.54. (*F*) Analytical stability regions of the neural and plasticity rule subsystems as a function of the free weights W_{EE} and W_{IE}. (Note that once W_{EE} and W_{IE} are set to generate self-sustained activity with specific E_{set} and I_{set} values, W_{EI} and W_{II} are fully determined by W_{EE} and W_{IE}, respectively). Here the stability plot is obtained by considering equal learning rates for all four plasticity rules (as used for panels *B*–*E*). Blue asterisk corresponds to the initial conditions shown in *E* (*Top*). (*G*) Similar to *F* but with $\alpha_E \gg \alpha_I$. (*H*) Similar to *F* but with but with $\alpha_E \ll \alpha_I$. To the right of the blue line, the network is in a paradoxical regime (defined by the condition $W_{EE*}g_E - 1 > 0$). (*I*) Condition of stability of the neural system and plasticity rule system when the learning rate on the inhibitory neuron dominates and an external excitatory current is applied to the excitatory neuron. The current produces an enlargement of the stability region of the neural subsystem. Right of blue line shows the area where the network is in a paradoxical regime.

from the setpoints (trial 500). It is possible to understand this instability by performing an analytical stability analysis. Specifically, a two-population network in which the weights undergo plasticity can be characterized as a dynamic system composed of two subsystems: the neural subsystem, composed of the two differential equations that define E and I dynamics, and the synaptic homeostatic plasticity rule subsystem, defined by the four plasticity rules (*SI Appendix*, Section 2.1). We use the two very different time scales of the neural (fast) and plasticity rule (slow) subsystems to perform a quasi-steady-state approximation of the neural subsystem; then we compute the eigenvalues of the four-dimensional plasticity rule subsystem and finally get an analytical expression for the stability condition of the plasticity rules (*SI Appendix*, Section 2.3). For the entire system to be stable, both the neural and plasticity rule subsystems have to be stable. For the results presented in Fig. 2 *B*–*E* we assumed the learning rates driving plasticity onto the excitatory (α_E) and inhibitory neurons (α_I) to be equal. Under these conditions, the standard homeostatic rules are mostly unstable for biologically meaningful parameter values in which the neural system is stable. The regions of stability can be seen in Fig. 2*F*. Critically, Fig. 2*F* shows that the stability region of the neural subsystem, that is, an inhibition-stabilized network (15, 19), is almost entirely within the region where the homeostatic plasticity rule system is unstable. The only region where a stable self-sustained dynamics can exist is the small triangle where the neural and synaptic stability regions overlap (*SI Appendix*, Fig. S1). Only when plasticity onto the excitatory neuron is significantly faster ($\alpha_E \gg \alpha_I$, resulting in very slow convergence to the inhibitory setpoints) is there a substantial region of overlap between the stability of the neural and plasticity rule subsystems (Fig. 2*G*; *SI Appendix*, Section 1.1).

Because inhibitory neurons seem to undergo homeostatic plasticity as quickly as or more quickly than excitatory neurons (40–42, 47, 48) we conclude that standard homeostatic rules by themselves do not account for the emergence of stable self-sustained and inhibition-stabilized dynamics. Similarly, a combination of analytical and numerical methods also indicates that variants of these homeostatic rules, such as synaptic scaling, are also stable only in a narrow region of parameter space (*SI Appendix*, Section 1.5). We next show that the inherent instability of standard homeostatic plasticity rules is related to the paradoxical effect.

The Paradoxical Effect Hampers the Ability of Homeostatic Rules to Lead to Self-Sustained Activity. The inability of the homeostatic plasticity rules to generate stable self-sustained activity is in part a consequence of the paradoxical effect, a counterintuitive yet well described property of two-population models of inhibition-stabilized networks (14, 15). Specifically, if during self-sustained activity one increases the excitatory drive to the inhibitory population, the net result is a decrease in the firing rate of the inhibitory units. This paradoxical effect can be understood in terms of the $I \rightarrow E \rightarrow I$ loop: The increased inhibitory drive leads to a lower steady-state rate for E, but this new steady-state value requires a decrease in the I firing rate to maintain an appropriate E/I balance (in effect, the decrease in E decreases the drive to I by more than the external increase to I). This paradoxical effect has profound consequences for plasticity rules that attempt to drive excitatory and inhibitory weights to an activity setpoint.

The relationship of the paradoxical effect and the homeostatic rule performance is presented in Fig. 2*H*. The region of stability for the homeostatic plasticity rules is shown in a parameter regime where inhibitory plasticity is much faster ($\alpha_E \ll \alpha_I$). Contrary to when excitatory plasticity dominates, the region of stability is small, and there is no overlap with the region of stability of the neural subsystem. Crucially, the boundary of the stability region of the plasticity rule coincides with the condition for the paradoxical effect to be present (right of the blue line in Fig. 2*H*, *SI Appendix*, Sections 2.2.4 and 2.3.6). Under these conditions, the rules can be stable only when the network is not in an inhibition-stabilized regime. If a network regime with nonzero E is forced to exist in that region (e.g., via a tonic external current, Fig. 2*I*), it would be stable only in the nonparadoxical region with the plasticity rules in place (*SI Appendix*, Section 2.5). Note that in the absence of a sufficiently strong external input it is not possible to have stable self-sustained activity that is not inhibition stabilized.

To understand the impact of the paradoxical effect on homeostatic plasticity rules, consider a network state in which the I rate falls significantly below its setpoint and the E rate is close to its setpoint (Fig. 3*A*). In order to reach the I setpoint, homeostatic plasticity in the inhibitory neuron would intuitively result in an increase of $W_{I\leftarrow E}$. However, because of the paradoxical effect, an increase in $W_{I\leftarrow E}$ actually makes I decrease (Fig. 3*B*), thus increasing the error term $I_{\text{set}} - I$. To increase the steady-state inhibitory rate, we can "anti-homeostatically" decrease the excitatory weight onto the inhibitory neurons (Fig. 3*C*). (Note that the converse is true for the $W_{I\leftarrow I}$ weight.) This simple example shows the complexity of designing a coherent set of rules in a strongly coupled system (an analysis of the paradoxical effect is in *SI Appendix*, Section 2.2.4). This analysis also explains why homeostatic plasticity rules can lead to self-sustained activity at the appropriate setpoints when $\alpha_E \gg \alpha_I$. Essentially, by allowing plasticity onto the E population to be faster, one overcomes the counterproductive homeostatic plasticity associated with the paradoxical effect.

The interaction between the paradoxical effect and homeostatic plasticity in inhibitory neurons leads to the question of whether anti-homeostatic plasticity rules may be more effective that standard homeostatic rules; for example, $\Delta W_{I\leftarrow E} \propto -(I_{\text{set}} - I_{\text{avg}})$. Thus, we also examined a number of hybrid families of plasticity rules with different combinations of homeostatic and anti-homeostatic rules. Indeed, some hybrid families exhibited large degrees of overlap between the stable regions of the network and plasticity rule subsystems. However, numerical simulations revealed that these rules were mostly ineffective in driving networks to self-sustained activity at the target setpoints (*SI Appendix*, Section 1.2 and Fig. S2). These two results are not inconsistent because the stability analysis speaks to cases when the network is initialized to weights that satisfy E_{set} and I_{set}, not whether the rules will drive network activity into these stable areas from any initial state, including an early developmental state. Thus, we interpret these results as meaning that while anti-homeostatic plasticity can contribute to stability of this dual dynamic system, anti-homeostatic plasticity is ineffective at driving the dynamics toward setpoints (in other words, that anti-homeostatic plasticity might allow for stable inhibition-stabilized dynamics but does not necessarily generate sizable basins of attraction around the fixed point).

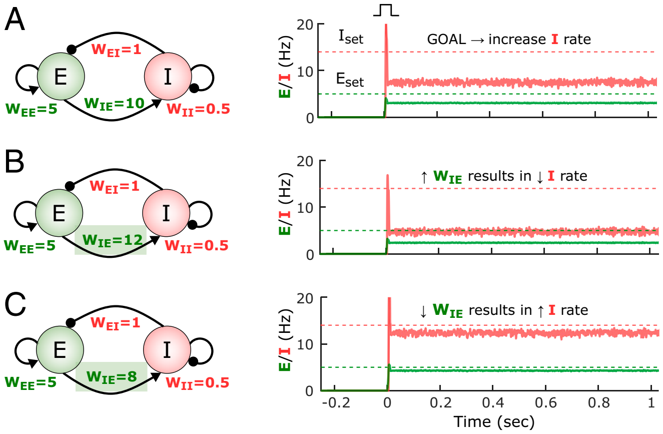


Fig. 3. The paradoxical effect constrains the learning rules that can lead to inhibition-stabilized dynamics. (*A*) Example of the self-sustained dynamics of a two-population model in the paradoxical regime with weight values shown in the diagram. Both the *E* and *I* firing rates fall below their respective setpoints. The objective is to adjust the weights so that the *E* and *I* activity match their setpoints. (*B*) An increase of W_{IE} from 10 to 12 results in a paradoxical decrease of the *I* rate. (*C*) Because of the paradoxical effect an effective way to increase the steady-state *I* firing rate is to decrease its excitatory drive (i.e., W_{IE}).

Cross-Homeostatic Rule Robustly Leads to the Emergence of Self-Sustained Dynamics. Given that a standard family of homeostatic plasticity rules did not robustly lead to stable dynamics, we explored alternative learning rules. By defining a loss function based on the sum of the excitatory and inhibitory errors, we analytically derived a set of learning rules using gradient descent (*SI Appendix*, Section 3). This approach led to mathematically complex and biologically implausible rules; however, approximations and simulations inspired a simple class of learning rules that we will refer to as cross-homeostatic (see *Methods*). The main characteristic of this set of rules is that the homeostatic setpoints are "crossed" (Fig. 4*A*). Specifically, the weights onto the excitatory neuron ($W_{E\leftarrow E}$ and $W_{E\leftarrow I}$) are updated to minimize the inhibitory error, while weights into the inhibitory neuron ($W_{I\leftarrow E}$ and $W_{I\leftarrow I}$) change to minimize the excitatory error. Although apparently nonlocal, from the perspective of an excitatory neuron these rules can be interpreted as cells having a setpoint for the total inhibitory input current onto the cell. Such inputs could be read by a cell as the activation of metabotropic receptors (e.g., gamma-aminobutyric acid B [GABA$_\text{B}$] and metabotropic glutamate; see *Discussion*). Indeed, a similar cross-homeostatic rule has been recently derived for $W_{I\leftarrow E}$ weights (49).

An example of the performance of the cross-homeostatic rules is shown in Fig. 4 *B* and *C*. After an initial phase with no self-sustained firing (trial 1), recurrent activity reaches stable self-sustained dynamics (trial 20), whose average rate continues to converge toward its defined setpoints (trial 100) until the learning rule system reaches steady state (trial 500). The average E and I rates of the network evolve asymptotically toward the defined setpoints, as the weights evolve and converge (Fig. 4*C*). Across different weight initializations the rules proved effective in driving the mean activity of the network to the target E and I setpoints and led to balanced, inhibition-stabilized dynamics (Fig. 4 *D* and *E*). The weight trajectory from its initial value to its final one is shown for 100 different simulations (Fig. 4*D*). Each line corresponds to individual experiments with different initializations (for visualization purposes weights were initialized around the final stable weights; *SI Appendix*, Fig. S3 includes broader initialization conditions). Circles indicate the

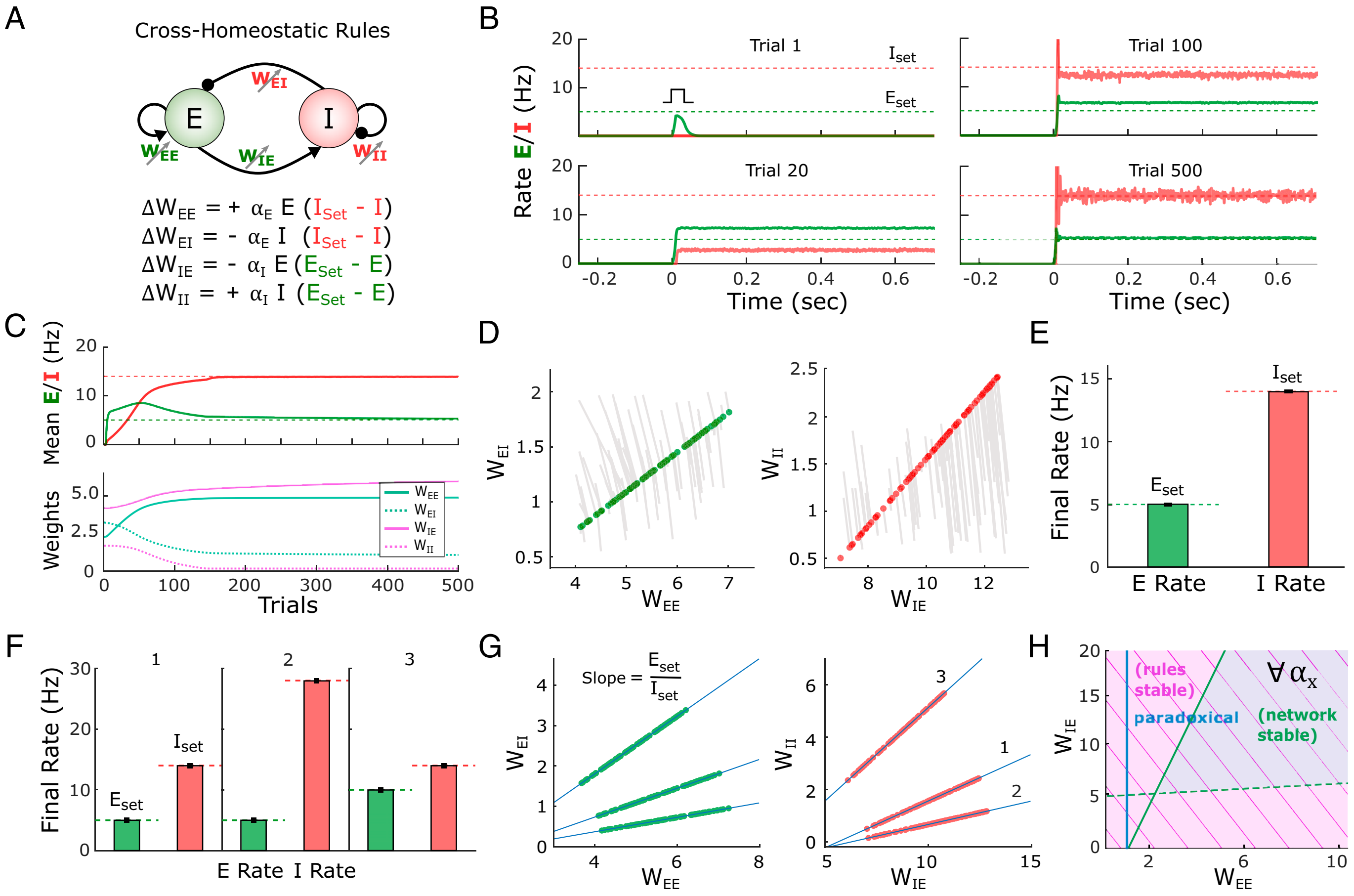


Fig. 4. A family of cross-homeostatic rules robustly lead to inhibition-stabilized dynamics at E_{set} and I_{set}. (*A*) Schematic of the network model and the family of cross-homeostatic plasticity rules. (*B*) Example network dynamics across simulated development. The network is initialized with weights that do not lead to self-sustained dynamics in response to an external input (trial 1, weights are initialized to $W_{EE} = 2.1$, $W_{EI} = 3$, $W_{IE} = 4$, and $W_{II} = 2$). By trial 20 stable self-sustained activity is observed but at firing rates far from the target setpoints (dashed lines). By trial 500 the network has converged to stable self-sustained activity in which *E* and *I* firing rates match their respective setpoints. The learning rate was set to $\alpha_E = \alpha_I = 5e^{-4}$. (*C*) Average rate across trials (*Top*) for the excitatory and inhibitory populations for the data shown in *B*. Weight dynamics (*Bottom*) induced by the cross-homeostatic rules across trials for the data shown in *B*. (*D*) Weight changes for 100 different simulations with random weight initializations (*SI Appendix*, *Supplementary Methods*). Lines show change from initial to final (circles) weight values. (*E*) Average final rates for 100 independent simulations with different weight initializations shown in *D*. Data represent mean ± SEM. (*F*) Final rates for the excitatory and inhibitory subpopulations after plasticity with same starting conditions as in *D* and *E* but for different setpoints. 1: $E_{set} = 5$, $I_{set} = 14$; 2: $E_{set} = 5$, $I_{set} = 28$; 3: $E_{set} = 10$, $I_{set} = 14$. Data shown in *D* and *E* correspond to 1. Data represent mean ± SEM. (*G*) Final weight values for homeostatic plasticity simulations for the three different pairs of setpoints shown in *F*. Blue lines correspond to the theoretical linear relationship between the excitatory and inhibitory weights at a fixed point obeying E_{set} and I_{set}. The slope of the line is defined by the ratio of the setpoints (see *Methods*). (*H*) Analytical stability regions of the neural subsystem and learning rule subsystem as a function of W_{EE} and W_{IE}. The stability condition holds for any possible combination of learning rates (*SI Appendix*, Section 1.3).

final values of the weights. Independently of the initial conditions, the weights converge to a line attractor (actually a two-dimensional plane attractor in four-dimensional weight space; *SI Appendix*, Section 2.1). Note that this attractor refers to the sets of weights that generate self-sustained dynamics where *E* and *I* activity matches E_{set} and I_{set}, respectively. That is, for a given pair of setpoints (E_{set}, I_{set}) the final values of the weights $W_{E\leftarrow I}$ and $W_{I\leftarrow I}$ are linear functions of the "free" weights $W_{E\leftarrow E}$ and $W_{I\leftarrow E}$, respectively. This is a direct consequence of the steady-state conditions for the nontrivial fixed-point of the two-population model (14, 15), where the slope of the line is defined by the setpoints E_{set}/I_{set} (see *Methods*). For example, to satisfy $dE/dt = 0$ at the neural activity fixed point, the net excitation and inhibition must obey a specific balance, meaning that once $W_{E\leftarrow E}$ or $W_{E\leftarrow I}$ is determined, the other weight is analytically constrained for a given set of setpoints and parameters. Once the weights reach this specific relationship, the *E* and *I* rates reach their corresponding E_{set} and I_{set} values (Fig. 4*E*). Numerical simulations confirm that the cross-homeostatic rule robustly guides self-sustained activity to different E_{set} and I_{set} setpoints (Fig. 4*F*), whose ratios define the slopes of the final relationship between the weights (Fig. 4*G*). The above implementations were trial-based, that is, the weights were updated at the end of every trial. An "online" implementation, in which weights were continuously updated, also led to convergence to the setpoints (*SI Appendix*, Fig. S4).

To further validate the effectiveness and stability of the cross-homeostatic rule, we again used analytic methods to determine the eigenvalues of the four-dimensional dynamic system describing the family of four cross-homeostatic rules. As above, stability is determined by the sign of the real part of the eigenvalues of the system. It can be shown (*SI Appendix*, Section 1.3) that these learning rules are stable for any set of parameter values, provided that the stability conditions of the neural subsystem are satisfied (Fig. 4*H*). Importantly, these results demonstrate that cross-homeostatic rules work in both paradoxical and nonparadoxical conditions. Furthermore, the stability of the rules is independent of the absence or presence of external input (*SI Appendix*, Section 2.5). Therefore, it is possible to formally establish that cross-homeostatic learning rules are inherently stable and can robustly account for the emergence and maintenance of self-sustained, inhibition-stabilized dynamics in the two-population model.

Cross-Homeostatic Rules Drive Average Activity to Setpoints in a Multiunit Model. The previous results demonstrate the robustness of the cross-homeostatic family of rules in driving a two-subpopulation rate model to a stable self-sustained, inhibition-stabilized regime. We next examined whether these rules are also effective for a multiunit model in which there are many excitatory and inhibitory units. The firing-rate model was composed of 80 excitatory and 20 inhibitory recurrently connected neurons (Fig. 5*A*). In this case, individual neurons adjust their weights to minimize the average error of their presynaptic partners (*SI Appendix*, *Supplementary Methods*). Starting with normally distributed weights, the network reaches stable self-sustained dynamics (Fig. 5 *B* and *C*). However, individual units converge to different final rate values, satisfying the defined setpoints only as an average (green and red thick lines of Fig. 5*B*). This is a result of the nature of the cross-homeostatic rules: Neurons adjust their weights to minimize the error of the mean activity of its presynaptic partners. For this reason, although the network is globally balanced, single units do not converge to the same balanced *E*–*I* line attractor (Fig. 5 *D* and *E*). After cross-homeostatic plasticity, some differential structure is visible among the various weight classes (Fig. 5*F*). Simulations across 400 different initialization conditions demonstrate that the rules lead the average excitatory and inhibitory population activity to E_{set} and I_{set}, respectively (Fig. 5 *G* and *H*). The cross-homeostatic rules are thus capable of driving a multiunit model to a stable self-sustained regime, but they do not guide individual units to local setpoints. Similar results are obtained when the network weights are initialized with log-normal distributions (*SI Appendix*, Fig. S5).

Learning Rule with Cross-Homeostatic and Homeostatic Terms Leads to Local Convergence to Setpoints. The above results demonstrate a potential limitation of the cross-homeostatic family of rules: The target setpoints are reached only at the population level. An additional and potentially more serious limitation is that cross-homeostatic rules predict that artificially altering the activity of a small number of excitatory neurons within a large network would not directly produce homeostatic plasticity in these neurons but would directly produce plasticity in their postsynaptic inhibitory neurons. This prediction seems to conflict with homeostatic plasticity experiments that have targeted specific cell types rather than globally altered activity through pharmacological means (50, 51). We therefore assessed the scenario in which both cross-homeostatic and homeostatic rules operate in parallel, resulting in a two-term cross-homeostatic family of rules. Interestingly, this family of rules can be obtained from an approximation of a gradient descent derivation on a loss function that includes the difference between *E* and *I* and their respective setpoints (*SI Appendix*, Section 3). In a two-population model, we first confirmed that this two-term cross-homeostatic family is stable, assuming that the learning rate of the homeostatic term does not dominate (*SI Appendix*, Section 1.4).

Simulations with the same multiunit model as in Fig. 5 show that with the two-term cross-homeostatic rule all individual units converge to their respective E_{set} and I_{set} (Fig. 6 *A*–*C*). Importantly, in contrast to the single-term cross-homeostatic rule, the total excitatory and inhibitory weight of each individual unit converged to the *E*–*I* balance of the line attractor predicted by the network equations (Fig. 6 *D* and *E*), while some structure in the different weight classes is also observed in the connectivity matrices (Fig. 6*F*). The convergence to the setpoints was stable across a wide range of initial states (Fig. 6 *G* and *H*). Thus, a hybrid family of plasticity rules that includes both cross-homeostatic and homeostatic forces provides global network stability while also locally driving each unit to their setpoint and a balanced *E*–*I* regime.

