bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Complementary inhibitory receptive fields emerge from synaptic plasticity and create an attentional switch in sensory circuits

Everton J. Agnes,* **Andrea I. Luppi**,† and **Tim P. Vogels**

Centre for Neural Circuits and Behaviour, University of Oxford, OX1 3SR Oxford, United Kingdom

Cortical areas comprise multiple types of inhibitory interneurons with stereotypical connectivity motifs, but the combined effect of different inhibitory connectivity patterns on postsynaptic dynamics has been largely unexplored. Here, we analyse the response of a single postsynaptic neuron receiving tuned excitatory connections that are balanced by various combinations of inhibitory input profiles. Inhibitory tuning can be flat, share the same tuning preference as the excitation or, alternatively, it can feature counter-tuning such that non-preferred excitatory inputs receive large inhibition. When all inhibitory populations are active, the net inhibitory effect is the same regardless of the tuning profile. By modulating the activity of specific inhibitory populations, strongly correlated responses to preferred or non-preferred inputs, as well as uncorrelated responses emerge. Moreover, biologically inspired inhibitory plasticity rules produce the necessary connectivity profiles, indicating how plasticity rules in various cell types can interact to shape cortical circuit motifs and their dynamics.

Inhibitory neurons exhibit large variability in morphology, connectivity motifs, and electrophysiological properties[1–4]. One of inhibition's main function is to balance excitatory inputs, thus stabilising neuronal network activity[5,6], and allowing for a range of different functions to be implemented by the brain[7–11]. When both inhibitory and excitatory inputs share the same statistics, and their weight profiles (receptive fields) are similar[12], the resulting state of the postsynaptic neuron is one of precise balance of input currents[9]. Modulation of inhibition, e.g., decrease or increase in local inhibitory activity and, consequently, change in the balance between excitation and inhibition, is essential to control the activity of neuronal groups[13]. It is thus believed that disinhibition is an important mechanism for the implementation of high-level brain functions, such as attention[14,15], memory retrieval[6,16,17], signal gating[18,19], and rapid learning[11]. This type of modulation cannot be easily implemented, and may hint at why neuronal circuit motifs have so many interneuron types[1,2,7].

To reach a state of precise balance, a Hebbian-like inhibitory plasticity rule — increase in synaptic weights for correlated pre- and postsynaptic activity, as observed, e.g., in auditory cortex[20] — is required[6]. A form of anti-Hebbian inhibitory plasticity — decrease of synaptic weights for correlated pre- and postsynaptic activity — has also been reported[21–23], and such a rule has been proposed as a mechanism for memory formation and retrieval[24]. These two types of plasticity rules would form opposite receptive fields for inhibitory connections: synapses following a Hebbian-like inhibitory plasticity rule would mirror excitatory inputs while an anti-Hebbian plasticity rule would impose strong inhibitory inputs for weak excitatory ones, and vice-versa. In line with this hypothesis, intracellular recordings indicate that strong inhibitory postsynaptic potentials can be elicited by stimuli with preferred orientations[25,26], but also by stimuli with non-preferred orientations[27,28]. Additionally, postsynaptic responses in auditory cortex of mice have been shown to vary with sound intensity[29].

To understand the mechanistic origin of such varying responses from the same cells, we investigated the behaviour of a single postsynaptic neuron model receiving tuned excitatory inputs, and inhibition from two distinctly tuned populations. Tuning may correspond to preference to a specific sound frequency[12], orientation of visual cues[30], taste[31], whisker stimulation[32], or position in space[33]. We show that when the postsynaptic neuron is in a balanced state with respect to its excitatory and inhibitory inputs, preferred signals are transiently revealed, but steady state responses are indiscriminate of the stimulus preference[6] (i.e., its 'orientation', etc.), regardless of the inhibitory connectivity. We could substantially alter the response profile of the postsynaptic neuron by modulating the activity of either of the two presynaptic inhibitory populations, allowing for the propagation of facets of the input patterns that were previously quenched by inhibition. We thus introduce a mechanism to selectively filter stimuli according to, e.g., attentional cues. We show that the tuning profiles necessary for such attentional gating can be achieved by a set of biologically plausible plasticity rules. These rules are based on previous theoretical and experimental work, suggesting that distinct plasticity rules can harmoniously coexist in the brain. Our work proposes a simple biological implementation for an attentional switch at a mechanistic level, and provides a solution for how such a neuronal circuit can emerge with autonomous and unsupervised, biologically plausible plasticity rules.

Results

To study the effect of interacting populations of feedforward inhibition, we investigated the response of a single postsynaptic leaky integrate-and-fire neuron receiving tuned excitatory and inhibitory inputs. Excitatory inputs were organised into a single population, subdivided into 16 signal groups of 200 excitatory afferents. Inhibitory inputs initially formed a single population, mirroring the excitatory subdivision, but with 50 afferents per group. Later, we split the inhibitory inputs into two populations with 25 afferents per signal group (Fig. 1A, see Methods for details), allowing us to obtain two differently tuned populations (presumably *types*) of inhibition. Excitatory and inhibitory afferents belonging to the same group shared temporal fluctuations in firing rates, termed input patterns, even if they belonged to different populations. In our simulations, input patterns could either be *natural* or *pulse*. Natural inputs were generated through an inhomogeneous Poisson process based on a modified Ornstein-Uhlenbeck process (Fig. 1B,C), such that neurons of the same signal group also had temporally-correlated firing patterns (Fig. 1C, top; see also Ujfalussy *et al.*[34]). The resulting long-tail distribution of interspike-intervals (Fig. 1C, bottom) was similar to experimentally

* Correspondence: everton.agnes@gmail.com
† Present address: Department of Psychology, Harvard University, Cambridge MA, USA
8th August 2019

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

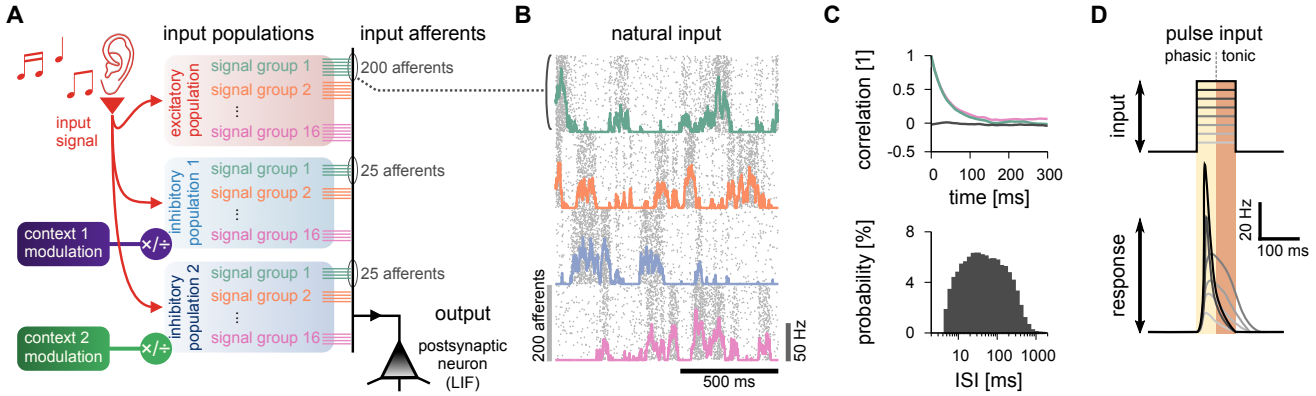

FIG. 1. Synaptic inputs. A, Schematic of the input organisation. An external signal (representing, e.g., sound) was delivered through three input populations (one excitatory and two inhibitory), with 16 input signals per population (representing, e.g., sound frequency). Each signal was simulated by 250 independent, but temporally correlated, spike trains (input afferents); 200 excitatory, and 50 inhibitory divided into two groups of 25. One postsynaptic neuron (black triangle) was the output of this system, simulated as a single-compartment leaky integrate-and-fire neuron (LIF). The firing rate of each of the inhibitory populations was modulated by a contextual cue (green and purple boxes). Excitatory and inhibitory input spike trains were generated as point processes (see Methods for details). **B**, Natural input statistics. Raster plot (grey dots) of 800 neurons that take part in 4 signal groups (200 neurons per signal group), each with firing-rate changing according to a modified Ornstein-Uhlenbeck process (coloured lines; Methods). **C**, Temporal autocorrelation (top) and distribution of the inter-spike intervals (ISI; bottom) of the pre-synaptic inputs. The autocorrelation of two groups are shown (green and pink), as well as the correlation between two different groups (black). Autocorrelation is computed as the Pearson coefficient with a delay (x-axis; Methods). **D**, Pulse input schematic. A step-like increase in the firing rate of a given input group lasting 100 ms with varying firing rates (grey scale). The postsynaptic response is separated in *phasic* (first 50 ms), and *tonic* (last 50 ms).

observed spike patterns *in vivo*[35,36]. We used this type of input to quantify steady-state (average) postsynaptic responses, and to train inhibitory synapses via plasticity rules.

In the alternative pulse input regime we analysed transient responses with 100-ms long pulses of varying amplitudes[6]. Pulses were delivered through a single signal group of excitatory and inhibitory afferents, while all other groups remained at baseline firing-rate (Methods). Responses were quantified according to postsynaptic firing rates during the first (phasic) and last (tonic) 50 ms stimulation (Fig. 1D), averaged over 100 trials. Separating responses in phasic and tonic allowed us to discriminate changes in output due to the input onset, and slower integration of the pulse, respectively.

Modulating a single inhibitory population. At first, we constructed a standard cortical circuit motif with one excitatory and one inhibitory population[6,37] (Fig. 2A, top). Both excitatory and inhibitory weights were tuned according to a receptive-field-like profile, providing precise balance between excitation and inhibition (Fig. 2A, bottom) and average post-synaptic firing rates of 5 Hz for natural inputs (Fig. 2B, left; Fig. 2C, control). We then changed the gain of all inhibitory afferents by modulating their firing rates, from 50% to 150% of control rates. This change of input balance translated into changes in output rates (Fig. 2C, bottom), and spike patterns (Fig. 2B, middle and right). When inhibition was equal or larger than excitation, the output was largely uncorrelated to any given input signal (Fig. 2D, top). When inhibitory firing rates fell below 90% of the control condition, the output first began to correlate with the preferred input signal. When inhibition became even weaker, the correlations increased, and even non-preferred signals were articulated in the postsynaptic firing patterns (Fig. 2D, bottom).

Transient presynaptic activity pulses caused strong phasic responses in the balance state when they were delivered through the afferents of the preferred inputs (Fig. 2E, top row). Stimuli from non-preferred afferents were largely ignored. This discriminability between transients of low or high amplitude pulses decreased when inhibition was down-regulated (Fig. 2E, middle row) such that pulse stimuli from all signal groups caused a response. Increased inhibition, on the other hand, completely abolished transient responses to non-preferred afferents (Fig. 2E, bottom row). In all three cases (balanced control, weak and strong inhibition), the postsynaptic neuron elicited most of its spikes within the phasic period of the total 100 ms input step (Fig. 2E).

Co-tuned and flat inhibitory populations. Next we incorporated two types of inhibition into the circuit motif (Fig. 3A, top). One population mirrored the synaptic weight profile of the excitation. We tuned the inhibitory weights of this *co-tuned population* so that the output weights of non-preferred signal populations (far from the peak of the receptive field) decreased to zero. The other inhibitory population featured an un-tuned response profile (we termed this the *flat population*; Fig. 3A, bottom). We ensured that the sum of the inhibitory weights for each signal group balanced the excitatory inputs by default (Methods). We tested three conditions: both inhibitory populations active (control); co-tuned population inactive; and flat population inactive (Fig. 3B-E). To maintain the same average output firing rate of 5 Hz in the modulated conditions, we increased the activity of the remaining active inhibitory population (Fig. 3C).

