Inhibitory stabilization and cortical computation

Sadra Sadeh and Claudia Clopath ✉

Abstract | Neuronal networks with strong recurrent connectivity provide the brain with a powerful means to perform complex computational tasks. However, high-gain excitatory networks are susceptible to instability, which can lead to runaway activity, as manifested in pathological regimes such as epilepsy. Inhibitory stabilization offers a dynamic, fast and flexible compensatory mechanism to balance otherwise unstable networks, thus enabling the brain to operate in its most efficient regimes. Here we review recent experimental evidence for the presence of such inhibition-stabilized dynamics in the brain and discuss their consequences for cortical computation. We show how the study of inhibition-stabilized networks in the brain has been facilitated by recent advances in the technological toolbox and perturbative techniques, as well as a concomitant development of biologically realistic computational models. By outlining future avenues, we suggest that inhibitory stabilization can offer an exemplary case of how experimental neuroscience can progress in tandem with technology and theory to advance our understanding of the brain.

Effective connectivity
Measure of how two neurons (or populations) influence each other's responses, which can be modulated in different regimes by the dynamic change in the efficacy of connections.

Receptive fields
Specific regions in the sensory space (such as the visual field) to which individual neurons are most responsive.

Spontaneous activity
The activity of the brain in the absence of external stimuli, governed by the intrinsic dynamics of the brain.

Pattern completion
The process of reactivation of a pattern of activity in neuronal responses, even when the input is not completely provided.

Bioengineering Department, Imperial College London, London, UK.

✉*e-mail: c.clopath@imperial.ac.uk*

https://doi.org/10.1038/s41583-020-00390-z

A central question in neuroscience is which special features of the brain's circuitry enable it to perform such a vast range of functions from sensation to cognition. Single neurons are versatile and span a broad range of morphological and electrophysiological features[1–7] that endow them with complex and non-linear computational mechanisms[8–11]. However, what gives brains their special capacity is arguably the ability of such single units to connect, communicate and organize in functional neuronal networks. These networks are dynamic in turn. A complex repertoire of activity dynamics arises from the interaction of single neuron properties and network connectivity[12–17]. However, another level of dynamism emerges when plasticity of neuronal properties and synaptic weights are considered[18–20]. This enables the circuitry to dynamically reconfigure its effective connectivity[21,22] and provides it with a unique capacity to learn and adapt across multiple timescales.

Given the importance of such effective connectivity, an important question is which connectivity motifs enable different cortical circuits to perform their specialized function. A key dichotomy that has emerged regards the contribution of feedforward versus recurrent pathways. Original studies, in both experimental and computational neuroscience[23–25] as well as artificial intelligence algorithms[26–30], focused on the role of feedforward motifs in the formation of receptive fields and their hierarchical structure. However, later developments highlighted the importance of recurrent and feedback interactions. For example, computational studies have acknowledged the limitations of feedforward models that are based solely on bottom-up architectures, especially when facing challenging tasks and performing in more realistic conditions, such as those involving limited visibility, signal degradation and heavy occlusion[31–35].

In terms of neuroanatomy, although cortical neurons receive a substantial feedforward input drive[36–38], most of their input connections originate from within their cortical networks[39,40]. Recurrent interactions (originating in the same region) and feedback connections (from top-down sources) heavily modulate the activity of neurons at different stages of processing[41–48]. As a result, even in 'idle' brain states, in which external input is weak or absent, the brain generates intrinsic patterns of spontaneous activity that resemble evoked patterns[49–54]. Indeed, the modulation of ongoing activity by a stimulus has been reported to be smaller than what would be expected from a system primarily driven by external inputs, suggesting that internally generated activity within cortical networks is strong and dominant[55,56].

Theoreticians have indeed suggested that there could be several prominent roles for strongly connected recurrent networks[57–61]. High-gain excitatory networks can support important computations such as amplifying weak and noisy signals and enhancing incomplete feedforward inputs through pattern completion and temporal binding. Evidence suggests that recurrent input in primary sensory cortices can amplify the feedforward signal[37,38,62]. Likewise, subnetworks with strong recurrent interactions in the somatosensory cortex have

been implicated in input-specific amplification during behaviour[63]. Subnetworks of cortical neurons with specific recurrent connections are also involved in response recovery after input deprivation, arguing for a role of recurrent connectivity in the reliability and stability of neural coding[64]. These advantages of recurrent connectivity can, however, be lost if strongly coupled networks experience pathological reverberations and runaway activity (as manifest in epileptic modes of neuronal networks, for example in the hippocampus or the prefrontal cortex[65–70]). Efficient computations could therefore arise only in functional networks in which feedforward and recurrent motifs work in tandem, while the system as a whole is kept in the stable regime.

In this Review, we focus on cortical networks that have strong recurrent connections, in terms of both excitation and inhibition. Strong recurrent inhibitory connectivity protects such networks from the runaway activity that arises from strong recurrent excitation, in a phenomenon known as inhibitory stabilization. We describe the theoretical and empirical evidence for the existence of such inhibition-stabilized networks (ISNs) and their signatures (BOX 1), as well as factors contributing to their identification in the brain. We then review advances in computational modelling of ISNs, and discuss how inhibitory stabilization can be linked to different types of computation in cortical networks. By highlighting the role of inhibitory stabilization in normal and pathological states of neuronal activity, we aim to shed light on ISNs as a fundamental network motif in the cortex.

Inhibitory stabilization

The main feature of networks with strong recurrent excitatory connections is their capacity to maintain and amplify input patterns. Such high-gain networks, however, can make the brain more susceptible to instability. A mechanism is therefore needed to prevent unbounded amplification, ideally by tuning the network to have maximum gain while still in the stable regime (in randomly connected recurrent neural networks, this regime is known as the 'edge of chaos'[71,72]; but see REFS[73–75]) (BOX 2). Notably, networks with feedforward inhibitory motifs (FIG. 1a) cannot guarantee the operation of the network in the stable regime when instability arises from recurrent excitation; this is because inhibitory neurons do not receive the necessary feedback from the excitatory neurons that have the potential to become excessively excited. Thus, although feedforward inhibition is useful to control excessive activity in response to strong external stimuli[76,77], maintaining network stability in the face of intrinsic instability requires a recurrent inhibitory mechanism.

Indeed, theoretical work has suggested that such recurrent stabilization can arise dynamically and intrinsically from the interaction of excitation and inhibition in neuronal networks[13,78,79]. Excitatory–inhibitory balance emerges naturally in these networks, without the need for fine-tuning the parameters or adding extra metaparameters. Excitatory–inhibitory-balanced networks can, in principle, be in, or switch between, two distinct regimes. The first is a regime in which the recurrent excitation is of high gain and controlled by recurrent inhibition but would not necessarily lead to runaway instability if inhibition is weak or absent (FIG. 1b). In the second regime, recurrent excitation is so strong that it would be unstable in the absence of recurrent inhibition; thus, inhibition is necessary for stabilization of the network (FIG. 1c).

Networks in the latter regime, which necessitates a dynamic stabilization of unstable excitatory subnetworks by inhibition, are called 'inhibition-stabilized networks' (ISNs) (BOX 1; FIG. 1c). A prominent dynamical signature emerges in this regime, whereby increasing direct excitatory inputs to inhibitory neurons leads to a paradoxical decrease in the latter's activity[80]. By contrast, reducing the feedforward excitatory drive to inhibitory neurons increases their activity, because the recurrent excitatory–inhibitory interactions oppose the feedforward effects. Owing to the peculiar nature of this dynamical effect, the 'paradoxical effects' of perturbing inhibitory neurons — either by activating them or by suppressing them — has been suggested as a useful signature to probe the presence of ISNs in the brain.

Early experimental evidence

The first experimental hint of the presence of ISNs in the brain was inspired by modelling of the hippocampus[80]. It resulted from an unexpected observation during analysis of how theta rhythms modulate the activity of excitatory and inhibitory neurons. This modulation was modelled as an inhibitory input to inhibitory neurons, mimicking the strong inhibitory projections from the medial septum onto inhibitory neurons in the hippocampus[81]. Such an external drive (inhibition of inhibition) would be expected to result in the lowest level of inhibitory activity, and hence the highest level of excitatory activity, being aligned with the peak of a theta cycle. Therefore, excitatory and inhibitory activity should be out of phase (showing a 180° phase shift). In contrast to this straightforward prediction, however, the model showed in-phase excitation and inhibition, with both being out of phase with the external drive. In line with the model, experimental data from rats revealed a similar relation between the phases of theta modulation[80].

Further modelling and mathematical analysis revealed that this paradoxical effect can be explained only if the effective recurrent coupling of excitation is so high that the excitatory subnetwork is unstable in the absence of recurrent inhibition[80]. Under such a condition, the increase in the activity of excitatory neurons when they are disinhibited would be very large. In turn, inhibitory neurons would receive a strong recurrent excitatory drive, and hence show higher activity at the peak of a theta cycle, despite being inhibited by the external drive. It was therefore suggested that external perturbations that modulate the inhibitory subnetworks of other ISNs should result in similar paradoxical effects, and that such effects can distinguish ISNs from networks with feedforward-dominated structures and weakly coupled recurrent subnetworks.

The next experimental evidence for the presence of such paradoxical effects in the brain was provided about a decade later. It was proposed, following modelling and

Temporal binding
The process of binding isolated events in time to form a coherent temporal sequence (for example, by extending brief neuronal responses).

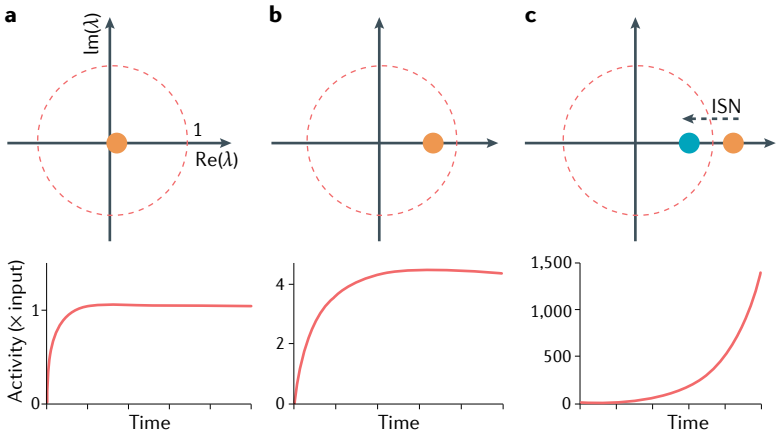
Box 1 | **Inhibition-stabilized networks: definition and signature**

Inhibition-stabilized networks (ISNs) can be described in terms of coupled rate-based equations of excitatory and inhibitory populations:

$$\tau d\mathbf{r}/dt = -\mathbf{r} + [W\mathbf{r} + \mathbf{s}]_+$$

where r is a vector of excitatory (E) and inhibitory (I) rates, $\mathbf{r} = \begin{pmatrix} r_E \\ r_I \end{pmatrix}$, and $\mathbf{s}$ is a vector of feedforward input to the two populations, $\mathbf{s} = \begin{pmatrix} s_E \\ s_I \end{pmatrix}$. The activation function can take different non-linear shapes; here we assume a half-wave rectification ($[\,.\,]_+$) for simplicity, reflecting the fact that neuronal activity cannot be negative. τ is the time constant of network integration, and W is a 2×2 weight matrix of coupling between subpopulations,

$$W = \begin{pmatrix} w_{E \leftarrow E} & w_{E \leftarrow I} \\ w_{I \leftarrow E} & w_{I \leftarrow I} \end{pmatrix}.$$



The equilibrium point of the network dynamics is obtained by solving the equation above for steady-state responses of excitatory and inhibitory populations, obtained by letting $dr_E/dt = 0$ and $dr_I/dt = 0$:

$$r_E = \left[W_{E \leftarrow E} r_E + W_{E \leftarrow I} r_I + s_E\right]_+,$$

$$r_I = \left[W_{I \leftarrow E} r_E + W_{I \leftarrow I} r_I + s_I\right]_+.$$

The equations describe how excitatory and inhibitory activity change for possible combinations of excitatory and inhibitory responses (excitatory and inhibitory nullclines in FIG. 1), with the equilibrium point of the system obtained by their crossing (the fixed point).