Spiking Neural Network Model with Sparse Connectivity Converges to an Inhibition-Stabilized Regime at the Setpoints. The previous results demonstrate the ability of the two-term cross-homeostatic plasticity rule to guide firing-rate-based models to inhibition-stabilized regimes at the target setpoints. We next examined the effectiveness of this family of learning rules in a sparsely connected spiking neural network (Fig. 7*A*). In a sparsely connected network, the cross-homeostatic component of the learning rule can be implemented globally (e.g., excitatory plasticity onto an excitatory unit is based on the mean error of the entire population of inhibitory units) or locally (e.g., excitatory plasticity onto an excitatory neuron is based on the mean error of its presynaptic inhibitory partners). Here we used a local implementation, in which, for example, an excitatory neuron has a setpoint interpreted as a target for the total amount of $\mathrm{GABA_B}$ receptor activation it should receive (see *Discussion*).

Starting from a developmental scenario with weak weights (i.e., that do not support any self-sustaining activity), the rules successfully drive the network to stable self-sustained and asynchronous spiking activity near the setpoints (Fig. 7 *B*–*E*; $\mathrm{CV_{ISI}}\sim1$). As seen in the firing rate multiunit model, the weights self-organize from an unstructured initial condition (Fig. 7 *F* and *G*) to a state where a balance of excitation and inhibition emerges (Fig. 7 *H*–*K*). After training, the firing rates of the excitatory and inhibitory neurons distribute around their setpoints (Fig. 7*H*). As expected, the emergent network dynamics exhibits the paradoxical effect; specifically, when inhibitory neurons are transiently activated with an external current, a net decrease in the mean firing rate is observed (Fig. 7*I*). Finally, convergence holds across networks initialized with different weights (Fig. 7*L*). These results demonstrate that the theoretical and computational results in rate networks translate to more complex and biologically realistic scenarios.

Discussion

Elucidating the learning rules that govern the connectivity within neural circuits is a fundamental goal in neuroscience, in part because learning rules establish unifying principles that span molecular, cellular, systems, and computational levels of analyses. Elucidation of Hebbian associative synaptic learning, for example, linked simple computations at the level of single proteins (the *N*-methyl-D-aspartate receptor) with higher-order computations at the system and computational levels (52–56). However, it remains the case that because most studies have focused on learning rules at one or two synapse classes, little is known about the learning rules that give rise to complex neural dynamic regimes. Here we have taken steps toward exploring families of learning rules that operate in parallel at four different synapse classes, and starting from a silent state, capture the experimentally observed emergence of self-sustained, inhibition-stabilized dynamics in cortical networks.

We first explored whether standard formulations of homeostatic plasticity can account for the unsupervised emergence of self-sustained, inhibition-stabilized regimes. Based on experimental

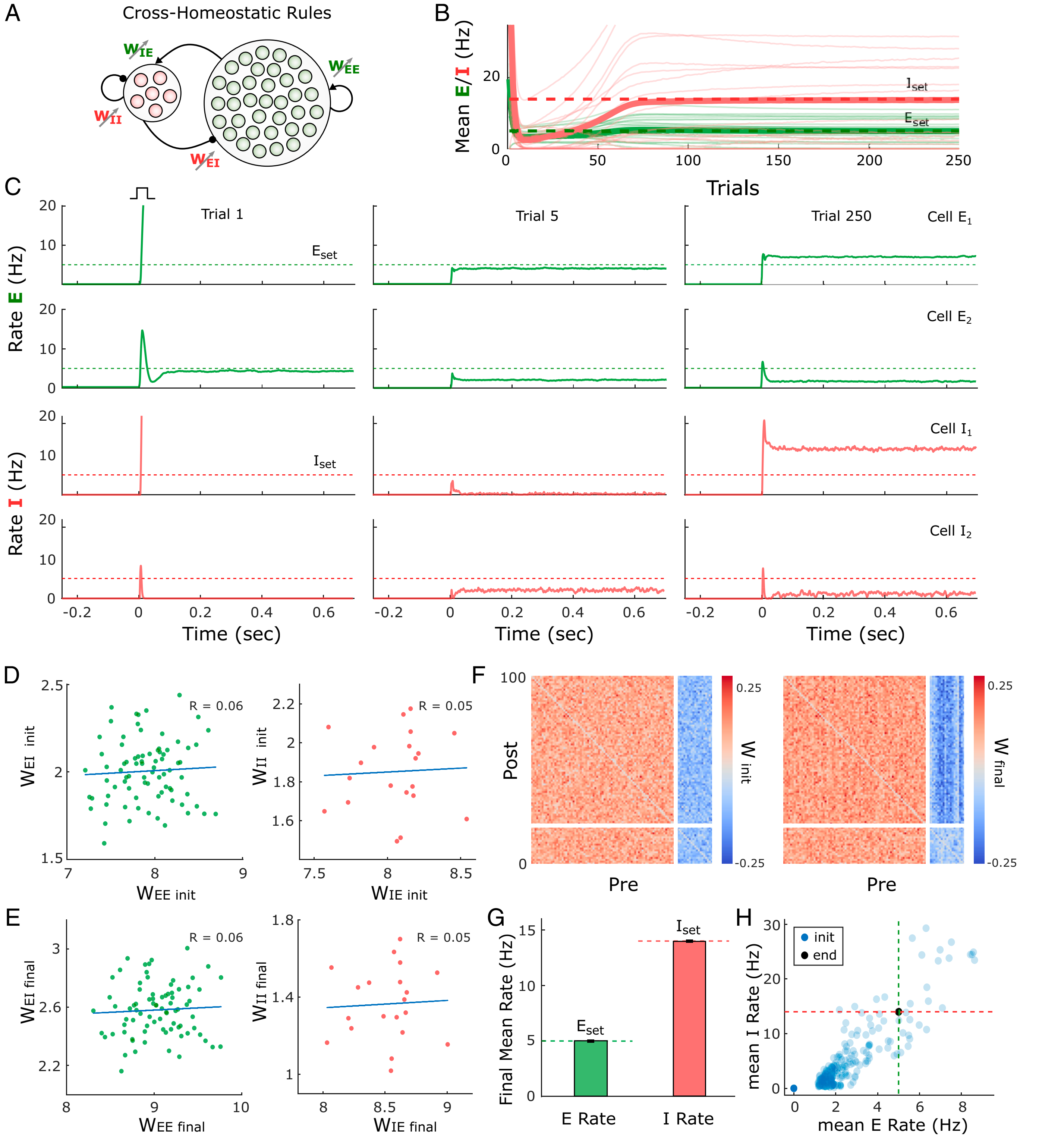


Fig. 5. Cross-homeostatic rules drive a multiunit firing rate model to a global network balance. (*A*) Schematic (*Left*) of the multiunit rate model. The network is composed of 80 excitatory and 20 inhibitory units recurrently connected. The four weight classes are governed by cross-homeostatic plasticity rules (*Right*). *SI Appendix, Supplementary Methods* includes a detailed explanation of the implementation. (*B*) Evolution of the average rate across trials of 20 excitatory and inhibitory units in an example simulation. The network is initialized with random weights (*SI Appendix, Supplementary Methods*), and so neurons present diverse initial rates. E_{set} = 5 and I_{set} = 14 represent the target homeostatic setpoints. Red and green lines represent the individual (thin lines) and average (thick lines) firing rate of inhibitory and excitatory population, respectively. The learning rate was set to $\alpha = 2e^{-5}$. (*C*) Example of the firing rates of two excitatory and two inhibitory units at different points in *B*. The evolution of the firing rates of the excitatory and inhibitory populations within a trial in response to a brief external input is shown in every plot. Individual units converge to stable self-sustained dynamics but not to the defined setpoint. (*D*) *E–I* weight relationships at the beginning of the simulation. Every dot represents the total presynaptic weight onto a single unit. *Left,* excitatory neurons; *Right,* inhibitory neurons. (*E*) Same plot as in *D* at the end of the simulation. (*F*) Weight matrix for the multiunit model at the beginning (*Left*) and end (*Right*) of the simulation. Inhibitory weights are shown in blue, excitatory weights in red. (*G*) Average firing rate of the units of the multiunit model and for different initializations of weights (n = 400). The network converges to the setpoints in average. Data represent mean ± SEM. (*H*) Same data as in *G* but showing the average initial rate of the network for the multiple initializations (blue dots) and the average rate at the end (black). Target rates are shown in dotted lines (green, E_{set} = 5, red I_{set} = 14).

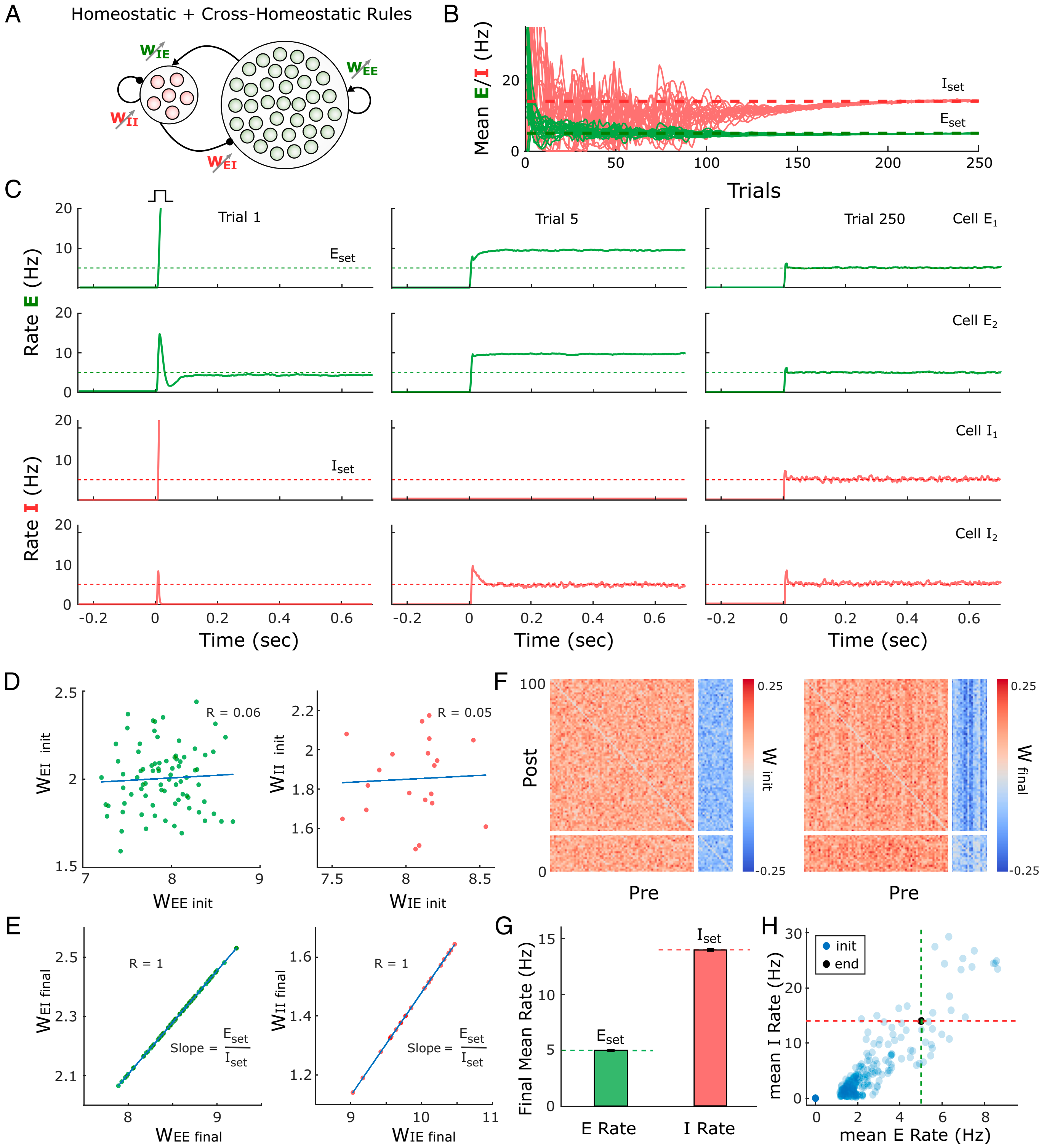


Fig. 6. Adding cross-homeostatic influences to homeostatic rules leads to global and local convergence to setpoints. (*A*) Schematic (*Left*) of the multiunit rate model. The network is composed of 80 excitatory and 20 inhibitory units recurrently connected. The four weight classes are governed by homeostatic rules with cross-homeostatic influences (*Right*). *SI Appendix, Supplementary Methods* includes a detailed explanation of the implementation. (*B*) Evolution of the average rate across trials in an example simulation (20 excitatory and inhibitory units). The network is initialized with random weights (same as in Fig. 5, *SI Appendix, Supplementary Methods*) and so neurons present diverse initial rates. $E_{set} = 5$ and $I_{set} = 14$ Hz represent the target homeostatic setpoints. The learning rate was set to $\alpha = 1e^{-5}$. (*C*) Example of the firing rate of two excitatory and two inhibitory units at different points in *B*. The evolution of the firing rate of the excitatory and inhibitory population within a trial in response to a brief external input is shown in every plot. Units converge to stable self-sustained activity and at an individual setpoint. (*D*) *E–I* weight relationships at the beginning of the simulation. Every dot represents the total presynaptic weight onto a single unit. *Left,* excitatory neurons; *Right,* inhibitory neurons. (*E*) Same plot as in *D* at the end of the simulation. The network has reached a stable state and weights converge to single *E–I* balance defined by a line attractor. (*F*) Weight matrix at the beginning (*Left*) and end (*Right*) of the simulation. Inhibitory weights are shown in blue, excitatory weights in red. (*G*) Average firing rate of the units of the multiunit model and for different initializations of weights ($n = 400$). Data represent mean ± SEM. (*H*) Same data as in *G* but showing the average initial rate of the network for the multiple initializations (blue dots) and the average rate at the end (overlapping black circles). Target rates are shown in dotted lines (green, E_{set}; red, I_{set}).

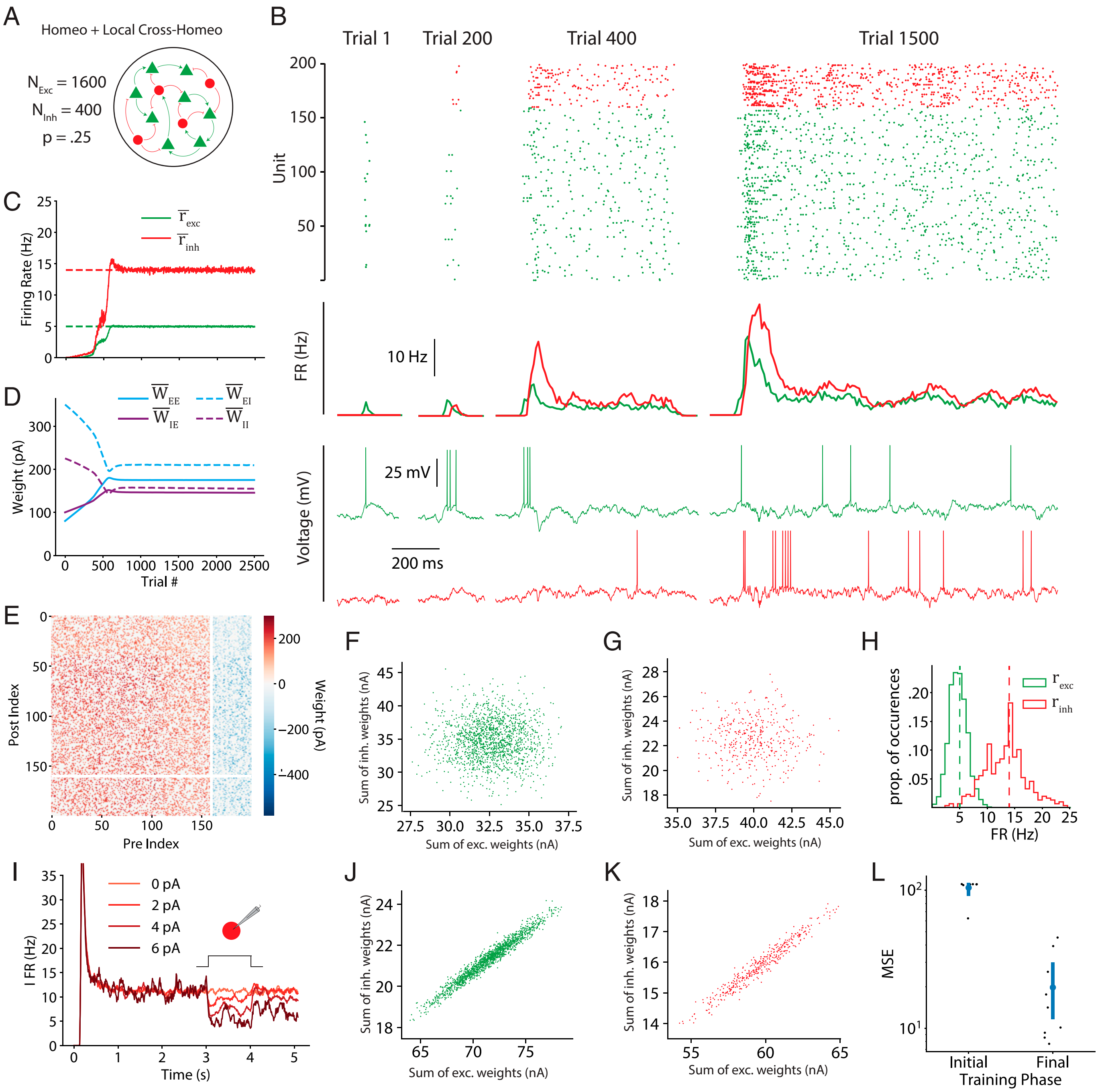


Fig. 7. Two-term cross-homeostatic learning rules guide a sparse spiking network model to an inhibition-stabilized regime at the target setpoints. (*A*) Schematic of the network. Two thousand leaky adaptive integrate-and-fire units (1,600 excitatory and 400 inhibitory) were connected with a 25% probability. (*B*) Spike rasters, population PSTHs (peristimulus time histograms), and sample voltage traces for excitatory (green) and inhibitory (red) units, across four stages over the course of training. (*C*) Population average firing rates over the course of training (moving average with a width of five trials). (*D*) Mean weights for each synaptic class over the course of training. Weights were initialized in an early developmental regime reflecting a silent network. $\overline{W_{EE}} = 80\ pA$, $\overline{W_{IE}} = 100\ pA$, $\overline{W_{EI}} = 350\ pA$, $\overline{W_{II}} = 225\ pA$. (*E*) Weight matrix at the end of training. Due to the size of the network, only a 10% subset of the full weight matrix is shown. (*F*) Initial *E*/*I* balance onto excitatory units, visualized as a scatterplot of the sum of incoming excitatory synaptic weights versus the sum of incoming inhibitory weights onto each excitatory unit. (*G*) Same as *F* but for the initial *E*/*I* balance onto inhibitory units. (*H*) Histograms of the firing ratess for each excitatory and inhibitory population at the end of training (dashed lines represent the setpoints). (*I*) At the end of training, the network exhibits the paradoxical effect. When a positive current is injected into all inhibitory units, the mean inhibitory firing rate decreases. This effect is visualized with PSTHs of the inhibitory population firing rate across 40 trials at each of the injected current values. (*J*) Final *E*/*I* balance onto excitatory units, visualized as a scatterplot of the sum of incoming excitatory synaptic weights versus the sum of incoming inhibitory weights onto each excitatory unit. (*K*) Same as *J* but for the final *E*/*I* balance onto inhibitory units. (*L*) Robustness of convergence to weight initialization for nine networks initialized with different mean weights, we show the initial and final MSE of the unit FRs with respect to their homeostatic setpoints after a 6,000-trial training session.

data we assumed that both excitatory and inhibitory neurons have an ontogenetically programmed activity setpoint during self-sustained activity and that plasticity in the four weight classes is driven by standard formulations of homeostatic plasticity. Numerical simulations and analytical stability analyses revealed that while some initial conditions and parameter regimes led to self-sustained dynamics, they occupied a narrow region of parameter space, when the rate of synaptic plasticity onto inhibitory neurons is much lower than that onto excitatory neurons (Fig. 2*G* and *SI Appendix*). When the rates of inhibitory and excitatory plasticity are comparable, analytical stability analyses confirmed that the region of stability of the network dynamics overlapped only in a narrow region. Such a narrow stability area seems incompatible with the robustness necessary in biological systems and with experimental data showing that inhibitory neurons exhibit homeostatic plasticity as fast as or faster than excitatory neurons (40–42, 47, 48). We thus conclude that a family of standard homeostatic plasticity rules operating in all four synapse classes is not sufficient to account for the experimentally observed emergence of self-sustained dynamics in cortical circuits.

Cross-Homeostatic Plasticity. Analyses of approximations of a gradient-descent-derived learning rule suggested, somewhat counterintuitively, that adjusting the E population based on the error of the I population (and vice versa) may prove to be an effective family of learning rules. Indeed, numerical simulations and analytical stability analyses revealed that this cross-homeostatic rule was robustly stable (Fig. 4). However, the convergence to the excitatory and inhibitory setpoints in a multiunit network occurred only at the population level, not at the level of individual units. This observation is not inconsistent with experimental data, which show that in vivo neurons do exhibit a wide range of variability in their apparent setpoints (57, 58). However, a significant concern with this single-term cross-homeostatic rule is that it predicts that selectively increasing activity in a subpopulation of excitatory neurons would first induce plasticity in inhibitory neurons ($W_{I \leftarrow E}$ and $W_{I \leftarrow I}$), which could in turn lead to plasticity in the manipulated excitatory neurons ($W_{E \leftarrow E}$ and $W_{E \leftarrow I}$). Most homeostatic plasticity studies do not speak to this prediction because they have used pharmacological manipulations of both excitatory and inhibitory neurons. However, some studies have used cell-specific manipulations—such as cell-specific overexpression of potassium channels (50, 51)—that strongly support the notion that synaptic plasticity is guided at least in part by their own deviation from setpoint.

In our opinion, and although we have explored alternative rules (*SI Appendix*, Section 1.6), the most biologically plausible set of plasticity rules that lead to stable self-sustained dynamics comprises a hybrid rule that includes both standard homeostatic and cross-homeostatic terms. Such a two-term cross-homeostatic rule robustly led to a self-sustained, inhibition-stabilized network, with all units converging to their setpoints, and is directly consistent with current experimental data.

Biological Plausibility of Cross-Homeostatic Plasticity. While the neural mechanisms underlying homeostatic plasticity remain to be elucidated, it is generally assumed that an individual neuron can maintain a running average of their firing rate over the course of hours as a result of Ca^{2+}-activated sensors. Based on the deviation of this value from an ontogenetically determined setpoint, neurons up- or down-regulate the density of postsynaptic receptors accordingly (38, 58–60). Two-term cross-homeostatic plasticity would require additional, and apparently nonlocal information about the error in a given neuron's presynaptic partners. Importantly however, this rule can be implemented locally because any postsynaptic neuron has access to the mean activity of its presynaptic partners simply as a result of its postsynaptic receptor activation. Indeed, a plasticity rule for $W_{I \leftarrow E}$ weights with a similar cross-homeostatic error term has also been recently proposed and implemented based on the mean activation of postsynaptic receptors—more specifically the net postsynaptic currents, which provide a coupled measure of average presynaptic firing and synaptic weights (49).