When both populations were active, the input-output correlation remained unchanged (Fig. 3D, top), because the two populations (co-tuned and flat) were mimicking the effect of the single (co-tuned) population. Deactivating either population (while increasing the firing-rate of the other) had pronounced effects on postsynaptic responses. The average postsynaptic firing rate remained unchanged, but fluctuations in firing rate and membrane potential increased in both cases (Fig. 3B,C). When the co-tuned inhibitory population was turned off, the emerging imbalance of excitation and inhibition unmasked the excitatory tuning curve, increasing the chance of action potential generation when preferred signal populations were active (Fig. 3B, middle). The compensatory increase in the activity of the flat population further quenched non-preferred excitatory signals, leading to anti-correlated responses for non-preferred

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

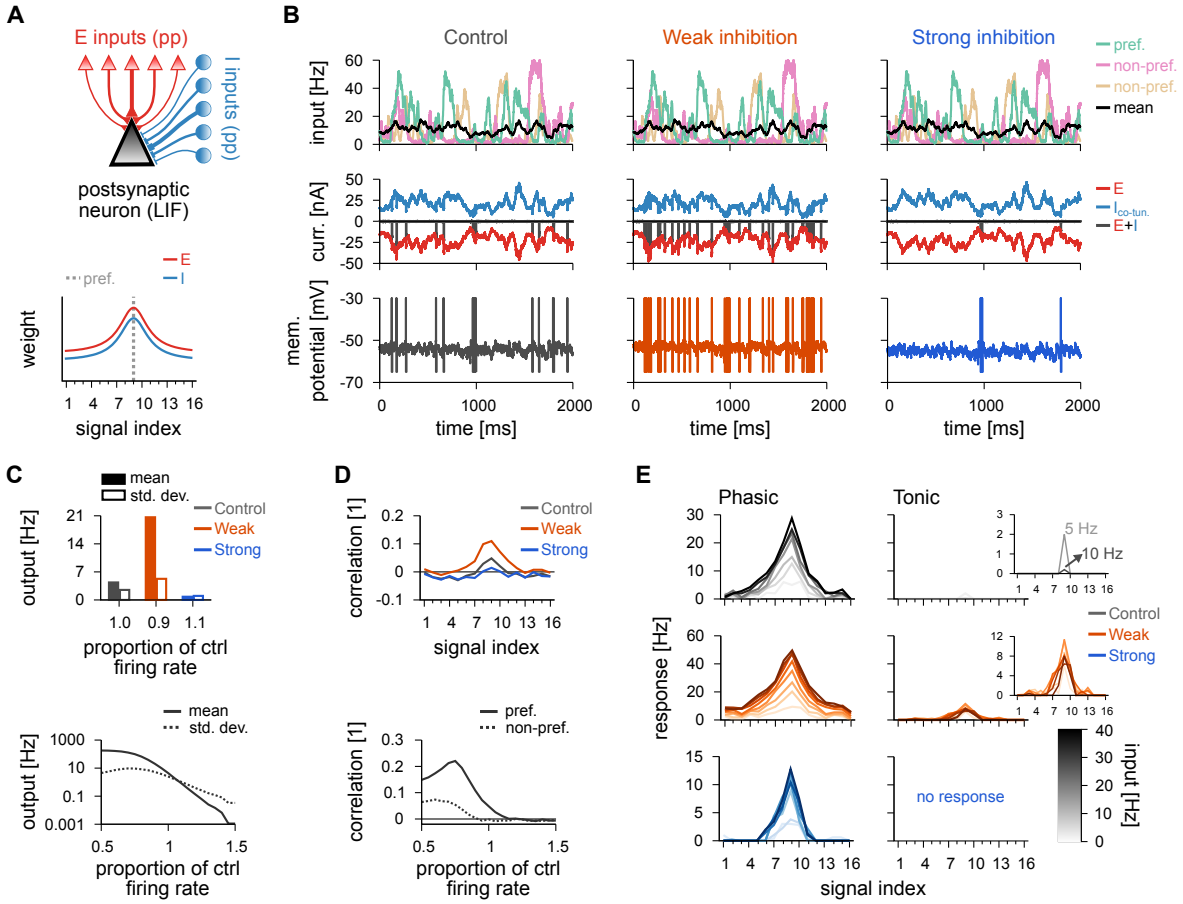

FIG. 2. Postsynaptic response for a model with a single inhibitory population. **A**, Schematic of the circuit with a single inhibitory population (top). Pre-synaptic spikes were generated as point-processes (pp), for both excitatory (red; 16 signals) and inhibitory (blue; 16 signals) inputs, and fed into a single-compartment leaky integrate-and-fire neuron (LIF). Receptive field profile (bottom). Average weight (y-axis) for different input signals (x-axis); preferred signal is pathway no. 9 (grey dashed line). **B**, Average firing-rate of the preferred, and two non-preferred inputs and mean of all inputs (top row), excitatory and inhibitory input currents (middle row), and membrane potentials (bottom row), for control (left), decreased (middle) and increased (right) inhibition. Control case is hand-tuned for postsynaptic firing-rates of ∼ 5 Hz. Decreased (increased) inhibition lowered (raised) inhibitory firing-rates by 10%, respectively. **C**, Average and standard deviation of the postsynaptic firing-rate in response to natural input for the three explored cases (top), and as a function of the inhibitory firing-rate (bottom). **D**, Pearson correlation between postsynaptic firing-rate and excitatory input firing-rates for different input signals for the three conditions in B (top). Correlation between output activity and preferred (continuous line) or non-preferred (dashed line) inputs as a function of the inhibitory firing-rate (bottom). **E**, Response to a pulse input in the phasic (left; first 50 ms), and tonic (right; last 50 ms) periods. Firing rate computed as the average number of spikes (for 100 trials) normalised by the bin size (50 ms). Each line corresponds to a different input strength; from light (low amplitude pulse) to dark (high amplitude pulse) colours. Insets show tonic response for control and decreased inhibitory firing-rates.

input signals (Fig. 3D, purple). The opposite effect could be observed when the flat population was deactivated. In this case, the lack of inhibition for non-preferred signals gave rise to input/output correlations for non-preferred signals, while preferred signals saw no response (Fig. 3B, right and Fig. 3D, green).

Compared to the control case (Fig. 3E, top), transient responses were drastically increased for preferred inputs when the co-tuned population was deactivated, and the response to non-preferred signals was completely diminished (Fig. 3E; middle). When the flat population was deactivated, the postsynaptic neuron responded strongly to the non-preferred inputs, but not to preferred inputs (Fig. 3E; bottom).

Co- and counter-tuned inhibitory populations. We also analysed the dynamics of two inhibitory populations which were tuned such that one population mirrored the tuning of the excitation (as before, *co-tuned*), while the other population was tuned such that the weakest excitatory signals received the strongest inhibition (Fig. 4A). We called this population *counter-tuned*. Counter-tuning has been observed as a consequence of anti-Hebbian inhibitory plasticity in experimental work in hippocampus[21–23] and used in theoretical work on associative memory networks[24]. In the balanced state, output behaviour is almost identical to the scenarios described above (Fig. 4B-D; control). The main distinction between the models with counter-tuned or flat inhibitory profiles is how they complemented the co-tuned inhibitory currents: the flat inhibition produced currents that tracked the co-tuned current, whereas counter-tuned inhibition produced inhibitory currents that were largely uncorrelated to the co-tuned inhibitory currents (Fig. 4B, left; compare with Fig. 3B, left).

When either the co- or the counter-tuned inhibitory populations were inactivated, fluctuations in both firing rate and membrane potential increased considerably (Fig. 4B, middle and right). Deactivation of the co-tuned population resulted in positive correlation between postsynaptic activity and preferred signals, and negative correlation between output and non-preferred signals (Fig. 4D, purple).

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

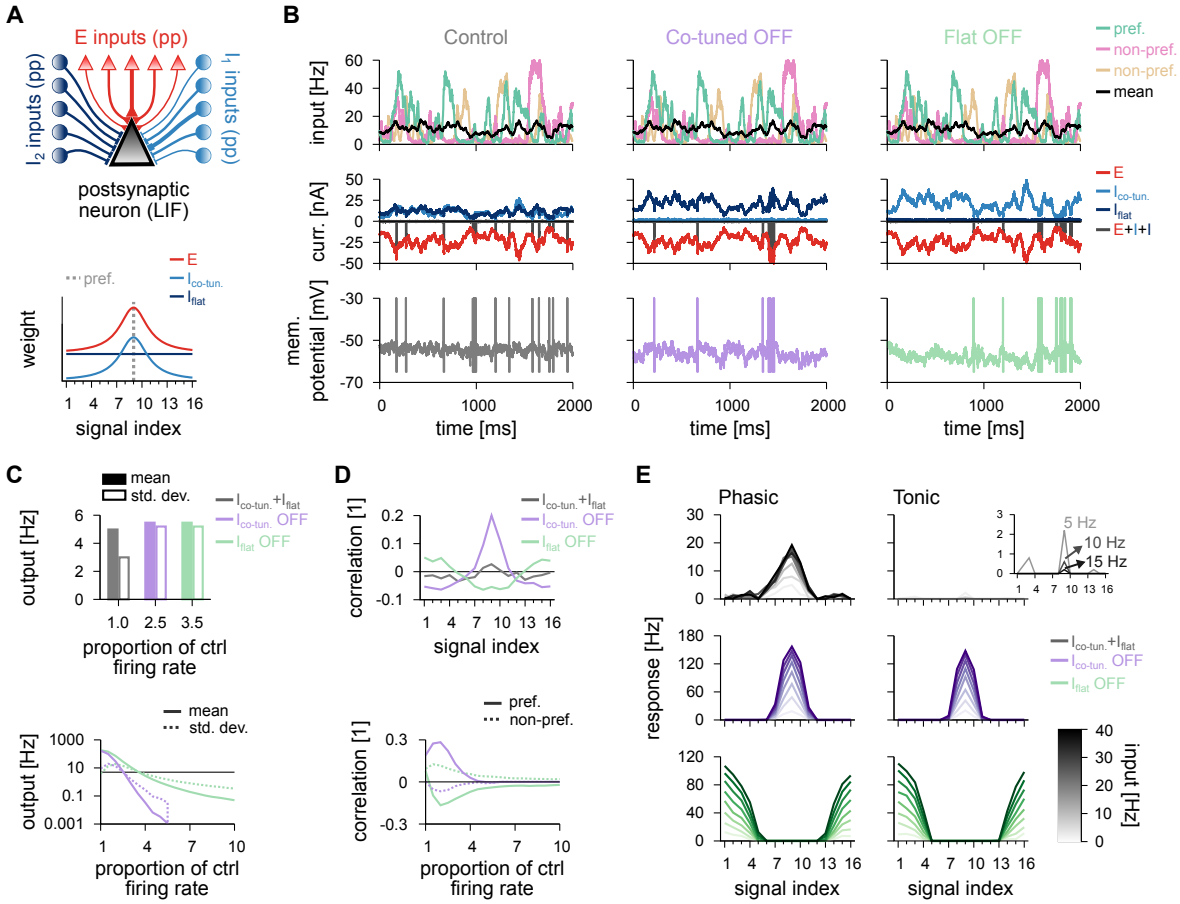

FIG. 3. Postsynaptic response a the model with co-tuned and flat inhibitory populations. **A**, Schematic of the circuit with two inhibitory populations (top); I_1 corresponds to the co-tuned population and I_2 to the flat population. Pre-synaptic spikes were generated as point-processes (pp) and fed into an LIF. Receptive field profile (bottom). Average weight (y-axis) for different input signals (x-axis); preferred signal is pathway no. 9 (grey dashed line). **B**, Average firing-rate of the preferred and two non-preferred inputs and mean of all inputs (top row), total excitatory current and inhibitory currents of both populations (middle row), and membrane potential (bottom row), for control (left), co-tuned (middle) and flat (right) population inactive. **C**, Average and standard deviation of the postsynaptic firing-rate in response to natural input for the three cases (top), and as a function of the inhibitory firing-rate (bottom). **D**, Pearson correlation between postsynaptic firing-rate and excitatory input firing-rates for different input signals for the three conditions in B. Correlation between output activity and preferred (continuous line) or non-preferred (dashed line) as a function of the inhibitory firing-rate of each inhibitory population (bottom). **E**, Response to a pulse input in the phasic (left; first 50 ms), and tonic (right; last 50 ms) periods. Firing rate computed as the average number of spikes (for 100 trials) normalised by bin size (50 ms). Each line corresponds to a different input strength; from light (low amplitude pulse) to dark (high amplitude pulse) colours. Insets show tonic response for control firing-rates.

For pulse stimulation, there was no discernible difference to the model with flat inhibition in the control state (Fig. 4E, top). Turning off counter-tuned inhibition (Fig. 4B-E) also had similar results in the postsynaptic response as turning off the flat inhibition (cf. Fig. 3B-E). Unlike before, turning off co-tuned inhibition produced elevated firing-rate responses also for transient stimuli from signals directly neighbouring the preferred input (Fig. 4E, middle row, compare with Fig. 3E, middle).

Quantitative differences of inhibitory profiles. For a quantitative evaluation of the differences between the three models with their varying inhibitory receptive fields, we introduced the parameter $\Delta C = 0.5(C_{\text{pref}} - C_{\text{non-pref}})$, i.e., 50% of the difference in input/output correlation between preferred, C_{pref}, and non-preferred, $C_{\text{non-pref}}$, signals (Methods). In the control condition, we observed similar $\Delta C \approx 0$ in all models (Fig. 5A, grey), reflecting low levels of correlation between output and input signals (Fig. 5B, top). With down-regulated inhibition, ΔC increased slightly in the model with one homogeneous inhibitory population. ΔC increased more considerably in a two-population model in which the co-tuned population was inactive (Fig. 5A, purple), indicating an increased correlation between preferred signal and output (Fig. 5B, middle). When the flat or the counter-tuned inhibitory populations were inactivated, we observed postsynaptic responses even to non-preferred input signals (Fig. 5B, bottom), which led to negative ΔC. Inactivating the counter-tuned inhibition resulted in a slightly better discrimination of non-preferred input signals (Fig. 5A, green).

To quantify pulse responses, we considered how many pulse signals could be recovered across the receptive field (Fig. 5C). We counted only those responses with more than 50% of the maximum firing-rate (Fig. 5D). The single inhibitory population model could only produce responses to preferred input signals, while co-modulation of two inhibitory populations could promote responses to non-preferred input signals, as well. Counter-tuned population achieved better postsynaptic control than flat inhibition (Fig. 5C).