If we change the input to the network around this equilibrium point by infinitesimal perturbations ($\delta\mathbf{s}$), the changes in the output ($\delta\mathbf{r}$) can be written as $\delta\mathbf{r} = (I - W)^{-1}\delta\mathbf{s}$. If we perturb only inhibitory neurons — that is, $\delta\mathbf{s} = \begin{pmatrix} 0 \\ \delta s_I \end{pmatrix}$ — the change of inhibitory activity is obtained as

$$\delta r_I = \frac{1 - W_{E \leftarrow E}}{\Delta} \delta s_I,$$

where Δ is the determinant of $(I - W)^{-1}$. If excitatory–excitatory recurrent coupling is weak — that is, $W_{E \leftarrow E} \ll 1$ — the change in inhibitory activity will mirror the change in the input (FIG. 1a). Stronger but still stable excitatory–excitatory coupling ($W_{E \leftarrow E} < 1$) would dampen the external perturbations (FIG. 1b); however, the input and output still change in the same direction (that is, increasing the input increases the output, as normally expected). For unstable excitatory–excitatory coupling ($W_{E \leftarrow E} > 1$), a transition happens whereby changes in the input and output of inhibitory neurons will have opposite signs. That is, increasing the input to inhibitory neurons decreases their output activity: a paradoxical effect. The paradoxical effect of perturbing inhibitory neurons is therefore a dynamical signature of the instability of the excitatory subnetwork ($W_{E \leftarrow E} > 1$).

Note, however, that the excitatory–excitatory coupling is not the same as the connectivity. As the perturbations are performed around an equilibrium point of the linearized recurrent equations, W can be interpreted as the 'effective' connectivity. Therefore, it is a state-dependent parameter that can be modulated by background activity, attention or neuromodulation[218–220]. Effective connectivity can also be modulated by the number of active neurons. Theoretically, the instability of the excitatory network can be best assessed by spectral analysis. The spectrum of eigenvalues (characteristic scalars along the principal axes, or eigenvectors) are obtained from the linearized response gains around a certain operating regime. The largest eigenvalue (λ, orange in the figure) determines the stability of the excitatory subnetwork (the activity of which is schematized in the lower graphs in the figure), with feedforward networks having $\lambda \approx 0$ (see the figure, part a), strong but stable excitatory subnetworks described by $0 < \lambda < 1$ (see the figure, part b) and unstable excitatory subnetworks happening at $\lambda > 1$ (see the figure, part c). Inhibitory stabilization brings this eigenvalue back to the stable region. The instability of the eigenvalue can happen for the global mode of connectivity, where the non-specific component of synaptic weights between excitatory neurons is considered, or for specific modes of excitatory–excitatory connectivity.

experiments, that surround suppression in the cat visual cortex can be explained only if the visual networks operate as ISNs[82]. Surround suppression has conventionally been thought to be mediated by lateral inhibition, whereby increasing the size of a stimulus recruits networks of neurons activated by the 'surround' stimulus, which suppress laterally the activity of neurons responding to the central part of the stimulus. If such a mechanism governed surround suppression, a stimulus extending to the surround would be expected to elicit an increase in the total inhibitory input and a decrease in the total excitatory drive. However, experimental results have shown that both excitatory and inhibitory inputs are decreased during surround suppression[82].

Modelling suggested that this can be explained by a mechanism based on inhibitory stabilization[82]. If local networks operate as ISNs, the excitatory drive from surround neurons onto inhibitory neurons at the centre would decrease the activity of both excitatory and inhibitory neurons in the centre network. This would be another manifestation of the paradoxical effect of modulating inhibitory neurons, in which the modulation is now obtained by an increase in the external drive to inhibitory neurons (as opposed to the decrease in the external drive, arising from theta oscillations in the hippocampus[80]). It was therefore concluded that, rather than resulting from an increase in lateral inhibition, surround suppression is mediated by the withdrawal of excitation from central ISNs[83].

Surround suppression
Suppression of neuronal responses by enlarging the stimulus, or adding a sensory stimulus outside the classical receptive field that would not activate the neurons if presented alone.

Probing through perturbation

In the studies described in the previous section, an internal mechanism in the brain (such as theta oscillations or surround suppression) was recruited to indirectly

Box 2 | **Tuning high-gain networks by plasticity rules**

Which plasticity rules can enable cortical networks to operate in their high-gain regimes? Classic Hebbian-type plasticity rules can enhance the recurrent coupling of a network by potentiating the synaptic weights between co-active excitatory neurons, but this can easily lead to runaway instability[221,222]. Efficient tuning of neuronal networks must therefore entail a balance of potentiation and homeostasis that regulates the network in functional and stable regimes. This tuning can apply to global modes of connectivity, whereby the general excitatory-to-excitatory connections are regulated, but also to specific modes, comprising subnetworks of neurons involved in performing specific computations and tasks. Such specific subnetworks have, for example, been reported in the mouse visual cortex[136–138,223], where connection weights between excitatory neurons with similar receptive fields are particularly strong[136]. Just a few such connections, if activated simultaneously, are therefore enough to bring a neuron to its firing threshold, and this would soon lead to instability as a result of the positive feedback between highly connected cells in a subnetwork. Simulations of large-scale excitatory networks with experimentally reported weights corroborate such unstable regimes, in the absence of stabilizing mechanisms[139].

How can such strong and specific subnetworks be formed with local plasticity mechanisms to be simultaneously high-gain and stable? One problem to address is the rapid onset of dynamic instability resulting from runaway excitation, which is typically much faster than the timescale of plasticity and homeostasis[224]. If not fast enough, plasticity mechanisms could lead to oscillations between unstable activity and silent states. Another problem is the distributed nature of network-wide instability, as unstable modes emerge at the network level. Therefore, local information available to biological neurons cannot fully advise a tuning mechanism to operate effectively on the relevant dimensions of activity and connectivity. For example, homeostatic mechanisms that scale inputs to individual neurons and regulate their activity[225] may not be able to distinguish between the inputs that are the source of the specific instability and other connections from non-specific sources. A robust stabilizing mechanism may therefore need to regulate the synapses along specific distributed modes that become unstable (for example, within a specific subnetwork).

Inhibitory stabilization can potentially offer a solution to these problems. It is fast, as it is governed by the dynamics of excitatory–inhibitory interactions. Moreover, it can stabilize distributed modes of activity by matching inhibition to control respective unstable modes[142,143,183] (for example, by potentiating specific excitatory–inhibitory connections in a specific excitatory subnetwork). Notably, this approach is different from scaling synaptic weights or limiting single-neuron activity. Here the weights of specific excitatory and inhibitory subnetworks can grow without an immediate upper bound. Controlling instability is instead achieved by balancing excitation with inhibition, either globally or within subnetworks, by forming functionally specific inhibition-stabilized networks[110]. The strong and specific connectivity of excitatory–inhibitory connections reported in the same visual circuits described above[139] and the task specificity of excitation and inhibition during learning[184,185] might indeed be evidence of such a tuning strategy.

How exactly different excitatory–inhibitory plasticity rules[142–144,226–228] can work in large-scale networks to provide such balanced states and shape inhibition-stabilized network regimes remains to be studied. Controlling excitation is necessary to avoid unstable regimes, but to achieve stable high-gain networks it is also important to let the excitatory subnetworks grow in the first place. Proper growth–control balance might require specific learning strategies at different stages of development. Consistent with this hypothesis, a switch from anti-Hebbian (depression) to Hebbian (potentiation) learning of inhibitory–excitatory synapses has been reported at the onset of a critical period of development in rats[229], and local networks in the mouse hippocampus have been shown to switch between high-inhibition and low-inhibition configurations during learning[230]. Future experimental and theoretical work could cast light on how plasticity can shape functional networks in inhibition-stabilized network regimes, for example by studying how excitatory–inhibitory subnetworks emerge sequentially during learning[184,185] and development[136,139].

infer the presence of ISNs through observations of the paradoxical effects of modulating inhibitory neurons. However, until the advent of optogenetics[84–86], direct testing of the effects of perturbing inhibitory neurons remained elusive. The first piece of evidence in this line was presented in passing, when optical perturbation was used to study how parvalbumin-expressing (PV^+) interneurons modulate the gain of pyramidal cells in the mouse visual cortex[87]. The study reported no paradoxical effects resulting from perturbing inhibition[87]. This led to the conclusion, for a while, that the operating regime of cortical networks in rodents might be different from that in other species, such as cats or monkeys.

However, later studies that focused more specifically on the detection of paradoxical effects reached a different conclusion. Two contemporaneous studies in awake, behaving mice combined optogenetic stimulation with subthreshold measurements of the net excitatory and inhibitory inputs to pyramidal cells, and reported paradoxical effects in the auditory cortex and visual cortex, respectively[88,89]. In the auditory cortex, perturbing PV^+ or somatostatin-expressing (SOM^+) interneurons could evoke paradoxical effects[88]. Optogenetic inhibition of either subtype transiently decreased the total inhibitory input to pyramidal cells, but this was rapidly followed by a much greater increase in the total inhibitory input[88] (FIG. 2a). This implies that the paradoxical effect arises as a result of recurrent interactions and reverberating activity, which is delayed relative to the initial feedforward response. Such recurrent interactions were shown to exert a potent lateral inhibition, whereby 'surround' (that is, non-preferred) tones suppress both excitatory and inhibitory synaptic inputs to the centre, which is indeed important at the network level for functional processing of auditory signals[88]. Thus, as in the cat visual cortex, surround suppression in the auditory cortex of mice might also be governed by ISN dynamics.

Inhibitory stabilization and its contribution to surround suppression was also observed in the mouse visual cortex[89]. Optogenetically suppressing the activity of SOM^+ interneurons led to an increase in the net inhibitory current impinged on pyramidal cells, demonstrating the paradoxical effect at the synaptic level[80,90]. Moreover, increasing the size or contrast of a visual stimulus increased both the excitatory and the inhibitory inputs to pyramidal cells, but remarkably decreased the excitatory-to-inhibitory ratio (that is, driving inhibition dominance), as predicted by some ISN models[83,91]. Thus, direct optogenetic perturbation of inhibitory neurons, combined with subthreshold measurements, revealed the ISN dynamics underlying sensory processing in awake mice.

Several subsequent studies provided further evidence for the presence of ISNs and paradoxical effects in mouse cortical networks. One study reported paradoxical effects of perturbing inhibitory neurons in the somatosensory cortex of awake mice, but only using photoinhibition protocols with the right strength and spatial scale of photostimulation[92]. This result corroborated the prediction of previous theoretical work that suggested that a large fraction of inhibitory neurons needs to be perturbed to observe the paradoxical effects[93,94]. Another study[95] found paradoxical effects across multiple areas (in the visual, somatosensory and motor cortices) and different states (evoked as well as spontaneous regimes of activity) (FIG. 2b). This study, too, confirmed the importance of how widely inhibitory neurons are perturbed for revealing ISNs: paradoxical effects were observed only in transgenic

Optogenetics
Use of light to increase or decrease the activity of neurons through optical stimulation of light-sensitive ion channels.

mice expressing the light-sensitive hyperpolarizing channel in PV$^+$ inhibitory neurons, but not in mice in which the channel was expressed through use of a targeted virus, which showed a lower level of channel-expressing PV$^+$ cells[95]. However, the study did not find any evidence for the transition from a non-ISN regime of activity (for weak drives) to ISN regimes (for strong feedforward inputs), as suggested by some models[83,91].