Here we propose that cross-homeostatic plasticity could be implemented through postsynaptic metabotropic receptors (e.g., mGlu and $GABA_B$). Such receptors would provide a mechanism for postsynaptic neurons to maintain a running average of the activity of its presynaptic partners that is decoupled from the synaptic weights. Metabotropic receptors are G protein coupled receptors that provide a low-pass filtered measure of presynaptic activity and are involved in a large number of incompletely understood neuromodulatory roles (61, 62). Since metabotropic receptors appear to undergo less homeostatic and associative plasticity, they provide a measure of presynaptic activity that is naturally decoupled from the ionotropic receptors (e.g., AMPA and $GABA_A$) that are being up- and down-regulated.

Further support for the notion that individual neurons have access to global network activity emerges from studies suggesting that neurons might not homeostatically regulate activity at the individual neuron level but rather at the global population level (63). Such a global-level homeostasis could be achieved by nonsynaptic paracrine transmission. Indeed, retrograde messenger systems are ideally suited for this role, as they have already been implicated in signaling mean activity levels to local capillaries, driving the activity-dependent vasodilation that underlies functional MRI (64).

Paradoxical Effect and Standard Homeostatic Rules. The paradoxical effect is one of the defining features of inhibition-stabilized networks, and a growing body of evidence suggests that the cortex operates in this particular dynamic regime (20, 65–67). As recent work has begun to hint (68), here we formally prove that the paradoxical effect applies important constraints to the potential learning rules that lead to the emergence of inhibition-stabilized networks. In the simplified case in which there is only homeostatic plasticity onto the inhibitory neurons, we can immediately see why the paradoxical effect renders standard homeostatic rules ineffective. If the I population is below its setpoint, standard homeostatic rules would increase $W_{I \leftarrow E}$, which paradoxically would further decrease I (Fig. 3), thus further increasing the error instead of decreasing it (Fig. 3). This reasoning is related to why, when using the standard family of homeostatic rules, the rate of plasticity onto the inhibitory neurons has to be much smaller—in effect making the "paradoxical homeostatic plasticity effect" much slower. Furthermore, our analytical stability analyses show that in the limit of vanishingly small excitatory learning rates ($\alpha_{EE,EI} \ll \alpha_{IE,II}$) the stability region of the weight subsystem is bounded by the paradoxical condition. This means that the only allowed stable states with nonzero E activity will occur in the nonparadoxical regime, if any, and they will not be proper inhibition-stabilized, self-sustained regimes.

Future Directions and Experimental Predictions. While we have taken the approach of implementing homeostatic plasticity rules at all four synapse classes in our model, it is important to stress that we have omitted other well-characterized forms of synaptic plasticity. In particular, we did not include associative long-term potentiation or spike-timing dependent plasticity. These forms of plasticity are generally considered to capture the correlation structure in networks that are driven by structured inputs. Arguably, because there is evidence that self-sustained forms of activity such as up-states develop in the absence of any structured external input (23, 25, 26) and because all excitatory and inhibitory neurons synchronously shift between quiescent and active states, associative forms of plasticity may not contribute significantly to these regimes. Nevertheless, we envision the cross-homeostatic rules we propose working hand in hand with associative forms of plasticity that impose high-dimensional structure on the top of the inhibition-stabilized dynamics. In fact, preliminary observations reveal that cross-homeostatic rules are capable of stabilizing recurrent networks while preserving imposed Hebbian-like structure in the weight matrix (*SI Appendix*, Fig. S6). Future work should explore the computational advantage of cross-homeostatic plasticity in models performing complex computational tasks, such as working memory, sensory timing, or motor control (5, 69–72).

An important implication of our results is that neuronal and network properties can operate in fundamentally different modes. That is, while homeostatic plasticity can lead to single neurons to reach their target setpoints in simple feedforward circuits, those same rules can be highly unstable when the neurons are placed even in the simplest of recurrent excitatory/inhibitory circuits with emergent dynamics. Furthermore, because emergent neural dynamic regimes are highly nonlinear, and in particular that stable self-sustained dynamic regimes exhibit a paradoxical effect, it is likely that the brain exhibits paradoxical or counterintuitive learning rules to generate self-sustained dynamic regimes.

Methods

Computational Model. A two-population firing-rate model was implemented based on a previous inhibition-stabilized network model (19). The firing rate of the excitatory (E) and inhibitory (I) population obeyed Wilson and Cowan dynamics (73).

$$\tau_E \frac{dE}{dt} = -E(t) + f_E\Big(W_{EE}E(t) - W_{EI}I(t) + \eta_E(t)\Big), \qquad [1]$$

$$\tau_I \frac{dI}{dt} = -I(t) + f_I\Big(W_{IE}E(t) - W_{II}I(t) + \eta_I(t)\Big), \qquad [2]$$

where W_{XY} represents the weight between the presynaptic unit Y and postsynaptic unit X. The parameters τ_X and η_X represent a time constant and an independent noise term, respectively. The time constants were set to $\tau_E = 10$ ms for the excitatory and $\tau_I = 2$ ms for the inhibitory subpopulations. The noise term was an Ornstein–Uhlenbeck process with mean $\mu_x = 0$, a time constant $1/\Theta_x = 1$ ms, and a sigma parameter of $\sigma_x = 10$. To elicit self-sustained activity, a step current was injected at the beginning of each trial on the excitatory population.

The function $f_Y(x)$ represents the intrinsic excitability of the neurons, and it is modeled as a threshold-linear function with threshold Θ_Y and gain g_Y.

$$f_Y(x) = \begin{cases} 0 \text{ if } x < \theta_Y \\ g_Y(x - \theta_Y) \text{ if } x \geq \theta_Y' \end{cases} \quad Y = \{E, I\}. \qquad [3]$$

The thresholds were set to $\Theta_E = 4.8$ and $\Theta_I = 25$, and the gains to $g_E = 1$ and $g_I = 4$. The higher thresholds in PV neurons are consistent with experimental findings (46).

The linear relationship between excitatory and inhibitory weights (Fig. 4) corresponds to the steady-state solution of the neural subsystem when the inhibitory and excitatory rates are at their target setpoints. The solution can be obtained by setting the left side of Eqs. **1** and **2** to zero and substituting the steady-state E and I values with E_{set} and I_{set}:

$$W_{EI} = \frac{W_{EE}E_{Set}}{I_{Set}} - \frac{\theta_E g_E + E_{Set}}{I_{Set} g_E}, \qquad [4]$$

$$W_{II} = \frac{W_{IE}E_{Set}}{I_{Set}} - \frac{\theta_I g_I + I_{Set}}{I_{Set} g_I}. \qquad [5]$$

Thus, the slope of the E/I balance line in Fig. 4 corresponds to $E_{\text{set}}/I_{\text{set}}$. We chose $W_{E \leftarrow E}$ and $W_{I \leftarrow E}$ as the "free" weights. See details and analytical results in *SI Appendix*, Section 2.2.

Synaptic Plasticity. Plasticity at all four weight classes ($W_{E \leftarrow E}$, $W_{E \leftarrow I}$, $W_{I \leftarrow E}$, and $W_{I \leftarrow I}$) was governed by different families of homeostatic-based plasticity rules, all driven by the deviation of the actual excitatory and inhibitory rates from their target setpoints (E_{set} and I_{set}). Three different learning rules are presented in the main text of this article.

Standard homeostatic family of rules.

$$\begin{aligned} \Delta W_{EE} &= +\alpha_E E(E_{set} - E) \\ \Delta W_{EI} &= -\alpha_E I(E_{set} - E) \\ \Delta W_{IE} &= +\alpha_I E(I_{set} - I) \\ \Delta W_{II} &= -\alpha_I I(I_{set} - I), \end{aligned} \qquad [6]$$

where α_E and α_I are the learning rates onto the excitatory and inhibitory units, respectively. The setpoints were based on empirically measured values in ex vivo cortical circuits (46), $E_{\text{set}} = 5$ and $I_{\text{set}} = 14$ Hz and follow a classic homeostatic formulation (33, 34, 43, 44). As outlined in *SI Appendix*, Section 1.5, we also examined variants of these rules, such as standard synaptic scaling (which includes the weight as a factor).

We prove that these rules are stable only in a narrow parameter regime: when excitatory plasticity dominates (*SI Appendix*, Section 2).

Cross-homeostatic family of rules.

$$\begin{aligned} \Delta W_{EE} &= +\alpha_E E(I_{set} - I) \\ \Delta W_{EI} &= -\alpha_E I(I_{set} - I) \\ \Delta W_{IE} &= -\alpha_I E(E_{set} - E) \\ \Delta W_{II} &= +\alpha_I I(E_{set} - E). \end{aligned} \qquad [7]$$

These rules differ from the standard homeostatic formulation in that the setpoints are "crossed," meaning that the weights onto the excitatory (inhibitory) population change in order to minimize the inhibitory (excitatory) error. We prove that these rules are stable for any set of parameters (*SI Appendix*, Section 1.3).

Two-term cross-homeostatic family of rules.

$$\begin{aligned} \Delta W_{EE} &= +\alpha_E E(E_{set} - E) + \alpha_E E(I_{set} - I) \\ \Delta W_{EI} &= -\alpha_E I(E_{set} - E) - \alpha_E I(I_{set} - I) \\ \Delta W_{IE} &= +\alpha_I E(I_{set} - I) - \alpha_I E(E_{set} - E) \\ \Delta W_{II} &= -\alpha_I I(I_{set} - I) + \alpha_I I(E_{set} - E). \end{aligned} \qquad [8]$$

The two-term rules combine homeostatic and cross-homeostatic terms. This exact formulation can be obtained after an approximation of a gradient descent derivation on the following loss function:

$$L = \frac{1}{2}(E - E_{set})^2 + \frac{1}{2}(I - I_{set})^2. \qquad [9]$$

The mathematical derivation can be found in the *SI Appendix*, Section 3. We prove that these rules are stable for a biologically meaningful set of parameter values, as long as the homeostatic part does not dominate (*SI Appendix*, Section 1.4).

For all other methods, including the implementation of the multiunit firing rate and spiking models, numerical and analytical methods, proofs, and derivation of the two-term cross-homeostatic rule, see the *SI Appendix*.

Data, Materials, and Software Availability. Software necessary to reproduce all data shown throughout this manuscript is fully available in the public repository GitHub (see links below). Computational analysis for the rate-based model was performed in custom-written MATLAB R2020a software (https://www.mathworks.com/). Brian2 was used to simulate the spiking model (https://briansimulator.org/). SageMath was used for the analytical proofs (https://www.sagemath.org/) (*SI Appendix*). The MATLAB source code that reproduces Figs. 2, 4–6 and *SI Appendix*, Fig. S4 is available at https://github.com/saraysoldado/Paradoxical2022 (74). The spiking model can be found at https://github.com/mikejseay/spiking-upstates/tree/Paradoxical2022 (75). The Jupyter notebooks with SageMath code to reproduce all analytical results are available at https://github.com/SMDynamicsLab/Paradoxical2022 (76).

ACKNOWLEDGMENTS. We thank Ben Liu, Shanglin Zhou, Juan Romero-Sosa, and Helen Motanis for technical assistance and helpful discussions. We thank Joana Soldado-Magraner for comments on the manuscript. D.V.B. and M.J.S. were supported by the National Institutes of Health grant no. NS116589. R.L. was supported by Universidad Nacional de Quilmes, CONICET, and The Pew Charitable Trusts grant no. 2009-000360-006. S.S.-M. was supported by the Swiss National Science Foundation grant nos. P2ZHP3-187943 and P500PB-203133.

1. J. M. Fuster, J. P. Jervey, Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* **212**, 952–955 (1981).
2. P. S. Goldman-Rakic, Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).
3. X.-J. Wang, Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* **24**, 455–463 (2001).
4. M. M. Churchland *et al.*, Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).
5. G. Hennequin, T. P. Vogels, W. Gerstner, Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron* **82**, 1394–1406 (2014).
6. C. van Vreeswijk, H. Sompolinsky, Chaotic balanced state in a model of cortical circuits. *Neural Comput.* **10**, 1321–1371 (1998).
7. N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* **8**, 183–208 (2000).
8. A. Destexhe, S. W. Hughes, M. Rudolph, V. Crunelli, Are corticothalamic 'up' states fragments of wakefulness? *Trends Neurosci.* **30**, 334–342 (2007).
9. A. Renart *et al.*, The asynchronous state in cortical circuits. *Science* **327**, 587–590 (2010).
10. S. Ostojic, Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nat. Neurosci.* **17**, 594–600 (2014).
11. M. Steriade, D. A. McCormick, T. J. Sejnowski, Thalamocortical oscillations in the sleeping and aroused brain. *Science* **262**, 679–685 (1993).
12. D. A. McCormick, GABA as an inhibitory neurotransmitter in human cerebral cortex. *J. Neurophysiol.* **62**, 1018–1027 (1989).
13. M. Steriade, D. Contreras, Spike-wave complexes and fast components of cortically generated seizures. I. Role of neocortex and thalamus. *J. Neurophysiol.* **80**, 1439–1455 (1998).
14. M. V. Tsodyks, W. E. Skaggs, T. J. Sejnowski, B. L. McNaughton, Paradoxical effects of external modulation of inhibitory interneurons. *J. Neurosci.* **17**, 4382–4388 (1997).
15. H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, D. Ferster, Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* **62**, 578–592 (2009).
16. D. B. Rubin, S. D. Van Hooser, K. D. Miller, The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* **85**, 402–417 (2015).
17. U. Rutishauser, J.-J. Slotine, R. Douglas, Computation in dynamically bounded asymmetric systems. *PLOS Comput. Biol.* **11**, e1004039 (2015).
18. A. Litwin-Kumar, R. Rosenbaum, B. Doiron, Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes. *J. Neurophysiol.* **115**, 1399–1409 (2016).
19. D. Jercog *et al.*, UP-DOWN cortical dynamics reflect state transitions in a bistable network. *eLife* **6**, e22425 (2017).
20. A. Sanzeni *et al.*, Inhibition stabilization is a widespread property of cortical networks. *eLife* **9**, e54875 (2020).
21. D. Plenz, S. T. Kitai, Up and down states in striatal medium spiny neurons simultaneously recorded with spontaneous activity in fast-spiking interneurons studied in cortex-striatum-substantia nigra organotypic cultures. *J. Neurosci.* **18**, 266–283 (1998).
22. J. K. Seamans, L. Nogueira, A. Lavin, Synaptic basis of persistent activity in prefrontal cortex in vivo and in organotypic cultures. *Cereb. Cortex* **13**, 1242–1250 (2003).
23. H. A. Johnson, D. V. Buonomano, Development and plasticity of spontaneous activity and Up states in cortical organotypic slices. *J. Neurosci.* **27**, 5915–5925 (2007).
24. P. Golshani *et al.*, Internally mediated developmental desynchronization of neocortical network activity. *J. Neurosci.* **29**, 10890–10899 (2009).
25. H. Motanis, D. Buonomano, Delayed in vitro development of Up states but normal network plasticity in Fragile X circuits. *Eur. J. Neurosci.* **42**, 2312–2321 (2015).
26. H. Motanis, D. Buonomano, Decreased reproducibility and abnormal experience-dependent plasticity of network dynamics in Fragile X circuits. *Sci. Rep.* **10**, 14535 (2020).
27. F. Donato, S. B. Rompani, P. Caroni, Parvalbumin-expressing basket-cell network plasticity induced by experience regulates adult learning. *Nature* **504**, 272–276 (2013).
28. R. C. Froemke, Plasticity of cortical excitatory-inhibitory balance. *Annu. Rev. Neurosci.* **38**, 195–219 (2015).
29. G. Hennequin, E. J. Agnes, T. P. Vogels, Inhibitory plasticity: Balance, control, and codependence. *Annu. Rev. Neurosci.* **40**, 557–579 (2017).
30. Y. Ahmadian, K. D. Miller, (2019) What is the dynamical regime of cerebral cortex? *arXiv:1908.10101 [q-bio]*.
31. Y.-H. Chen *et al.*, PV network plasticity mediated by neuregulin1-ErbB4 signalling controls fear extinction. *Mol. Psychiatry* **27**, 896–906 (2021).
32. X. He *et al.*, Gating of hippocampal rhythms and memory by synaptic plasticity in inhibitory interneurons. *Neuron* **109**, 1013–1028.e1019 (2021).
33. G. G. Turrigiano, K. R. Leslie, N. S. Desai, L. C. Rutherford, S. B. Nelson, Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature* **391**, 892–896 (1998).
34. M. C. van Rossum, G. Q. Bi, G. G. Turrigiano, Stable Hebbian learning from spike timing-dependent plasticity. *J. Neurosci.* **20**, 8812–8821 (2000).
35. V. Kilman, M. C. van Rossum, G. G. Turrigiano, Activity deprivation reduces miniature IPSC amplitude by decreasing the number of postsynaptic GABA(A) receptors clustered at neocortical synapses. *J. Neurosci.* **22**, 1328–1337 (2002).
36. G. G. Turrigiano, S. B. Nelson, Homeostatic plasticity in the developing nervous system. *Nat. Rev. Neurosci.* **5**, 97–107 (2004).
37. Y.-R. Peng *et al.*, Postsynaptic spiking homeostatically induces cell-autonomous regulation of inhibitory inputs via retrograde signaling. *J. Neurosci.* **30**, 16220–16231 (2010).
38. K. Pozo, Y. Goda, Unraveling mechanisms of homeostatic synaptic plasticity. *Neuron* **66**, 337–351 (2010).
39. G. G. Turrigiano, The self-tuning neuron: Synaptic scaling of excitatory synapses. *Cell* **135**, 422–435 (2008).
40. K. B. Hengen, M. E. Lambo, S. D. Van Hooser, D. B. Katz, G. G. Turrigiano, Firing rate homeostasis in visual cortex of freely behaving rodents. *Neuron* **80**, 335–342 (2013).
41. Z. Ma, G. G. Turrigiano, R. Wessel, K. B. Hengen, Cortical circuit dynamics are homeostatically tuned to criticality in vivo. *Neuron* **104**, 655–664.e654 (2019).
42. S. J. Kuhlman *et al.*, A disinhibitory microcircuit initiates critical-period plasticity in the visual cortex. *Nature* **501**, 543–546 (2013).
43. J. K. Liu, D. V. Buonomano, Embedding multiple trajectories in simulated recurrent neural networks in a self-organizing manner. *J. Neurosci.* **29**, 13172–13181 (2009).
44. T. P. Vogels, H. Sprekeler, F. Zenke, C. Clopath, W. Gerstner, Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* **334**, 1569–1573 (2011).
45. G. T. Neske, S. L. Patrick, B. W. Connors, Contributions of diverse excitatory and inhibitory neurons to recurrent network activity in cerebral cortex. *J. Neurosci.* **35**, 1089–1105 (2015).
46. J. L. Romero-Sosa, H. Motanis, D. V. Buonomano, Differential excitability of PV and SST neurons results in distinct functional roles in inhibition stabilization of up states. *J. Neurosci.* **41**, 7182–7196 (2021).
47. T. Keck *et al.*, Loss of sensory input causes rapid structural changes of inhibitory neurons in adult mouse visual cortex. *Neuron* **71**, 869–882 (2011).
48. M. A. Gainey, J. W. Aman, D. E. Feldman, Rapid disinhibition by adjustment of PV intrinsic excitability during whisker map plasticity in mouse S1. *J. Neurosci.* **38**, 4749–4761 (2018).
49. O. Mackwood, L. B. Naumann, H. Sprekeler, Learning excitatory-inhibitory neuronal assemblies in recurrent networks. *eLife* **10**, e59715 (2021).
50. J. Burrone, M. O'Byrne, V. N. Murthy, Multiple forms of synaptic plasticity triggered by selective suppression of activity in individual neurons. *Nature* **420**, 414–418 (2002).
51. M. Xue, B. V. Atallah, M. Scanziani, Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature* **511**, 596–600 (2014).
52. D. O. Hebb, *The Organisation of Behaviour: A Neuropsychological Theory* (Science Editions, New York, 1949).
53. K. D. Miller, J. B. Keller, M. P. Stryker, Ocular dominance column development: Analysis and simulation. *Science* **245**, 605–615 (1989).
54. D. V. Buonomano, M. M. Merzenich, Cortical plasticity: From synapses to maps. *Annu. Rev. Neurosci.* **21**, 149–186 (1998).
55. S. J. Martin, P. D. Grimwood, R. G. Morris, Synaptic plasticity and memory: An evaluation of the hypothesis. *Annu. Rev. Neurosci.* **23**, 649–711 (2000).
56. S. Song, K. D. Miller, L. F. Abbott, Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* **3**, 919–926 (2000).
57. K. B. Hengen, A. Torrado Pacheco, J. N. McGregor, S. D. Van Hooser, G. G. Turrigiano, Neuronal firing rate homeostasis is inhibited by sleep and promoted by wake. *Cell* **165**, 180–191 (2016).
58. N. F. Trojanowski, J. Bottorff, G. G. Turrigiano, Activity labeling in vivo using CaMPARI2 reveals intrinsic and synaptic differences between neurons with high and low firing rate set points. *Neuron* **109**, 663–676.e665 (2021).
59. Z. Liu, J. Golowasch, E. Marder, L. F. Abbott, A model neuron with activity-dependent conductances regulated by multiple calcium sensors. *J. Neurosci.* **18**, 2309–2320 (1998).
60. A. Joseph, G. G. Turrigiano, All for one but not one for all: Excitatory synaptic scaling and intrinsic excitability are coregulated by CaMKIV, whereas inhibitory synaptic scaling is under independent control. *J. Neurosci.* **37**, 6778–6785 (2017).
61. S. Blein, E. Hawrot, P. Barlow, The metabotropic GABA receptor: Molecular insights and their functional consequences. *Cell. Mol. Life Sci.* **57**, 635–650 (2000).
62. C. M. Niswender, P. J. Conn, Metabotropic glutamate receptors: Physiology, pharmacology, and disease. *Annu. Rev. Pharmacol. Toxicol.* **50**, 295–322 (2010).
63. E. Slomowitz *et al.*, Interplay between population firing stability and single neuron dynamics in hippocampal networks. *eLife* **4**, e04378 (2015).
64. P. J. Drew, Vascular and neural basis of the BOLD signal. *Curr. Opin. Neurobiol.* **58**, 61–69 (2019).
65. S. Zucca *et al.*, An inhibitory gate for state transition in cortex. *eLife* **6**, e26177 (2017).
66. A. Mahrach, G. Chen, N. Li, C. van Vreeswijk, D. Hansel, Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation. *eLife* **9**, e49967 (2020).
67. S. Sadeh, C. Clopath, Inhibitory stabilization and cortical computation. *Nat. Rev. Neurosci.* **22**, 21–37 (2021).
68. S. Sadeh, C. Clopath, Excitatory-inhibitory balance modulates the formation and dynamics of neuronal assemblies in cortical networks. *Sci. Adv.* **7**, eabg8411 (2021).
69. D. Sussillo, L. F. Abbott, Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63**, 544–557 (2009).
70. R. Laje, D. V. Buonomano, Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* **16**, 925–933 (2013).
71. V. Goudar, D. V. Buonomano, Encoding sensory and motor patterns as time-invariant trajectories in recurrent neural networks. *eLife* **7**, e31134 (2018).
72. C. J. Cueva *et al.*, Low-dimensional dynamics for working memory and time encoding. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23021–23032 (2020).
73. H. R. Wilson, J. D. Cowan, Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12**, 1–24 (1972).
74. S. Soldado-Magraner, Paradoxical2022. GitHub. https://github.com/saraysoldado/Paradoxical2022. Deposited 7 July 2022.
75. M. J. Seay, Spiking-upstates. GitHub. https://github.com/mikejseay/spiking-upstates/tree/Paradoxical2022. Deposited 29 June 2022.
76. R. Laje, Paradoxical2022. GitHub. https://github.com/SMDynamicsLab/Paradoxical2022. Deposited 30 June 2022.