Plasticity shapes inhibitory receptive fields. We have shown that complementary tuning of two inhibitory populations allows the retrieval of various facets of mixed and coincident

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

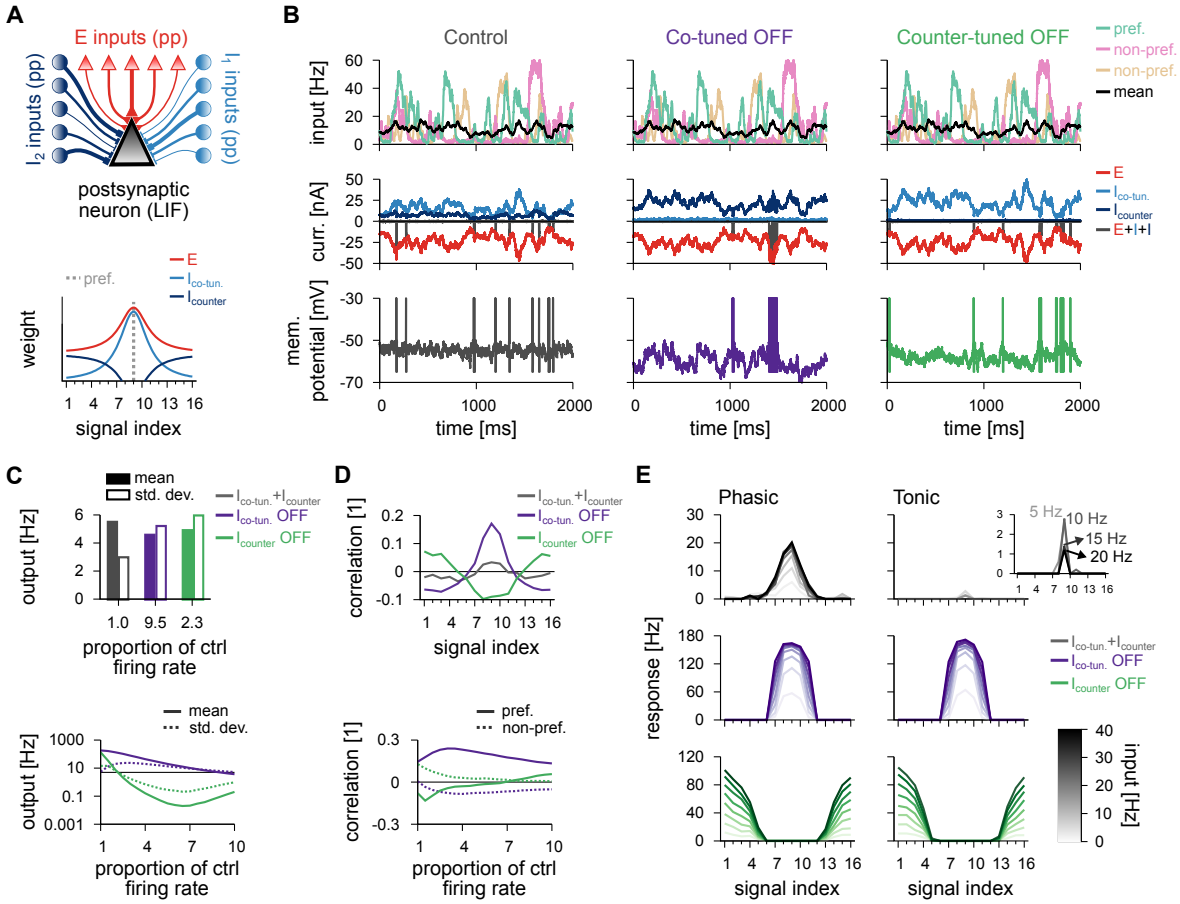

FIG. 4. Postsynaptic response for the model with co- and counter-tuned inhibitory populations. A, Schematic of the circuit with two inhibitory populations (top); I_1 corresponds to co-tuned and I_2 to counter-tuned population. Pre-synaptic spikes were generated as point-processes (pp) and fed into an LIF. Receptive field profile (bottom). Average weight (y-axis) for different input signals (x-axis); preferred signal is pathway no. 9 (grey dashed line). **B**, Average firing-rate of the preferred and two non-preferred inputs and mean of all inputs (top row), total excitatory current and inhibitory currents of both populations (middle row), and membrane potentials (bottom row), for control (left), co-tuned (middle) and counter-tuned (right) population inactive. **C**, Average and standard deviation of the postsynaptic firing-rate due to natural input for the three cases (top), and as a function of the inhibitory firing-rate (bottom). **D**, Pearson correlation between postsynaptic firing-rate and excitatory input firing-rates for different input signals for the three conditions in B. Correlation between preferred (continuous line) or non-preferred (dashed line) with the output activity as a function of the inhibitory firing-rate of each inhibitory population (bottom). **E**, Response to a pulse input in the phasic (left; first 50 ms), and tonic (right; last 50 ms) periods. Firing rate computed as the average number of spikes (for 100 trials) normalised by the bin size (50 ms). Each line corresponds to a different input strength; from light (low amplitude pulse) to dark (high amplitude pulse) colours. Insets show tonic response for control firing-rates.

stimuli. We wondered how such tuning patterns could emerge from naïve connectivity. To study how plasticity can shape the emergence of opposite receptive fields, we incorporated inhibitory synaptic plasticity mechanisms into our model with two inhibitory populations. We first implemented a Hebbian rule (that potentiated synaptic weights for coincident pre- and postsynaptic spikes and depressed them for sole presynaptic spikes[6], Fig. S1A), in one of the two inhibitory populations while the synapses of the excitatory and the other inhibitory population remained fixed (Fig. 6). This learning rule has previously been shown to generate inhibitory weight profiles that mirror the excitatory receptive field of a postsynaptic neuron, imposing a firing-rate fixing-point (target; Fig. S1B) by balancing excitation and inhibition. Simulations began with tuned excitatory synapses and flat inhibitory weight profiles in both inhibitory populations (Fig. 6A).

After 30 minutes of stimulation with natural inputs (cf. Fig. 1B), inhibitory weights of the plastic population stabilised (Fig. 6D-G). Whether the target firing rate (Fig. 6B,C) was reached depended on the synaptic strength of the other, static population of inhibitory synapses. If the static weights were weak, the plastic synapses increased their strength until the target firing rate was reached (Fig. 6C). If the static population provided strong inhibition (and thus kept postsynaptic firing below the target rate), weights from the plastic population would eventually vanish — before the target firing-rate could be reached (Fig. 6C,G).

The effective receptive field could be interpreted as the cumulative difference between excitatory and inhibitory weights. Consequently, the shape of the static population determined the shape of the plastic population (Fig. 6D,E). As expected, the input/output correlation of the postsynaptic responses followed the effective receptive field profile (Fig. 7A, cf. Fig. 6E). The best performance, i.e., distinct responses to preferred or non-preferred stimuli according to the modulatory state, and flat responses otherwise (Fig. 7B), emerged when the synapses of the static population were constrained to a narrow band of synaptic weights (Fig. 7C, shaded region).

Next, we introduced plasticity to the second population of inhibitory afferents. We tried two different rules. We began

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

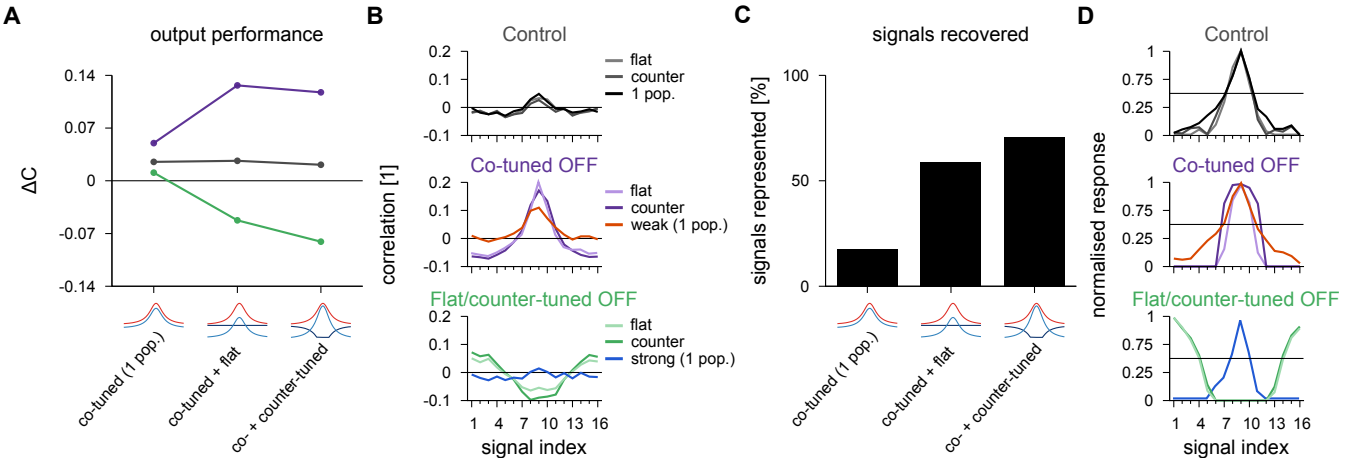

FIG. 5. Comparison of postsynaptic responses receiving co-tuned & flat or co-tuned & counter-tuned inhibitory populations. **A**, Performance index as the difference in input/output correlation between preferred and non-preferred signal. Ideal outcome is $\Delta C = 0$ for control case (grey), $\Delta C > 0$ for co-tuned population inactive (purple), and $\Delta C < 0$ for flat or counter-tuned populations inactive (green). We added the values for a single inhibitory population with control (grey), weak (purple) and strong (green) inhibitory inputs for comparison. **B**, Pearson correlation between postsynaptic firing-rate and excitatory input firing-rates for different signal indices from Figs. 2D, 3D and 4D, replotted for reference. **C**, Signals recovered in the pulse input paradigm. Signals represented are calculated as the percentage of signal afferents that activate the postsynaptic neuron with more that half the spikes of the maximum response for the three cases considered for the circuits analysed together. **D**, Normalised phasic response to a pulse input of 40 Hz, from Figs. 2E, 3E and 4E, replotted for reference. Horizontal line indicates 50% of maximum response.

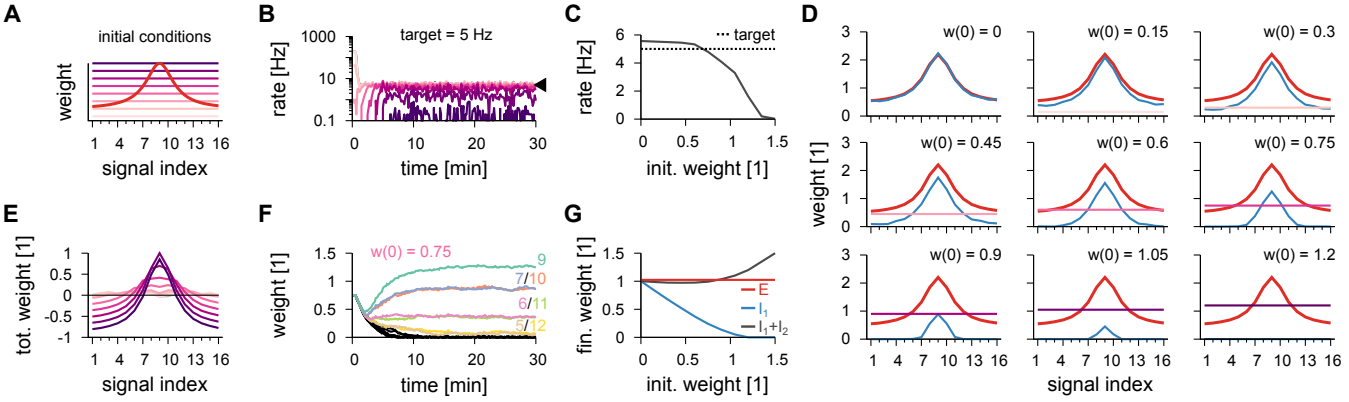

FIG. 6. Inhibitory plasticity acting on one inhibitory population compensates global inhibition from second inhibitory population. **A**, Schematic of the receptive profile for excitatory synapses (red) and different initial conditions for inhibitory synapses (pink to purple colour-code). Inhibitory population 1 has its inhibitory synapses changing according to a plasticity mechanism while population 2 remains fixed. **B**, Time-course of the postsynaptic firing-rate for different initial conditions (colours as in A). Inhibitory plasticity on population 1 is set to achieve a balanced state with target of 5 Hz (arrowhead). **C**, Stabilised postsynaptic firing-rate as a function of the initial inhibitory synaptic weight. **D**, Individual receptive field profiles for excitatory (red), inhibitory population 1 (blue, after synaptic stabilisation), and inhibitory population 2 (colour coded as A). **E**, Total synaptic weight per signal (excitatory minus inhibitory) for different initial conditions after stabilisation of synapses from population 1. **F**, Example of synaptic dynamics of inhibitory population 1 for a given initial condition. Colours represent different signal groups. **G**, Final weights as a function of initial inhibitory weights. Plotted are excitatory (red), plastic inhibitory (blue) and sum of total inhibitory synapses (grey).

with a homeostatic plasticity rule (Methods) which scaled synapses to reach a fixed-point in the postsynaptic firing-rates (Fig. S1C), achieving results that matched the handed-tuned solution. Notably, this plasticity rule was purely local, taking only presynaptic weights and postsynaptic firing rates into account, similar to experimentally observed scaling of inhibitory synapses[38,39]. With the homeostatic rule co-active, the *Hebbian* synapses — connections changing according to the Hebbian plasticity rule — developed a co-tuned profile from initially random weights (Fig. 8A, top; Fig. 8C, left), while the synapses following the scaling rule collapsed to a single value (Fig. 8A, bottom; Fig. 8C, right). Consequently, the postsynaptic neuron received precisely balanced inputs (Fig. 8B), autonomously arriving in a regime of optimal performance (cf. Fig. 7C).