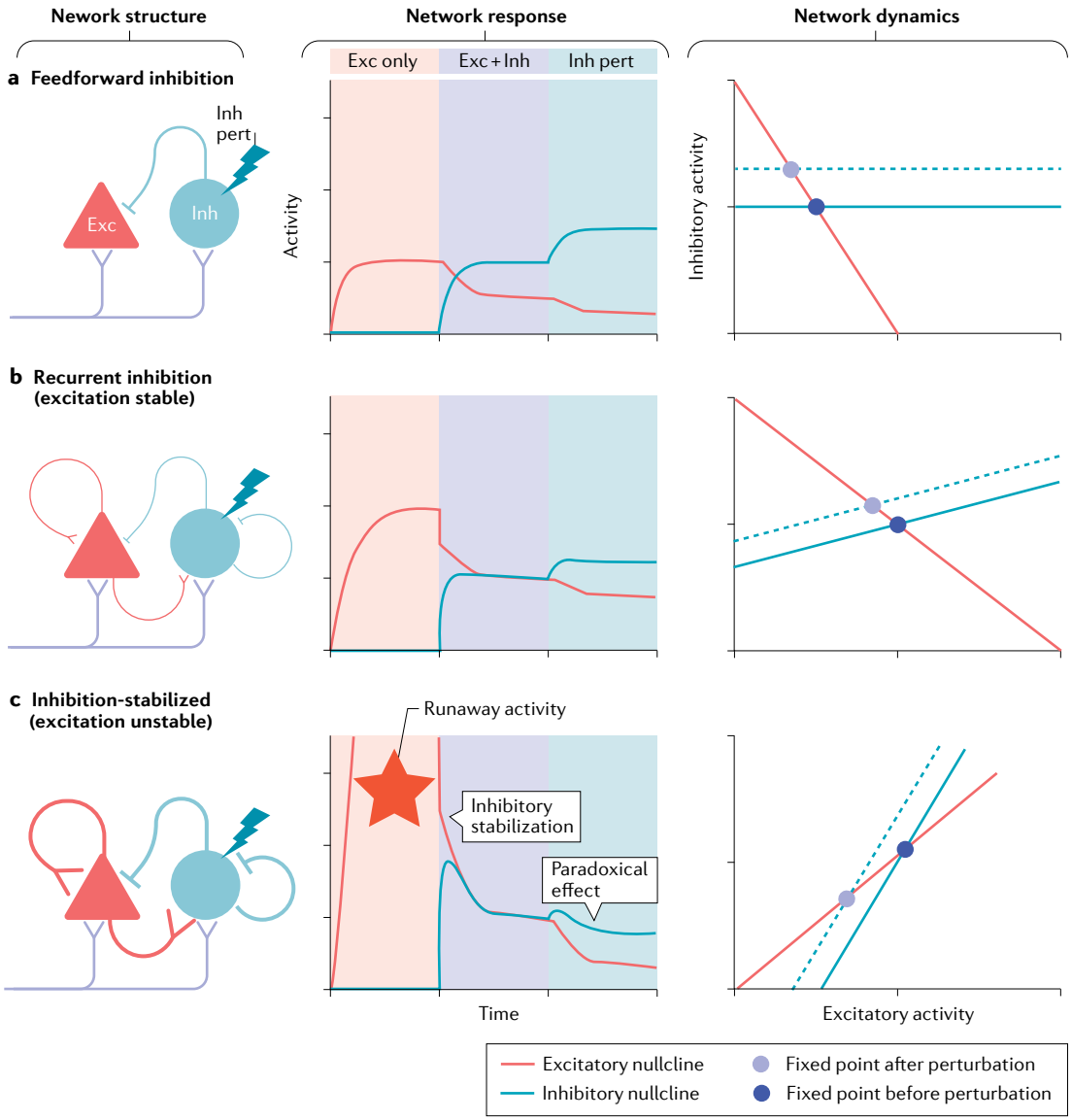
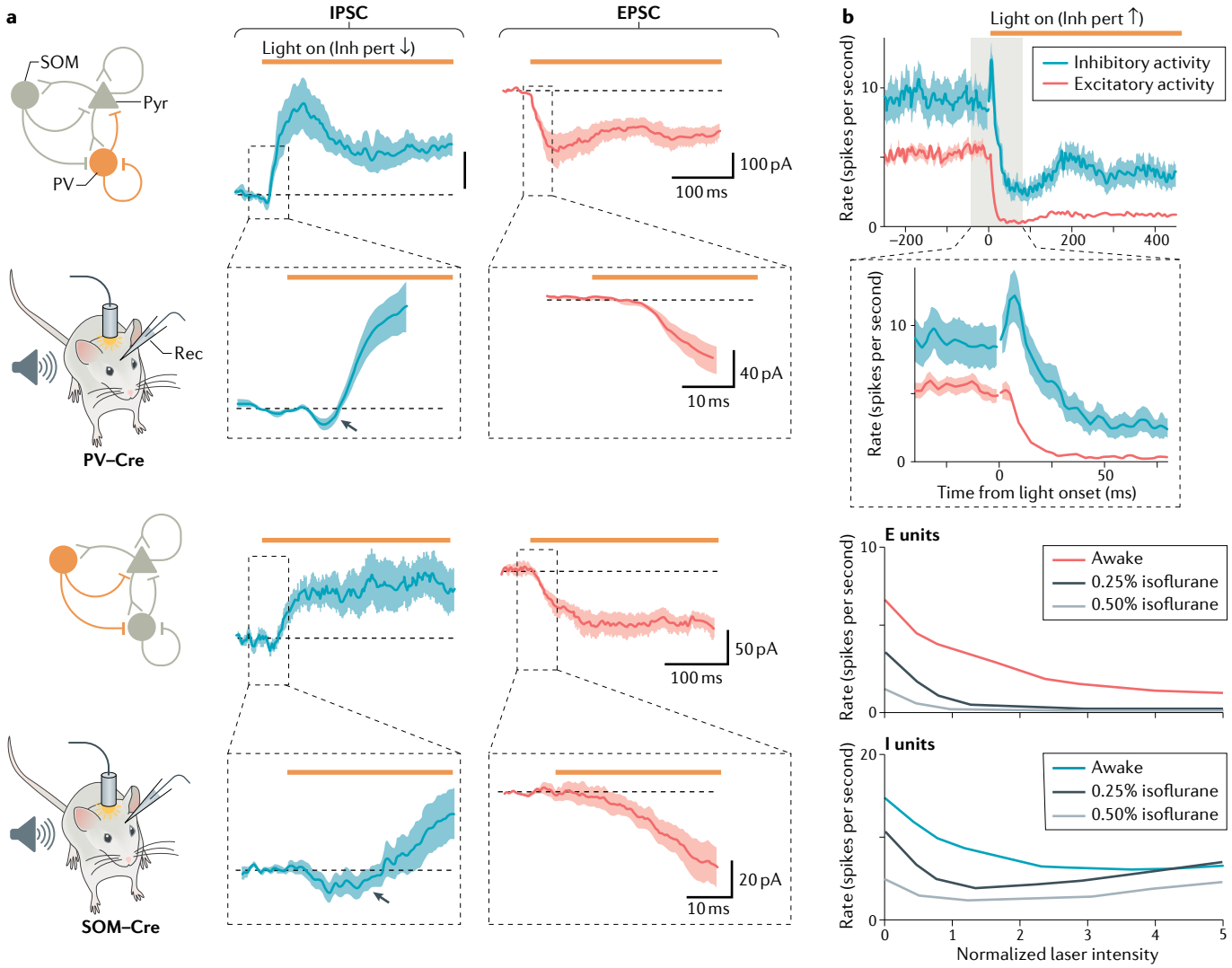
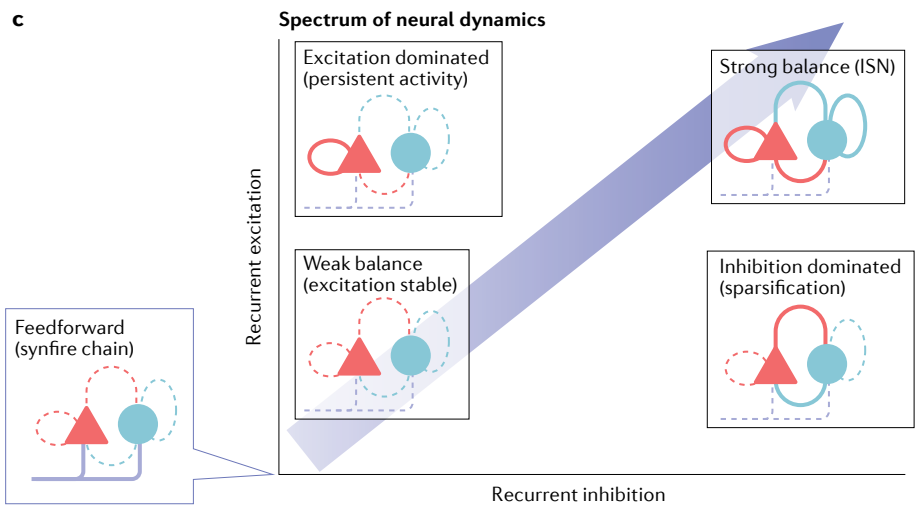
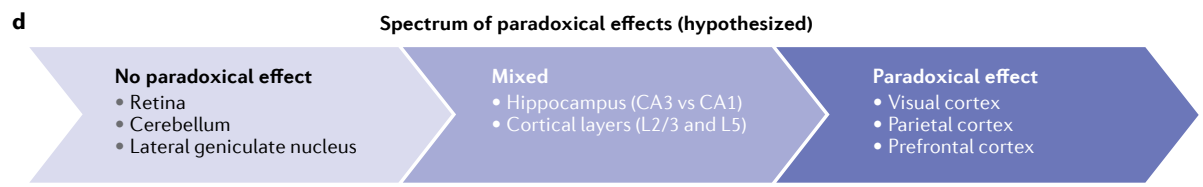


Fig. 1 | **Inhibition-stabilized dynamics emerge in networks with strong recurrent excitation and inhibition. a** | Left: Schematic of a network with feedforward excitation (to excitatory (Exc; red) and inhibitory (Inh; blue) subnetworks) and feedforward inhibition (mediated via Inh to Exc). Middle: the response dynamics of excitatory (red line) and inhibitory (blue line) subnetworks, for three conditions: response dynamics of the excitatory subnetwork in isolation, in the absence of the inhibitory subnetwork (Exc only); when the inhibitory subnetwork and its interaction with the excitatory subnetwork is added to the network (Exc + Inh); and when an external perturbation increases the activity of the inhibitory subnetwork (Inh pert). Right: analysis of response dynamics in the network based on the fixed point of the system (solution of the excitatory and inhibitory rate equations in the steady state) before and after perturbation. The excitatory and inhibitory nullclines (dimensions along which the activity of excitatory and inhibitory subnetworks remains constant, respectively) are shown by the solid lines. The inhibitory nullcline changes after inhibitory neurons have been perturbed (shown by the dashed line), leading to a new fixed point. **b** | Same as part **a** but for a network with weak recurrent interaction of excitation and inhibition. Weak recurrent excitation means that the activity of the excitatory subnetwork is stable even in the absence of inhibition, and that increasing the activity of inhibitory neurons with the perturbation decreases the activity of excitatory neurons without any paradoxical effect (middle and right). **c** | Same as part **a** but for a network with strong recurrent connectivity. In the absence of inhibition, the excitatory network shows unstable activity that is stabilized only when the inhibitory subnetwork is included. Perturbing inhibitory neurons transiently increases the activity of the inhibitory subnetwork, but when recurrent interactions occur, the activities of both the excitatory subnetwork and the inhibitory subnetwork decrease in the steady state (middle). The activity of the inhibitory subnetwork thus changes in a paradoxical direction. This can be explained in terms of the change in the slope of the excitatory nullcline (right), compared with networks with weak balance (in part **b**) or feedforward inhibition (in part **a**).


a
IPSC
EPSC
Light on (Inh pert ↓)
SOM
Pyr
PV
Rec
PV–Cre
100 pA
100 ms
40 pA
10 ms
50 pA
100 ms
20 pA
10 ms
SOM–Cre
b
Light on (Inh pert ↑)
Inhibitory activity
Excitatory activity
Rate (spikes per second)
Time from light onset (ms)
E units
Awake
0.25% isoflurane
0.50% isoflurane
I units
Normalized laser intensity


c
Spectrum of neural dynamics
Excitation dominated (persistent activity)
Strong balance (ISN)
Weak balance (excitation stable)
Inhibition dominated (sparsification)
Feedforward (synfire chain)
Recurrent excitation
Recurrent inhibition


d
Spectrum of paradoxical effects (hypothesized)
No paradoxical effect
• Retina
• Cerebellum
• Lateral geniculate nucleus
Mixed
• Hippocampus (CA3 vs CA1)
• Cortical layers (L2/3 and L5)
Paradoxical effect
• Visual cortex
• Parietal cortex
• Prefrontal cortex

◀ Fig. 2 | **Inhibitory stabilization in the brain. a** | Paradoxical effects of perturbing inhibitory neurons in the auditory cortex of awake mice. Inhibitory neurons expressing parvalbumin (in PV–Cre transgenic mice, top) or somatostatin (in SOM–Cre transgenic mice, bottom) are optogenetically suppressed (Inh pert ↓) while the inhibitory postsynaptic currents (IPSCs) and excitatory postsynaptic currents (EPSCs) of pyramidal cells (Pyr) are measured by subthreshold recordings (Rec). Note that an initial, transient reduction in inhibitory input after perturbation (arrows in the blow-up panels) precedes the paradoxical increase in IPSC (cf. FIG. 1c). **b** | Top: average activity of excitatory and inhibitory units following perturbation of inhibitory neurons by optogenetic activation (Inh pert ↑). The blow-up panel highlights the transient increase in inhibitory activity (relative to the baseline) before paradoxical suppression. Bottom: the average activity of excitatory units (E units) and inhibitory units (I units) in response to different normalized laser intensities and in different states: awake and under two levels of anaesthesia (0.25% isoflurane and 0.5% isoflurane). An inhibitory rate less than the value without optogenetic activation (at zero laser intensity) suggests the observation of a paradoxical effect at the respective laser intensity. Detection of inhibition-stabilized networks (ISNs) thus becomes more difficult for deeper levels of anaesthesia (such as 0.5% isoflurane). **c** | The spectrum of neural dynamics and paradoxical effects. Neuronal networks in different brain regions can have different combinations of feedforward and recurrent strengths. Networks with dominant feedforward architectures (as in synfire chains and perceptrons) are not expected to show paradoxical effects, as networks with weak balance of excitation and inhibition, where recurrent excitation is stable by itself. Strengthening the recurrent interaction between excitation and inhibition brings about inhibition-dominated dynamics, which can enable computations such as lateral inhibition and sparsification in the network, but still would not lead to paradoxical effects if the recurrent excitation is weak. By contrast, strengthening recurrent excitation on its own can lead to excitation-dominated dynamics, which can support the emergence of persistent activity, but would still not manifest itself in paradoxical effects if stabilization is not provided by recurrent excitatory–inhibitory interaction. Paradoxical effects of perturbing inhibition emerge only in ISNs with strong, unstable excitation that is stabilized by a strong recurrent inhibition. **d** | The spectrum of neural dynamics in different brain regions can be tested according to a hypothesized spectrum of paradoxical effects. Part **a** is adapted with permission from REF.[88], Elsevier. Part **b** is adapted from REF.[95], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Factors affecting ISN detection

What factors contribute to the detection of ISNs, and why did the initial optogenetic study[87] not detect the paradoxical effect?