PNAS

Supplementary Information for

Paradoxical Self-Sustained Dynamics Emerge from Orchestrated Excitatory and Inhibitory Homeostatic Plasticity Rules

Saray Soldado-Magraner, Michael J. Seay, Rodrigo Laje, Dean V. Buonomano

This PDF file includes:

-Supplementary Figures

- Supplementary Figures S1 to S7

-Supplementary Methods

-Supplementary Material (Analytical results and derivations)

Supplementary Figures

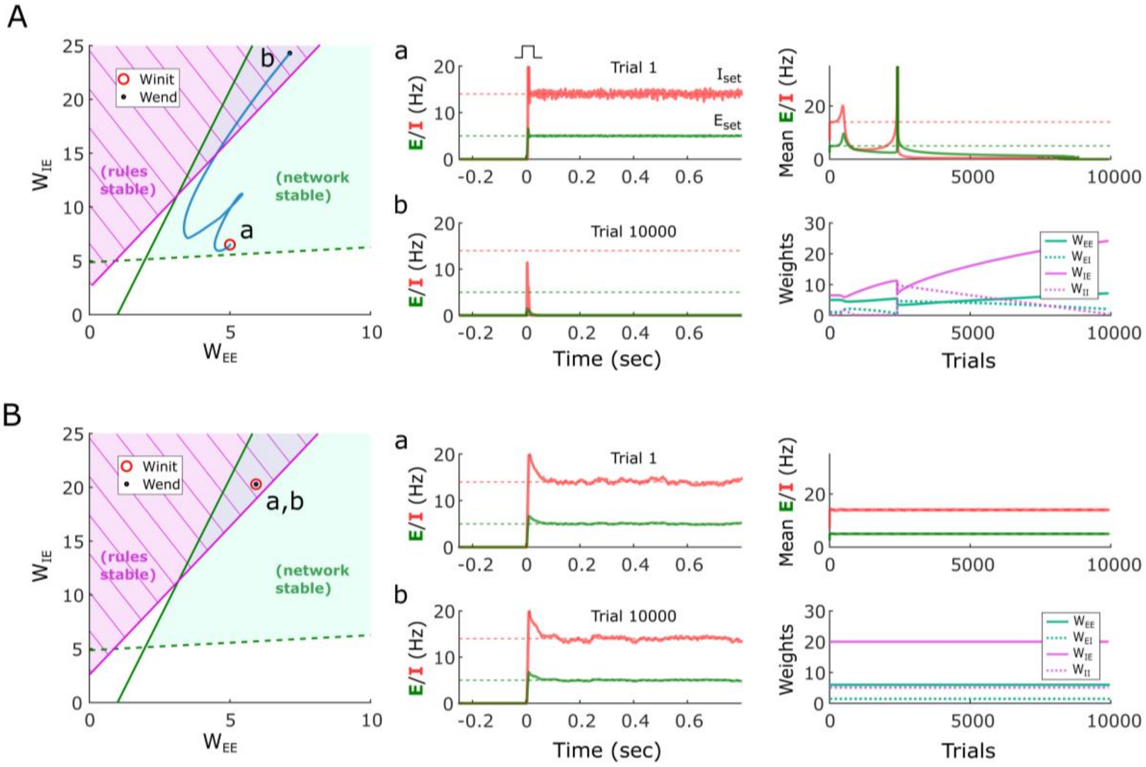
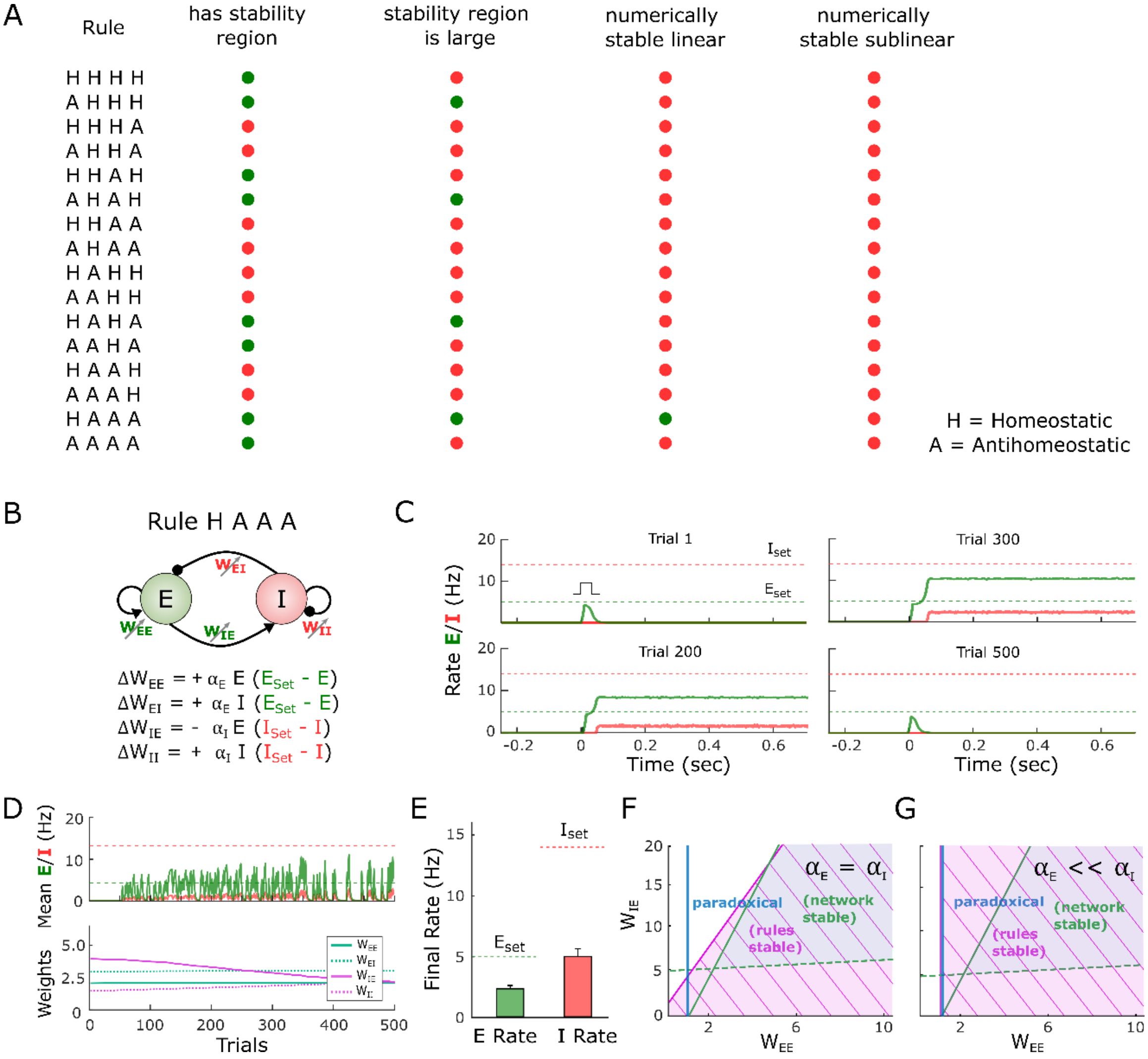


Figure S1. Stability is achieved only in the region where both the neural and synaptic plasticity rule subsystems are stable.

(A) Example of a network initialized within the stability region of the neural subsystem, but within the unstable region of the plasticity rule subsystem. At the beginning of the simulation the network is in a stable self-sustained activity (a). Specifically, in addition to the “free” W_{EE} and W_{IE} weights shown in the plot, the W_{EI} and W_{II} weights are set according to the steady state solution for activity fixed point (loosely speaking W_{EE}/W_{EI} and W_{IE}/W_{II} are “balanced”, see Eqs. 4 and 5). Under the standard homeostatic rules, the weights of the network diverge (blue trajectory), and with time the stable self-sustained activity is no longer observed (b). Right panels show the evolution of the average firing rate and weights across trials. The firing rate diverges under the presence of the rules, and explodes (until reaching saturation, see Methods), finally settling in a Down-state. The weights keep evolving unbounded. The initial weights are $W_{EE}=5$ and $W_{EI}=6.5$ (W_{EI} and W_{II} are set according to Eqs. 4 and 5, see Methods). Note that during plasticity the W_{EE} and W_{IE} weights may cross the overlapping region in which the plasticity rule (hatched) and network (green) subsystems are stable, but still not converge because the W_{EI} and W_{II} weights have also been evolving and are no longer “balanced” with W_{EE} and W_{IE}, respectively.

(B) Example of a network initialized within the stability region of the neural subsystem, and within the stable region of the plasticity rule subsystem. At the beginning of the simulation the network is on a stable self-sustained activity (a). In this case, despite the presence of the homeostatic rules, the weights of the network do not diverge and the network remains in a stable self-sustained regime (b). Right panels show the evolution of the average firing rate and weights across trials. The firing rate and weights remain stable across trials. The initial weights are $W_{EE}=6$ and $W_{EI}=20$ (W_{EI} and W_{II} are set according to equations 4 and 5, see Methods).



Rule	has stability region	stability region is large	numerically stable linear	numerically stable sublinear
H H H H	🟢	🔴	🔴	🔴
A H H H	🟢	🟢	🔴	🔴
H H H A	🔴	🔴	🔴	🔴
A H H A	🔴	🔴	🔴	🔴
H H A H	🟢	🔴	🔴	🔴
A H A H	🟢	🟢	🔴	🔴
H H A A	🔴	🔴	🔴	🔴
A H A A	🔴	🔴	🔴	🔴
H A H H	🔴	🔴	🔴	🔴
A A H H	🔴	🔴	🔴	🔴
H A H A	🟢	🟢	🔴	🔴
A A H A	🟢	🔴	🔴	🔴
H A A H	🔴	🔴	🔴	🔴
A A A H	🔴	🔴	🔴	🔴
H A A A	🟢	🟢	🟢	🔴
A A A A	🟢	🔴	🔴	🔴

H = Homeostatic
A = Antihomeostatic

Rule H A A A

$\Delta W_{EE} = +\, \alpha_E\, E\, (E_{Set} - E)$
$\Delta W_{EI} = +\, \alpha_E\, I\, (E_{Set} - E)$
$\Delta W_{IE} = -\, \alpha_I\, E\, (I_{Set} - I)$
$\Delta W_{II} = +\, \alpha_I\, I\, (I_{Set} - I)$

Figure S2. Homeostatic and anti-homeostatic combinations of plasticity rules also fail to drive the emergence of self-sustained dynamics.

(A) Sixteen variations of the standard homeostatic rules presented in **Fig. 2** were assessed for stability. The plasticity governing each of the four weight types, W_{EE}, W_{EI}, W_{IE}, W_{II} was set to be either homeostatic (H) or antihomeostatic (A). The first rule on the table (HHHH) corresponds to the standard homeostatic rules presented in **Fig. 2**, where all weights obey homeostatic plasticity. All rules were tested for stability analytically and numerically. A red dot implies that the listed condition is not satisfied, while a green dot means that it does. The condition on the first column indicates whether a stability region for the plasticity rule is present. The second column indicates whether such region has a large overlap with the region of stability of the neural subsystem. The third column indicates whether the rule is successful, using numerical simulations, at driving the network to a stable self-sustained activity when starting from regimes with self-sustained activity already present (meaning the network is initialized in the linear regime). The fourth column indicates the same as the former, but with the network initialized in the sub-linear regime, where activity is not initially present (e.g., as observed early in developmental conditions).

(B) Schematic (top) of the population rate model in which the four weights are governed by the HAAA rule in panel (A).

(C) Example simulation of the HAAA rule over the course of simulated development. The evolution of the firing rate of the excitatory and inhibitory population within a trial in response to a brief external input is shown in every plot. $E_{set}=5$ and I_{set}=14 represent the target homeostatic setpoints. Weights were initialized to W_{EE}=2.1, W_{EI}=3, W_{IE}=4, and W_{II}=2 as in **Fig. 2**. Note that while an evoked self-sustained activity emerges by Trial 200 the firing rates do not converge to their setpoints, and by Trial 500 the stable activity is no longer observed.

(D) Average rate across trials (upper plot) for the excitatory and inhibitory populations for the data shown in **C**. Weight dynamics (bottom plot) produced by the homeostatic rules across trials for the data shown in **C**.

(E) Average final rate for 100 independent HAAA simulations with different weight initializations. Those initializations included cases in which the network starts in the sublinear regime (where the initial *E* firing rate was zero or very low). The weights were initialized uniformly between the following ranges: W_{EE}[1,3], W_{EI}[0.5,1.5], W_{IE}[4,8], W_{II}[0.2,0.8]. Data represents mean ± SEM.

(F) Analytical stability regions of the neural and HAAA plasticity rule subsystems as a function of the free weights W_{EE} and W_{IE}. Here the stability plot is obtained by considering equal learning rates for all four plasticity rules (as used for panels **C-E**).

(G) Similar to **F** but with but with $\alpha_E << \alpha_I$. Right of blue line shows the area where the network is in a paradoxical regime (defined by the condition $W_{EE} * g_E – 1 > 0$). Contrary to standard homeostatic rules (**Fig. 2**), the HAAA rule is only stable in the paradoxical region of parameter space (i.e., $W_{EE}*g_E – 1 > 0$; note white area to the left of the blue line). This may explain why the rule fails at driving the network to a stable self-sustained activity when starting with developmental-like conditions.

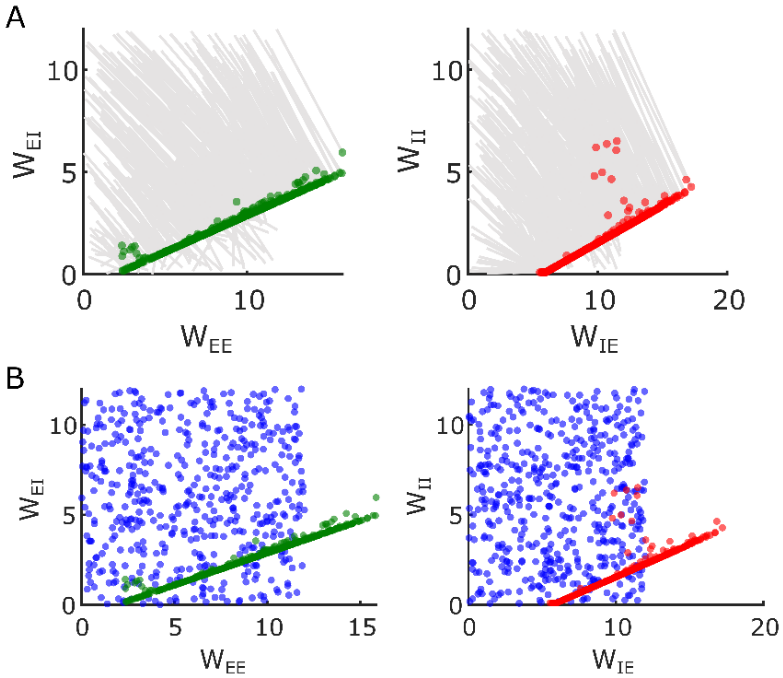


Figure S3. A broader weight initialization also leads to converge of the cross-homeostatic rules in the two-population model.

(A) Same as in **Fig.4D** weight changes for 100 different simulations with random weight initializations are shown. Lines show change from initial to final (circles) weight values. Weights are initialized uniformly between the following ranges: W_{EE}[0,12], W_{EI}[0,12], W_{IE}[0,12], W_{II}[0,12].

(B) Same data as in (A) but only displaying the initial weight values (blue circles) and the final ones (green and red circles).

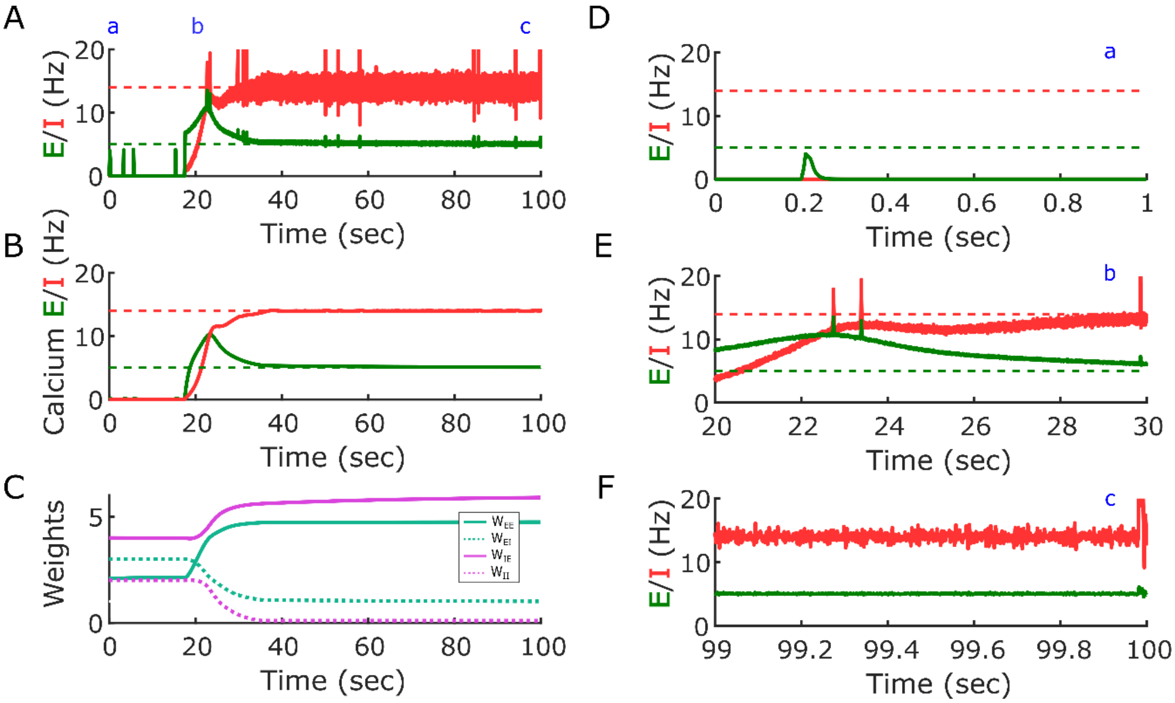


Figure S4. An online implementation of cross-homeostatic plasticity in the two-population model also converges to inhibition-stabilized dynamics at the setpoints.

(A) Firing rate of the excitatory and inhibitory population over time. Cross-homeostatic plasticity implemented in an online or continuous fashion (as opposed to trial-based updates of the weights) also drives the network from a silent state to its setpoints at E_{set}*=5* and I_{set}=14 (dashed lines). Random Poisson input arrives at a frequency of 0.1 Hz to engage recurrent activity (I_{extt}=7, I_{dur}=10 ms).

(B) A calcium sensor in both, the excitatory and inhibitory population continuously integrates their activity with $\tau_{Ca^{2+}} = 1000\, ms$.

(C) The weights are updated at every time step based on the instantaneous calcium sensor value. Weights are initialized as in **Fig.4** W_{EE}=2.1, W_{EI}=3, W_{IE}=4, and W_{II}=2. Learning rate is set to $\alpha = 5e^{-7}$. The rest of the network parameters are the same as in **Fig.4.**

(D) Snippet of the first second of simulation time shown in **A.** This time point is shown as a blue a). The network starts in a silent state as in **Fig.4B.** The first external input to the network is set manually for comparison. The rest of the inputs arrive at random Poisson times with a frequency of 0.1 Hz.

(E) Snippet of the activity of the network between seconds 20 and 30 of simulation time (b). The firing rate of the excitatory and inhibitory connectivity raises towards its setpoints. Note the Poisson external input is still present but it does not disturb the network stability.

(F) Last second of simulation time (c). The activity of the network has converged to its setpoints.