Instead of a purely homeostatic scaling rule, we also tried an experimentally observed[21–23] anti-Hebbian rule in the second inhibitory population (Fig. S1D). The Hebbian plasticity rule in one synapse population leads to stable receptive fields and imposes a firing-rate set point for the postsynaptic neuron. The anti-Hebbian rule, on the other hand, increases the firing rate of the postsynaptic neuron indefinitely, because correlated activity decreases synaptic weights (only sole presynaptic spikes increase synaptic weights; Methods). Anti-Hebbian plasticity rules is thus unstable (Fig. S1E). To overcome this problem without incorporating additional, complex dynamics, we set the learning rate of the anti-Hebbian rule to decrease exponentially over time (Fig. S1F; see Discussion). With the anti-Hebbian rule active, initially random weights evolved into co-tuned and counter-tuned receptive field profiles (Fig. 8D). As learning slowed down, the *anti-Hebbian* synapses — connec-

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

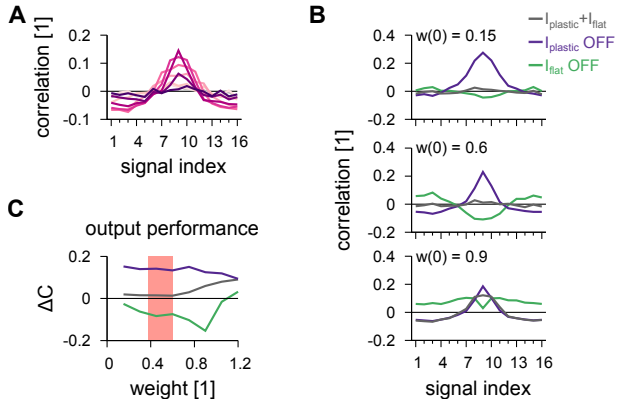

FIG. 7. Postsynaptic response after stabilisation of synapses from one population. A, Pearson correlation between postsynaptic firing-rate and excitatory input firing-rates with both inhibitory populations active. Colour code as in Fig. 6A. **B**, Pearson correlation between postsynaptic firing-rate and excitatory input firing-rates for different signals with both inhibitory populations active, population 1 inactive, and population 2 inactive. Three examples are shown. **C**, Performance index as the difference between preferred signal and non-preferred signal as in Fig. 5A, but for the whole range of different initial conditions (inhibitory synaptic weights). Pink shaded area corresponds to best performance, which maximises the difference between preferred and non-preferred correlation responses oppositely for inactivation of different populations, and minimising difference between preferred and non-preferred correlation when both inhibitory populations are active.

tions changing according to the anti-Hebbian plasticity rule — stabilised, and Hebbian synapses ceased to change once the target firing rate was reached (Fig. 8E; Fig. 8F). The final shape of both receptive fields was nearly identical to the hand-tuned profiles (cf. Fig. 4A), with equivalent postsynaptic performances (data not shown; see also Fig. 3 and Fig. 4).

Discussion

We investigated how several distinctly tuned inhibitory and excitatory synaptic populations interact in a receptive field-like paradigm. In our model, we aimed for precise balance of excitation and inhibition, accordant with evidence of excitatory and inhibitory co-tuning in cat visual cortex[40], rodent auditory cortex[12,41,42] and rodent hippocampus[43], and temporal correlations in neighbouring excitatory and inhibitory synapses[44].

Consistent with earlier work, we could modulate the efficacy of a single inhibitory population to enhance the output correlation with the preferred input[18,19], but the flexibility of the control mechanism was very limited. Non-preferred signals never evoked faithful responses (Fig. 2). We wondered if additional, independently tuned synapse populations would grant more flexible control over the signal stream.

We constructed models with two types of inhibitory interneurons that could be modulated independently and carried a diverse set of input signals. Each signal group of excitatory synapses had two corresponding signal groups of inhibitory synapses, all three displaying correlated activity patterns, with their combined efficacy balancing the net input current in the baseline (control) state. To emphasise different aspects of the overall input, we allowed for the ratio of inhibitory currents delivered by the two inhibitory populations to vary dynamically over time, consistent with evidence from macaque V1 indicating that the receptive fields of neurons are dynamic, rapidly switching their selectivity[45]. Our results suggest that a discrimination of preferred and non-preferred stimuli, such as occurs in selective attention, may be achieved by dynam-

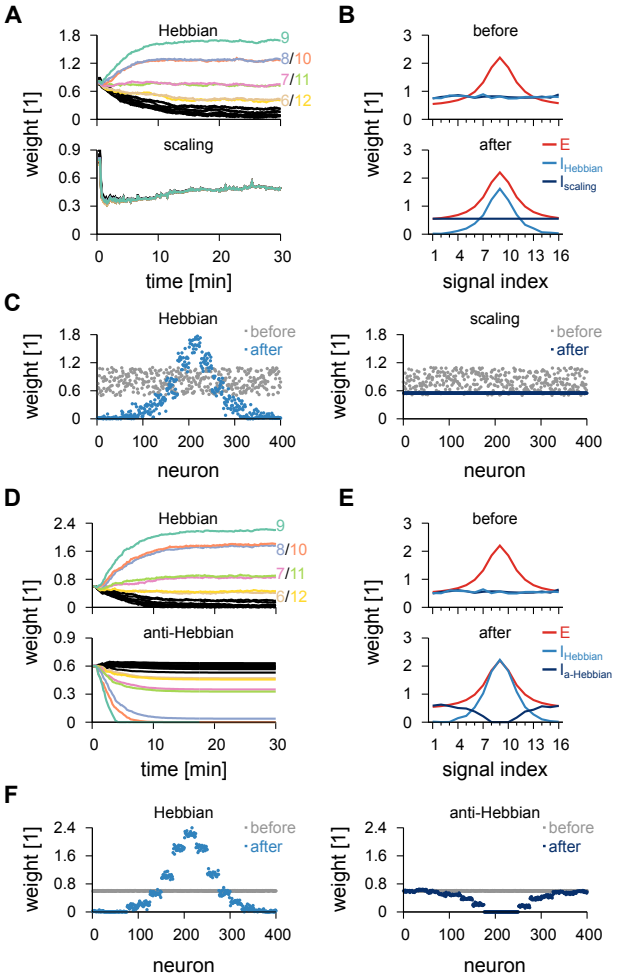

FIG. 8. Simultaneous learning of two inhibitory profiles. A, Temporal evolution of inhibitory synaptic weights when one inhibitory population follows a Hebbian plasticity rule (top) and the other population follows a synaptic scaling plasticity rule (bottom). **B**, Initial (top) and final (bottom) weight profiles from A with excitatory weights for reference. **C**, Individual synaptic weights before and after learning for synapses following the Hebbian plasticity rule (left) and synapses following the scaling plasticity rule (right). **D**, Same as A, but with Hebbian (top) and anti-Hebbian (bottom) plasticity rules acting concurrently. **E**, Same as B, but for Hebbian (top) and anti-Hebbian (bottom) plasticity rules. **F**, Same as C, but for Hebbian (left) and anti-Hebbian (right) plasticity rules.

ically modulating inhibition. Such modulation of stimulus-selectivity could correspond to mechanisms of top-down selective attention[46–48].

Attending to a specific stimulus is known to increase the firing rate of sensory cells selective for this stimulus[49–51]. Such sharpening of stimulus-selectivity was also observed in our simulations (Fig. 3 and Fig. 4; purple). We also observed the opposite: enhanced postsynaptic responses to non-preferred and suppressed responses to preferred stimuli (Fig. 3 and Fig. 4; green). We interpreted this as selectively ignoring a specific stimulus, rather than attending to it. Thus, the two aspects of selective attention – enhancing response to targets, and suppressing the response to distractors – were implemented in our model by the two types of disinhibition.

The architecture of our model fits well with the diversity of interneuron types and connectivity profiles in the neocortex[1,4]. Co-tuned inhibitory afferents may originate from

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

parvalbumin-positive (PV+) interneurons. As the main source of inhibition to principal cells, PV+ interneurons tend to target postsynaptic neurons with similar preferred orientation[52]. Moreover, activation of these neurons leads to a flattened selectivity curve[52] (but see Lee *et al.* [53]). Flat (or counter-tuned) inhibitory afferents may arrive from somatostatin-positive (SOM+) interneurons, with their less selective connectivity pattern[52]. This interpretation is in line with recent evidence suggesting that top-down visual attention relies on local inhibitory circuitry in primary visual cortex[42]. In this scheme, PV+ and SOM+ neurons inhibit pyramidal cells, while vasoactive intestinal peptide-positive (VIP+) neurons suppress other inhibitory interneurons, acting as a source of disinhibition. Direct manipulation of SOM+, PV+ and VIP+ neurons confirms these respective roles in inhibition and disinhibition in both visual[54] and auditory cortices[15,55]. Additionally, Zhou *et al.* [42] reported that VIP+ neurons received excitatory top-down inputs from the rodent cingulate cortex, allowing for a narrow selectivity profile of pyramidal cells when active and the opposite when inactive. Finally, blocking cortical inhibition reduces the stimulus-selectivity of cortical neurons[56,57] (but see Nelson *et al.* [58]).

To show how these inhibitory synaptic profiles could develop, we implemented a Hebbian-like inhibitory plasticity rule[6,20], co-active with either a homeostatic scaling[38,39] (Fig. 8A-C) or an anti-Hebbian[21–23] (Fig. 8D-F) plasticity rule. The scaling plasticity rule we implemented decreases synaptic weights proportionally to their current weight (multiplicative), and strengthens weights by a constant amount (additive), relative to the firing rate of the postsynaptic neuron. Consequently, all synapses tend to converge to the same value, without being enforced by a global rule. There is currently no direct evidence for this specific implementation, but it would provide a parsimonious explanation for the un-tuned, *blanket* inhibition often encountered in experiments[59], and would allow modulating postsynaptic responses independently of the presynaptically tuned receptive field.

The anti-Hebbian rule we used had the opposite shape of the Hebbian plasticity rule, and was unstable, i.e., it could lead to infinite strengthening of presynaptic weights. A silent feedforward network would be the result of such connectivity (but see Hendin *et al.* [24]). To obtain a plausible receptive field, we thus had to use an additional mechanism. The simplest solution was to include a modulatory term to the learning rate of the anti-Hebbian plasticity rule. This process mimics modulatory control of plasticity through the activity of other neuronal types[12,60,61] and allowed us to balance the effect of the opposing rules. Other mechanisms could be used to control the unstable nature of the Anti-Hebbian plasticity rule, but the generality of the results would likely not be altered.

To explore the interaction between two distinct inhibitory plasticity rules without confounds, we made the simplifying assumption that the excitatory synapses would remain fixed (but see Clopath *et al.* [62], Litwin-Kumar and Doiron [63], and Zenke *et al.* [64]). Further work could explore how inhibitory plasticity rules interact not only with each other, but also with rules for excitatory synaptic plasticity — thus getting closer to the realistic conditions of biological neural circuits. Additionally, incorporating more types of inhibitory interneurons is a possible next step to be explored, and thus the study of the interaction of more inhibitory plasticity rules[9,65,66].

Our model only considered a single postsynaptic neuron, with no feedback connectivity, which is thought to be the source of most orientation-selective input to cortical cells[67]. Further theoretical work should extend the combined use of these distinct plasticity rules for inhibitory synaptic connectivity to larger architectures, involving multiple neurons and incorporating feedback connections. Further synergies may emerge, such as multiplicative and additive modulation of receptive fields, and surround suppression[68]. Other possible functionalities that have been explored in this context are gating of different signal streams[47,48], one-shot learning in hippocampus[11], and consequences to brain disorders[69].

Interestingly, artificial networks have been shown to develop similar receptive field profiles to the ones explored here when they are trained to solve multiple tasks[70]. Yang *et al.* [70] have shown that clusters of neurons can acquire co-tuned or flat connectivity, which are controlled by context-encoding signals. These results hint at the possibility that biological and artificial systems may utilise similar strategies to solve context-dependent filtering tasks.

Finally, we predict that various GABAergic interneurons in the same cortical region must obey a range of different inhibitory synaptic plasticity rules, to restore or reverse neuronal stimulus-selectivity as appropriate and necessary. Such evidence would provide empirical validation to the theoretical results presented here, and in turn inform future computational modelling.

Materials and Methods

Detailed methods are given in the supplementary materials.

Software and code availability

Simulations were run in Fortran, compiled with Intel Fortran Compiler 19.0 on an Intel-based Linux computer (Debian 9; i9-9900X processor; 32 GB memory). Codes will be made available online upon publication[71]. Individual plots were generated with Gnuplot. Figures were generated with Inkscape.

Acknowledgements

We thank the members of the Vogels group, and particularly Georgia Christodoulou and William Podlaski, for fruitful discussions. This work was supported by a Research Project Grant by the Leverhulme Trust (RPG-2016-446; EJA), a Sir Henry Dale Fellowship by the Wellcome Trust and the Royal Society (WT100000; EJA, TPV), a Wellcome Trust Senior Research Fellowship (214316/Z/18/Z; EJA, TPV), and a Claredon Scholarship from the University of Oxford (AIL).

Author contributions

EJA and TPV designed research; EJA and AIL carried out the simulations and analysis; EJA, AIL and TPV wrote the manuscript.

Rerences

[1] X Jiang, S Shen, CR Cadwell, P Berens, F Sinz, AS Ecker, S Patel, and AS Tolias, "Principles of connectivity among morphologically defined cell types in adult neocortex," Science **350**, aac9462 (2015).

[2] NW Gouwens, SA Sorensen, J Berg, C Lee, T Jarsky, J Ting, SM Sunkin, D Feng, CA Anastassiou, E Barkan, K Bickley, N Blesie, T Braun, K Brouner, A Budzillo, S Caldejon, T Casper, D Castelli, P Chong, K Crichton, [...], and C Koch, "Classification of electrophysiological and morphological neuron types in the mouse visual cortex," Nature Neuroscience **22**, 1182–1195 (2019).

[3] J-S Jouhanneau, J Kremkow, and JFA Poulet, "Single synaptic inputs drive high-precision action potentials in parvalbumin expressing gaba-ergic cortical neurons in vivo," Nature Communications **9**, 1540 (2018).

[4] H Markram, M Toledo-Rodriguez, Y Wang, A Gupta, G Silberberg, and C Wu, "Interneurons of the neocortical inhibitory system," Nature Reviews Neuroscience **5**, 793–807 (2004).

[5] C Van Vreeswijk and H Sompolinsky, "Chaos in neuronal networks with balanced excitatory and inhibitory activity," Science **274**, 1724–1726 (1996).