Anaesthetized versus awake states. The earlier optogenetic study of the effects of suppressing PV^+ neurons, which did not observe a paradoxical effect, was performed in mice anaesthetized with a high concentration (2%) of isoflurane[87]. By contrast, all of the more recent optogenetic experiments that reported paradoxical effects in the mouse cortex were performed in awake animals[88,89,92,95]. One of the latter studies[95] reported similar paradoxical effects during wake and under light anaesthesia (0.25% isoflurane), but a much weaker effect with deeper anaesthesia (0.5% isoflurane), especially with higher laser intensities (that is, strong perturbations) (FIG. 2b). The transition from the anaesthetized state to the wake state might therefore be an important parameter contributing to the detection of ISNs, as anaesthesia can modulate their effective connectivity[96,97]. Theoretically, effective excitatory–inhibitory coupling can be approximated by small perturbations of the linearized dynamics of the system around an equilibrium point (see BOX 1); given the dependence of this approximation on the actual equilibrium point, effective coupling can thus heavily depend on factors such as neuromodulation, behavioural state and background activity, which modulate the operating regime.

Transfer functions
Functions describing how the output activity of a neuron changes with different input values.

Spontaneous versus evoked activity. Dependence of the effective connectivity on the operating regime implies that ISN detection can also be affected by spontaneous and evoked regimes of activity. If more neurons are active during evoked states, owing to higher input drives, the effective size and ultimately the effective connectivity of the network increases, which can potentially improve the detectability of ISNs. Another factor that affects the detection of ISNs is change in the gain of neurons. If neurons respond with higher gains at higher levels of activity — for example, as might be expected from supralinear transfer functions[83,91] — the effective coupling of the network increases in the evoked state. This increase in effective coupling can also cause a transition to ISN-like behaviour. Experimental results, however, show that paradoxical effects can be observed in both the spontaneous state and the evoked state[95], suggesting that the transition to the ISN regime may happen at lower levels of activity, at least for the global mode arising from non-specific stimulation of the network.

Nevertheless, an argument can be made for why spontaneous regimes of activity, or evoked responses with weak stimuli (such as low-contrast gratings or natural images), can potentially help the detection of ISNs. Excitatory neurons in sensory cortices show sparse and highly selective responses, whereby they are activated by only a select range of preferred stimuli[98,99]. Therefore, if the network is driven by strong stimuli (such as high-contrast visual gratings), a specific subnetwork of excitatory neurons is activated during evoked states. This selectivity leaves the rest of the network almost inactive, which can in turn reduce the effective size of the excitatory subnetwork (that is, equivalent to a smaller network with less active neurons) (BOX 1) and thus reduce the chance of observing ISN behaviour. However, this picture might become more complicated in networks with strong functional subnetworks, as the enhancement of effective connectivity within individual subnetworks may compensate for the suppression of weakly connected synapses that originate from other subnetworks. Further work is therefore needed to assess the relative contribution of these two opposing factors (number of synapses versus their strength) in cortical networks with specific connectivity.

Pattern of perturbation. Similarly to input strength and stimulus pattern, the pattern of experimental perturbations could also be an important factor in probing inhibitory stabilization. As discussed above, modelling and experimental studies have already demonstrated that the strength and spatial extent of perturbations of inhibition, and the proportion of inhibitory cells affected, can contribute considerably to the detection of ISNs[92,93,95,100]. Whereas in classical models, the entire inhibitory cell population was treated as one entity and perturbed as a whole, theoretical work showed that, in more detailed models of ISNs, the emergence of paradoxical effects can depend on the fraction of perturbed inhibitory neurons[93]. When a set of inhibitory neurons is perturbed (I_p), the effect of perturbation depends on not only the recurrent interaction between the perturbed inhibitory population and excitation ($I_p \Leftrightarrow E$ loop) but also more

complex motifs involving the non-perturbed inhibitory subpopulation (I_{NP}), including disinhibitory loops (I_P–I_{NP}) and trisynaptic motifs (I_P–E–I_{NP})[101]. Mathematical analysis[93] shows that, if a minimum fraction of inhibition is not perturbed, the recurrent feedback recruited as a result of overall excitatory–inhibitory interaction cannot dominate the change in the feedforward drive, and hence no paradoxical effect is observed. In the mouse visual cortex, for instance, a perturbation of approximately 70% of inhibitory neurons in the local network (or perturbations with a minimum spatial width of approximately 250 μm) was estimated to be needed[93]. It is therefore plausible that the effective perturbation of inhibitory neurons has been the main difference between experiments with different results, and hence proper tuning of the perturbation protocols is necessary when one is assessing paradoxical effects in ISNs.

This dependence of the perturbation experiments on the strength and pattern of perturbations had been observed in other studies too[102,103]. It was suggested as an explanation for why different laboratories obtained inconsistent results when probing the contributions of different subtypes of inhibitory neurons to cortical computation[104–106]. Given the operating regime of sensory cortices, with strong inhibition dominance[107,108] and sparse activity of pyramidal cells[98,109], strong activation of inhibitory neurons can soon lead to the so-called floor effect, whereby many neurons in the network are silenced. This could affect neural computations[104–106], but could also mask the presence of ISNs, by decreasing the number of active neurons in the network. This can, in turn, decrease the effective coupling of the network, and therefore the remaining excitatory subnetwork might not be unstable any more.

Decreasing the excitatory input to inhibitory neurons might be a better strategy to probe ISNs in regions with a sparse regime of activity, as it would disinhibit the activity of excitatory neurons without an immediate upper bound. This in turn recruits the recurrent interaction within the network more effectively, which is crucial to reveal excitatory reverberation and subsequent inhibitory stabilization. It is interesting that the initial experimental evidence for ISNs came from theta oscillations, which involve inhibition of inhibition[80]. Indeed, modelling shows that negative perturbations of inhibitory interneurons (that is, by decreasing their excitatory inputs or increasing their inhibitory inputs) can more reliably reveal paradoxical effects compared with perturbation protocols that activate interneurons[93,110]. Care should be taken, however, not to induce negative perturbations too strongly, as potent disinhibition can destabilize an ISN regime and lead to pathological oscillations.

Subtypes of inhibitory neurons. Lastly, the presence of multiple subtypes of inhibition in the cortex[2,111–114] may facilitate, but can also complicate, the detection of ISNs. The potent inhibition provided by PV^+ inhibitory neurons and other subtypes, such as SOM^+ neurons or vasoactive intestinal polypeptide-expressing (VIP^+) neurons, can enhance inhibitory stabilization. The crucial role of SOM^+ interneurons, in addition to PV^+ cells, has been evident in recent experimental studies that argue for the presence of ISNs in sensory cortices[88,89]. Similarly to the previous results from the cat visual cortex[82], both studies in mice[88,89] investigated ISNs in the context of surround suppression, in which the role of SOM^+ inhibitory neurons is central[115]. However, the complex interaction of different subtypes of interneurons, especially the inhibition of PV^+ neurons by SOM^+ interneurons[116], can complicate the interpretation of paradoxical effects. Indeed, a recent study argued that the paradoxical effects of perturbing inhibitory interneurons can be explained by networks without unstable excitatory subnetworks, hence raising the possibility that paradoxical effects may coexist with non-ISN regimes of activity dynamics, owing to the complex interaction of more than one type of inhibitory subnetwork[117]. This picture can be further complicated by the strong interaction of excitation and inhibition across cortical layers[118]. Further experimental and theoretical work is thus needed to shed light on the contribution of inhibitory stabilization and disinhibitory mechanisms to the emergence of paradoxical effects (BOX 3).

Extending computational models

As described above, theoretical modelling has been crucial to delineate the dynamics of inhibitory stabilization in complex networks of the brain. The original models of ISNs have been very useful in providing the main dynamical insights into what to expect in simplified circuit scenarios. However, more biologically realistic models are needed to interpret the results of perturbation studies performed in real neuronal networks. Such models can in turn help us to test different hypotheses (such as whether paradoxical effects arise from inhibitory stabilization or disinhibition (BOX 3)), and provide novel predictions to be tested in new experiments.

Non-linear transfer function. Theoretical studies have already expanded the original ISN models in several directions. A prominent example is the family of supralinear stabilized networks (SSNs), which expand ISN models by accounting for the non-linear transfer function of neurons[83,91]. They have been proposed to explain various non-linear cortical computations[83], including multi-input integration and normalization. The non-linearity of the input–output function gives SSNs extra capacity to transition between and operate in two distinct regimes: a non-ISN regime for weak inputs, and an ISN regime for stronger feedforward drives. Such an input-dependent behaviour of SSNs enables them to especially model computational and functional properties that depend on contextual modulation (for example, surround suppression or normalization)[83,89,119,120]. Later studies used this modelling framework to shed light on various other properties, including top-down modulation[121], suppression of variability upon stimulation[122] and the generation of oscillatory and persistent activity[123].

Cell-type diversity. Another useful extension of ISN models is to go beyond a single inhibitory subpopulation and account for multiple subtypes of interneurons, as found in cortical circuits[112]. Different subtypes of

Disinhibition
Inhibition of the suppressive effect of inhibitory neurons, which can lead to an effective increase in the activity of target neurons.

Supralinear stabilized networks
(SSNs). Inhibition-stabilized network models with non-linear neuronal responses, whereby neurons respond with higher gains to stronger inputs (supralinear input–output transfer functions).

Normalization
A non-linear computation whereby the activity of a neuron is normalized by another parameter (for instance, divided by the population activity).

Persistent activity
Neuronal activity that persists even when the stimulus is not present (for example, to help working memory retain the information).

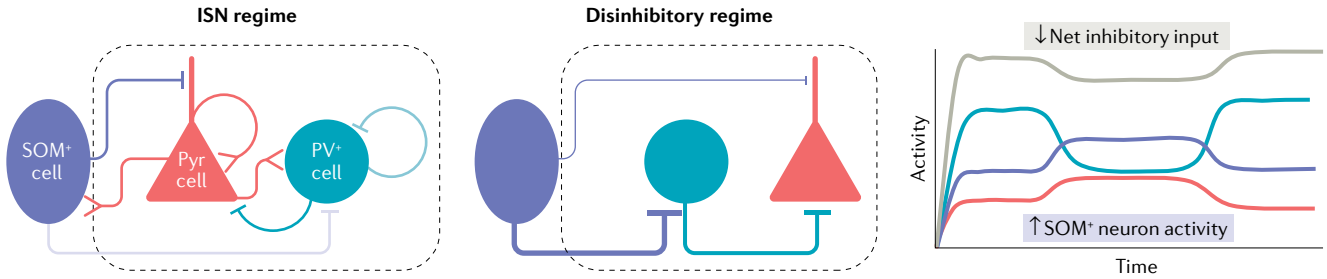
Box 3 | **Inhibitory stabilization versus disinhibition**

To safely infer an inhibition-stabilized regime of operation in the cortex, we need to rule out that alternative mechanisms are not responsible for the observed paradoxical effects. One such mechanism is 'disinhibition', or the inhibition of inhibition, which can arise from the complex pattern of connectivity between multiple subtypes of interneurons[1,116]. The local network, which is composed of balanced pyramidal neurons (Pyr cells) and parvalbumin-expressing (PV^+) inhibitory neurons, is in turn under the control of other inhibitory neurons, such as somatostatin-expressing (SOM^+) neurons (see the figure). Disinhibition in such a canonical network (SOM^+ neuron → (Pyr cell ↔ PV^+ neuron)) arises as a result of the $SOM^+ \rightarrow PV^+$ motif. If the local Pyr cell–PV^+ neuron network is weakly coupled and the connectivity from the surround is strong, disinhibition (SOM^+ neuron → PV^+ neuron) governs the dynamical properties of the circuitry (see figure). Notably, it is possible to observe paradoxical effects in such a network configuration, in the following manner. Increasing the input to SOM^+ neurons suppresses the activity of PV^+ neurons. If the decrease in PV^+ neuron → Pyr cell inhibition is more than the increase in SOM^+ neuron → Pyr cell inhibition, the total inhibition that Pyr cells receive can decrease, in a 'paradoxical' manner. Thus, paradoxical effects of perturbing inhibitory neurons can in principle be observed, even in the absence of unstable excitation[117].