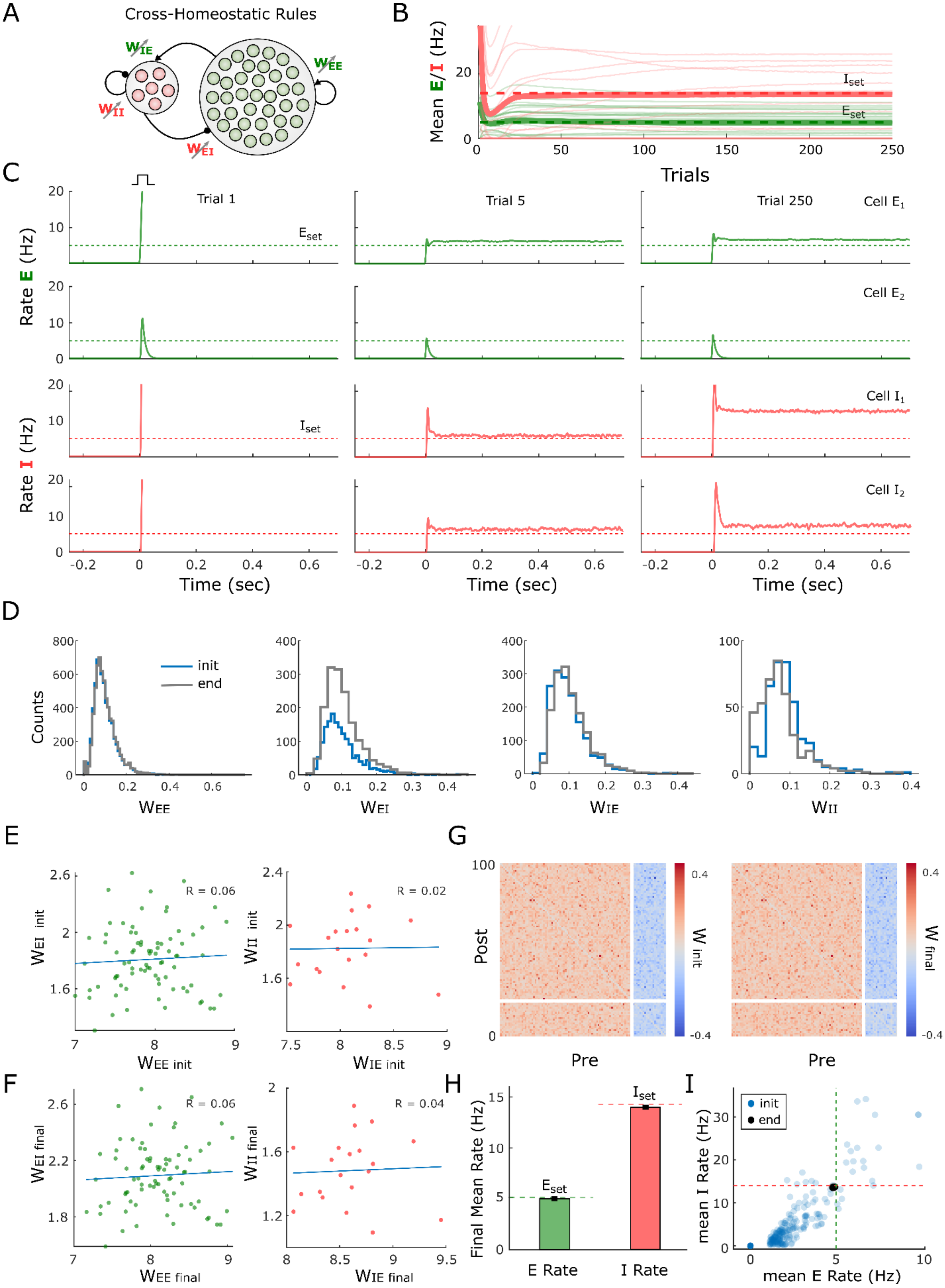

A
Cross-Homeostatic Rules
W_{IE}
W_{EE}
W_{II}
W_{EI}
B
Mean E/I (Hz)
20
0
I_{set}
E_{set}
0
50
100
150
200
250
Trials
C
Trial 1
Trial 5
Trial 250
Cell E_1
Cell E_2
Cell I_1
Cell I_2
E_{set}
I_{set}
Rate E (Hz)
Rate I (Hz)
20
10
0
-0.2
0
0.2
0.4
0.6
Time (sec)
D
Counts
init
end
800
600
400
200
0
0.2
0.4
0.6
WEE
400
300
200
100
0.1
0.2
0.3
0.4
WEI
WIE
100
50
WII
E
R = 0.06
R = 0.02
WEI init
WEE init
WII init
WIE init
2.6
2
1.6
2.4
7
8
9
7.5
8.5
F
R = 0.06
R = 0.04
WEI final
WEE final
WII final
WIE final
2.6
2.2
1.8
1.6
1.2
9.5
G
Post
Pre
100
0
0.4
-0.4
W init
W final
H
Final Mean Rate (Hz)
15
10
5
0
E_{set}
I_{set}
E Rate
I Rate
I
init
end
mean I Rate (Hz)
mean E Rate (Hz)
30
20
10
0
5
10

Figure S5. Log-normal initialization of weights also leads to convergence of the cross-homeostatic rules in a multi-unit model.

(A) Schematic (left) of the multi-unit rate model. The network is composed of 80 excitatory and 20 inhibitory units recurrently connected. The four weight classes are governed by cross-homeostatic plasticity rules (right). See Methods for a detailed explanation of the implementation.

(B) Evolution of the average rate across trials of 20 excitatory and inhibitory units in an example simulation. The network is initialized with random log-normal weights (see Methods) and so neurons present diverse initial rates. E_{set}=5 and I_{set}=14 represent the target homeostatic setpoints. Red and green lines represent the individual (thin lines) and average (thick lines) firing rate of inhibitory and excitatory population, respectively.

(C) Example of the firing rate of two excitatory and two inhibitory units at different points in **B**. The evolution of the firing rate of the excitatory and inhibitory population within a trial in response to a brief external input is shown in every plot. Individual units converge to stable self-sustained dynamics but not to the defined setpoint.

(D) Initial distribution of weights at the beginning (blue) and end of the simulation (grey).

(E) E-I weight relationships at the beginning of the simulation. Every dot represents the total presynaptic weight onto a single unit. Left excitatory neurons. Right inhibitory neurons.

(F) Same plot as in **E** at the end of the simulation.

(G) Weight matrix for the multi-unit model at the beginning (left) and end (right) of the simulation. Inhibitory weights are shown in blue, and excitatory weights in red.

(H) Average firing rate of the units of the multi-unit model and for different initializations of weights (n=400). The network converges to the setpoints in average. Data represents mean ± SEM.

(I) Same data as in **H** but showing the average initial rate of the network for the multiple initializations (blue dots) and the average rate at the end (black). Target rates are shown in dotted lines (green, E_{set}=5, red I_{set}=14).

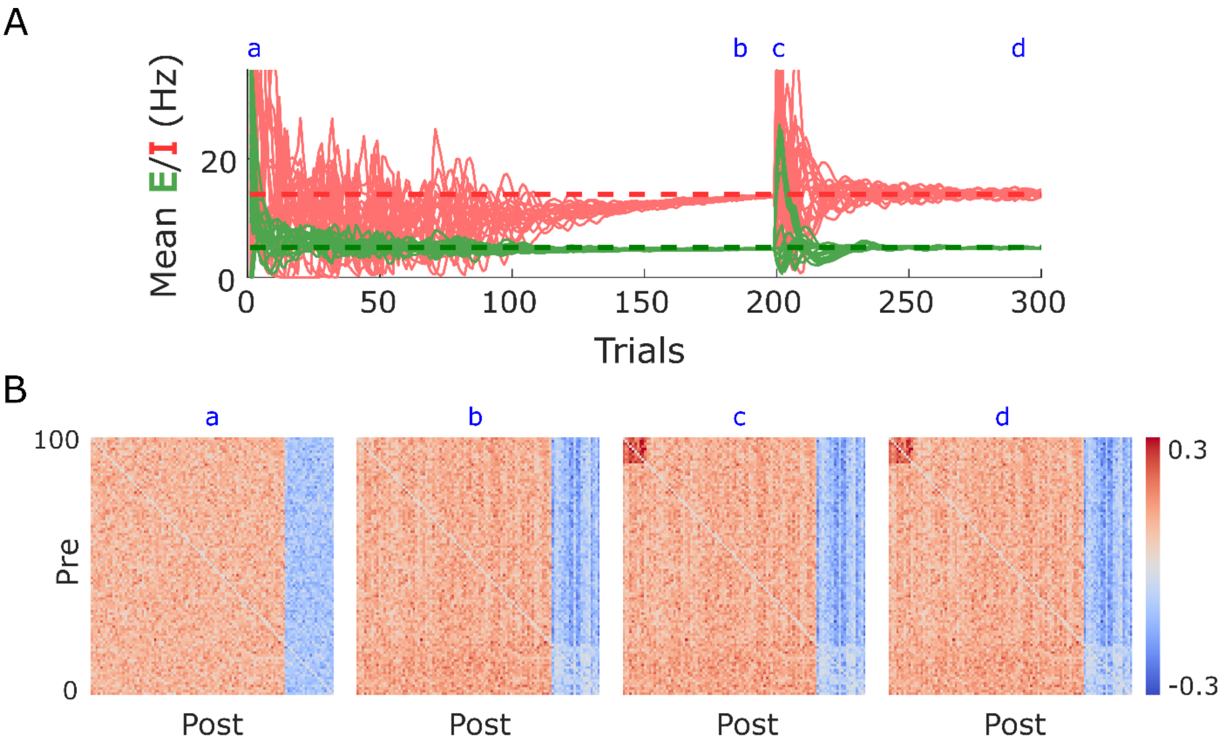


Figure S6. Hebbian-like changes in the connectivity matrix are preserved in the presence of two-term cross-homeostatic plasticity.

- **(A)** Same example as in **Fig. 6B**, where a Hebbian change in the connectivity matrix is introduced at Trial 200. A ‘memory’ is imprinted in the first 10 excitatory neurons by increasing their recurrent weights by a constant factor. This change drives the network away from its setpoints and reengages the two-term cross-homeostatic rules. The rules successfully bring the network rates back to the setpoints, while preserving much of the differential connectivity between the altered weights. Setpoints are shown in dashed lines, E_{set}=5 and I_{set}=14.
- **(B)** Weight matrix of the network at four different time points labeled in **A**. a) Initial random weight matrix. b) The weight matrix after two-term cross homeostatic plasticity has driven the network rates to setpoints. c) A Hebbian-like change in connectivity is imposed into the recurrent weights of the first 10 excitatory neurons. d) Weight matrix after two-term cross-homeostatic plasticity re-stabilizes the firing rates. Note that the ‘memory’ imposed into the weight matrix is preserved.

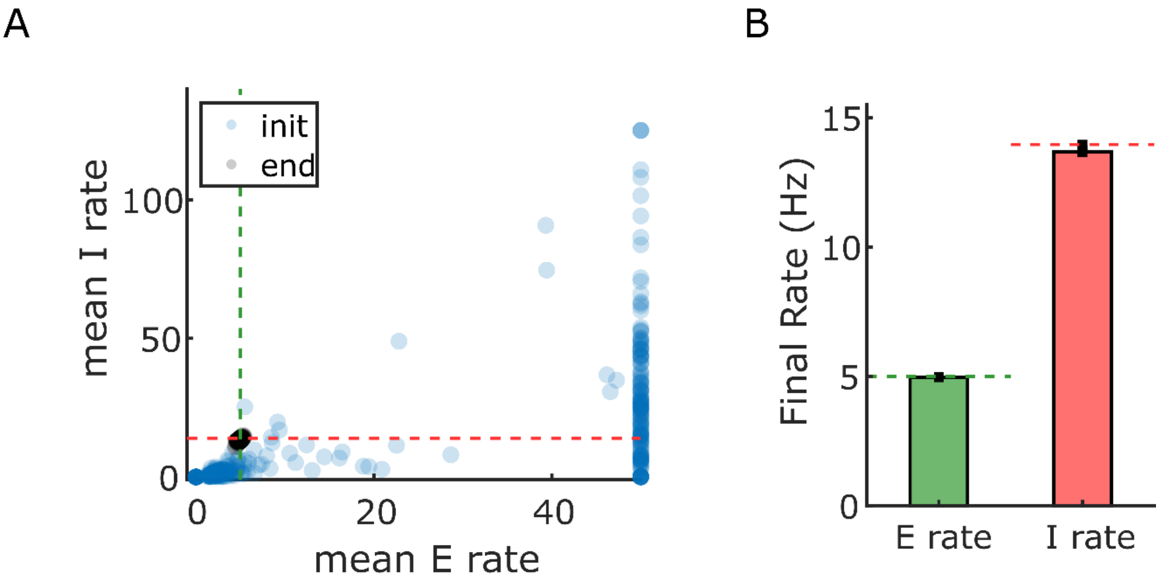


Figure S7. Broader weight initializations for the multi-unit model also lead to convergence of the cross-homeostatic rules.

(A) Average initial rate of the network for multiple weight initializations (blue dots) and the average rate at the end (black dots) after cross-homeostatic plasticity. Target setpoints are shown in dotted lines (green, E_{set}=5, red I_{set}=14). Networks have the same parameters as in **Fig.5**. Mean weights are initialized as W_{EE}[0,10], W_{EI}[0,10], W_{IE}[0,10], W_{II}[0,10], with a normally distributed standard deviation of 10% around the means (n=400).

(B) Same simulations as in **A** showing the final average firing rate of the units of the multi-unit model for the different networks. The networks converge to the setpoints on average. Data represents mean ± SEM.

Supplementary Methods

Numerical simulations of the firing rate model

For all simulations, the weights were updated after the completion of every trial. The trials lasted 2 seconds. Note that the value of E and I on every rule are implemented as average firing rates. The average of E and I is computed after every trial and then is low pass filtered by a process with a time constant $\tau_{trial} = 2$. The numerical integration time step was 0.1 ms. A minimum weight of 0.1 was set for all weights.

A saturation to the excitatory and inhibitory firing rate (100 and 250 Hz, respectively) was added to prevent the nonbiological scenario in which activity could diverge towards infinity under unstable conditions. Note that at the fixed point the saturation is not necessary for the cross-homeostatic rule because it is inherently stable as proved in the Supplementary Material (**Section 1.3**).

In **Fig. 2D** and **4D-G** we initialize the weights uniformly between the following ranges: W_{EE}[4,7], W_{EI}[0.5,2], W_{IE}[7,13], W_{II}[0.5,2]. Simulations were run for 3000 trials to assess stability and convergence. Note that this initialization was chosen for visualization purposes in order to represent a range of initial values surrounding the E-I balance line attractor, but the rules are robust to much broader and equal initializations for all weights, W_{EE}[0,12], W_{EI}[0,12], W_{IE}[0,12], W_{II}[0,12], (**Supplementary Figure S3**). We emphasize that many of these initializations resulted in starting conditions with exploding network rates, which were held in check by the saturation imposed on the transfer function. Despite this initial instability, the rules successfully brought the rates to the setpoint values.

Analytical stability analyses of the firing rate model

We analyzed the entire dynamical system (composed of the neural subsystem and the learning rule subsystem) for every synaptic learning rule considered in this work, and analyzed its stability. In every case, the general prescription is:

a) Take the combined neural and learning rule subsystems and nondimensionalize all variables, so that the two different time scales are evident (fast neural, slow synaptic plasticity). For the description of the learning rule subsystem we switch from discrete-time dynamics to continuous-time dynamics: $\Delta W \rightarrow \tau_0\, dW/dt$
b) Make a quasi-steady state (QSS) approximation of the neural subsystem. This means we will consider the neural subsystem is fast enough so that it converges "instantaneously" (when compared to the synaptic plasticity subsystem) to its corresponding fixed point. For this we will require that the stability conditions of the neural subsystem are satisfied (see below).

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

d) Address (linear) stability. If both nonzero eigenvalues have negative real parts, then the fixed point is stable under the learning rule; if at least one of the nonzero eigenvalues has positive real part, then the fixed point is unstable. (A note on abuse of notation: we might say indistinctly "the fixed point is stable/unstable" and "the learning rule is stable/unstable".)

For a detailed explanation see **Section 2** in the Supplementary Material.

Implementation of the Multi-unit firing rate model

A rate-based recurrent network model containing N_e = 80 excitatory and N_i = 20 inhibitory neurons was implemented with all-to-all connectivity (without self-connections). The activation of the neurons followed equations (1), (2) and (3) in the main Methods. The same parameters as for the population model were used, where W_{XY} represents now a matrix of synaptic weights from population X to population Y. A minimum weight of $0.1/N_x$ for W_{EI} and W_{IE} and $0.1/(N_x\text{-}1)$ for W_{EE} and W_{II} was set for all weights.

The synaptic plasticity rules were implemented as follows.

Cross-homeostatic family of rules:

$$(10)\ \Delta W_{ij}^{EE} = +\alpha E_j \sum_{k=1}^{N_I} (I_{set} - I_k)/N_I$$

$$\Delta W_{ij}^{EI} = -\alpha I_j \sum_{k=1}^{N_I} (I_{set} - I_k)/N_I$$

$$\Delta W_{ij}^{IE} = -\alpha E_j \sum_{k=1}^{N_E} (E_{set} - E_k)/N_E$$

$$\Delta W_{ij}^{II} = +\alpha I_j \sum_{k=1}^{N_E} (E_{set} - E_k)/N_E\ ,$$

where *i* and *j* represent the post- and presynaptic neurons, respectively, and *k* denotes the presynaptic inhibitory neurons targeting the excitatory neurons (or the presynaptic excitatory neurons targeting an inhibitory neuron). *N_E* and *N_I* denote the total number of excitatory and inhibitory neurons, respectively. The weights are therefore updated following the *average* presynaptic error of the crossed E/I population classes. Note as stated above that this formulation can be implemented in a local manner (see Discussion). A learning rate of $\alpha = 0.00002$ was used for all simulations.

Two-term cross-homeostatic family of rules:

$$(11)\ \Delta W_{ij}^{EE} = +\alpha E_j(E_{set} - E_i) + \alpha E_j \sum_{k=1}^{N_I} (I_{set} - I_k)/N_I$$

$$\Delta W_{ij}^{EI} = -\alpha I_j(E_{set} - E_i) - \alpha I_j \sum_{k=1}^{N_I} (I_{set} - I_k)/N_I$$

$$\Delta W_{ij}^{IE} = +\alpha E_j(I_{set} - I_i) - \alpha E_j \sum_{k=1}^{N_E} (E_{set} - E_k)/N_E$$

$$\Delta W_{ij}^{II} = -\alpha I_j(I_{set} - I_i) + \alpha I_j \sum_{k=1}^{N_E} (E_{set} - E_k)/N_E\ ,$$

here the first term represents the standard homeostatic rule, and the second term cross-homeostatic plasticity (as implemented above). A learning rate of $\alpha = 0.00001$ was used for all simulations.

In **Fig. 5G-H** and **6G-H** we initialize the mean weights of the population uniformly in between the following ranges: W_{EE}[1,6], W_{EI}[0.5,2], W_{IE}[5,7], W_{II}[0.5,2]. The weights within each class were then normally distributed around that mean (and normalized by the number of neurons) with a standard deviation of 10% of the mean. Note that this initialization led to multiple initial conditions with exploding network rates (which were held in check by the saturation cutoff of the neurons). Those initial rates are not displayed in **Fig. 5-6H** for visualization purposes, but the rules successfully brought all those cases to the corresponding setpoints (final rates are displayed). Simulations were run for 1000 trials to assess stability of the convergence. Note that broader initializations for all weights W_{EE}[0,10], W_{EI}[0,10], W_{IE}[0,10], W_{II}[0,10] also result in convergence of the rules, although many more of those combinations display initial exploding network rates (**Supplementary Fig. S7)**. In the example shown in **Fig. 5A-F** and **6A-F** individual weights were normally distributed with equal mean across weights classes and standard deviation (0.1±0.04). For **Supplementary Fig. S5G-H** a lognormal distribution of weights was used. The mean weights were distributed uniformly as in Figure **5G-H**, and then weights within each class were distributed log-normally, with arithmetic standard deviation of 0.05. We note that while increasing the standard deviation of the log-normal led to more complex time-varying neural dynamics (1), the rules still converged to the average set-points.

Implementation of the Spiking model

The spiking model was designed based on previous work (2). Units in the model were leaky integrate-and-fire neurons with a spike adaptation current. The membrane potential of each unit was represented as

$$(12)\ C_m \frac{dV(t)}{dt} = g_L(E_L - V(t)) + I_{syn}(t) - I_{adapt}(t) + \sigma\sqrt{\tau_m}\eta(t)$$

$$(13)\ \frac{dI_{adapt}(t)}{dt} = \frac{-I_{adapt}(t)}{\tau_{adapt}}.$$

The noise term $\sigma\sqrt{\tau_m}\eta(t)$ represents an Ornstein-Uhlenbeck process with zero mean, standard deviation σ, and a time constant equal to the membrane time constant $\tau_m = C_m/g_L$. When $V(t) \geq V_{thresh}$, the unit emitted a spike, its voltage was reset to V_{reset}, and its adaptation current I_{adapt} was incremented by β/τ_{adapt}. After spiking, the unit entered an absolute refractory period $\tau_{refractory}$. Default values for unit parameters can be found in **Table S1**.

Synapses were current-based, and the total synaptic current $I_{syn}(t)$ was summed across each unit's incoming synapses with distinct synaptic weights determined by the matrices W_{EE}, W_{IE}, W_{EI}, and W_{II}. Total synaptic current to a postsynaptic excitatory or inhibitory unit was given by each of the following two equations, respectively

$$(13)\ I_{syn}(x,t) = \sum_{y=1}^{N_{exc}} W_{EE}(x,y)\, s_{syn}(x,y,t) + \sum_{y=1}^{N_{inh}} W_{EI}(x,y)\, s_{syn}(x,y,t)$$

$$(14)\ I_{syn}(x,t) = \sum_{y=1}^{N_{exc}} W_{IE}(x,y)\, s_{syn}(x,y,t) + \sum_{y=1}^{N_{inh}} W_{II}(x,y)\, s_{syn}(x,y,t)\,.$$

The kinetics of the synaptic currents were determined by the function $s_{syn}(x,y,t)$ for each presynaptic unit y and postsynaptic unit x. When a presynaptic spike occurred in unit y at time t^*, $s_{syn}(x,y,t)$ was incremented by an amount described by a delayed difference of exponentials equation (3)

$$(15)\ \Delta s_{syn}(x,y,t) = \frac{\tau_m}{\tau_d - \tau_r}\left[\exp\left(-\frac{t-\tau_l-t^*}{\tau_d}\right) - \exp\left(-\frac{t-\tau_l-t^*}{\tau_r}\right)\right],$$

where τ_m indicates the postsynaptic membrane time constant. Thus, synaptic kinetics were determined by the delay τ_l, the rise time τ_r, and the decay time τ_d.

The synaptic delay τ_l was uniformly distributed between 0 and 1 ms across all excitatory (inhibitory) synapses. Synaptic parameters can be found in **Table S2**.

Networks consisted of 1600 *E* units and 400 *I* units with probability of connection $p_{conn} = 0.25$ and no autapses (self-connections). Connectivity was uniformly random, and weights for each synaptic class were initialized from normal distributions with a coefficient of variation equal to 0.2. For the example training session shown in **Figure 7**, the initial weights were $\overline{W_{EE}} = 80\,pA,\ \overline{W_{IE}} = 100\,pA,$ $\overline{W_{EI}} = 350\,pA,\ \overline{W_{II}} = 225\,pA$. Network simulations were evaluated using forward Euler integration using a time step of 0.1 ms.

During each trial (1.5 s) of training a brief external current large enough to cause a spike ($I_{syn} \Rightarrow I_{syn} + 0.98\,nA$) was injected into 100 *E* units. This constituted a "kick" (2, 4) that provided the possibility for recurrent excitation to ignite a self-sustaining Up-state. After each trial, the contiguous time period of nonzero FR was detected, and each unit's FR during that time period was calculated. FRs for each unit in each trial contributed to a moving average vector with a time constant of 2 trials, which we refer to as $\vec{r}_{exc}$ and $\vec{r}_{inh}$. Accordingly, we refer to the firing rate setpoints as r_{excSet} and r_{inhSet}, which were set to 5 and 14 Hz respectively, just as in the firing rate model.