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[6] TP Vogels, H Sprekeler, F Zenke, C Clopath, and W Gerstner, "Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks," Science **334**, 1569–1573 (2011).

[7] A Maffei, "Fifty shades of inhibition," Current Opinion in Neurobiology **43**, 43–47 (2017).

[8] H Sprekeler, "Functional consequences of inhibitory plasticity: homeostasis, the excitation-inhibition balance and beyond," Current Opinion in Neurobiology **43**, 198–203 (2017).

[9] G Hennequin, EJ Agnes, and TP Vogels, "Inhibitory plasticity: balance, control, and codependence," Annual Review of Neuroscience **40**, 557–579 (2017).

[10] CQ Chiu, A Barberis, and MJ Higley, "Preserving the balance: diverse forms of long-term gabaergic synaptic plasticity," Nature Reviews Neuroscience **20**, 272–281 (2019).

[11] W Nicola and C Clopath, "A diversity of interneurons and hebbian plasticity facilitate rapid compressible learning in the hippocampus," Nature Neuroscience **22**, 1168–1181 (2019).

[12] RC Froemke, MM Merzenich, and CE Schreiner, "A synaptic memory trace for cortical receptive field plasticity," Nature **450**, 425–429 (2007).

[13] JJ Letzkus, SBE Wolff, and A Lüthi, "Disinhibition, a circuit mechanism for associative learning and memory," Neuron **88**, 264–276 (2015).

[14] S Zhang, M Xu, T Kamigaki, JPH Do, W-C Chang, S Jenvay, K Miyamichi, L Luo, and Y Dan, "Long-range and local circuits for top-down modulation of visual cortex processing," Science **345**, 660–665 (2014).

[15] KV Kuchibhotla, JV Gill, GW Lindsay, ES Papadoyannis, RE Field, TA Sten Hindmarsh, KD Miller, and RC Froemke, "Parallel processing by cortical inhibition enables context-dependent behavior," Nature Neuroscience **20**, 62–71 (2017).

[16] D Vallentin, G Kosche, D Lipkind, and MA Long, "Inhibition protects acquired song segments during vocal learning in zebra finches," Science **351**, 267–271 (2016).

[17] HC Barron, TP Vogels, UE Emir, TR Makin, J O'Shea, S Clare, S Jbabdi, RJ Dolan, and TEJ Behrens, "Unmasking latent inhibitory connections in human cortex to reveal dormant cortical memories," Neuron **90**, 191–203 (2016).

[18] TP Vogels and LF Abbott, "Gating multiple signals through detailed balance of excitation and inhibition in spiking networks," Nature Neuroscience **12**, 483–491 (2009).

[19] J Kremkow, A Aertsen, and A Kumar, "Gating of signal propagation in spiking neural networks by balanced and correlated excitation and inhibition," Journal of Neuroscience **30**, 15760–15768 (2010).

[20] JA D'amour and RC Froemke, "Inhibitory and excitatory spike-timing-dependent plasticity in the auditory cortex," Neuron **86**, 514–528 (2015).

[21] MA Woodin, K Ganguly, and M-m Poo, "Coincident pre-and postsynaptic activity modifies gabaergic synapses by postsynaptic changes in cl- transporter activity," Neuron **39**, 807–820 (2003).

[22] J Ormond and MA Woodin, "Disinhibition mediates a form of hippocampal long-term potentiation in area ca1," PLoS One **4**, e7224 (2009).

[23] J Ormond and MA Woodin, "Disinhibition-mediated ltp in the hippocampus is synapse specific," Frontiers in Cellular Neuroscience **5**, 17 (2011).

[24] O Hendin, D Horn, and MV Tsodyks, "The role of inhibition in an associative memory model of the olfactory bulb," Journal of Computational Neuroscience **4**, 173–182 (1997).

[25] RJ Douglas and KA Martin, "A functional microcircuit for cat visual cortex," Journal of Physiology **440**, 735–769 (1991).

[26] D Ferster, "Orientation selectivity of synaptic potentials in neurons of cat primary visual cortex," Journal of Neuroscience **6**, 1284–1301 (1986).

[27] X Pei, TR Vidyasagar, M Volgushev, and OD Creutzfeldt, "Receptive field analysis and orientation selectivity of postsynaptic potentials of simple cells in cat visual cortex," Journal of Neuroscience **14**, 7130–7140 (1994).

[28] M Volgushev, Xing Pei, TR Vidyasagar, and OD Creutzfeldt, "Excitation and inhibition in orientation selectivity of cat visual cortex neurons revealed by whole-cell recordings in vivo," Visual Neuroscience **10**, 1151–1155 (1993).

[29] B Bathellier, L Ushakova, and S Rumpel, "Discrete neocortical dynamics predict behavioral categorization of sounds," Neuron **76**, 435–449 (2012).

[30] SL Smith, IT Smith, T Branco, and M Häusser, "Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo," Nature **503**, 115–120 (2013).

[31] MS Haley, A Fontanini, and A Maffei, "Laminar-and target-specific amygdalar inputs in rat primary gustatory cortex," Journal of Neuroscience **36**, 2623–2637 (2016).

[32] L Estebanez, I Férézou, V Ego-Stengel, and DE Shulz, "Representation of tactile scenes in the rodent barrel cortex," Neuroscience **368**, 81–94 (2018).

[33] CD Harvey, F Collman, DA Dombeck, and DW Tank, "Intracellular dynamics of hippocampal place cells during virtual navigation," Nature **461**, 941–946 (2009).

[34] BB Ujfalussy, JK Makara, M Lengyel, and T Branco, "Global and multiplexed dendritic computations under in vivo-like conditions," Neuron **100**, 579–592 (2018).

[35] DS Reich, F Mechler, KP Purpura, and JD Victor, "Interspike intervals, receptive fields, and information encoding in primary visual cortex," Journal of Neuroscience **20**, 1964–1974 (2000).

[36] KB Hengen, ME Lambo, SD Van Hooser, DB Katz, and GG Turrigiano, "Firing rate homeostasis in visual cortex of freely behaving rodents," Neuron **80**, 335–342 (2013).

[37] M Diesmann, M-O Gewaltig, and A Aertsen, "Stable propagation of synchronous spiking in cortical neural networks," Nature **402**, 529–533 (1999).

[38] D Stellwagen and RC Malenka, "Synaptic scaling mediated by glial tnf-α," Nature **440**, 1054–1059 (2006).

[39] LR Zhong, X Chen, E Park, TC Südhof, and L Chen, "Retinoic acid receptor rarα-dependent synaptic signaling mediates homeostatic synaptic plasticity at the inhibitory synapses of mouse visual cortex," Journal of Neuroscience **38**, 10454–10466 (2018).

[40] JS Anderson, M Carandini, and D Ferster, "Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex," Journal of Neurophysiology **84**, 909–926 (2000).

[41] M Wehr and AM Zador, "Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex," Nature **426**, 442–446 (2003).

[42] M Zhou, F Liang, XR Xiong, L Li, H Li, Z Xiao, HW Tao, and LI Zhang, "Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex," Nature Neuroscience **17**, 841–850 (2014).

[43] A Bhatia, S Moza, and US Bhalla, "Precise excitation-inhibition balance controls gain and timing in the hippocampus," eLife **8**, e43415 (2019).

[44] M Okun and I Lampl, "Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities," Nature Neuroscience **11**, 535–537 (2008).

[45] DL Ringach, MJ Hawken, and R Shapley, "Dynamics of orientation tuning in macaque primary visual cortex," Nature **387**, 281–284 (1997).

[46] PE Haenny and PH Schiller, "State dependent activity in monkey visual cortex," Experimental Brain Research **69**, 225–244 (1988).

[47] R Naud and H Sprekeler, "Sparse bursts optimize information transmission in a multiplexed neural code," Proceedings of the National Academy of Sciences **115**, E6329–E6338 (2018).

[48] L Hertäg and H Sprekeler, "Amplifying the redistribution of somato-dendritic inhibition by the interplay of three interneuron types," PLoS Computational Biology **15**, e1006999 (2019).

[49] SJ Luck, L Chelazzi, SA Hillyard, and R Desimone, "Neural mechanisms of spatial selective attention in areas v1, v2, and v4 of macaque visual cortex," Journal of Neurophysiology **77**, 24–42 (1997).

[50] AA Disney, C Aoki, and MJ Hawken, "Gain modulation by nicotine in macaque v1," Neuron **56**, 701–713 (2007).

[51] J Krueger and AA Disney, "Structure and function of dual-source cholinergic modulation in early vision," Journal of Comparative Neurology **527**, 738–750 (2019).

[52] NR Wilson, CA Runyan, FL Wang, and M Sur, "Division and subtraction by distinct cortical inhibitory networks in vivo," Nature **488**, 343–348 (2012).

[53] S-H Lee, AC Kwan, S Zhang, V Phoumthipphavong, JG Flannery, SC Masmanidis, H Taniguchi, ZJ Huang, F Zhang, ES Boyden, K Deisseroth, and Y Dan, "Activation of specific interneurons improves v1 feature selectivity and visual perception," Nature **488**, 379–383 (2012).

[54] Y Fu, JM Tucciarone, JS Espinosa, N Sheng, DP Darcy, RA Nicoll, ZJ Huang, and MP Stryker, "A cortical circuit for gain control by behavioral state," Cell **156**, 1139–1152 (2014).

[55] H-J Pi, B Hangya, D Kvitsiani, JI Sanders, ZJ Huang, and A Kepecs, "Cortical interneurons that specialize in disinhibitory control," Nature **503**, 521–524 (2013).

[56] AM Sillito, "Inhibitory mechanisms influencing complex cell orientation selectivity and their modification at high resting discharge levels," Journal of Physiology **289**, 33–53 (1979).

[57] AM Sillito, JA Kemp, JA Milson, and N Berardi, "A re-evaluation of the mechanisms underlying simple cell orientation selectivity," Brain Research **194**, 517–520 (1980).

[58] S Nelson, L Toth, B Sheth, and M Sur, "Orientation selectivity of cortical neurons during intracellular blockade of inhibition," Science **265**, 774–777 (1994).

[59] MM Karnani, M Agetsuma, and R Yuste, "A blanket of inhibition: functional inferences from dense inhibitory connectivity," Current Opinion in Neurobiology **26**, 96–102 (2014).

[60] E Abs, RB Poorthuis, D Apelblat, K Muhammad, MB Pardi, L Enke, D Kushinsky, D-L Pu, MF Eizinger, K-K Conzelmann, I Spiegel, and JJ Letzkus, "Learning-related plasticity in dendrite-targeting layer 1 interneurons," Neuron **100**, 684–699 (2018).

[61] WC Abraham and MF Bear, "Metaplasticity: the plasticity of synaptic plasticity," Trends in Neurosciences **19**, 126–130 (1996).

[62] C Clopath, TP Vogels, RC Froemke, and H Sprekeler, "Receptive field formation by interacting excitatory and inhibitory synaptic plasticity," BioRxiv , 066589 (2016).

[63] A Litwin-Kumar and B Doiron, "Formation and maintenance of neuronal assemblies through synaptic plasticity," Nature Communications **5**, 5319 (2014).

[64] F Zenke, EJ Agnes, and W Gerstner, "Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks," Nature Communications **6**, 6922 (2015).

[65] CD Holmgren and Y Zilberter, "Coincident spiking activity induces long-term changes in inhibition of neocortical pyramidal cells," Journal of Neuroscience **21**, 8270–8277 (2001).

[66] JS Haas, T Nowotny, and HDI Abarbanel, "Spike-timing-dependent plas-

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ticity of inhibitory synapses in the entorhinal cortex," Journal of Neurophysiology **96**, 3305–3313 (2006).

[67] B Ahmed, JC Anderson, RJ Douglas, KAC Martin, and JC Nelson, "Polyneuronal innervation of spiny stellate neurons in cat visual cortex," Journal of Comparative Neurology **341**, 39–49 (1994).

[68] A Litwin-Kumar, R Rosenbaum, and B Doiron, "Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes," Journal of Neurophysiology **115**, 1399–1409 (2016).

[69] C O'Donnell, JT Goncalves, C Portera-Cailliau, and TJ Sejnowski, "Beyond excitation/inhibition imbalance in multidimensional models of neural circuit changes in brain disorders," eLife **6**, e26724 (2017).

[70] GR Yang, MR Joglekar, HF Song, WT Newsome, and X-J Wang, "Task representations in neural networks trained to perform many cognitive tasks," Nature Neuroscience **22**, 297–306 (2019).

[71] EJ Agnes, "attentional_switch. github," `https://github.com/ejagnes/attentional_switch/` (2019).

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary material of

Complementary inhibitory receptive fields emerge from synaptic plasticity and create an attentional switch in sensory circuits

Everton J. Agnes, Andrea I. Luppi, and Tim P. Vogels

Codes for all results are openly available at GitHub, repository `https://github.com/ejagnes/attentional_switch`.