By contrast, the paradoxical effect resulting from the perturbation of SOM^+ neurons in two recent experimental studies[88,89] was attributed to inhibitory stabilization of unstable excitatory subnetworks. The connectivity reported in the visual cortex[116] seems to support this inhibition-stabilized network scenario too, as the connections between Pyr cells and PV^+ neurons in the local network are as strong as the surround connections to SOM^+ neurons (see figure, ISN regime). Thus the reciprocal SOM^+ neuron–Pyr cell connectivity[116,231] and PV^+ neuron–Pyr cell connectivity[116,232] combined with strong Pyr cell–Pyr cell coupling can be a source of paradoxical effects when either SOM^+ neurons or PV^+ neurons are perturbed[88].

For the disinhibition scenario to give rise to paradoxical effects, we must allow for artificial connectivity profiles, with weak recurrent connections between Pyr cells and SOM^+ neurons, and strong SOM^+ neuron → PV^+ neuron weights (see figure, disinhibitory regime). Increasing the activity of SOM^+ neurons in this scenario can lead to a decrease in the net inhibition to Pyr cells, as a result of SOM^+ neuron → PV^+ neuron disinhibition. However, decreased inhibitory input increases the activity of Pyr cells, and hence we would not observe a concomitant decrease in the activity of PV^+ neurons and Pyr cells, as expected in inhibition-stabilized networks.

To uncover the precise mechanism underlying paradoxical effects unequivocally, future studies would need to consider measuring the actual connectivity between major subtypes of neurons[1,112,116,231,233,234]; differentially perturbing each inhibitory subtype[88,89,102,103,132] (especially PV^+ neurons and SOM^+ neurons); and simultaneously recording neuronal activity and inputs to neurons.



interneurons have distinct morphological, electrophysiological and connectivity profiles[1,114,116,124,125], show different levels of stimulus selectivity[126–129] and contribute differently to cortical computation in local and global networks[102,103,115,130]. How inhibitory stabilization is linked to each subtype, and how this relates to their specific computational and functional properties, can be addressed only in models with detailed modelling of inhibitory subpopulations.

The first question to answer is how to detect inhibitory stabilization in such networks and with which kind of paradoxical effects. A modelling study addressed this issue by analysing the dynamics of networks composed of pyramidal cells and PV^+ and SOM^+ neurons[90], with connectivity profiles resembling those reported in the mouse visual cortex[116]. Rather than focusing on the activity of a single inhibitory subpopulation, the study described a new criterion for ISNs in terms of the concomitant changes in the excitatory activity and the total inhibitory input to excitatory neurons. This could be obtained by combined recordings of the suprathreshold responses of pyramidal cells and of intracellular measures of their inputs, as was later achieved in experiments[88,89].

In the absence of such recordings, the detection of ISNs must necessarily become more sophisticated, owing to the complex interaction of inhibitory neurons (BOX 3). A new study found that networks in different areas (anterior lateral motor cortex and barrel cortex (S1)) and layers (L5 and L2/3) may or may not show paradoxical effects, and that the differences between them cannot be accounted for by models with a single inhibitory neuron subtype[117]. Using population models with more than a single inhibitory neuron subtype, the study explored different configurations of connectivity of pyramidal cell–PV^+ neuron–SOM^+ neuron networks and showed that paradoxical effects can be observed even under non-ISN scenarios, when the excitatory subnetwork is not unstable[117] (for instance, by disinhibitory mechanisms; BOX 3).

This picture becomes more complex, and more computationally diverse, when additional inhibitory subtypes are added to the circuit[7,112,113,116,131]. In addition to PV^+ neurons and SOM^+ neurons, VIP^+ interneurons are another major subclass of inhibitory neurons; they gate behavioural state modulation and top-down feedback[131–135]. One study showed that network models involving the interaction of three interneuron subtypes (PV^+, SOM^+ and VIP^+), combined with non-linear input–output transfer functions, can explain the paradoxical effects of top-down modulation of cortical circuitries as reported in experiments[121]. Mirroring the

Table 1 | **Current and potential future inhibition-stabilized network models**

Model	Extension	Refs
Current models		
Supralinear stabilized network	Extends classical ISNs to include non-linear transfer functions of neuronal responses	83,91
ISN with multiple forms of inhibition	Models ISN regimes in networks with multiple subtypes of inhibitory interneurons	90,117
ISN with partial perturbation	Models extended populations and explains the effects of partial perturbation of inhibition	93
Patterned perturbation and specific ISN	Studies the consequences of specific, rather than random recurrent, connectivity	110
Potential future extensions		
Dendritic processing	Could account for the role of different dendritic compartments and cast light on the contribution of dendritic non-linearity	–
Synaptic plasticity	Could explain how plasticity rules can organize ISNs in the right regimes and how ISN regimes can shape synaptic plasticity	–
Layered ISNs	Could extend the model of ISNs with different layers, and include interlaminar versus intralaminar interactions	–
Networks of networks	Could model the collective interaction and organization of ISN networks and the flow of information in networks of networks	–

ISN, inhibition-stabilized network.

transition from the non-ISN to the ISN regime in SSN models, the study described a transition from disinhibition to a 'response reversal' regime, depending on different levels of the visual input. In this response reversal regime, the activity of SOM^+ neurons is first suppressed by VIP^+ neurons, but eventually reverses to values higher than the baseline due to an increase in the recurrent excitation (as expected in ISNs).

All the models described above reduce each population of neurons to a single node, ignoring the complexity that can arise from differential activation of neurons within each subpopulation. To account for such differences, computational models with detailed modelling of neuronal populations are needed. For example, perturbing a small number of inhibitory neurons may mask the presence of paradoxical effects in mouse neocortical networks, as has been discussed before[93]. Accounting for more biological realism in computational models might thus be crucial to explain seemingly contradictory results obtained in different experiments[87–89,92,95,117], and to put constraints on the design of future experiments to detect ISNs in different brain regions.

Functional properties of neurons. Another way to account for the diversity and heterogeneity of neuronal responses within a subpopulation is to consider the difference in their functional properties (such as their selectivity for certain stimulus features). This is especially important, as the connectivity in many cortical networks is organized according to functional similarity of neuronal pairs — that is, neurons with similar feature selectivities are wired together more strongly and with higher connection probabilities than are neurons with dissimilar selectivities[111,136–140]. A theoretical study showed that such feature-specific connectivity can lead to functionally specific ISNs, which can be identified by a specific type of paradoxical effect[110]: instead of a paradoxical change in the mean activity of inhibitory neurons, the paradoxical effect is reflected in the paradoxical slope of the response changes as a function of input perturbations. This paradoxical effect would be revealed only if the inhibitory subnetworks were probed with patterned perturbations that are delivered according to functional properties of the neurons (for example, neurons with similar visual receptive fields are perturbed with similar levels of perturbations — or a lack thereof). Network models with feature-specific inhibitory stabilization also showed spontaneous transitions between selective activity patterns, as found in experiments[49,50]. Importantly, the specific paradoxical effects observed in these networks would be masked if typical perturbation protocols that perturb all inhibitory neurons uniformly or with random variability were applied. This again highlights the importance of the design of perturbation protocols to uncover ISNs and their various computational properties.

Future directions. Despite modelling efforts so far, there are still some critical biological details missing from ISN models that can be incorporated in future computational studies (TABLE 1). All current models ignore differences between cell types in terms of their anatomical and electrophysiological properties[1–6] and assume compact neurons, whereby the intricate structure of cells is reduced to a single point. Real neurons, however, have different compartments, with detailed dendritic structures and non-linear mechanisms[8–11]. It will be important to see how compartmentalized signalling and non-linear mechanisms at the cellular level work together with the non-linear properties of ISNs that emerge at the network level[100]. For example, different interneurons target different parts of pyramidal cells, with PV^+ neurons and SOM^+ neurons specifically targeting perisomatic and dendritic regions, respectively[111–113,141]. It will be interesting to see how such differential patterns of projections may affect paradoxical effects, and how such potentially different involvements in ISN behaviour are linked to cortical computation.

Another important aspect of biological neuronal networks that has remained understudied in ISN models is plasticity. It is crucial to study how — and which — plasticity mechanisms can organize neuronal networks to operate as ISNs[142–144] (BOX 2) and, in turn, how operating in ISN regimes can modulate plasticity in cortical networks[145].

Last, the interactions between local networks operating as ISNs can be studied within and across cortical layers, and as networks of networks that underlie the communication and coordination of different brain regions across larger scales[146]. It will be interesting to see whether the complex interlaminar excitatory–inhibitory interactions[7,44,111,147] or coupling between different regions (for example, between sensory and motor areas[148], or between the thalamus and the cortex[45,149,150]) can lead to long-range ISNs, whereby paradoxical effects emerge in the long-range interaction of excitatory–inhibitory

networks. Indeed, a recent study has suggested that ISN-like dynamics can be observed when the cortico-thalamic loop involving the visual cortex is considered[151]. Accounting for biological realism in these directions, combined with more detailed experiments, can enhance our understanding of the operating regimes of the cortex, and the role of inhibitory stabilization in cortical computation across multiple scales.

From neural dynamics to computation

Assessing inhibitory stabilization can provide a systematic way to shed light on the dynamical regimes of neuronal computation in different brain circuitries[122,152,153]. The spectrum of neural dynamics from feedforward architectures to ISNs (FIG. 2c) can be probed by perturbing inhibitory neurons in each region and assaying the presence or absence of paradoxical effects (FIG. 2d). Would cortical networks with more recurrent dynamics (for example, different visual circuits[152], the parietal cortex and the prefrontal cortex[154,155]) manifest prominent paradoxical effects that are absent in neuronal networks with feedforward-dominated structures (such as the retina or the cerebellum)? Could the dynamical signature of ISNs be used to reveal differences between different regions of the hippocampus (such as CA3 with its extensive recurrent connections[156,157] versus other subregions such as CA1)? A systematic assessment of inhibitory stabilization may reveal a spectrum of paradoxical effects and neural dynamics that can be further linked to the differential contribution of these brain regions to specialized computation and function[158–163].

In the neocortex, for example, a gradient of excitation–inhibition balance has been reported across cortical layers, with more superficial layers (such as L2/3) showing more inhibition-dominated connectivity, and deeper layers (such as L5) expressing more excitation-dominated dynamics[164]. It would be interesting to probe the spectrum of neural dynamics across these layers by assessing the corresponding gradient of inhibitory stabilization. Another interesting direction will be to map inhibitory stabilization at a more global scale, and compare it with the macroscopic gradients of excitation–inhibition balance in cortical networks[165]. Applying a standardized protocol of ISN detection across these global regions could reveal the main dynamical trends that could in turn serve as an intermediate level of description between structure and function.

Different computations for different regimes. The biggest gap in our knowledge of this field is in our understanding of the relationships between dynamics and computations. Progress in uncovering different dynamical regimes of neuronal networks should therefore be accompanied in the next step by unravelling how each regime is poised for different neural computations. Of special interest is to study which particular computations are better supported by inhibition-stabilized regimes (FIG. 2c). Recurrent computations can span a broad range, from input amplification (as reported in sensory cortices[37,38,62,63]), multi-input integration (as in contextual modulation and normalization[83,89,115]) and network multistability[123] (in the form of either network synchrony or persistent activity as in attractor networks[150,166–171]) to associative memory and temporal binding (as in the hippocampus and posterior parietal cortex[172–174]). Mapping specific dynamics and computations at the same time will provide us with tools to map which neural computations emerge in neural circuits in different excitation–inhibition regimes (FIG. 2c,d).