The combined plasticity equations used in the spiking model were each formulated as a sum of homeostatic (first) and local cross-homeostatic (second) terms

$$(16)\ \Delta W_{EE} = +\alpha \cdot \vec{r}_{exc}^{\ \mathrm{T}} \cdot (r_{excSet} - \vec{r}_{exc}) + \alpha \cdot \vec{r}_{exc}^{\ \mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inhCross})$$
$$\Delta W_{EI} = -\alpha \cdot \vec{r}_{inh}^{\ \mathrm{T}} \cdot (r_{excSet} - \vec{r}_{exc}) - \alpha \cdot \vec{r}_{inh}^{\ \mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inhCross})$$
$$\Delta W_{IE} = +\alpha \cdot \vec{r}_{exc}^{\ \mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inh}) - \alpha \cdot \vec{r}_{exc}^{\ \mathrm{T}} \cdot (r_{excSet} - \vec{r}_{excCross})$$
$$\Delta W_{II} = -\alpha \cdot \vec{r}_{inh}^{\ \mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inh}) + \alpha \cdot \vec{r}_{inh}^{\ \mathrm{T}} \cdot (r_{excSet} - \vec{r}_{excCross})\,,$$

where α is a learning rate constant set to $0.0025\frac{pA}{Hz^2}$. For the local cross-homeostatic term, each element of $\vec{r}_{popCross}$ represents the average FR of the units in the opposite population that synapse onto that unit. This is calculated by multiplying the unit FRs by the connectivity matrix A_{XY} and dividing by the vector that results from summing its columns, which we refer to as $\vec{a}_{XY}$. Note that the $\oslash$ symbol refers to element-wise division

$$(17)\ \vec{r}_{inhCross} = A_{EI}\vec{r}_{inh} \oslash \vec{a}_{EI}$$
$$\vec{r}_{excCross} = A_{IE}\vec{r}_{exc} \oslash \vec{a}_{IE}\,.$$

For a vector of presynaptic FRs ($\vec{r}_{exc}$, $\vec{r}_{inh}$) we imposed a minimum value such that each element was at least 1 Hz (otherwise networks can get stuck in the down-state). Additionally, all synaptic weights were bounded to stay within minimum and maximum weight values of 10 pA and 750 pA respectively.

For the paradoxical effect analysis in **Figure 7I**, the adaptation current was disabled for all units in order to allow for a long and stable Up-state. In each trial, a kick was given at 100 ms. From 3 to 4 s, a small positive current was injected into all inhibitory units. 40 trials were conducted at each value of the injected current, and a PSTH for each value was constructed using the inhibitory population spiking activity across all trials.

For the analysis in **Figure 7L**, nine networks were trained from distinct mean weight values for the four synaptic classes as shown in **Table S3**. We measured the error of unit firing rates with respect to their setpoints at the beginning and end of training, as quantified by the mean-squared error (MSE) across all individual excitatory and inhibitory units at each trial

$$(18)\ \ MSE(trial) = \frac{1}{2000}\sum_{i=1}^{2000}\left(\vec{r}_i - r_{popSet}\ \right)^2,$$

where $\vec{r}_i$ represented the FR of a unit in that population (excitatory or inhibitory) and r_{popSet} represented the corresponding set-point.

Table S1. Unit parameters

Cell Parameter	**Symbol**	**Value (E)**	**Value (I)**	**Unit**
Resting potential	E_L	7.6	6.5	mV
Reset potential	V_{reset}	14	14	mV
Spike threshold	V_{thresh}	20	20	mV
Refractory period	$\tau_{refractory}$	5	2	ms
Membrane capacitance	C_m	200	100	pF
Leak conductance	g_L	10	10	nS
Membrane time constant	τ	20	10	ms
Adaptation strength	β	3	0	nA·ms
Adaptation time constant	τ_a	500	n/a	ms
Noise standard deviation	σ	2.5	2.5	mV

Model parameters defining intrinsic properties of excitatory (E) and inhibitory (I) units.

Table S2. Synaptic parameters

Synaptic Parameter	Symbol	Value (E)	Value (I)	Unit
Rise time	τ_r	8	1	ms
Fall time	τ_d	23	1	ms
Mean synaptic delay	τ_l	1	0.5	ms

Model parameters defining kinetics of excitatory (E) and inhibitory (I) synapses.

Table S3. Initial weight means for robustness analysis

Network	$\overline{W_{EE}}$	$\overline{W_{IE}}$	$\overline{W_{EI}}$	$\overline{W_{II}}$
1	50	75	300	300
2	100	75	400	300
3	125	75	500	300
4	75	100	300	200
5	100	100	200	200
6	125	100	300	200
7	75	100	300	100
8	100	125	200	100
9	125	125	100	100

Initial means for each synaptic class in robustness analysis (all in units of pA).

REFERENCES

1. Khajeh R, Fumarola F, & Abbott L (2022) Sparse balance: Excitatory-inhibitory networks with small bias currents and broadly distributed synaptic weights. *PLoS computational biology* 18(2):e1008836.
2. Jercog D, *et al.* (2017) UP-DOWN cortical dynamics reflect state transitions in a bistable network. *eLife*.
3. Brunel N & Wang X-J (2003) What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. *Journal of neurophysiology* 90(1):415-430.
4. DeWeese MR & Zador AM (2006) Non-Gaussian membrane potential dynamics imply sparse, synchronous activity in auditory cortex. *J. Neurosci.* 26(47):12206-12218.

Supplementary Material

Paradoxical Self-sustained Dynamics Emerge from Orchestrated Excitatory and Inhibitory Homeostatic Plasticity Rules

Soldado-Magraner, Seay, Laje & Buonomano 2022

June 30, 2022

Contents

1 Summary of results

In this section we describe the general results of the analytical stability analyses of the joint neural and synaptic plasticity subsystems. We express results in terms of the “free weights” W_{EE} and W_{IE}. Subscript “up” is used to identify values at the nontrivial fixed point where E and I are larger than zero (as opposed to “down” where $E = I = 0$ which is the other possible solution). In Section 2 we provide a detailed description of the approach.

1.1 *Homeostatic* plasticity

In continuous-time dynamics, the equations for the Homeostatic plasticity rule are

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

The condition for the fixed point to be stable (i.e., the two nonzero eigenvalues to have negative real parts, see Section 2) under this rule is:

$$\begin{aligned}&(E_{set}^2\alpha_{IE} + I_{set}^2\alpha_{II})I_{set}(W_{EEup}g_E - 1) <\\ &(E_{set}^2\alpha_{EE} + I_{set}^2\alpha_{EI})(E_{set}W_{IEup}g_E - \Theta_I g_E)\end{aligned} \tag{2}$$

It is difficult to determine whether the stability condition of Eq. 2 is satisfied for a general set of parameter values (see numerical analysis below). However, this condition can be re-expressed in a more useful form in terms of W_{EE} and W_{II}:

$$\begin{aligned}(R^2\alpha_3{+}\alpha_4)&(W_{EEup}\, g_E - 1)g_I\\ &< (R^2 + \alpha_2)(W_{II\,up}\, g_I + 1)g_E\end{aligned} \tag{3}$$

where

$$\begin{aligned}R &= E_{set}/I_{set}\\ \alpha_2 &= \alpha_{EI}/\alpha_{EE}\\ \alpha_3 &= \alpha_{IE}/\alpha_{EE}\\ \alpha_4 &= \alpha_{II}/\alpha_{EE}\end{aligned}$$

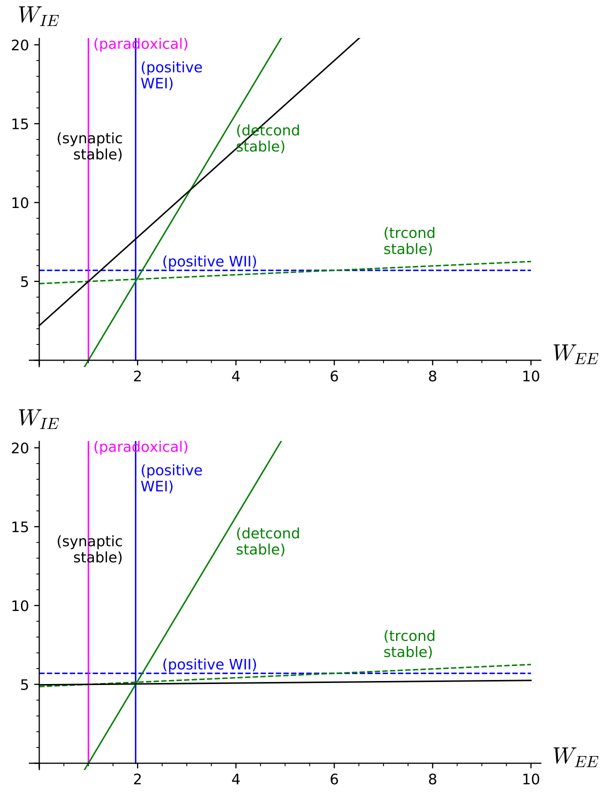


Figure S8: Regions of stability, *Homeostatic* plasticity. (Top) For biologically backed parameter values (Table 1) and learning rates of the same value ($\alpha_{XY} = 0.02$), the stability region of the Homeostatic plasticity (left of black line) has little overlap with the region where the neural subsystem is stable (triangle between the two green lines in the top-left quadrant). (Bottom) Setting $\alpha_{EE} = \alpha_{EI} = 0.02$ and $\alpha_{IE} = \alpha_{II} = 0.0002$ enlarges the stability region of the plasticity rule and makes it overlap with the stability region of the neural subsystem. Every label is on the side where the corresponding condition holds (synaptic stable: Eq. 2; detcond stable: Eq. 22; trcond stable: Eq. 23; positive W_{EI}: Eq. 25; positive W_{II}: Eq. 26; paradoxical: Eq. 27).

Note that learning rate values of the same order lead to $\alpha_{2,3,4} \sim 1$ and that biologically backed parameter values satisfy:

$$I_{set} > E_{set}$$
$$g_I > g_E$$

both likely preventing the condition to hold. On the other hand, if α_{IE} and α_{II} are small enough (slow dynamics of the weights onto the inhibitory neuron) the rule can be stable. See the step-by-step derivation of this stability condition in Section 2.3.

As an illustration of the results above, in Figure S8(top) we plot the stability condition Eq. 2 with parameter values as in Table 1 and learning rates $\alpha_{XY} = 0.02$. It is clear that the plasticity rule is stable in a region with little overlap with the stability region of the neural subsystem. The stability region can be enlarged by making the dynamics of the weights onto the inhibitory neuron slower, as in Figure S8(bottom) where $\alpha_{EE} = \alpha_{EI} = 0.02$ and $\alpha_{IE} = \alpha_{II} = 0.0002$.

See Section 2.3 for a detailed analysis.

1.2 Homeo-antiHomeo variations

The stability condition in the previous section was obtained by assuming all learning rates are positive. Interestingly, if some of them are negative then the fixed point may still be stable. A negative learning rate can be interpreted as the corresponding equation being *anti*-homeostatic, i.e. if the neural activity (E or I) departs from its setpoint then the rule will drive it even farther away. While this kind of behavior would be usually deemed undesired, it is worth considering due to its relationship with the paradoxical regime.

In this section we consider the Homeostatic rule, Eq. 28, and let the learning rates α_{XY} be either positive or negative. The particular case where all learning rates are positive corresponds to the original Homeostatic plasticity rule.

Once we free the signs of the learning rates, the fixed point needs two conditions to be stable:

$$(R^2\alpha_3+\alpha_4)(W_{EE\,up}\,g_E - 1)g_I < (R^2 + \alpha_2)(W_{II\,up}\,g_I + 1)g_E \tag{4}$$

$$(R^2\alpha_3+\alpha_4)(R^2 + \alpha_2) > 0 \tag{5}$$

where

$$\begin{aligned} R &= E_{set}/I_{set} \\ \alpha_2 &= \alpha_{EI}/\alpha_{EE} \\ \alpha_3 &= \alpha_{IE}/\alpha_{EE} \\ \alpha_4 &= \alpha_{II}/\alpha_{EE} \end{aligned}$$

Eq. 4 is equal to the stability condition of the original Homeostatic rule (Eq. 3). The additional condition Eq. 5 is very interesting in that it allows the fixed point to be stable, for instance, under full anti-Homeo plasticity where all four learning rates are negative (leading to $\alpha_{2,3,4}$ all positive).

See details in the corresponding section of the SageMath-Jupyter notebook:
`upstates-Homeostatic stability.ipynb`

1.3 *Cross-Homeostatic* plasticity

In continuous-time dynamics, the equations for the Cross-Homeostatic plasticity rule are

$$\begin{aligned} \frac{dW_{EE}}{dt} &= +\alpha_{EE} E(I_{set} - I) \\ \frac{dW_{EI}}{dt} &= -\alpha_{EI} I(I_{set} - I) \\ \frac{dW_{IE}}{dt} &= -\alpha_{IE} E(E_{set} - E) \\ \frac{dW_{II}}{dt} &= +\alpha_{II} I(E_{set} - E) \end{aligned} \tag{6}$$

and its stability condition in terms of the free weights W_{EE} and W_{IE} reads:

$$\begin{aligned} &(E_{set}^2 \alpha_{EE} + I_{set}^2 \alpha_{EI}) I_{set} W_{IEup} g_E \\ &\quad > -(E_{set}^2 \alpha_{IE} + I_{set}^2 \alpha_{II}) \\ &\qquad ((W_{EEup} g_E - 1) E_{set} - \Theta_E g_E) \end{aligned} \tag{7}$$

This stability condition can be put in a simpler form by switching to W_{EI} and W_{IE}:

$$(R^2 \alpha_3 + \alpha_4) W_{EIup} + (R^2 + \alpha_2) W_{IEup} > 0 \tag{8}$$

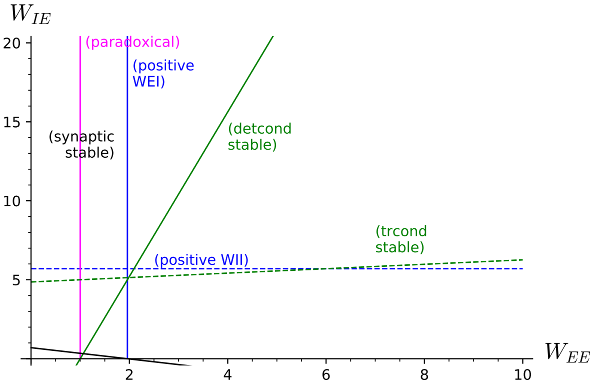


Figure S9: Stability of the *Cross-Homeostatic* rule. The rule is stable for any parameter value; the fixed point is thus stable where the neural subsystem is stable, i.e. in the upper right region between the two green lines. Every label is on the side where the corresponding condition holds (synaptic stable: Eq. 7; detcond stable: Eq. 22; trcond stable: Eq. 23; positive W_{EI}: Eq. 25; positive W_{II}: Eq. 26; paradoxical: Eq. 27). Parameter values as in Table 1.

(where R and $\alpha_{2,3,4}$ are defined as in the previous subsection). This condition is always satisfied because the weights and parameters are positive definite and thus the rule is stable for any choice of parameter values (as long as the neural subsystem is). Fig. S9 shows the stability region of the neural subsystem for the set of parameter values of Table 1. Any choice of values for the weights W_{EE} and W_{IE} within the stability region of the neural subsystem will lead to a stable fixed point.

See Section 2.4 for a detailed analysis.

1.4 *Two-Term* plasticity

The equations for the Two-Term plasticity rule in continuous-time dynamics are

$$\begin{aligned} \frac{dW_{EE}}{dt} &= +\alpha E(I_{set} - I) + \beta E(E_{set} - E) \\ \frac{dW_{EI}}{dt} &= -\alpha I(I_{set} - I) - \beta I(E_{set} - E) \\ \frac{dW_{IE}}{dt} &= -\alpha E(E_{set} - E) + \beta E(I_{set} - I) \\ \frac{dW_{II}}{dt} &= +\alpha I(E_{set} - E) - \beta I(I_{set} - I) \end{aligned} \tag{9}$$

and its stability condition in terms of the free weights

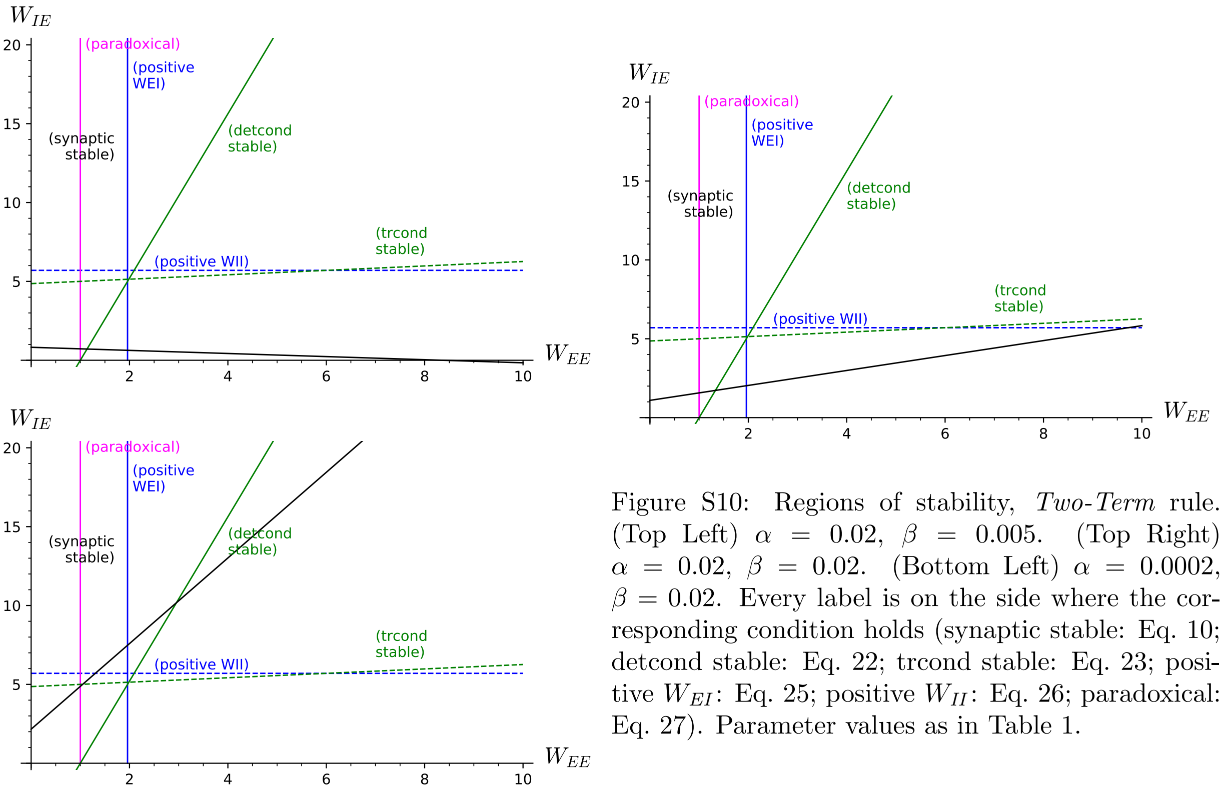


Figure S10: Regions of stability, *Two-Term* rule. (Top Left) $\alpha = 0.02$, $\beta = 0.005$. (Top Right) $\alpha = 0.02$, $\beta = 0.02$. (Bottom Left) $\alpha = 0.0002$, $\beta = 0.02$. Every label is on the side where the corresponding condition holds (synaptic stable: Eq. 10; detcond stable: Eq. 22; trcond stable: Eq. 23; positive W_{EI}: Eq. 25; positive W_{II}: Eq. 26; paradoxical: Eq. 27). Parameter values as in Table 1.

W_{EE} and W_{IE} is

$$\begin{aligned}(I_{set}\alpha + E_{set}\beta)W_{IE\,up}g_E \\ &> (I_{set}\beta - E_{set}\alpha)W_{EE\,up}g_E \\ &+ (\Theta_E g_E + E_{set})\alpha + (\Theta_I g_E - I_{set})\beta\end{aligned} \tag{10}$$

In Figure S10 we plot the stability condition of this rule, Eq. 10, for three different parameter values: $\alpha \gg \beta$ (the "Cross-Homeostatic" terms dominate over the "Homeostatic" terms, and the rule is stable with the largest stability region); $\alpha = \beta$ (the two terms are of comparable size); and $\alpha \ll \beta$ (the "Homeostatic" terms dominate instead, and the stability region of the rule is as small as that of the Homeostatic plasticity).

In order to determine the validity of the stability condition, Eq. 10, in a more general situation, we rewrite it in a more useful form:

$$(a - b)\beta < (a' + b' + c)\alpha \tag{11}$$

where

$$\begin{aligned}a &= (W_{EE\,up}g_E - 1)E_{set}I_{set}g_I \\ a' &= (W_{EE\,up}g_E - 1)E_{set}^2 g_I \\ b &= (W_{II\,up}g_I + 1)E_{set}I_{set}g_E \\ b' &= (W_{II\,up}g_I + 1)I_{set}^2 g_E \\ c &= (I_{set}\Theta_I - E_{set}\Theta_E)g_E g_I\end{aligned}$$

Note that the following is satisfied for a biologically backed set of parameter values:

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

and thus it is likely that $c > 0$. In addition, b and b' are positive definite, and $a, a' > 0$ in the paradoxical regime ($W_{EE}g_E - 1 > 0$). All this makes the stability condition likely satisfied, and thus the plasticity rule stable. Finally, a small enough β would make the condition more likely to hold.