Neuron model. To investigate changes in neuronal response due to specific inhibitory connectivity motif we simulated a postsynaptic leaky integrate-and-fire neuron (LIF) receiving excitatory and inhibitory afferents. Postsynaptic neuronal membrane potential dynamics is governed by

$$\tau_{\mathrm{m}} \frac{\mathrm{d}u(t)}{\mathrm{d}t} = -\left[u(t) - u_{\mathrm{rest}}\right] - g_{\mathrm{E}}(t)\left[u(t) - E_{\mathrm{E}}\right] - g_{\mathrm{I}}(t)\left[u(t) - E_{\mathrm{I}}\right], \tag{1}$$

where $u(t)$ is the somatic voltage at time t, $\tau_{\mathrm{m}} = RC$ is the membrane time constant (membrane resistance, R, times membrane conductance, C), u_{rest} is the resting membrane potential, and E_{E} and E_{I} are the reversal potential for excitatory and inhibitory synapses, respectively. Synaptic conductances, $g_{\mathrm{E}}(t)$ and $g_{\mathrm{I}}(t)$, evolve according to

$$\frac{\mathrm{d}g_{\mathrm{E}}(t)}{\mathrm{d}t} = -\frac{g_{\mathrm{E}}(t)}{\tau_{\mathrm{E}}} + \sum_{j=1}^{N_{\mathrm{E}}} w_j(t) S_j(t) \tag{2}$$

and

$$\frac{\mathrm{d}g_{\mathrm{I}}(t)}{\mathrm{d}t} = -\frac{g_{\mathrm{I}}(t)}{\tau_{\mathrm{I}}} + \sum_{j=N_{\mathrm{E}}+1}^{N} w_j(t) S_j(t). \tag{3}$$

Both excitatory and inhibitory conductances decay exponentially to zero with time constants τ_{E} and τ_{I}, respectively. Presynaptic action potentials trigger increase in synaptic conductances through the sum of Dirac delta functions,

$$S_j(t) = \sum_k \delta\left(t - t_{kj}\right), \tag{4}$$

where t_{kj} is the time of the kth spike of presynaptic afferent j. The contribution of a given presynaptic afferent j to changes in conductances is given by the synaptic weight $w_j(t)$, which was fixed for excitatory synapses and could change over time due to plasticity mechanisms for inhibitory synapses. The total number of presynaptic afferents is $N = N_{\mathrm{E}} + N_{\mathrm{I}}$, with N_{E} being the number of excitatory and N_{I} of inhibitory presynaptic afferents.

An action potential is triggered at the postsynaptic neuron once its membrane potential $u(t)$ crosses the spiking threshold u_{th} from below. The membrane potential is then instantaneously reset to u_{reset}, being clamped at this value for the duration of the refractory period, τ_{ref}. The postsynaptic spike train is described here as a sum of Dirac deltas,

$$S_{\mathrm{post}}(t) = \sum_k \delta(t - t_k), \tag{5}$$

where t_k is the time of the kth spike of the postsynaptic neuron, or the time the membrane potential crosses the spiking threshold from below. Parameters used for the postsynaptic neuron are detailed in Table I.

Inputs. To mimic receptive field input-like, we divided the synaptic inputs into P signal groups ($\mu = 1, ..., P$) that share the same fluctuation in firing rate. We tested two cases: (*i*), natural input, and (*ii*), pulse input. Both are described below.

Natural input. For presynaptic activity mimicking a natural input, activity follows an inhomogeneous Poisson process that changes according to a modified Ornstein-Uhlenbeck (OU) process. We first defined an auxiliary variable for each pattern, $y_\mu(t)$, that follows a stochastic first-order differential equation given by

$$\tau_{\mathrm{OU}} \frac{\mathrm{d}y_\mu(t)}{\mathrm{d}t} = -y_\mu(t) + \xi_\mu(t), \tag{6}$$

where μ is the signal group index, τ_{OU} is the time constant for the decaying process that changes due to a Gaussian noise term $\xi_\mu(t)$ with unitary standard deviation. The mean value of the variable y_μ is zero, and thus it assumes positive and negative values with same probability (for long periods).

The spike train of an afferent in a given signal group μ is given by the variable $\nu_{\mathrm{X}\mu}(t)$ which is a rectified version of the auxiliary variable plus a term to generate background firing rate, ν_{Xbg}, where X indicates the presynaptic population; $X = \mathrm{E}$ for excitatory and $X = \mathrm{I}$ for inhibitory. The spike trains of the afferents of signal group μ are generated by

$$\nu_{\mathrm{X}\mu}(t) = \nu_{\mathrm{X}0}\left[y_\mu(t)\right]_+ + \nu_{\mathrm{Xbg}}, \tag{7}$$

where $\nu_{\mathrm{X}0}$ is the amplitude of the modulated firing rate fluctuations, and $[\cdot]_+$ is a rectifying function,

$$[y]_+ = \begin{cases} y, & \text{if } y > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Note that due to the symmetry of $y_\mu(t)$, an afferent is half the time in the background state and half the time in the active state.

Presynaptic action potentials were generated as an inhomogenous Poisson process according to the modified OU process described above and a fixed background firing rate. Additionally, we implemented a refractory period, τ_{Eref} for excitatory and τ_{Iref} for inhibitory inputs. Given the time step of the simulation Δt, spikes of a presynaptic afferent that is part of the signal group μ are generated with a probability $p_{\mathrm{X}\mu}(t) = \nu_{\mathrm{X}\mu}(t)\Delta t$ if there was no spike elicited during the refractory period beforehand, and thus the average firing rate of a $X = \mathrm{E}$ (excitatory) or $X = \mathrm{I}$ (inhibitory) afferent that is part of the signal group μ becomes

$$F_{\mathrm{X}\mu}(t) = \frac{1}{\Delta t} p_{\mathrm{X}\mu}(t)\left(1 - p_{\mathrm{X}\mu}(t)\right)^{\tau_{\mathrm{Xref}}/\Delta t}. \tag{9}$$

Pulse input. To test transient responses to brief changes in presynaptic activity we also quantified postsynaptic responses to pulse inputs. In this case, we simulated the postsynaptic neuron receiving inputs with constant background firing-rate. For 100 ms we increased the probability of presynaptic spikes for a given signal group μ by a factor $k\nu^*$, with k being an integer larger or equal than zero, and $\nu^* = 5$ Hz. Thus presynaptic spikes are generated by

$$\nu_{\mathrm{X}\mu}(t) = \alpha_{\mathrm{X}\nu} k \nu^* + \nu_{\mathrm{Xbg}}, \tag{10}$$

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

during the 100 ms step and by

$$\nu_{X\mu}(t) = \nu_{\text{Xbg}} \tag{11}$$

during only background activity. Parameter $\alpha_{X\nu}$ is a scalar that sets the ratio of excitatory and inhibitory firing rate.

Responses to the pulse input were divided in two bins: *phasic* and *tonic*. Phasic responses were defined as the postsynaptic activity elicited in the first 50 ms of the pulse input. Tonic activity was correspondingly defined as having occurred in the last 50 ms of the stimulus. We simulated 100 trials per input strength $k\nu^*$, and defined the response (for both phasic and tonic) as the average number of spikes on the period for the strength $k\nu^*$ minus the average number of spikes on the same period without extra input, multiplied by 20 to convert to Hz. We subtracted background spikes to ascertain that we quantified the response to the extra step input alone. Parameters used for inputs are detailed in Table II.

Synaptic tuning. Based on Vogels *et al.* [6], we used a receptive field profile given by

$$r(\mu) = \left(\frac{1}{1+r_0}\right) + \left(\frac{r_0}{1+r_0}\right)\left(\frac{1}{1+b\,(\mu-\mu_0)^c}\right), \tag{12}$$

where r_0, b and c are parameters defining the shape of the receptive field and μ_0 defines the preferred signal group, which maximises $r(\mu)$; $r(\mu_0) = 1$. Note that $r_0 \geqslant 1$, $b \leqslant 1$, $\mu_0 > 0$, and c is an even positive integer.

For simplicity we define ζ_j as the signal group that afferent j is part of. Thus excitatory synapses are set as

$$w_j = w_{\text{E0}} r(\zeta_j) + \epsilon_j, \quad j = 1, \ldots, N_\text{E}, \tag{13}$$

where w_{E0} is a normalisation factor for excitatory weights, and ϵ_j is a noise term drawn from a uniform random distribution between $-\epsilon^*_\text{E}$ and ϵ^*_E.

First we simulated a single inhibitory population with a tuned profile (Fig. 2), following Eq. 12 such that,

$$w_j = w_{\text{I0}} r(\zeta_j) + \epsilon_j, \quad j = N_\text{E}+1, \ldots, N, \tag{14}$$

where w_{I0} is a normalisation factor for inhibitory weights, and ϵ_j is a noise term drawn from a uniform random distribution between $-\epsilon^*_\text{I}$ and ϵ^*_I. The parameter w_{I0} was chosen so that a state of balance was enforced, with postsynaptic firing-rate of 5 Hz. Due to the small number of inhibitory afferents compared to the excitatory ones, and the difference in driving force, inhibitory weights were much larger than excitatory ones. Thus, to plot excitatory and inhibitory weights on the same scale we computed the *correcting* factor, α_w, from

$$\alpha_w = \frac{N_\text{E} \sum_{j=N_\text{E}+1}^{N} w_j}{N_\text{I} \sum_{j=1}^{N_\text{E}} w_j}. \tag{15}$$

In all plots with excitatory and inhibitory weights, we plotted excitatory weights multiplied by the parameter α_w.

Next, inhibitory afferents were divided in two types that connect to the postsynaptic neuron following different receptive field profiles. We used two approaches to sculpt the inhibitory synaptic weight profiles. In our first approach we defined them based on Eq. 12, i.e., we *hand-tuned* their shape. For our second approach we started them randomly and applied plasticity rules so that they produced the previously hand-tuned shapes through unsupervised learning.

We combined a co-tuned population with either a flat (Fig. 3) or a counter-tuned (Fig. 4) population. In both cases we based co-tuned weights on a modified version of Eq. 12,

$$r_\text{I}(\mu) = \frac{1}{1+b\,(\mu-\mu_0)^c}, \tag{16}$$

with b, μ_0 and c the same as for Eq. 12. Inhibitory weights for the co-tuned profiles are chosen so that

$$w_j = w_{\text{Ico}} r_\text{I}(\zeta_j) + \epsilon_j, \quad j = N_\text{E}+1, \ldots, N_\text{E} + (N_\text{I}/2), \tag{17}$$

where w_{Ico} is a normalisation factor for inhibitory weights following the co-tuned receptive profile, and are different when combined with either the flat or the counter-tuned populations.

We set the flat population such that

$$w_j = w_{\text{If}} + \epsilon_j, \quad j = N_\text{E} + (N_\text{I}/2) + 1, \ldots, N, \tag{18}$$

where w_{If} is the average for the flat population. The shape of the counter-tuned population was defined by

$$\bar{r}_\text{I}(\mu) = \frac{3}{2} r(1) - \frac{1}{2} r(\mu), \tag{19}$$

and synapses were hence tuned such that

$$w_j = \left[w_{\text{Icounter}} \bar{r}_\text{I}(\zeta_j) + \epsilon_j\right]_+, \quad j = N_\text{E} + (N_\text{I}/2) + 1, \ldots, N, \tag{20}$$

where $[\cdot]_+$ is a rectifier (Eq. 8), used to enforce only positive synaptic weights. The parameters w_{Icounter} is a normalisation factor for the counter-tuned inhibitory populations.

When plasticity was simulated, initial conditions for all plastic inhibitory populations were flat with noise (Fig. 6, Fig. 7 and Fig. 8),

$$w_j(0) = w_{\text{IF}} + \epsilon_j, \quad j = N_\text{E}+1, \ldots, N. \tag{21}$$

Parameters used for the tuning curves are detailed in Table II, and for synaptic weights in Table III. Both the average of the weights for flat population, w_{If}, and the noise term, ϵ^*_I, were distinct for different simulations.

Plasticity models. In this work we used three different inhibitory synaptic plasticity (ISP) rules. We termed them *Hebbian*, *scaling*, and *anti-Hebbian*. Both Hebbian and anti-Hebbian plasticity rules are triggered by pre- and postsynaptic spikes, and depend on a low-pass filter of these spike trains. The presynaptic trace (low-pass filter) is given by

$$\frac{\mathrm{d}x_j(t)}{\mathrm{d}t} = -\frac{x_j(t)}{\tau_\text{STDP}} + S_j(t), \tag{22}$$

where $x_j(t)$ is the value of the trace of the spike train of presynaptic afferent j at time t; τ_STDP is the time constant of the trace, and $S_j(t)$ is a sum of Dirac delta functions (Eq. 4) representing the spike train of afferent j. The same is considered for the postsynaptic neuron,

$$\frac{\mathrm{d}x_\text{post}(t)}{\mathrm{d}t} = -\frac{x_\text{post}(t)}{\tau_\text{STDP}} + S_\text{post}(t), \tag{23}$$

where $x_\text{post}(t)$ is the postsynaptic trace, and $S_\text{post}(t)$ is the spike train of the postsynaptic neuron (Eq. 5). Note that we used the same time constant for both pre- and postsynaptic traces.

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hebbian inhibitory plasticity. Precise balance of excitatory and inhibitory inputs was learned by a Hebbian inhibitory plasticity rule[6]. The weight of the jth inhibitory synapse changes according to

$$\frac{\mathrm{d}w_j(t)}{\mathrm{d}t} = \eta_{\mathrm{H}}\left[x_j(t)S_{\mathrm{post}}(t) + x_{\mathrm{post}}(t)S_j(t) - \alpha_{\mathrm{H}}S_j(t)\right], \tag{24}$$

where η_{H} is the learning rate, and α_{H} is a parameter that defines the postsynaptic firing-rate. The first two terms on the right-hand side of Eq. 24 are Hebbian terms that increase the weights when both pre- and postsynaptic activities are correlated. The last term on the right-hand side of Eq. 24 is a penalty term for inhibitory spikes alone, which creates a homeostatic set-point for the postsynaptic firing-rate given by

$$\rho_0 \approx \frac{\alpha_{\mathrm{H}}}{2\tau_{\mathrm{STDP}}}. \tag{25}$$

Later we describe how to arrive at this approximation.