These regimes and the computations within them do not need to be fixed, however, and can be modulated by the state of the network. Controlling the level of inhibition is a fast and effective way by which ISNs can transition from weakly coupled to strongly coupled and even unstable regimes (for example, in their transient dynamics). This control can be achieved by top-down disinhibitory mechanisms, as, for example, mediated by VIP$^+$ neurons[111,131–135] (BOX 3). ISNs would therefore have the flexibility of performing different recurrent computations, ranging from those that can be equally obtained by stable or unstable excitatory connections (such as amplification) to others that depend on an unstable excitation[73–75]. In networks with specific excitatory subnetworks, for example, a sudden transition from linear amplification to non-linear processing happens as an excitatory subnetwork becomes unstable (that is, when the corresponding eigenvalue becomes larger than 1 (BOX 1)). This transition can lead to the emergence of non-linear computations such as cross-subnetwork suppression, or selective activation of specific subnetworks during spontaneous activity[49,50,110], which can help sensory representation or memory consolidation. It would be interesting to see whether such non-linear computations — and others, such as response normalization[175], feature-specific competition[176] and winner-take-all dynamics[177] — emerge in ISN regimes with strong recurrent excitation and inhibition, and to link them to the presence or absence of paradoxical effects in each circuit.

Spatio-temporal patterns of activity

Ultimately, to perform behaviourally relevant functions, neuronal networks need to generate complex spatio-temporal patterns of activity. Consider a simple task in which the animal needs to initiate movement either in the left direction or in the right direction (such as in an arm reach task). Learning to form and follow the proper trajectory of movements over time is governed by dynamically changing neuronal networks that represent different stages of the sequence (for example, networks representing specific joints or muscles). If this were driven by networks with feedforward architectures, neuronal representations would become weaker, more susceptible to noise and hence less reliable as the activity propagates through the layers (FIG. 3a). One proposed solution to overcome these problems is synchronous cascades of activity, or synfire chains[178–181]. Networks with excitation-dominated dynamics can provide another solution, by regenerating the activity at each stage via recurrent excitation (FIG. 3b). This results in persistent activity, which would amplify the signal, but inevitably leads to some overlap in time (with network activity in the previous and next layers) and space (by activating other streams of information flow, such as the left trajectory), thus causing spatio-temporal non-specificity.

Eigenvalue
The scalar factor by which a specific mode of activity (eigenvector) in a network is amplified, with values greater than 1 denoting unstable amplification.

Synfire chains
Multilayered feedforward networks that govern the flow of information via the synchronous cascade of activity from one layer to another.

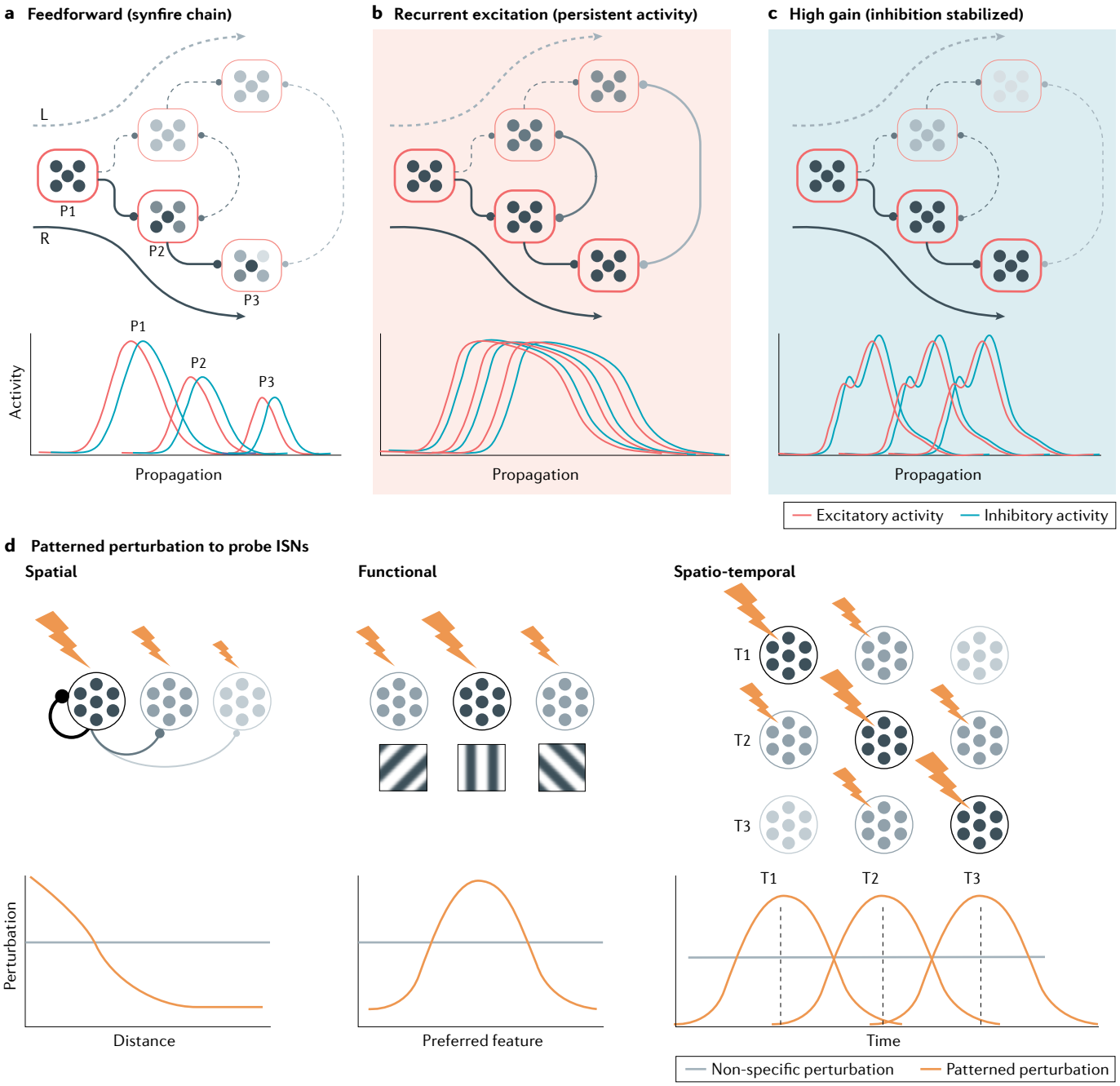


Fig. 3 | **Spatio-temporal patterns of neural activity and inhibitory stabilization. a–c** | Propagation of spatio-temporal patterns of activity in networks of neurons underlying the execution of a behavioural task involving the choice between left (L) and right (R). **a** | Propagation of activity when choosing 'right' in networks with predominantly feedforward connectivity. The activity becomes weaker and noisier as it propagates to downstream regions (denoted by the sequence of populations P1–P3). Schematics of the activity of neuronal populations in the right trajectory at different stages of propagation are shown at the bottom (red, excitation; blue, inhibition). **b** | Propagation of activity in networks with dominant recurrent excitation. The activity remains strong at all stages of propagation, as it is amplified by reverberating activity generated within each recurrent network. However, this can lead to reduced specificity, both within and across the trajectories. Along the right trajectory, persistent activity extends the responses in time (bottom), which makes the transition to different stages of propagation more ambiguous. Across trajectories, strong persistent activity activates the other channel of propagation (the left trajectory), even via weak connections. **c** | Propagation of activity via networks with inhibition-stabilized dynamics. Amplification and denoising can be combined with selectivity and precise timing, if the initial excitatory amplification resulting from recurrent excitation is curbed in time and space by stabilizing inhibition (bottom). **d** | Patterned perturbation of inhibition to reveal inhibition-stabilized networks (ISNs). Patterned perturbations can be delivered in space (left), whereby proximal inhibitory neurons are perturbed more strongly than distal ones. Patterned perturbations would therefore depend on the spatial distance from a target focal point, unlike non-specific perturbations, which lack such a distance dependence (grey line in bottom graph). Patterned perturbations can be functional, whereby neurons with similar functional properties (such as neurons preferring similar orientations) are perturbed more strongly than dissimilar neurons (middle). To probe inhibitory stabilization in spatio-temporal patterns of activity, patterned perturbations should be delivered along the trajectory of propagation. Largest perturbations at each time (T) should be delivered to the neuronal population that is expected to be the most active at that time, with other populations receiving perturbations proportionate to their distance from the trajectory (the farther away, the weaker the perturbation).

By quenching transient excitatory amplifications, networks with ISN dynamics can mitigate this problem and provide a higher signal-to-noise ratio and reliability, while maintaining response selectivity (FIG. 3c). This would reveal a signature in the specific dynamics of interaction of excitation and inhibition, whereby the initial amplification of the signal by unstable excitation is stabilized by a delayed (but still rapid) recurrent inhibition (FIG. 3c). Transient amplification of input patterns has indeed been proposed to emerge in ISNs[182], in which fast amplification of the input can arise as a result of effective connectivity between specific activity patterns. Transient dynamics have also been suggested to support optimal motor control, which is governed by strong excitation that is balanced by fine-tuned inhibition to stabilize specific modes of limb movement[183]. Such dynamics can emerge during learning in networks with specific excitatory–inhibitory tuning, as recently reported in experiments[184,185]. Indeed, a recent study has suggested that inhibition-dominated transients can emerge in recurrent excitatory–inhibitory networks optimized to perform probabilistic inference[120].

Spatio-temporal patterns of inhibitory stabilization should also manifest themselves in neural manifolds, in which propagating patterns of neuronal responses are projected into the low-dimensional space of neural activity[186,187]. As our technological toolkits progress[188–193], it will be interesting to record — and perturb — inhibitory activity along specific spatio-temporal sequences (as designated by, for example, the principal components that describe neural activity) (FIG. 3d). Such task-specific, low-dimensional projections of inhibitory stabilization[110], rather than the classical projections over recorded neurons or simpler features (such as the one-dimensional space of orientation selectivity[110]), could provide us with a more informative account of neural activity[186,187], pertinent to particular functional (and dysfunctional) properties of the brain.

Dysfunctional ISNs

Disruption of the proper operation of ISNs may, in turn, underlie brain dysfunction. In the absence of healthy inhibitory feedback, neuronal networks can experience pathological states of runaway activity[68,69]. Transition to such epileptic states might be due to the instability of the global mode of activity (for example, when inhibition is generally weak or reduced). However, it is likely that epileptic modes arise from certain unstable modes in specific regions or subnetworks with broken inhibitory stabilization. Recording the distributed activity of excitatory and inhibitory neurons, in combination with large-scale modelling, can shed light on the specific sources of epileptogenesis[194,195]. Detailed ISN models comprising dendritic and plasticity mechanisms (TABLE 1) can also delineate the contribution of different subtypes of inhibitory neurons, for example by teasing apart the role of perisomatic versus dendritic inhibition[68,196].

Impaired inhibitory stabilization may not always lead to extreme consequences such as runaway activity or pathological oscillations. In more moderate forms, it may cause hyperactivity and increased pairwise correlations[197–199], which can in turn decrease the coding capacity of the brain[200,201]. It would be interesting to see whether similar mechanisms contribute to the early-stage neuronal hyperactivity observed in Alzheimer disease[202–204] (and during ageing in general[205,206]), and how they link to the deficits of computation and learning in such networks[202]. That the risk of epilepsy is increased in Alzheimer disease[204,207,208], and that some antiepileptic drugs show effectiveness in models of Alzheimer disease[209], might hint at a common mechanism involving (impaired) ISNs[210,211]. As disrupted excitatory–inhibitory balance has been implicated in many brain disorders ranging from schizophrenia to autism[212–217], studying (dys)functional ISNs may have broad relevance for the healthy operation of the brain.

Open questions and future directions

Despite our progress, many questions remain to be answered with regard to the dynamical regimes of cortical networks and their operation as ISNs. On the modelling front, despite considerable developments in making more detailed models of ISNs, a gap between the models and biology still exists (TABLE 1). For example, most ISN models ignore the spiking dynamics of neurons and focus on rate-based equations. Can the properties reported in rate-based models be reproduced in more realistic models? Which type of non-linearity is necessary for different features of ISNs (such as persistent activity or normalization), and which type is more consistent with biology? Can the non-linearities assumed in models arise in spiking networks and, if so, under which regimes? How can other sources of non-linearity such as dendritic amplification in turn be incorporated into models? Increasing the biological plausibility of the models could enable a better interaction between theory and experiment.