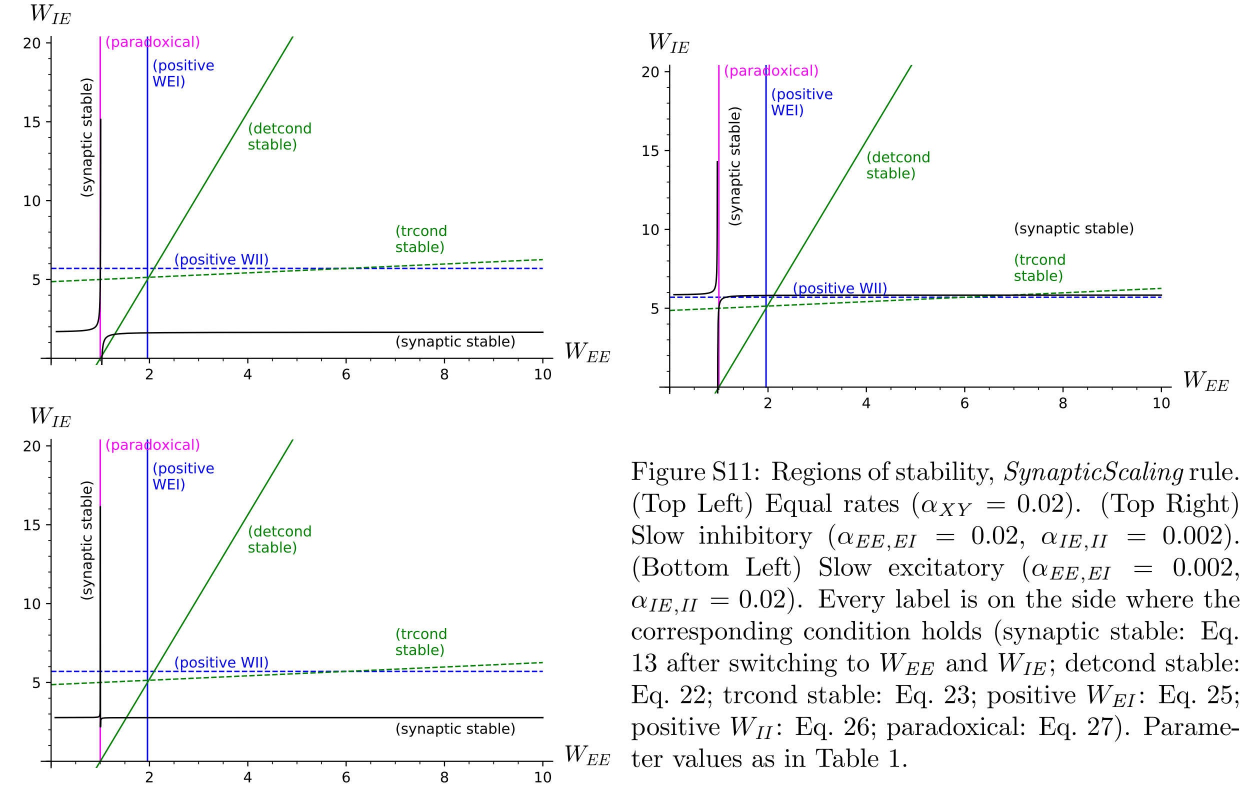


Figure S11: Regions of stability, *SynapticScaling* rule. (Top Left) Equal rates ($\alpha_{XY} = 0.02$). (Top Right) Slow inhibitory ($\alpha_{EE,EI} = 0.02$, $\alpha_{IE,II} = 0.002$). (Bottom Left) Slow excitatory ($\alpha_{EE,EI} = 0.002$, $\alpha_{IE,II} = 0.02$). Every label is on the side where the corresponding condition holds (synaptic stable: Eq. 13 after switching to W_{EE} and W_{IE}; detcond stable: Eq. 22; trcond stable: Eq. 23; positive W_{EI}: Eq. 25; positive W_{II}: Eq. 26; paradoxical: Eq. 27). Parameter values as in Table 1.

See Section 2.4 for a detailed analysis.

1.5 *SynapticScaling* plasticity

The equations for the SynapticScaling plasticity rule in continuous-time dynamics are

$$\begin{aligned}
\frac{dW_{EE}}{dt} &= +\alpha_{EE}(E_{set} - E)W_{EE} \\
\frac{dW_{EI}}{dt} &= -\alpha_{EI}(E_{set} - E)W_{EI} \\
\frac{dW_{IE}}{dt} &= +\alpha_{IE}(I_{set} - I)W_{IE} \\
\frac{dW_{II}}{dt} &= -\alpha_{II}(I_{set} - I)W_{II}
\end{aligned} \tag{12}$$

and the condition for the fixed point to be stable under this rule is

$$(W_{EE\,up}g_E - 1)a < (W_{II\,up}g_I + 1)b \tag{13}$$

where

$$\begin{aligned}
a &= (I_{set}W_{II}\alpha_4 + \Theta_I\alpha_3)g_I \\
b &= E_{set}W_{EE\,up}g_E \\
&\quad + ((W_{EE\,up}g_E - 1)E_{set} - \Theta_E g_E)\alpha_2 \\
&\quad - (W_{EE\,up}g_E - 1)I_{set}\alpha_3
\end{aligned}$$

(where $\alpha_{2,3,4}$ are defined as in previous subsections). This stability condition does not hold for biologically backed parameter values unless the dynamics of the weights onto the inhibitory neuron are slow enough (and in a few fine-tuned cases). To show this, we express the stability condition in terms of the free weights W_{EE} and W_{IE} and plot it with parameter values as in Table 1 and equal rates ($\alpha_{XY} = 0.02$; Figure S11 top left). The stability condition is a homographic function (i.e. a hyperbola) with stability regions in its upper-left and lower-right quadrants—entirely outside the stability region of the neural subsystem. If the dynamics of the weights onto the excitatory neuron are made slower, the homographic function is even steeper (bottom left); if the weights onto the inhibitory neuron are made slower instead, the stability regions switch and overlap with the stability region of the neural subsystem, making the fixed point stable (top right).

It is illustrative to consider the particular case where all learning rates are equal. In this case the stability condition, Eq. 13, doesn't depend on the learning rates and takes the simpler form:

$$\begin{aligned}(W_{IIup}g_I{+}1)a > {} & (W_{EEup}g_E - 1)a' \\ & + (W_{EEup}g_E - 1)(W_{IIup}g_I + 1)b\end{aligned} \tag{14}$$

where

$$\begin{aligned}a &= (E_{set}W_{EEup} - \Theta_E)g_E \\ a' &= (I_{set}W_{IIup} + \Theta_I)g_I \\ b &= I_{set} - E_{set}\end{aligned}$$

Note that in a biologically backed set of parameter values the following is true:

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

This makes $b > 0$ and likely $a' > a$ (in addition, a' is a sum of positive terms while a is a difference). Then in the paradoxical regime ($W_{EE}g_E - 1 > 0$) it seems likely that the stability condition is not satisfied, because the right-hand side is a sum of positive terms and one of them is likely greater than the left-hand side. The SynapticScaling rule is then likely unstable when the learning rates are equal.

A more general case with different learning rates can be analyzed by grouping terms in the following way:

$$\begin{aligned}(I_{set}W_{IIup}\alpha_4 + \Theta_I\alpha_3)g_I(W_{EEup}g_E - 1) \\ <(((W_{EEup}g_E - 1)E_{set} - \Theta_E g_E)\alpha_2 \\ - (W_{EEup}g_E - 1)I_{set}\alpha_3 \\ + E_{set}W_{EEup}g_E)(W_{IIup}g_I + 1)\end{aligned}$$

If $(W_{EE}g_E - 1) > 0$ (paradoxical regime), then decreasing α_3 and/or α_4 (slow dynamics of the weights onto the inhibitory neuron) helps satisfying the condition and thus making the rule stable.

See Section 2.4 for a detailed analysis.

1.6 *ForcedBalance* plasticity

The equations for the ForcedBalance plasticity rule are

$$\begin{aligned}\frac{dW_{EE}}{dt} &= +\alpha_1 g_E E(E_{set} - E) \\ \frac{dW_{EI}}{dt} &= \frac{1}{\tau_0}(W_{EIup} - W_{EI}) \\ \frac{dW_{IE}}{dt} &= +\alpha_3 g_I E(I_{set} - I) \\ \frac{dW_{II}}{dt} &= \frac{1}{\tau_0}(W_{IIup} - W_{II})\end{aligned} \tag{15}$$

and the conditions for the fixed point to be stable under this rule are

$$\begin{aligned}a_1 + b_1(W_{IIup}\, g_I + 1) &< b_1'(W_{EEup}\, g_E - 1) \\ a_2 + b_2(W_{IIup}\, g_I + 1) &< b_2'(W_{EEup}\, g_E - 1)\end{aligned} \tag{16}$$

where

$$\begin{aligned}a_1 &= (I_{set}\Theta_E\Theta_I\,\alpha_1\, g_E g_I + E_{set}^3\alpha_3)\, g_E g_I \\ b_1 &= I_{set}^2\Theta_E\,\alpha_1 g_E^2 g_I - E_{set}^2 I_{set}\,\alpha_1\, g_E^2 \\ b_1' &= E_{set}I_{set}\Theta_I\,\alpha_1\, g_E g_I^2 + E_{set}^2 I_{set}\,\alpha_3\, g_I^2 \\ a_2 &= 2\Theta_E\Theta_I\,\alpha_1\, g_E^2 g_I^2 \\ b_2 &= 2I_{set}\Theta_E\,\alpha_1\, g_E^2 g_I - E_{set}^2\,\alpha_1\, g_E^2 \\ b_2' &= 2E_{set}\Theta_I\,\alpha_1\, g_E g_I^2 + E_{set}^2\,\alpha_3\, g_I^2\end{aligned}$$

In Figure S12 we plot the stability condition of this rule, Eq. 16, for three different parameter values: $\alpha_1 = \alpha_3$, $\alpha_1 \gg \alpha_3$ (inhibitory plasticity slower); and $\alpha_1 \ll \alpha_3$ (excitatory plasticity slower).

In order to decide whether conditions Eq. 16 are satisfied in a more general case, note that b_1 and b_2 on the left-hand side are subtractions whereas b_1' and b_2' on the right-hand side are sums of positive definite terms, which helps satisfying the condition. On the other hand, one of the stability conditions of the neural subsystem might counter the effect: $(W_{IIup}\, g_I + 1)\tau_E > (W_{EEup}\, g_E - 1)\tau_I$ (see Section 2.2 below) but for biologically backed parameter values it is $\tau_E > \tau_I$ thus leaving room for the condition to hold. See Section 2.4 for a detailed analysis.

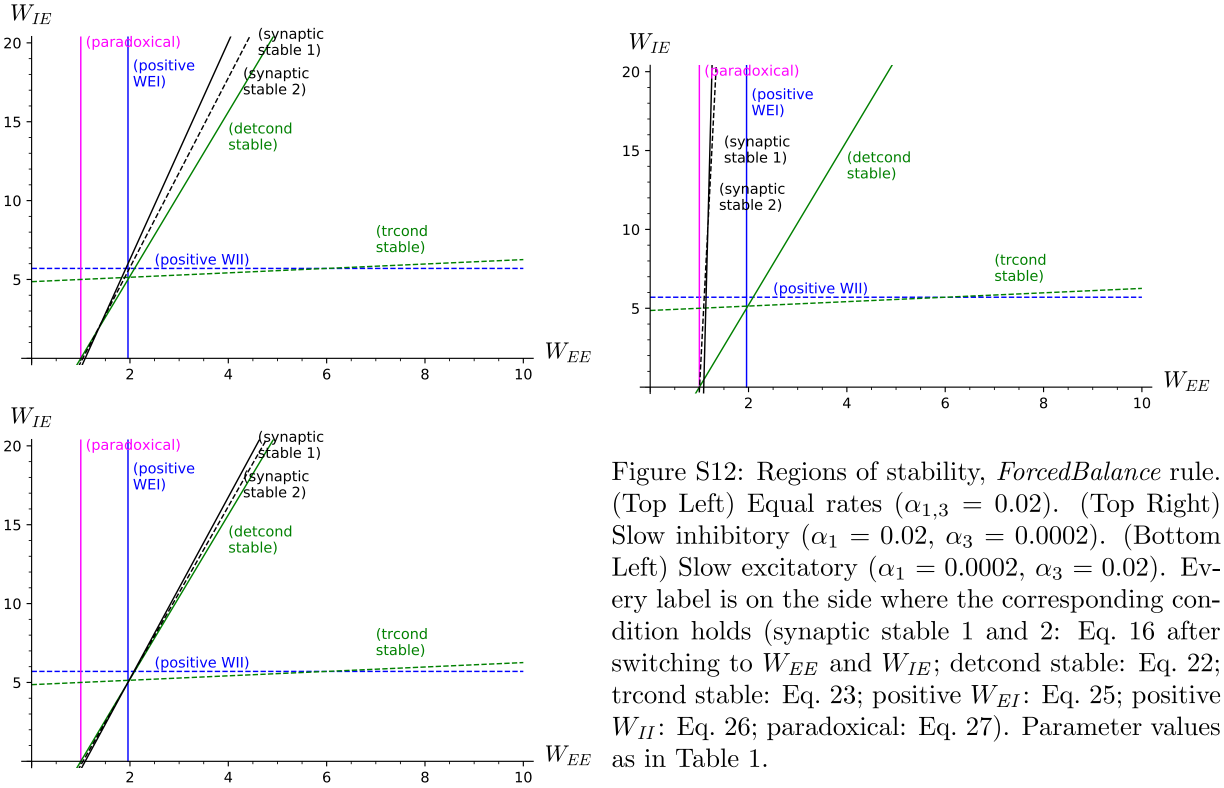


Figure S12: Regions of stability, *ForcedBalance* rule. (Top Left) Equal rates ($\alpha_{1,3} = 0.02$). (Top Right) Slow inhibitory ($\alpha_1 = 0.02$, $\alpha_3 = 0.0002$). (Bottom Left) Slow excitatory ($\alpha_1 = 0.0002$, $\alpha_3 = 0.02$). Every label is on the side where the corresponding condition holds (synaptic stable 1 and 2: Eq. 16 after switching to W_{EE} and W_{IE}; detcond stable: Eq. 22; trcond stable: Eq. 23; positive W_{EI}: Eq. 25; positive W_{II}: Eq. 26; paradoxical: Eq. 27). Parameter values as in Table 1.

2 Detailed calculations

2.1 Overview

We analyze the whole neural+synaptic system for every synaptic plasticity 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 [see Sections 1.2 and 1.4 of Ref. 1][see Section 3.5 of Ref. 2], 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 fixed point; compute the Jacobian of the synaptic subsystem at the fixed point; compute the eigenvalues of the Jacobian [2, 3]. Two out of the four eigenvalues are expected to be zero because the fixed point 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 fixed point is stable under the plasticity rule; if at least one of the nonzero eigenvalues has positive real part, then the fixed point is unstable [2, 3]. (A note on abuse of notation: we might say indis-

tinctly “the fixed point is stable/unstable” and “the plasticity 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 the eigenvalues' critical value in these references is 1 instead of zero because they consider eigenvalues of the connectivity matrix alone, whereas we consider eigenvalues of the whole linear part). As the fixed point 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 ΔW applied every certain time interval) to a continuous-time description (where the weights are continuously evolving albeit with a long time scale τ_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 plasticity rule in detail (following subsections).

Paradoxical regime. In this text we show detailed calculations of the stability conditions for the Homeostatic plasticity in the paradoxical regime only; see Section 2.5 for the non-paradoxical case.

2.2 Neural dynamics

For the neural+synaptic system in the QSS approximation to be stable under a specific synaptic plasticity 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: `upstates-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 Θ_X, with $X = E, I$). The dynamics for synaptic currents above threshold is given by:

$$\begin{aligned} \frac{dE}{dt} &= \frac{1}{\tau_E}(-E + g_E(W_{EE}E - W_{EI}I - \Theta_E)) \\ \frac{dI}{dt} &= \frac{1}{\tau_I}(-I + g_I(W_{IE}E - W_{II}I - \Theta_I)) \end{aligned} \tag{17}$$

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

Neural fixed point. The fixed point of the neural subsystem in the supreathreshold regime is the solution of $dE/dt = dI/dt = 0$:

$$\begin{aligned} E_{up} &= (W_{EI}\, g_I\, \Theta_I - (W_{II}\, g_I + 1)\, \Theta_E)\, g_E/C \\ I_{up} &= ((W_{EE}\, g_E - 1)\, \Theta_I - W_{IE}\, g_E\, \Theta_E)\, g_I/C \end{aligned} \tag{18}$$

where

$$C = W_{EI} W_{IE}\, g_E\, g_I - (W_{II}\, g_I + 1)(W_{EE}\, g_E - 1) \tag{19}$$

We named it with the subscript “up” to distinguish it from the trivial solution “down” where E and I are zero (and the neural subsystem is below threshold).

The activity of the excitatory and inhibitory subpopulations at the nontrivial fixed point, 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 the nontrivial neural fixed point

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. 17) is

$$J = \begin{pmatrix} (W_{EE} g_E - 1)/\tau_E & -W_{EI} g_E/\tau_E \\ W_{IE} g_I/\tau_I & -(W_{II} g_I + 1)/\tau_I \end{pmatrix} \tag{20}$$

Its eigenvalues can be expressed as:

$$\lambda_{1,2} = \frac{1}{2}\left(Tr \pm \sqrt{Tr^2 - 4Det}\right) \tag{21}$$

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 fixed point is stable) when $Det > 0$ and $Tr < 0$, that is:

$$W_{EI} W_{IE} g_E g_I > (W_{EE} g_E - 1)(W_{II} g_I + 1) \tag{22}$$

$$(W_{II} g_I + 1)\tau_E > (W_{EE} g_E - 1)\tau_I \tag{23}$$

Note that the positive determinant condition, Eq. 22, is equivalent to $C > 0$ (Eq. 19).

In the following, we will require that the stability conditions of the neural subsystem, Eqs. 22 and 23, are satisfied.

2.2.3 Weight values consistent with the neural fixed point

The fixed point relationships, Eq. 18, are expressed as the E and I values resulting from a given set of weight values. If we set instead E and I to their target values E_{set} and I_{set} and solve for the weights, we get the weight values that are consistent with a given fixed point activity:

$$\begin{aligned} W_{EIup} &= \frac{(E_{set} W_{EEup} - \Theta_E)\, g_E - E_{set}}{I_{set}\, g_E} \\ W_{IIup} &= \frac{(E_{set} W_{IEup} - \Theta_I)\, g_I - I_{set}}{I_{set}\, g_I} \end{aligned} \tag{24}$$

Note first that any stable plasticity rule for the evolution of the weights for the neural subsystem (Eq. 17) must converge to weight values in accordance with these relationships (either in the form Eq. 24 or Eq. 18).

Second, note that the system is underdetermined and that is why two of the weights are free (chosen to be W_{EE} and W_{EI}). Note also that all weight values must be positive; specifically, requiring $W_{EIup} > 0$ and $W_{IIup} > 0$ leads to

$$W_{EEup} > \frac{\Theta_E\, g_E + E_{set}}{E_{set}\, g_E} \tag{25}$$

$$W_{IEup} > \frac{\Theta_I\, g_I + I_{set}}{E_{set}\, g_I} \tag{26}$$

We refer to these expressions as the "positive W_{EI}" and the "positive W_{II}" conditions, respectively.

2.2.4 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 Θ_I, thus there is a paradoxical effect whenever the coefficient of Θ_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

$$W_{EE}\, g_E - 1 > 0 \tag{27}$$

The paradoxical effect can also be seen in a plot of the fixed point values E_{up} and I_{up} (Eq. 18) 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 S13).

I_{set}	=	14	E_{set}	=	5
g_I	=	4	g_E	=	1
Θ_I	=	25	Θ_E	=	4.8
τ_I	=	2	τ_E	=	10

Table 1: Parameter values throughout the Supplementary Material. This set of parameter values makes the neural subsystem to be in the paradoxical regime (i.e. the fixed point is an inhibition-stabilized fixed point [6]). For non-paradoxical conditions, see Section 2.5.

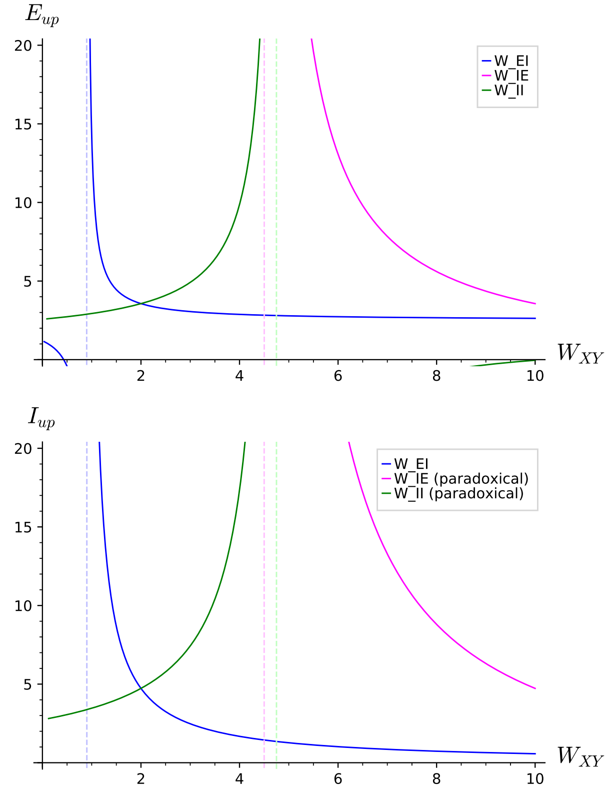


Figure S13: Paradoxical effect in the neural subsystem ($W_{EE} = 5$; parameter values as in Table 1). E_{up} behaves as expected when each weight is varied. I_{up}, however, shows paradoxical behavior when either W_{IE} or W_{II} are varied. Dashed lines are the vertical asymptote of every case.

2.3 *Homeostatic* plasticity: Detailed calculation

In this section we show in detail the calculation of the stability condition for the Homeostatic plasticity rule.

(SageMath code in the Supplementary Material: `upstates-Homeostatic stability.ipynb`)

2.3.1 Definition of the plasticity rule

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

$$\begin{aligned} \frac{dW_{EE}}{dt} &= +\alpha_{EE}\, E(E_{set} - E) \\ \frac{dW_{EI}}{dt} &= -\alpha_{EI}\, I(E_{set} - E) \\ \frac{dW_{IE}}{dt} &= +\alpha_{IE}\, E(I_{set} - I) \\ \frac{dW_{II}}{dt} &= -\alpha_{II}\, I(I_{set} - I) \end{aligned} \tag{28}$$

where α_{XY} $(X, Y = E, I)$ are the learning rates (with appropriate units) setting the time scales 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: it is the set of weight values such that:

$$\begin{aligned} E_{up} &= E_{set} \\ I_{up} &= I_{set} \end{aligned} \tag{29}$$

The values of the weights corresponding to the nontrivial neural fixed point are given by the (underdetermined) system defined by equating Eqs. 29 and 18. 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 values of the other two are given by Eq. 24. This means that the fixed point is actually a continuous set of nonisolated fixed points forming a 2D plane in 4D weight space. In other words, there is an infinite number of weight values compatible with the nontrivial neural fixed point (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, τ, 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. 17 and 28) and choose the values of the scaling parameters such that all nondimensional variables are of order 1 (see attached SageMath code). With this, the full system reads:

$$\begin{aligned}\epsilon_E \frac{de}{d\tau} &= -e + Rew_{EE} - \frac{iw_{EI}}{R} - \theta_E \\ \epsilon_I \frac{di}{d\tau} &= -i + \frac{Rew_{IE}}{g} - \frac{iw_{II}}{Rg} - \theta_I \\ \frac{dw_{EE}}{d\tau} &= -e(e-1) \\ \frac{dw_{EI}}{d\tau} &= +\alpha_2 i(e-1) \\ \frac{dw_{IE}}{d\tau} &= -\alpha_3 e(i-1) \\ \frac{dw_{II}}{d\tau} &= +\alpha_4 i(i-1)\end{aligned} \tag{30}$$

where we defined the new parameters

$$\begin{aligned}\epsilon_E &= \tau_E/\tau_0 \\ \epsilon_I &= \tau_I/\tau_0 \\ \tau_0 &= 1/(\alpha g_E E_{set} I_{set}) \\ R &= E_{set}/I_{set} \\ g &= g_E/g_I \\ \alpha_2 &= \alpha_{EI}/\alpha_{EE} \\ \alpha_3 &= \alpha_{IE}/\alpha_{EE} \\ \alpha_4 &= \alpha_{II}/\alpha_{EE} \\ \theta_E &= (g_E/E_{set})\Theta_E \\ \theta_I &= (g_I/I_{set})\Theta_I\end{aligned}$$

2.3.3 Quasi-steady state approximation

Neural dynamics evolves in a much shorter time scale (τ_E and τ_I) than synaptic dynamics (τ_0):

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

which implies

$$\begin{aligned}\epsilon_E \frac{de}{d\tau} &\sim 0 \\ \epsilon_I \frac{di}{d\tau} &\sim 0\end{aligned} \tag{31}$$

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 plasticity rule. They are determined by applying Eq. 31 to the first two rows of Eq. 30; solving for e and i leads to

$$\begin{aligned}e_{qss} &= (g\theta_I w_{EI} - (w_{II} + Rg)\theta_E)/c \\ i_{qss} &= (Rg\theta_I(Rw_{EE} - 1) - R^2\theta_E w_{IE})/c\end{aligned} \tag{32}$$

where

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

The full system in the QSS approximation reads

$$\begin{aligned}\frac{dw_{EE}}{d\tau} &= -e_{qss}(e_{qss} - 1) \\ \frac{dw_{EI}}{d\tau} &= +\alpha_2 i_{qss}(e_{qss} - 1) \\ \frac{dw_{IE}}{d\tau} &= -\alpha_3 e_{qss}(i_{qss} - 1) \\ \frac{dw_{II}}{d\tau} &= +\alpha_4 i_{qss}(i_{qss} - 1)\end{aligned} \tag{33}$$

where e_{qss} and i_{qss} are nonlinear functions of the weights as defined by Eq. 32.