Inhibitory synaptic scaling for flat tuning. One of the receptive profiles we used for inhibitory synapses was flat, i.e., every synapse group had the same strength. To learn the flat profile from random initial weights we implemented a scaling plasticity rule, partially based on experimental work that observed synaptic scaling on inhibitory synapses[38,39]. Weights are increased if the postsynaptic firing-rates are too high, and decreased otherwise,

$$\begin{aligned}\frac{\mathrm{d}w_j(t)}{\mathrm{d}t} =& \eta_{\mathrm{s}} w_{\mathrm{Is}}\left[y_{\mathrm{post}}(t) - \rho_0\right]\Theta\left(y_{\mathrm{post}}(t) - \alpha_{\mathrm{s}}\rho_0\right) \\ &- \eta_{\mathrm{s}} w_j(t)\left[\rho_0 - y_{\mathrm{post}}(t)\right]\Theta\left(\frac{\rho_0}{\alpha_{\mathrm{s}}} - y_{\mathrm{post}}(t)\right) +\end{aligned} \tag{26}$$

where η_{s} is a learning rate, ρ_0 is a firing-rate reference value, chosen to be the same as the one for Hebbian plasticity rule, $\Theta(\cdot)$ is the Heaviside function and α_{s} is a term that sets the firing-rate range for which synapses do not change. Postsynaptic neuron's firing-rate is computed with a slow averaging of the postsynaptic activity through

$$\frac{\mathrm{d}y_{\mathrm{post}}(t)}{\mathrm{d}t} = -\frac{y_{\mathrm{post}}(t)}{\tau_{\mathrm{scaling}}} + \frac{1}{\tau_{\mathrm{scaling}}}S_{\mathrm{post}}(t), \tag{27}$$

where τ_{scaling} is the time constant for the postsynaptic activity and $S_{\mathrm{post}}(t)$ is the postsynaptic spike train (Eq. 5). Note that the last term on the right-hand side of the equation above is divided by τ_{scaling} so that $y_{\mathrm{post}}(t)$ is in units of rate. Synaptic depression is weight dependent while synaptic potentiation is not, which ensures that all synaptic weights tend to the same value. When the postsynaptic neuron is firing below a threshold $\rho_0/\alpha_{\mathrm{s}}$, all inhibitory synapses in the flat group have their weights decreased proportionally to the difference between the target firing-rate and the average firing-rate, but also proportional to the current weight value. This way, strong synapses undergo stronger decrease than weak ones. Conversely, when the postsynaptic neuron is firing above a threshold $\alpha_{\mathrm{s}}\rho_0$, the same synapses increase in value by the same amount. Intuitively, these mechanism ensures that all synapses converge to the same value for a long run.

Anti-Hebbian inhibitory plasticity. The third inhibitory plasticity rule we used is an anti-Hebbian rule based on experimental data[21–23] and theoretical work on recurrent networks[24]. Synaptic weights change according to

$$\frac{\mathrm{d}w_j(t)}{\mathrm{d}t} = -\eta_{\mathrm{aH}}(t)\left[x_j(t)S_{\mathrm{post}}(t) + x_{\mathrm{post}}(t)S_j(t) - \alpha_{\mathrm{aH}}S_j(t)\right], \tag{28}$$

where $\eta_{\mathrm{aH}}(t)$ is a variable learning rate and α_{aH} is a parameter to counterbalance the anti-Hebbian term. The resulting rule dictates that coincident events decrease inhibitory synapses, while non-coincident ones increase synaptic weights. Due to the unstable nature of this plasticity rule (see details below), we implemented a time-varying learning rate which evolves according to

$$\frac{\mathrm{d}\eta_{\mathrm{aH}}(t)}{\mathrm{d}t} = -\frac{\eta_{\mathrm{aH}}(t)}{\tau_{\mathrm{aH}}} + M_{\mathrm{aH}}(t), \tag{29}$$

where τ_{aH} is the decay time constant for the learning rate, and $M_{\mathrm{aH}}(t)$ is an external signal to transiently activate plasticity. We speculate that such signal could come from modulatory neurons such as dopaminergic or cholinergic and assumed that the external signal peaks at a time t_0 (beginning of the simulation), so that

$$M_{\mathrm{aH}}(t) = \eta^*_{\mathrm{aH}}\delta(t - t_0), \tag{30}$$

where η^*_{aH} is the maximum learning rate before decaying to zero, and t_0 is the time when plasticity at these synapses are initiated. Parameters used for plasticity models are detailed in Table IV.

Mean-field analysis of the plasticity rules. We were interested in plasticity rules with stable dynamics. For a better intuition on fixed-point dynamics and stability we consider here a simplified dynamics of a mean-field model for both the Hebbian[6] and the anti-Hebbian models. We define the postsynaptic firing-rate as $\nu_{\mathrm{post}}(t)$ and the presynaptic firing-rates as $\nu_j(t)$. The traces of both presynaptic afferent and postsynaptic neuron thus have an average of $\tau_{\mathrm{STDP}}\nu_j(t)$ and $\tau_{\mathrm{STDP}}\nu_{\mathrm{post}}(t)$, respectively[64]. Neglecting any correlation between pre- and postsynaptic spikes, the average weight change for Hebbian synapses is given by

$$\left\langle\frac{\mathrm{d}w_j(t)}{\mathrm{d}t}\right\rangle = \eta_{\mathrm{H}}\left[2\tau_{\mathrm{STDP}}\nu_j(t)\nu_{\mathrm{post}}(t) - \alpha_{\mathrm{H}}\nu_j(t)\right], \tag{31}$$

where $\langle\cdot\rangle$ represents average over time. Intuitively, the postsynaptic firing-rate, $\nu_{\mathrm{post}}(t)$, changes negatively with changes in inhibitory weights — increased inhibition generates fewer postsynaptic spikes and vice-versa for decreased inhibition. This means that average firing rates are inversely linked to average inhibitory weights, i.e.,

$$\left\langle\frac{\mathrm{d}\nu_{\mathrm{post}}(t)}{\mathrm{d}t}\right\rangle \propto -\left\langle\frac{\mathrm{d}w_j(t)}{\mathrm{d}t}\right\rangle = 2\eta_{\mathrm{H}}\nu_j(t)\tau_{\mathrm{STDP}}\left[\frac{\alpha_{\mathrm{H}}}{2\tau_{\mathrm{STDP}}} - \nu_{\mathrm{post}}(t)\right]. \tag{32}$$

The steady state is computed by considering the vanishing point of the equation above (we assume that the presynaptic activity is nonzero), thus

$$\nu_{\mathrm{post}}(t) = \frac{\alpha_{\mathrm{H}}}{2\tau_{\mathrm{STDP}}} \equiv \rho_0. \tag{33}$$

This means that the postsynaptic activity $\nu_{\mathrm{post}}(t)$ increases (via decreases in inhibitory efficacy) when below ρ_0 and decreases when above ρ_0, creating a stable fixed-point for the postsynaptic firing-rate.

The opposite is true for the anti-Hebbian plasticity rule. Changes in postsynaptic firing-rate (with the same assumption as for the Hebbian plasticity rule) follow

$$\left\langle\frac{\mathrm{d}\nu_{\mathrm{post}}(t)}{\mathrm{d}t}\right\rangle \propto \nu_{\mathrm{post}}(t) - \frac{\alpha_{\mathrm{aH}}}{2\tau_{\mathrm{STDP}}} = \nu_{\mathrm{post}}(t) - \rho_1. \tag{34}$$

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Because postsynaptic activity increases when it is above threshold ρ_1 and decreases when it is below, this rule is unstable. The postsynaptic firing-rate eventually explodes or vanishes. We chose the simplest way to overcome these problems by setting a time-varying learning-rate. Other intricate mechanisms could be implemented, but this is not the scope of our work.

Correlation. We quantified the response of the postsynaptic neuron to natural inputs with the Pearson correlation between postsynaptic firing-rate and input firing-rate fluctuations, per signal group. We computed the firing rate of a signal groups as the low-pass filter of the spike trains of its excitatory afferents,

$$\tau_Z \frac{\mathrm{d}Z_\mu(t)}{\mathrm{d}t} = -Z_\mu(t) + \sum_{j\subset\mu} S_j(t), \tag{35}$$

where $Z_\mu(t)$ is the firing rate of the signal group μ at time t, filtered with a time constant τ_Z. The postsynaptic activity is also computed through a low-pass filter of its spike train,

$$\tau_Y \frac{\mathrm{d}Y(t)}{\mathrm{d}t} = -Y(t) + S_{\mathrm{post}}(t), \tag{36}$$

where $Y(t)$ is the activity of the postsynaptic neuron at time t filtered with a time constant τ_Y. The correlation is then computed as

$$C_\mu = \frac{cov(Z_\mu, Y)}{\sigma_{Z_\mu}\sigma_Y} = \frac{\langle (Z_\mu - \langle Z_\mu\rangle)(Y - \langle Y\rangle)\rangle}{\sqrt{\langle (Z_\mu - \langle Z_\mu\rangle)^2\rangle\langle (Y - \langle Y\rangle)^2\rangle}}, \tag{37}$$

where $cov(z, y)$ is the covariance between variables z and y, σ_z is the standard deviation of variable z, and $\langle\cdot\rangle$ represents time average.

Subsequently we computed a *performance index* ΔC as the difference between the correlation measure for preferred ($\mu = 9$) and non-preferred ($\mu = 1$) input signals,

$$\Delta C = \frac{1}{2}\left(C_9 - C_1\right). \tag{38}$$

Maximum positive performance index, $\Delta C = 1$, means that the preferred signal group has maximum correlation ($C_9 = 1$) while the non-preferred signal group has maximum anti-correlation ($C_1 = -1$), indicating that the postsynaptic neuron is responding solely to the preferred signal group. Consequently, $\Delta C = -1$, indicates that the postsynaptic neuron is responding solely to the non-preferred signal group. A flat response is indicated by $\Delta C = 0$. Note that maximum ΔC (either positive and negative) is only achievable if there is no overlap between activation of preferred and non-preferred input signals, which is never the case here. We define as *best performance* when $\Delta C = 0$ for all inhibitory inputs active (control), $\Delta C = 1$ (or $\Delta C > 0$) for one inhibitory population inactive, and $\Delta C = -1$ (or $\Delta C < 0$) when the other inhibitory population is inactive. Parameters used for computing correlations are detailed in Table V.

Implementation. Models were simulated with a time-step Δt, with either analytical or *semi-analytical* solution of the corresponding differential equation. All codes were written in Fortran, compiled with Intel Fortran Compiler 19.0, running on an Intel-based Linux computer (Debian 9; i9-9900X processor; 32 GB memory). Bellow we describe how each equation was implemented, with the parameter values in tables in the end of this section.

When not in the refractory period (see below), the leaky integrate-and-fire neuron is updated as

$$u^{n+1} = u^n_\infty + \left[u^n - u^n_\infty\right]\exp\left[-\frac{\Delta t}{\tau^n_{\mathrm{eff}}}\right], \tag{39}$$

where n is the iteration index, u^n_∞ and τ^n_{eff} are auxiliary variables described by

$$u^n_\infty = \frac{u_{\mathrm{rest}} + g^n_{\mathrm{E}}E_{\mathrm{E}} + g^n_{\mathrm{I}}E_{\mathrm{I}}}{1 + g^n_{\mathrm{E}} + g^n_{\mathrm{I}}} \tag{40}$$

and

$$\tau^n_{eff} = \frac{\tau_{\mathrm{m}}}{1 + g^n_{\mathrm{E}} + g^n_{\mathrm{I}}}. \tag{41}$$

This is the analytical solution when considering that all variables apart from $u(t)$ are constant during a time-step, which we refer to as *semi-analytical*.

When the membrane potential crosses a threshold from below, the membrane potential is reset (because of a spike being triggered), and kept at the reset potential for the duration of the refractory period,

$$u^m = u_{\mathrm{reset}}, \text{ if } u^n > u_{\mathrm{th}}, \tag{42}$$

where

$$m = n+1, n+2, ..., n + \frac{\tau_{\mathrm{ref}}}{\Delta t}. \tag{43}$$

Synaptic conductances are implemented as

$$g^{n+1}_{\mathrm{E}} = g^n_{\mathrm{E}}\exp\left(-\frac{\Delta t}{\tau_{\mathrm{E}}}\right) + \sum_j^{N_{\mathrm{E}}} w_j S^n_j \tag{44}$$

$$g^{n+1}_{\mathrm{I}} = g^n_{\mathrm{I}}\exp\left(-\frac{\Delta t}{\tau_{\mathrm{I}}}\right) + \sum_{j=N_{\mathrm{E}}+1}^{N_{\mathrm{E}}+N_{\mathrm{I}}} w^n_j S^n_j. \tag{45}$$

Note that here S^n_j is equal to one when afferent j spiked at time-step n and zero otherwise.

For natural inputs, we updated the auxiliary variable $y_\mu(t)$ every 1 millisecond,

$$y^{n+1}_\mu = \begin{cases} y^n_\mu \exp\left(-\frac{1\,\mathrm{ms}}{\tau_{\mathrm{OU}}}\right) + \xi^n_\mu, \text{ if } mod\left(n, \frac{1\,\mathrm{ms}}{\Delta t}\right) = 0 \\ y^n_\mu, \text{ otherwise,} \end{cases} \tag{46}$$

where $mod(\cdot,\cdot)$ is the modulo operation and ξ^n_μ is a random number drawn from a gaussian distribution with zero mean and unitary standard deviation. Presynaptic spikes are generated as point processes, so that at each time-step the probability of a presynaptic afferent to spike is

$$p^n_{\mathrm{E}\mu} = \left(\nu_{\mathrm{E0}}\left[y^n_\mu\right]_+ + \nu_{\mathrm{Ebg}}\right)\Delta t \tag{47}$$

and $p^n_{\mathrm{E}\mu} = 0$ during the $\tau_{\mathrm{Eref}}/\Delta t$ iterations after a spike. The same is valid for an inhibitory afferent; the probability of firing an action potential is

$$p^n_{\mathrm{I}\mu} = \left(\nu_{\mathrm{I0}}\left[y^n_\mu\right]_+ + \nu_{\mathrm{Ibg}}\right)\Delta t \tag{48}$$

and $p^n_{\mathrm{I}\mu} = 0$ during the $\tau_{\mathrm{Iref}}/\Delta t$ iterations after a spike.