On the experimental front, we still need a consensus of brain regions and cortical networks that operate as ISNs. As we progress in understanding this basic question, we can ask further and more advanced questions, such as which brain regions show more ISN-like dynamics, and whether there exists a spectrum of ISNs in the brain. The biggest gulf, however, exists in our understanding of the links between neural dynamics and neural computation. Which specific computational and functional properties that are supported by ISNs cannot emerge in other regimes? And how do deficits in inhibitory stabilization affect these computations? Can inhibitory stabilization and its role in behaviourally relevant computations (as in perception, decision making or an arm reach task) be revealed by analysing it along the low-dimensional manifolds of neural activity? By studying inhibitory stabilization in cortical networks, future studies can cast light on these fundamental questions of neuroscience, which are becoming more amenable to experimental investigation as our technological and theoretical toolkits develop.

Concluding remarks

Cortical networks can recruit their recurrent interactions to go beyond feedforward processing and perform complex computations. Inhibitory stabilization offers a rapid and reliable mechanism to ensure the stability

Neural manifolds
Projections of the activity of neuronal populations in a reduced space composed of the main components of correlated activity.

of strongly coupled recurrent networks and to extend the repertoire of neuronal dynamics. The presence of such inhibitory stabilization mechanisms in the brain can be detected by perturbing their inhibitory neurons. The paradoxical effects of perturbing inhibitory neurons expected in ISNs can in turn be used as a signature with which to probe the operating regime and computational properties of cortical networks in different brain regions.

Recent technological advances in optogenetics have enabled a more precise interrogation of cortical circuitry, which is necessary to address various aspects of inhibitory stabilization in the brain. It has led to a new wave of experiments addressing the presence of inhibitory stabilization and its associated dynamical features in the cortex. This has, in turn, reinvigorated the theoretical endeavours to advance models of inhibitory stabilization, by making them more biologically realistic in order to relate them better to experimental results. Here we reviewed this recent resurgence in experimental and theoretical studies of ISNs, and showed how it can cast light on the computational properties of the brain and its operating regimes. These results, tools and models open avenues to study novel questions in the future, and to shed light on the fundamental principles of computation underlying brain function and dysfunction.

Published online 11 November 2020

1. Jiang, X. et al. Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* **350**, aac9462 (2015).
2. Markram, H. et al. Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.* **5**, 793–807 (2004).
3. DeFelipe, J. & Fariñas, I. The pyramidal neuron of the cerebral cortex: morphological and chemical characteristics of the synaptic inputs. *Prog. Neurobiol.* **39**, 563–607 (1992).
4. Le Be, J.-V., Silberberg, G., Wang, Y. & Markram, H. Morphological, electrophysiological, and synaptic properties of corticocallosal pyramidal cells in the neonatal rat neocortex. *Cereb. Cortex* **17**, 2204–2213 (2007).
5. DeFelipe, J. Cortical interneurons: from Cajal to 2001. *Prog. Brain Res.* **136**, 215–238 (2002).
6. Braitenberg, V. & Schüz, A. *Cortex: Statistics and Geometry of Neuronal Connectivity* (Springer, 1998).
7. Harris, K. D. & Shepherd, G. M. The neocortical circuit: themes and variations. *Nat. Neurosci.* **18**, 170–181 (2015).
8. London, M. & Häusser, M. Dendritic computation. *Annu. Rev. Neurosci.* **28**, 503–532 (2005).
9. Silver, R. A. Neuronal arithmetic. *Nat. Rev. Neurosci.* **11**, 474–489 (2010).
10. Poirazi, P., Brannon, T. & Mel, B. W. Pyramidal neuron as two-layer neural network. *Neuron* **37**, 989–999 (2003).
11. Poirazi, P. & Papoutsi, A. Illuminating dendritic function with computational models. *Nat. Rev. Neurosci.* **21**, 303–321 (2020).
12. Brunel, N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* **8**, 183–208 (2000).
13. van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).
14. Vogels, T. P., Rajan, K. & Abbott, L. F. Neural network dynamics. *Annu. Rev. Neurosci.* **28**, 357–376 (2005).
15. Wilson, H. R. & Cowan, J. D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12**, 1–24 (1972).
16. Wilson, H. R. & Cowan, J. D. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* **13**, 55–80 (1973).
17. Ermentrout, G. B. & Cowan, J. D. A mathematical theory of visual hallucination patterns. *Biol. Cybern.* **34**, 137–150 (1979).
18. Citri, A. & Malenka, R. C. Synaptic plasticity: multiple forms, functions, and mechanisms. *Neuropsychopharmacology* **33**, 18–41 (2008).
19. Pascual-Leone, A., Amedi, A., Fregni, F. & Merabet, L. B. The plastic human brain cortex. *Annu. Rev. Neurosci.* **28**, 377–401 (2005).
20. Goto, Y., Yang, C. R. & Otani, S. Functional and dysfunctional synaptic plasticity in prefrontal cortex: roles in psychiatric disorders. *Biol. Psychiatry* **67**, 199–207 (2010).
21. Aertsen, A. M. H. J., Gerstein, G. L., Habib, M. K. & Palm, G. Dynamics of neuronal firing correlation: modulation of 'effective connectivity'. *J. Neurophysiol.* **61**, 900–917 (1989).
22. Vaadia, E. et al. Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* **373**, 515–518 (1995).
23. Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154 (1962).
24. Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.* **28**, 229–289 (1965).
25. Ferster, D. & Miller, K. D. Neural mechanisms of orientation selectivity in the visual cortex. *Annu. Rev. Neurosci.* **23**, 441–471 (2000).
26. Riesenhuber, M. & Poggio, T. Neural mechanisms of object recognition. *Curr. Opin. Neurobiol.* **12**, 162–168 (2002).
27. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**, 1019–1025 (1999).
28. Giese, M. A. & Poggio, T. Neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.* **4**, 179–192 (2003).
29. Serre, T., Oliva, A. & Poggio, T. A feedforward architecture accounts for rapid categorization. *Proc. Natl Acad. Sci. USA* **104**, 6424–6429 (2007).
30. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73**, 415–434 (2012).
31. Tang, H. et al. Recurrent computations for visual pattern completion. *Proc. Natl Acad. Sci. USA* **115**, 8835–8840 (2018).
32. Wyatte, D., Curran, T. & O'Reilly, R. The limits of feedforward vision: recurrent processing promotes robust object recognition when objects are degraded. *J. Cogn. Neurosci.* **24**, 2248–2261 (2012).
33. Kietzmann, T. C. et al. Recurrence is required to capture the representational dynamics of the human visual system. *Proc. Natl Acad. Sci. USA* **116**, 21854–21863 (2019).
34. Kar, K., Kubilius, J., Schmidt, K., Issa, E. B. & DiCarlo, J. J. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat. Neurosci.* **22**, 974–983 (2019).
35. Rajaei, K., Mohsenzadeh, Y., Ebrahimpour, R. & Khaligh-Razavi, S.-M. Beyond core object recognition: recurrent processes account for object recognition under occlusion. *PLoS Comput. Biol.* **15**, e1007001 (2019).
36. Chung, S. & Ferster, D. Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. *Neuron* **20**, 1177–1189 (1998).
37. Lien, A. D. & Scanziani, M. Tuned thalamic excitation is amplified by visual cortical circuits. *Nat. Neurosci.* **16**, 1315–1323 (2013).
38. Li, Y., Ibrahim, L. A., Liu, B., Zhang, L. I. & Tao, H. W. Linear transformation of thalamocortical input by intracortical excitation. *Nat. Neurosci.* **16**, 1324–1330 (2013).
39. Peters, A. & Payne, B. R. Numerical relationships between geniculocortical afferents and pyramidal cell modules in cat primary visual cortex. *Cereb. Cortex* **3**, 69–78 (1993).
40. Peters, A., Payne, B. R. & Budd, J. A numerical analysis of the geniculocortical input to striate cortex in the monkey. *Cereb. Cortex* **4**, 215–229 (1994).
41. Przybyszewski, A. W. Vision: does top-down processing help us to see? *Curr. Biol.* **8**, R135–R139 (1998).
42. Briggs, F. & Usrey, W. M. Corticogeniculate feedback and visual processing in the primate. *J. Physiol.* **589**, 33–40 (2011).
43. Gilbert, C. D. & Li, W. Top-down influences on visual processing. *Nat. Rev. Neurosci.* **14**, 350–363 (2013).
44. Olsen, S. R., Bortone, D. S., Adesnik, H. & Scanziani, M. Gain control by layer six in cortical circuits of vision. *Nature* **483**, 47–52 (2012).
45. Sillito, A. M. & Jones, H. E. Corticothalamic interactions in the transfer of visual information. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **357**, 1739–1752 (2002).
46. Mumford, D. On the computational architecture of the neocortex. *Biol. Cybern.* **65**, 135–145 (1991).
47. Harris, J. A. et al. Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195–202 (2019).
48. Keller, A. J., Roth, M. M. & Scanziani, M. Feedback generates a second receptive field in neurons of the visual cortex. *Nature* **582**, 545–549 (2020).
49. Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A. & Arieli, A. Spontaneously emerging cortical representations of visual attributes. *Nature* **425**, 954–956 (2003).
50. Miller, J.-E. K., Ayzenshtat, I., Carrillo-Reid, L. & Yuste, R. Visual stimuli recruit intrinsically generated cortical ensembles. *Proc. Natl Acad. Sci. USA* **111**, E4053–E4061 (2014).
51. Luczak, A., Bartho, P., Marguet, S. L., Buzsaki, G. & Harris, K. D. Sequential structure of neocortical spontaneous activity in vivo. *Proc. Natl Acad. Sci. USA* **104**, 347–352 (2007).
52. Carrillo-Reid, L., Yang, W., Bando, Y., Peterka, D. S. & Yuste, R. Imprinting and recalling cortical ensembles. *Science* **353**, 691–694 (2016).
53. Reilly, K. T. Persistent hand motor commands in the amputees' brain. *Brain* **129**, 2211–2223 (2006).
54. Newbold, D. J. et al. Plasticity and spontaneous activity pulses in disused human brain circuits. *Neuron* **107**, 580–589.e6 (2020).
55. Arieli, A., Sterkin, A., Grinvald, A. & Aertsen, A. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* **273**, 1868–1871 (1996).
56. Fiser, J., Chiu, C. & Weliky, M. Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* **431**, 573 (2004).
57. Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. & Suarez, H. H. Recurrent excitation in neocortical circuits. *Science* **269**, 981–985 (1995).
58. Douglas, R. J. & Martin, K. A. C. Canonical cortical circuits. in *Handbook of Brain Microcircuits* (eds Shepherd, G. M. & Grillner, S.) 23–34 (Oxford University Press, 2017).
59. Douglas, R. J. & Martin, K. A. C. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* **27**, 419–451 (2004).
60. Douglas, R. J. & Martin, K. A. C. Recurrent neuronal circuits in the neocortex. *Curr. Biol.* **17**, R496–R500 (2007).
61. Mumford, D. On the computational architecture of the neocortex. *Biol. Cybern.* **66**, 241–251 (1992).
62. Li, L., Li, Y., Zhou, M., Tao, H. W. & Zhang, L. I. Intracortical multiplication of thalamocortical signals in mouse auditory cortex. *Nat. Neurosci.* **16**, 1179–1181 (2013).
63. Peron, S. et al. Recurrent interactions in local cortical circuits. *Nature* **579**, 256–259 (2020).
64. Barnes, S. J. et al. Subnetwork-specific homeostatic plasticity in mouse visual cortex in vivo. *Neuron* **86**, 1290–1303 (2015).
65. Schwartzkroin, P. A. Role of the hippocampus in epilepsy. *Hippocampus* **4**, 239–242 (1994).