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

$$\begin{aligned}e_{qss} &= 1 \\ i_{qss} &= 1\end{aligned} \tag{34}$$

which is the nondimensionalized version of Eq. 29. The weight values compatible with this condition are defined by equating Eqs. 32 and 34:

$$\begin{aligned}w_{EIup} &= R(Rw_{EEup} - 1) - R\theta_E \\ w_{IIup} &= R(Rw_{IEup} - g) - Rg\theta_I\end{aligned} \tag{35}$$

(w_{EEup} and w_{IEup} are free). This is the nondimensionalized version of Eq. 24.

2.3.4 Stability condition

The program for assessing linear stability of the fixed point is as follows: a) compute the Jacobian (the matrix of first derivatives) of Eq. 33 and evaluate it at the fixed point; 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 fixed point is stable; if at least one of the nonzero eigenvalue has positive real part then the fixed point is unstable.

Jacobian matrix. Let the full system in the QSS approximation (Eq. 33) 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. 32. By applying the chain rule the elements J_{ij} $(i, j = 1\ldots4)$ 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 fixed point, these expressions are to be substituted by Eqs. 32-35.

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

$$\lambda_{\pm} = \frac{A \pm \sqrt{A^2 - DC}}{C} \tag{36}$$

where

$$\begin{aligned}A &= R^2 g\theta_I + (R^2\alpha_3 + \alpha_4)Rw_{EEup}\\ &\quad - (R^2 + \alpha_2)Rw_{IEup} + \alpha_2 g\theta_I - R^2\alpha_3 - \alpha_4\\ C &= 2R(Rg\theta_I w_{EEup} - R\theta_E w_{IEup} - g\theta_I)\\ D &= 2(R^2\alpha_3 + \alpha_4)(R^2 + \alpha_2)/R\end{aligned} \tag{37}$$

Sign of the eigenvalues. To determine the sign of the real part of Eq. 36, first note that the factor D is positive definite. Second, C must be positive because it is related to one of the stability conditions of the neural subsystem (Eq. 22, 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 plasticity rule is then stable (both eigenvalues have negative real part) if $A < 0$, which in terms of the original parameters and free weights W_{EE} and W_{IE} reads:

$$\begin{aligned}&(E_{set}^2\alpha_{IE} + I_{set}^2\alpha_{II})I_{set}(W_{EEup}g_E - 1) <\\ &(E_{set}^2\alpha_{EE} + I_{set}^2\alpha_{EI})(E_{set}W_{IEup}g_E - \Theta_I g_E)\end{aligned} \tag{38}$$

2.3.5 Analysis of the stability condition

It is hard to determine whether the stability condition Eq. 38 is satisfied for a general set of parameter values (see numerical analysis below). However, by using the fixed point relationship Eq. 24, this condition can be re-expressed in a more useful form in terms of W_{EE} and W_{II}:

$$\begin{aligned}(R^2\alpha_3{+}\alpha_4)&(W_{EEup}\, g_E - 1)g_I\\ &< (R^2 + \alpha_2)(W_{IIup}\, g_I + 1)g_E\end{aligned} \tag{39}$$

Note that learning rates values of the same order lead to $\alpha_{2,3,4} \sim 1$ and that biologically backed parameter values satisfy:

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

both likely preventing the condition to hold.

On the other hand, small enough values of α_3 and α_4 (by making the dynamics of the weights onto the inhibitory neuron W_{IE} and W_{II} slower) would help satisfy the condition thus making the system stable.

2.3.6 Relationship between the synaptic stability and the paradoxical condition

The boundary of the stability condition for this plasticity rule, Eq. 38, is a linear function in the (W_{EE}, W_{IE}) space with a slope that tends to infinity as the excitatory learning rates $(\alpha_{EE,EI})$ tend to zero:

$$\text{slope} = \frac{(E_{set}^2\alpha_{IE} + I_{set}^2\alpha_{II})I_{set}}{(E_{set}^2\alpha_{EE} + I_{set}^2\alpha_{EI})E_{set}}$$

while its root is a complicated expression (see SageMath notebook) that tends to $W_{EE} = 1/g_E$. The region of stability is to the left of the line. Thus, the boundary of stability in this limit coincides exactly with the boundary of the paradoxical condition $(W_{EE} > 1/g_E)$. This can be construed as an inconsistency/contradiction between the stability of the rule and the existence of the paradoxical effect.

2.4 Detailed calculations for the other rules

The stability calculations for the rest of the rules follow very similar paths. They can be found in the corresponding SageMath-Jupyter notebooks:

`upstates-CrossHomeostatic stability.ipynb`

`upstates-TwoTerm stability.ipynb`

`upstates-SynapticScaling stability.ipynb`

`upstates-ForcedBalance stability.ipynb`

2.5 Stability of the rules in a non-paradoxical regime

All results above were developed with the neural subsystem set in the paradoxical regime—that is, the region in (W_{EE}, W_{IE}) leading to a stable fixed point was completely within the paradoxical region $(W_{EE}g_E > 1)$. In order to show the importance of the paradoxical behavior for the stability of the plasticity rules, we also computed the stability conditions of every plasticity rule in a more general setting where the excitatory subpopulation in the neural subsystem has an external, constant, excitatory input current I_{ext}. This allows the neural subsystem to display both paradoxical and non-paradoxical stable behavior (in the second case, at the expense of the fixed point not being an inhibition-stabilized fixed point; see `upstates-Neural subsystem stability-with Iext.ipynb`).

2.5.1 *Homeostatic* with I_{ext}

The stability condition doesn't depend on I_{ext} and it reads the same as Eq. 2:

$$\begin{aligned}(E_{set}^2\alpha_{IE} + I_{set}^2\alpha_{II})I_{set}(W_{EEup}g_E - 1) <\\ (E_{set}^2\alpha_{EE} + I_{set}^2\alpha_{EI})(E_{set}W_{IEup}g_E - \Theta_I g_E)\end{aligned} \tag{40}$$

(SageMath code in `upstates-Homeostatic stability-with Iext.ipynb`)

2.5.2 *CrossHomeostatic* with I_{ext}

The stability condition with I_{ext} is:

$$\begin{aligned}&(E_{set}^2\alpha_{EE} + I_{set}^2\alpha_{EI})I_{set}W_{IEup}g_E\\ &\quad > -(E_{set}^2\alpha_{IE} + I_{set}^2\alpha_{II})\\ &\qquad ((W_{EEup}g_E - 1)E_{set} - (\Theta_E - I_{ext})g_E)\end{aligned} \tag{41}$$

which is very similar to Eq. 7 except that it has $(\Theta_E - I_{ext})$ instead of just Θ_E. From this it should be evident that the condition will still hold for any positive value of I_{ext} (right-hand side decreases).

The validity of the condition can also be seen after switching to W_{IE} and W_{EI}, leading to exactly the same condition as Eq. 8:

$$(R^2\alpha_3 + \alpha_4)W_{EIup} + (R^2 + \alpha_2)W_{IEup} > 0 \tag{42}$$

which holds for any value of I_{ext}.

(SageMath code in `upstates-CrossHomeostatic stability-with Iext.ipynb`)

2.5.3 ***TwoTerm*** with I_{ext}

The stability condition with I_{ext} is:

$$\begin{aligned}(I_{set}\alpha + E_{set}\beta)W_{IEup}g_E& \\ &> (I_{set}\beta - E_{set}\alpha)W_{EEup}g_E \\ &+ ((\Theta_E - I_{ext})g_E + E_{set})\alpha + (\Theta_I g_E - I_{set})\beta\end{aligned} \tag{43}$$

which is very similar to Eq. 10 except that it has $(\Theta_E - I_{ext})$ instead of just Θ_E. From this it should be evident that the larger the value of I_{ext} (right-hand side decreases) the larger the stability region.

(SageMath code in `upstates-TwoTerm stability-with Iext.ipynb`)

2.5.4 ***SynapticScaling*** with I_{ext}

When I_{ext} is included in the dynamics of E, the stability condition for the rule reads:

$$(W_{EEup}g_E - 1)a < (W_{IIup}g_I + 1)b \tag{44}$$

where

$$\begin{aligned}a &= (I_{set}W_{II}\alpha_4 + \Theta_I\alpha_3)g_I \\ b &= E_{set}W_{EEup}g_E \\ &\quad + ((W_{EEup}g_E - 1)E_{set} - (\Theta_E - I_{ext})g_E)\alpha_2 \\ &\quad - (W_{EEup}g_E - 1)I_{set}\alpha_3\end{aligned}$$

which is very similar to Eq. 13 except that it has $(\Theta_E - I_{ext})$ instead of just Θ_E. From this it should be evident that including a positive I_{ext} will increase the chances that the condition holds (right-hand side increases).

(SageMath code in `upstates-SynapticScaling stability-with Iext.ipynb`)

2.5.5 ***ForcedBalance*** with I_{ext}

The stability conditions when I_{ext} is included in the neural subsystem are:

$$\begin{aligned}a_1 + b_1(W_{IIup}\, g_I + 1) &< b_1'(W_{EEup}\, g_E - 1) \\ a_2 + b_2(W_{IIup}\, g_I + 1) &< b_2'(W_{EEup}\, g_E - 1)\end{aligned} \tag{45}$$

where

$$\begin{aligned}a_1 &= (I_{set}(\Theta_E - I_{ext})\Theta_I\,\alpha_1\, g_E g_I + E_{set}^3\alpha_3)\, g_E g_I \\ b_1 &= I_{set}^2(\Theta_E - I_{ext})\,\alpha_1 g_E^2 g_I - E_{set}^2 I_{set}\,\alpha_1\, g_E^2 \\ b_1' &= E_{set}I_{set}\Theta_I\,\alpha_1\, g_E g_I^2 + E_{set}^2 I_{set}\,\alpha_3\, g_I^2 \\ a_2 &= 2(\Theta_E - I_{ext})\Theta_I\,\alpha_1\, g_E^2 g_I^2 \\ b_2 &= 2I_{set}(\Theta_E - I_{ext})\,\alpha_1\, g_E^2 g_I - E_{set}^2\,\alpha_1\, g_E^2 \\ b_2' &= 2E_{set}\Theta_I\,\alpha_1\, g_E g_I^2 + E_{set}^2\,\alpha_3\, g_I^2\end{aligned}$$

which are very similar to Eqs. 16 except that there is a $(\Theta_E - I_{ext})$ instead of just Θ_E. From this it should be evident that the larger the value of I_{ext} (left-hand side decreases) the larger the stability region.

(SageMath code in `upstates-ForcedBalance stability-with Iext.ipynb`)

3 Derivation of a plasticity rule from a loss function

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

Here we show how to compute a plasticity rule for the weights starting from a loss function. Then we make an approximation by considering that the weight values are close to the values corresponding to the fixed point.

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

$$\begin{aligned}E &= E_{up}(W_{EE}, W_{EI}, W_{IE}, W_{II}) \\ I &= I_{up}(W_{EE}, W_{EI}, W_{IE}, W_{II})\end{aligned} \tag{46}$$

where the functions E_{up} and I_{up} are defined by Eq. 18 (see [7] for a related discussion on quasi-steady state, synaptic plasticity, and gradient descent).

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

   $$\Delta\mathbf{W} = -\alpha\nabla L \tag{47}$$

   with a single learning rate α for simplicity. The unfolded plasticity rules, that is the equations that govern the weights' dynamics, are then

   $$\begin{aligned} \Delta W_{EE} &= -\alpha\frac{\partial L}{\partial W_{EE}} \\ \Delta W_{EI} &= -\alpha\frac{\partial L}{\partial W_{EI}} \\ \Delta W_{IE} &= -\alpha\frac{\partial L}{\partial W_{IE}} \\ \Delta W_{II} &= -\alpha\frac{\partial L}{\partial W_{II}} \end{aligned} \tag{48}$$

3. The partial derivatives of the loss function in Eq. 48 are:

   $$\begin{aligned} \frac{\partial L}{\partial W_{EE}} &= \frac{\partial L}{\partial E}\frac{\partial E}{\partial W_{EE}} + \frac{\partial L}{\partial I}\frac{\partial I}{\partial W_{EE}} \\ \frac{\partial L}{\partial W_{EI}} &= \frac{\partial L}{\partial E}\frac{\partial E}{\partial W_{EI}} + \frac{\partial L}{\partial I}\frac{\partial I}{\partial W_{EI}} \\ \frac{\partial L}{\partial W_{IE}} &= \frac{\partial L}{\partial E}\frac{\partial E}{\partial W_{IE}} + \frac{\partial L}{\partial I}\frac{\partial I}{\partial W_{IE}} \\ \frac{\partial L}{\partial W_{II}} &= \frac{\partial L}{\partial E}\frac{\partial E}{\partial W_{II}} + \frac{\partial L}{\partial I}\frac{\partial I}{\partial W_{II}} \end{aligned} \tag{49}$$

   or, in vector notation:

   $$\nabla L = \frac{\partial L}{\partial E}\nabla E + \frac{\partial L}{\partial I}\nabla I \tag{50}$$

   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 $\nabla E = \left(\frac{\partial E}{\partial W_{EE}}, \ldots\right)$ and $\nabla I = \left(\frac{\partial I}{\partial W_{EE}}, \ldots\right)$ etc. are to be taken from the quasi-steady-state values of E and I, Eq. 46. We will, however, compute the partial derivatives from the implicit expressions given by setting $dE/dt = dI/dt = 0$ in Eq. 17 without solving for E and I.

3.2 Detailed calculation

3.2.1 Exact plasticity rules

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

$$L(E, I) = \frac{1}{2}(E_{set} - E)^2 + \frac{1}{2}(I_{set} - I)^2 \tag{51}$$

This loss function is an elliptic paraboloid in (E, I) space with a global minimum at (E_{set}, I_{set}) so a gradient descend working on E and I 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 (particularly for a paradoxical system), 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

$$\begin{aligned} \frac{\partial L}{\partial E} &= -(E_{set} - E) \\ \frac{\partial L}{\partial I} &= -(I_{set} - I) \end{aligned} \tag{52}$$

Partial derivatives of E and I. We compute the partial derivatives $\partial X/\partial W_{XY}$ $(X, Y = E, I)$ by first equating the neural subsystem (Eq. 17) to zero:

$$\begin{aligned} E &= g_E(W_{EE}E - W_{EI}I - \Theta_E) \\ I &= g_I(W_{IE}E - W_{II}I - \Theta_I) \end{aligned} \tag{53}$$

then differentiating the implicit functions:

$$\begin{aligned}
\frac{\partial E}{\partial W_{EE}} &= g_E(E + W_{EE}\frac{\partial E}{\partial W_{EE}}) - g_E W_{EI}\frac{\partial I}{\partial W_{EE}} \\
\frac{\partial E}{\partial W_{EI}} &= g_E W_{EE}\frac{\partial E}{\partial W_{EI}} - g_E(I + W_{EI}\frac{\partial I}{\partial W_{EI}}) \\
\frac{\partial E}{\partial W_{IE}} &= g_E W_{EE}\frac{\partial E}{\partial W_{IE}} - g_E W_{EI}\frac{\partial I}{\partial W_{IE}} \\
\frac{\partial E}{\partial W_{II}} &= g_E W_{EE}\frac{\partial E}{\partial W_{II}} - g_E W_{EI}\frac{\partial I}{\partial W_{II}} \\
\frac{\partial I}{\partial W_{EE}} &= g_I W_{IE}\frac{\partial E}{\partial W_{EE}} - g_I W_{II}\frac{\partial I}{\partial W_{EE}} \\
\frac{\partial I}{\partial W_{EI}} &= g_I W_{IE}\frac{\partial E}{\partial W_{EI}} - g_I W_{II}\frac{\partial I}{\partial W_{EI}} \\
\frac{\partial I}{\partial W_{IE}} &= g_I(E + W_{IE}\frac{\partial E}{\partial W_{IE}}) - g_I W_{II}\frac{\partial I}{\partial W_{IE}} \\
\frac{\partial I}{\partial W_{II}} &= g_I W_{IE}\frac{\partial E}{\partial W_{II}} - g_I(I + W_{II}\frac{\partial I}{\partial W_{II}})
\end{aligned} \tag{54}$$

and then solving for the derivatives:

$$\begin{aligned}
\frac{\partial E}{\partial W_{EE}} &= -(EW_{II}\, g_E\, g_I + Eg_E)/C \\
\frac{\partial E}{\partial W_{EI}} &= (IW_{II}\, g_E\, g_I + Ig_E)/C \\
\frac{\partial E}{\partial W_{IE}} &= EW_{EI}\, g_E\, g_I/C \\
\frac{\partial E}{\partial W_{II}} &= -IW_{EI}\, g_E\, g_I/C \\
\frac{\partial I}{\partial W_{EE}} &= -EW_{IE}\, g_E\, g_I/C \\
\frac{\partial I}{\partial W_{EI}} &= IW_{IE}\, g_E\, g_I/C \\
\frac{\partial I}{\partial W_{IE}} &= (EW_{EE}\, g_E - E)g_I/C \\
\frac{\partial I}{\partial W_{II}} &= -(IW_{EE}\, g_E - I)g_I/C
\end{aligned} \tag{55}$$

where

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

Exact plasticity rules. Putting everything together, the plasticity rules Eq. 48 are:

$$\begin{aligned}
\Delta W_{EE} &= -\frac{\alpha}{C}((I_{set} - I)EW_{IE}\, g_e\, g_I \\
&\qquad + (E_{set} - E)E(W_{II}\, g_I + 1)g_E) \\
\Delta W_{EI} &= +\frac{\alpha}{C}((I_{set} - I)IW_{IE}\, g_e\, g_I \\
&\qquad + (E_{set} - E)I(W_{II}\, g_I + 1)g_E) \\
\Delta W_{IE} &= +\frac{\alpha}{C}((E_{set} - E)EW_{EI}\, g_e\, g_I \\
&\qquad + (I_{set} - I)E(W_{EE}\, g_E - 1)g_I) \\
\Delta W_{II} &= -\frac{\alpha}{C}((E_{set} - E)IW_{EI}\, g_e\, g_I \\
&\qquad + (I_{set} - I)I(W_{EE}\, g_E - 1)g_I)
\end{aligned} \tag{56}$$

Note that these are very complicated, nonlinear expressions because both E and I depend on all weights via Eq. 53. Also the denominator C depends on all weights (see previous paragraph).

3.2.2 Approximation

We want simpler expressions for the plasticity rules. Note that the exact expressions above all have a homeostatic factor (either $E - E_{set}$ or $I - I_{set}$) and a presynaptic factor (either E or I), while the rest are complicated expressions coming from the derivatives $\partial E/\partial W_{XY}$ and $\partial I/\partial W_{XY}$. We want to keep the homeostatic and presynaptic factors as they are while simplifying the rest of the expressions (explicit dependence on the weights including C) by performing a lowest-order Taylor series expansion of the explicit dependence of Eqs. 55 on the weights. Although this is not a textbook Taylor expansion of the full expressions, it is very informative because the results can be much more easily interpreted (for a similar approach see [7]).

We perform a zeroth-order approximation of the derivatives $\partial E/\partial W_{XY}$ and $\partial I/\partial W_{XY}$ as functions of the weights (i.e. while holding the presynaptic factors E and I constant) around the fixed point. In this approximation the weights are not small but close to their target values, represented by the relationships Eq. 24. By substituting the result in Eq. 49, we get the following approximated plasticity rules:

$$\begin{aligned}\Delta W_{EE} &= +\alpha_E E(I_{set} - I) + \beta_E E(E_{set} - E)\\ \Delta W_{EI} &= -\alpha_E I(I_{set} - I) - \beta_E I(E_{set} - E)\\ \Delta W_{IE} &= -\alpha_I E(E_{set} - E) + \beta_I E(I_{set} - I)\\ \Delta W_{II} &= +\alpha_I I(E_{set} - E) - \beta_I I(I_{set} - I)\end{aligned} \tag{57}$$

where

$$\begin{aligned}\alpha_E &= \alpha g_E E_{set} W_{IEup}/D\\ \alpha_I &= \alpha A/D\\ \beta_E &= \alpha g_E B/D\\ \beta_I &= \alpha I_{set}(1 - W_{EEup} g_E)/D\end{aligned}$$

and

$$\begin{aligned}A &= E_{set} W_{EEup} g_E - \Theta_E g_E - E_{set}\\ B &= E_{set} W_{IEup} - \Theta_I\\ D &= \Theta_I W_{EEup} g_E - \Theta_E W_{IEup} g_E - \Theta_I\end{aligned}$$

Analysis. Note that α_E, α_I, β_E, and β_I are all constant. Furthermore, note that

- $A > 0$ as it is equal to the "positive W_{EI}" condition, Eq. 25;
- $B > 0$ as it is part of the "positive W_{II}" condition, Eq. 26;
- $D > 0$ as it is equal to the numerator of I_{up}, Eq. 18 (up to a positive factor), which must be positive because the denominator is.

Interestingly, note that the learning rate β_I can be either negative or positive depending on whether the fixed point where the dynamics is converging to is paradoxical ($W_{EEup} g_E - 1 > 0$) or not ($W_{EEup} g_E - 1 < 0$).
Note that the terms with $\alpha_{E,I}$ in the approximated plasticity rules, Eq. 57, are exactly equal to the Cross-Homeostatic rules, Eq. 6. Additionaly, the terms with $\beta_{E,I}$ are exactly equal to the Homeostatic rules, Eq. 28, unless $\beta_I < 0$ which would make the plasticity rule a Cross-Homeo-antiHomeo hybrid.

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. Sadeh, S. & Clopath, C. Inhibitory stabilization and cortical computation. *Nature Reviews Neuroscience* **22,** 21–37 (2021).
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).