For pulse inputs, presynaptic afferents were set to fire at background firing-rate and had an elevated firing-rate during a 100-ms period, which was varied in 5 Hz steps. For the activated pattern

$$p^n_{\mathrm{E}\mu} = \begin{cases} \left(\alpha_{\mathrm{E}\nu}k\nu^* + \nu_{\mathrm{Ebg}}\right)\Delta t, \text{ for } n = \left[n_0, n_0 + \frac{100\,\mathrm{ms}}{\Delta t}\right] \\ \nu_{\mathrm{Ebg}}\Delta t, \text{ otherwise,} \end{cases} \tag{49}$$

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where $\alpha_{\mathrm{E}\nu}$ adjusts the excitatory firing-rate, k is an integer for varying the pulse intensity, and n_0 is the iteration in which the pulse starts. The same implementation was used for inhibitory afferents ($\alpha_{\mathrm{I}_n u}$ being the parameter to adjust the inhibitory firing-rate),

$$p^n_{\mathrm{I}\mu} = \begin{cases} \left(\alpha_{\mathrm{I}\nu} k \nu^* + \nu_{\mathrm{Ibg}}\right)\Delta t, \text{ for } n = \left[n_0, n_0 + \frac{100\,\mathrm{ms}}{\Delta t}\right] \\ \nu_{\mathrm{Ibg}}\Delta t, \text{ otherwise.} \end{cases} \tag{50}$$

An afferent in an inactive pattern fires action potentials with background frequency ($p^n_{\mathrm{E}\mu} = \nu_{\mathrm{Ebg}}\Delta t$ and $p^n_{\mathrm{I}\mu} = \nu_{\mathrm{Ibg}}\Delta t$), and there is no spike elicit in the refractory period ($p^n_{\mathrm{E}\mu} = 0$ and $p^n_{\mathrm{I}\mu} = 0$ during the $\tau_{\mathrm{Eref}}/\Delta t$ and $\tau_{\mathrm{Iref}}/\Delta t$ iterations after a spike, respectively).

Plasticity was implemented with spike triggered events. For the Hebbian and anti-Hebbian plasticity rules, auxiliary variables changed as

$$x^{n+1}_{\mathrm{post}} = x^n_{\mathrm{post}} \exp\left(-\frac{\Delta t}{\tau_{\mathrm{STDP}}}\right) + S^n_{\mathrm{post}} \tag{51}$$

$$x^{n+1}_j = x^n_j \exp\left(-\frac{\Delta t}{\tau_{\mathrm{STDP}}}\right) + S^n_j, \tag{52}$$

where $S^n_{\mathrm{post}} = 1$ if the postsynaptic neurons generated an action potential at iteration n and zero otherwise. Hebbian weights changed according to

$$w^{n+1}_j = w^n_j + \eta_{\mathrm{H}}\left(x^n_{\mathrm{post}} - \alpha_{\mathrm{H}}\right) S^n_j + \eta_{\mathrm{H}} x^n_j S^n_{\mathrm{post}}, \tag{53}$$

and anti-Hebbian to

$$w^{n+1}_j = w^n_j - \eta^n_{\mathrm{aH}}\left(x^n_{\mathrm{post}} - \alpha_{\mathrm{aH}}\right) S^n_j - \eta^n_{\mathrm{aH}} x^n_j S^n_{\mathrm{post}}, \tag{54}$$

with the learning rate varying as

$$\eta^{n+1}_{\mathrm{aH}} = \eta^n_{\mathrm{aH}} \exp\left(-\frac{\Delta t}{\tau_{\mathrm{aH}}}\right) + M^n_{\mathrm{aH}} \tag{55}$$

with $M^0_{\mathrm{aH}} = \eta^*_{\mathrm{aH}}$, and $M^n_{\mathrm{aH}} = 0$ for $n > 0$.

Scaling was implemented with a different trace,

$$y^{n+1}_{\mathrm{post}} = y^n_{\mathrm{post}} \exp\left(-\frac{\Delta t}{\tau_{\mathrm{scaling}}}\right) + S^n_{\mathrm{post}}, \tag{56}$$

with weight update following

$$w^{n+1}_j = \begin{cases} w^n_j + \Delta t \eta_{\mathrm{s}} w^n_j \left[y^n_{\mathrm{post}} - \rho_0\right], \text{ if } y^n_{\mathrm{post}} < \rho_0/\alpha_{\mathrm{s}} \\ w^n_j + \Delta t \eta_{\mathrm{s}} w_{\mathrm{Is}} \left[y^n_{\mathrm{post}} - \rho_0\right], \text{ if } y^n_{\mathrm{post}} > \alpha_{\mathrm{s}}\rho_0. \end{cases} \tag{57}$$

Correlation-related variables were updated as

$$Z^{n+1}_\mu = Z^n_\mu \exp\left(-\frac{\Delta t}{\tau_Z}\right) + \sum_{j\in\mu} S^n_j \tag{58}$$

and

$$Y^{n+1}_\mu = Y^n_\mu \exp\left(-\frac{\Delta t}{\tau_Y}\right) + S^n_{\mathrm{post}}. \tag{59}$$

Parameter	Symbol	Value	Figs.
Membrane time constant	τ_{m}	30 ms	2 – 8
Resting potential	u_{rest}	−65 mV	2 – 8
Excitatory reversal potential	E_{E}	0 mV	2 – 8
Inhibitory reversal potential	E_{I}	−80 mV	2 – 8
Excitatory time constant	τ_{E}	5 ms	2 – 8
Inhibitory time constant	τ_{I}	10 ms	2 – 8
Spiking threshold	u_{th}	−50 mV	2 – 8
Reset potential	u_{reset}	−60 mV	2 – 8
Refractory period	τ_{ref}	5 ms	2 – 8
Simulation time step	Δt	0.1 ms	2 – 8

TABLE I. Simulation parameters for postsynaptic neuron.

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Parameter	Symbol	Value	Figs.
Number of excitatory afferents	N_E	3200	2 – 8
Number of inhibitory afferents	N_I	800	2 – 8
Number of signal groups	P	16	2 – 8
Refractory period for excitatory afferents	τ_{Eref}	5 ms	All
Refractory period for inhibitory afferents	τ_{Iref}	2.5 ms	All
Ornstein-Uhlenbeck process (OU) time constant	τ_{OU}	50 ms	All
Excitatory firing rate amplityde for OU	ν_{E0}	250 Hz	All
Inhibitory firing rate amplitude for OU	ν_{I0}	500 Hz	All
Excitatory background firing-rate	ν_{Ebg}	2 Hz	All
Inhibitory background firing-rate	ν_{Ibg}	4 Hz	All
Pulse amplitude reference	ν^*	5 Hz	2 – 5
Excitatory ratio for pulse input	$\alpha_{E\nu}$	1	2 – 5
Inhibitory ratio for pulse input	$\alpha_{I\nu}$	2	2 – 5
Receptive field profile amplitude	r_0	4	2 – 8
Receptive field profile slope	b	0.25	2 – 8
Preferred pattern index	μ_0	9	2 – 8
Receptive field profile power	c	2	2 – 8
Simulation time step	Δt	0.1 ms	All

TABLE II. Simulation parameters for inputs.

Parameter	Symbol	Value	Figs.
Excitatory baseline weight	w_{E0}	0.5	2 – 8
Noise parameter for excitatory weights	ϵ_E^*	0.01	2 – 8
Inhibitory baseline weight (one inh. population)	w_{I0}	1.04	2, 5
Inhibitory baseline weight (co-tuned & flat)	w_{Ico}	1.58	3, 5
Inhibitory baseline weight (co-tuned & flat)	w_{If}	0.52	3, 5
Inhibitory baseline weight (co- & counter-tuned)	w_{Ico}	2.43	4, 5
Inhibitory baseline weight (co- & counter-tuned)	$w_{Icounter}$	0.53	4, 5
Noise parameter for inhibitory weights	ϵ_I^*	0.01	2 – 5
Inhibitory baseline weight	w_{IF}	Varying	6, 7
Noise parameter for inhibitory weights	ϵ_I^*	0.01	6, 7
Inhibitory baseline weight (Hebbian & scaling)	w_{IF}	0.8	8A-C
Noise parameter for inhibitory weights (Hebbian & scaling)	ϵ_I^*	0.3	8A-C
Inhibitory baseline weight (Hebbian & anti-Hebbian)	w_{IF}	0.6	8D-F
Noise parameter for inhibitory weights (Hebbian & anti-Hebbian)	ϵ_I^*	0.01	8D-F

TABLE III. Simulation parameters for weights.

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Parameter	Symbol	Value	Figs.
STDP time constant	τ_{STDP}	20 ms	6 – 8
Hebbian learning rate	η_{H}	10^{-3}	6 – 8
Hebbian decay term	α_{H}	0.2	6 – 8
Firing-rate set-point	ρ_0	5 Hz	6 – 8
Anti-Hebbian initial learning rate	η^*_{aH}	10^{-3}	8
Anti-Hebbian learning rate time constant	τ_{aH}	200 s	8
Anti-Hebbian increase term	α_{aH}	0.6	8
Anti-Hebbian peak time	t_0	0 ms	8
Scaling time constant	τ_{scaling}	1000 ms	8
Scaling learning rate	η_{s}	5×10^{-7}	8
Scaling learning rate weight	w_{Is}	0.6	8
Scaling threshold parameter	α_{s}	2	8
Simulation time	—	30 mins	6 – 8

TABLE IV. Simulation parameters for plasticity rules.

Parameter	Symbol	Value	Figs.
Presynaptic time constant	τ_{Z}	10 ms	2 – 4, 7
Postsynaptic time constant	τ_{Y}	250 ms	2 – 4, 7
Simulation time	—	30 mins	2 – 4, 7

TABLE V. Simulation parameters for correlation measure.

bioRxiv preprint doi: https://doi.org/10.1101/729988; this version posted August 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary figures of

Complementary inhibitory receptive fields emerge from synaptic plasticity and create an attentional switch in sensory circuits

Everton J. Agnes, Andrea I. Luppi, and Tim P. Vogels

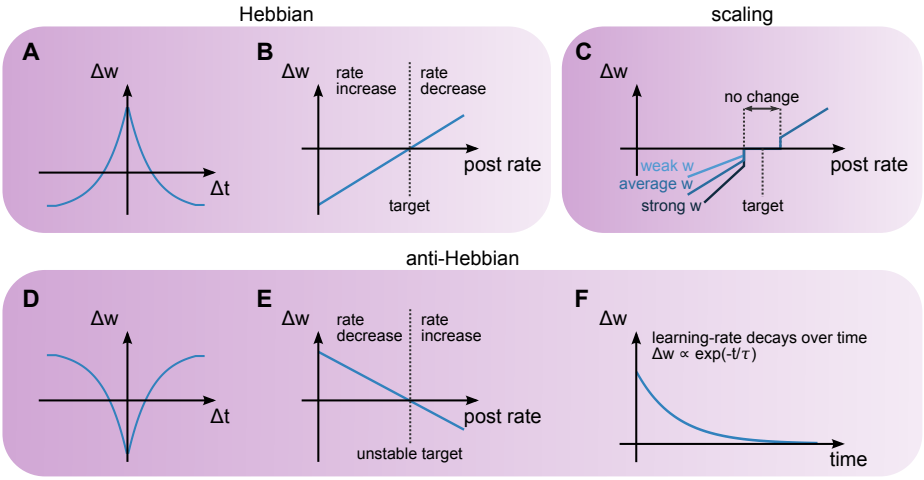

FIG. S1. Synaptic plasticity models. A, Spike-timing dependency of the Hebbian plasticity model. Δw indicates level of synaptic change, and Δt indicates interval between pre- and postsynaptic spikes. Coincident pre- and postsynaptic spikes elicit positive changes while presynaptic spikes alone elicit negative changes in synaptic strength (Eq. 24). **B**, Synaptic changes (Δw) as a function of postsynaptic firing-rate for the Hebbian plasticity model. When the postsynaptic neuron's firing-rate is above the target rate, inhibitory synapses increase in weight and, as a consequence, the postsynaptic neuron's firing-rate decreases. The opposite happens for when the postsynaptic neuron's firing-rate is lower than the target rate. See Eq. 31 to Eq. 33 for the mathematical analysis. **C**, Synaptic scaling model. Changes in synaptic strength (Δw) as a function of the postsynaptic neuron's firing-rate (Eq. 26). When the postsynaptic neuron's firing-rate is lower than a *lower bound* threshold, inhibitory synapses decrease, proportionally to their current strenght. When the postsynaptic neuron's firing-rate is higher than a *upper bound* threshold, inhibitory synapses increase. Because of the lower and upper bounds, a region with no change around the target rate is created. **D**, Spike-timing dependency of the anti-Hebbian plasticity model. Presynaptic spikes elicit positive changes, while coincident pre- and postsynaptic spikes elicit negative changes in synaptic weights (Eq. 28). **E**, Same as B for anti-Hebbian plasticity model. Because the anti-Hebbian plasticity model is a reversed version of the Hebbian plasticity one, the target rate becomes unstable. See Eq. 34 for mathematical analysis. **F**, Evolution of the learning-rate of the anti-Hebbian plasticity model. Due to its unstable nature, we set the learning-rate to decay exponentially over time (Eq. 29 and Eq. 30).