66. McNamara, J. O., Constant Byrne, M., Dasheiff, R. M. & Gregory Fitz, J. The kindling model of epilepsy: a review. *Prog. Neurobiol.* **15**, 139–159 (1980).
67. Sveinbjornsdottir, S. & Duncan, J. S. Parietal and occipital lobe epilepsy: a review. *Epilepsia* **34**, 493–521 (1993).
68. Maglóczky, Z. & Freund, T. F. Impaired and repaired inhibitory circuits in the epileptic human hippocampus. *Trends Neurosci.* **28**, 334–340 (2005).
69. McCormick, D. A. & Contreras, D. On the cellular and network bases of epileptic seizures. *Annu. Rev. Physiol.* **63**, 815–846 (2001).
70. Devinsky, O. et al. Epilepsy. *Nat. Rev. Dis. Prim.* **4**, 18024 (2018).
71. Bertschinger, N. & Natschläger, T. Real-time computation at the edge of chaos in recurrent neural networks. *Neural Comput.* **16**, 1413–1436 (2004).
72. Boedecker, J., Obst, O., Lizier, J. T., Mayer, N. M. & Asada, M. Information processing in echo state networks at the edge of chaos. *Theory Biosci.* **131**, 205–213 (2012).
73. Toyoizumi, T. & Abbott, L. F. Beyond the edge of chaos: amplification and temporal integration by recurrent networks in the chaotic regime. *Phys. Rev. E* **84**, 051908 (2011).
74. Laje, R. & Buonomano, D. V. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* **16**, 925–933 (2013).
75. Ostojic, S. Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nat. Neurosci.* **17**, 594–600 (2014).
76. Pouille, F., Marin-Burgin, A., Adesnik, H., Atallah, B. V. & Scanziani, M. Input normalization by global feedforward inhibition expands cortical dynamic range. *Nat. Neurosci.* **12**, 1577–1585 (2009).
77. Cruikshank, S. J., Lewis, T. J. & Connors, B. W. Synaptic basis for intense thalamocortical activation of feedforward inhibitory cells in neocortex. *Nat. Neurosci.* **10**, 462–468 (2007).
78. van Vreeswijk, C. & Sompolinsky, H. Chaotic balanced state in a model of cortical circuits. *Neural Comput.* **10**, 1321–1371 (1998).
79. Deneve, S. & Machens, C. K. Efficient codes and balanced networks. *Nat. Neurosci.* **19**, 375–382 (2016).
80. Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J. & McNaughton, B. L. Paradoxical effects of external modulation of inhibitory interneurons. *J. Neurosci.* **17**, 4382–4388 (1997).
By combining theory and experiments in the hippocampus, this study shows that paradoxical effects of perturbing inhibitory neurons can arise in biological networks.
81. Freund, T. F. & Antal, M. GABA-containing neurons in the septum control inhibitory interneurons in the hippocampus. *Nature* **336**, 170–173 (1988).
82. Ozeki, H., Finn, I. M., Schaffer, E. S., Miller, K. D. & Ferster, D. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* **62**, 578–592 (2009).
This study analyses mechanisms underlying surround suppression in the cat visual cortex to show that paradoxical effects can contribute to visual processing.
83. Rubin, D. B., Van Hooser, S. D. & Miller, K. D. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* **85**, 402–417 (2015).
This study shows that different computations, such as input normalization and surround suppression in the visual cortex, can be explained by SSNs.
84. Fenno, L., Yizhar, O. & Deisseroth, K. The development and application of optogenetics. *Annu. Rev. Neurosci.* **34**, 389–412 (2011).
85. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. Optogenetics in neural systems. *Neuron* **71**, 9–34 (2011).
86. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005).
87. Atallah, B. V., Bruns, W., Carandini, M. & Scanziani, M. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**, 159–170 (2012).
This study addresses inhibitory stabilization in cortical networks by optogenetically perturbing inhibitory neurons in the mouse visual cortex, and concludes that paradoxical effects could not be observed.
88. Kato, H. K., Asinof, S. K. & Isaacson, J. S. Network-level control of frequency tuning in auditory cortex. *Neuron* **95**, 412–423.e4 (2017).
This study combines intracellular recordings with optogenetic inactivation of inhibitory interneurons in the auditory cortex of awake behaving mice to show paradoxical effects.
89. Adesnik, H. Synaptic mechanisms of feature coding in the visual cortex of awake mice. *Neuron* **95**, 1147–1159.e4 (2017).
Using intracellular recordings in the visual cortex of awake behaving mice, this study shows paradoxical effects of perturbing inhibitory neurons and of surround modulation.
90. Litwin-Kumar, A., Rosenbaum, R. & Doiron, B. Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes. *J. Neurophysiol.* **115**, 1399–1409 (2016).
This theoretical study evaluates the presence of inhibitory stabilization in networks with multiple subtypes of inhibition.
91. Ahmadian, Y., Rubin, D. B. & Miller, K. D. Analysis of the stabilized supralinear network. *Neural Comput.* **25**, 1994–2037 (2013).
92. Li, N. et al. Spatiotemporal constraints on optogenetic inactivation in cortical circuits. *eLife* **8**, (2019).
The study tests several spatio-temporal protocols and shows that the strength and spatial scale of optogenetic perturbation matter for observing paradoxical effects.
93. Sadeh, S., Silver, R. A., Mrsic-Flogel, T. D. & Muir, D. R. Assessing the role of inhibition in stabilizing neocortical networks requires large-scale perturbation of the inhibitory population. *J. Neurosci.* **37**, 12050–12067 (2017).
This study extends ISN models to networks with extended populations and shows that detecting paradoxical effects depends on the fraction of perturbed inhibitory neurons.
94. Gutnisky, D. A. et al. Mechanisms underlying a thalamocortical transformation during active tactile sensation. *PLoS Comput. Biol.* **13**, e1005576 (2017).
95. Sanzeni, A. et al. Inhibition stabilization is a widespread property of cortical networks. *eLife* **9**, e54875 (2020).
The study shows that paradoxical effects can be observed in different cortices and across different brain states in mice, but only when perturbations affect a large number of inhibitory neurons.
96. Campagna, J. A., Miller, K. W. & Forman, S. A. Mechanisms of actions of inhaled anesthetics. *N. Engl. J. Med.* **348**, 2110–2124 (2003).
97. Richards, C. D. Actions of general anaesthetics on synaptic transmission in the CNS. *Br. J. Anaesth.* **55**, 201–207 (1983).
98. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008).
99. Niell, C. M. & Stryker, M. P. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* **65**, 472–479 (2010).
100. Gleeson, P. et al. Open source brain: a collaborative resource for visualizing, analyzing, simulating, and developing standardized models of neurons and circuits. *Neuron* **103**, 395–411.e5 (2019).
101. Sadeh, S. & Clopath, C. Theory of neuronal perturbome in cortical networks. *Proc. Natl Acad. Sci. USA* https://doi.org/10.1073/pnas.2004568117 (2020).
102. Lee, S.-H. et al. Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488**, 379–383 (2012).
103. Wilson, N. R., Runyan, C. A., Wang, F. L. & Sur, M. Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature* **488**, 343–348 (2012).
104. El-Boustani, S., Wilson, N. R., Runyan, C. A. & Sur, M. El-Boustani et al. reply. *Nature* **508**, E3–E4 (2014).
105. Lee, S.-H., Kwan, A. C. & Dan, Y. Interneuron subtypes and orientation tuning. *Nature* **508**, E1–E2 (2014).
106. Atallah, B. V., Scanziani, M. & Carandini, M. Atallah et al. reply. *Nature* **508**, E3 (2014).
107. Haider, B., Häusser, M. & Carandini, M. Inhibition dominates sensory responses in the awake cortex. *Nature* **493**, 97–100 (2013).
108. Rudolph, M., Pospischil, M., Timofeev, I. & Destexhe, A. Inhibition determines membrane potential dynamics and controls action potential generation in awake and sleeping cat cortex. *J. Neurosci.* **27**, 5280–5290 (2007).
109. Rochefort, N. L. et al. Sparsification of neuronal activity in the visual cortex at eye-opening. *Proc. Natl Acad. Sci. USA* **106**, 15049–15054 (2009).
110. Sadeh, S. & Clopath, C. Patterned perturbation of inhibition can reveal the dynamical structure of neural processing. *eLife* **9**, e52757 (2020).
The study shows that ISNs with functionally specific subnetworks express feature-specific paradoxical effects, which can be detected using patterned perturbations.
111. Harris, K. D. & Mrsic-Flogel, T. D. Cortical connectivity and sensory coding. *Nature* **503**, 51–58 (2013).
112. Tremblay, R., Lee, S. & Rudy, B. GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* **91**, 260–292 (2016).
113. Hattori, R., Kuchibhotla, K. V., Froemke, R. C. & Komiyama, T. Functions and dysfunctions of neocortical inhibitory neuron subtypes. *Nat. Neurosci.* **20**, 1199–1208 (2017).
114. Kepecs, A. & Fishell, G. Interneuron cell types are fit to function. *Nature* **505**, 318–326 (2014).
115. Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J. & Scanziani, M. A neural circuit for spatial summation in visual cortex. *Nature* **490**, 226–231 (2012).
116. Pfeffer, C. K., Xue, M., He, M., Huang, Z. J. & Scanziani, M. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat. Neurosci.* **16**, 1068–1076 (2013).
117. Mahrach, A., Chen, G., Li, N., van Vreeswijk, C. & Hansel, D. Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation. *eLife* **9**, e49967 (2020).
This study observes a mixture of paradoxical and non-paradoxical effects in the mouse, and argues that these results could be consistent with network models with disinhibition and without unstable excitation.
118. Moore, A. K., Weible, A. P., Balmer, T. S., Trussell, L. O. & Wehr, M. Rapid rebalancing of excitation and inhibition by cortical circuitry. *Neuron* **97**, 1341–1355.e6 (2018).
119. Liu, L. D., Miller, K. D. & Pack, C. C. A unifying motif for spatial and directional surround suppression. *J. Neurosci.* **38**, 989–999 (2018).
120. Echeveste, R., Aitchison, L., Hennequin, G. & Lengyel, M. Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. *Nat. Neurosci.* **23**, 1138–1149 (2020).
121. Garcia del Molino, L. C., Yang, G. R., Mejias, J. F. & Wang, X.-J. Paradoxical response reversal of top-down modulation in cortical circuits with three interneuron types. *Elife* **6**, e29742 (2017).
The study suggests that paradoxical effects of top-down modulation arise in network models with multiple inhibitory subtypes and non-linear input–output transfer functions.
122. Hennequin, G., Ahmadian, Y., Rubin, D. B., Lengyel, M. & Miller, K. D. The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron* **98**, 846–860.e5 (2018).
This study shows that the reduction of variability after stimulus presentation in cortical networks can be explained by SSNs.
123. Kraynyukova, N. & Tchumatchenko, T. Stabilized supralinear network can give rise to bistable, oscillatory, and persistent activity. *Proc. Natl Acad. Sci. USA* **115**, 3464–3469 (2018).
This article shows that working memory, persistent activity and rhythms can emerge from the bistable dynamics of SSNs.
124. Gupta, A., Wang, Y. & Markram, H. Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science* **287**, 273–278 (2000).
125. Galarreta, M. & Hestrin, S. Electrical and chemical synapses among parvalbumin fast-spiking GABAergic interneurons in adult mouse neocortex. *Proc. Natl Acad. Sci. USA* **99**, 12438–12443 (2002).
126. Bock, D. D. et al. Network anatomy and in vivo physiology of visual cortical neurons. *Nature* **471**, 177–182 (2011).
127. Ma, W. et al. Visual representations by cortical somatostatin inhibitory neurons–selective but with weak and delayed responses. *J. Neurosci.* **30**, 14371–14379 (2010).
128. Runyan, C. A. & Sur, M. Response selectivity is correlated to dendritic structure in parvalbumin-expressing inhibitory neurons in visual cortex. *J. Neurosci.* **33**, 11724–11733 (2013).
129. Li, L.-Y. et al. Differential receptive field properties of parvalbumin and somatostatin inhibitory neurons in mouse auditory cortex. *Cereb. Cortex* **25**, 1782–1791 (2015).

130. El-Boustani, S. & Sur, M. Response-dependent dynamics of cell-specific inhibition in cortical networks in vivo. *Nat. Commun.* **5**, 5689 (2014).
131. Rudy, B., Fishell, G., Lee, S. & Hjerling-Leffler, J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011).
132. Wood, K. C., Blackwell, J. M. & Geffen, M. N. Cortical inhibitory interneurons control sensory processing. *Curr. Opin. Neurobiol.* **46**, 200–207 (2017).
133. Fu, Y. et al. A cortical circuit for gain control by behavioral state. *Cell* **156**, 1139–1152 (2014).
134. Reimer, J. et al. Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron* **84**, 355–362 (2014).
135. Jackson, J., Ayzenshtat, I., Karnani, M. M. & Yuste, R. VIP$^+$ interneurons control neocortical activity across brain states. *J. Neurophysiol.* **115**, 3008–3017 (2016).
136. Cossell, L. et al. Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* **518**, 399–403 (2015).
137. Ko, H. et al. Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87–91 (2011).
138. Ko, H. et al. The emergence of functional microcircuits in visual cortex. *Nature* **496**, 96–100 (2013).
139. Znamenskiy, P. et al. Functional selectivity and specific connectivity of inhibitory neurons in primary visual cortex. *bioRxiv* https://doi.org/10.1101/294835 (2018).
140. Ko, H., Mrsic-Flogel, T. D. & Hofer, S. B. Emergence of feature-specific connectivity in cortical microcircuits in the absence of visual experience. *J. Neurosci.* **34**, 9812–9816 (2014).
141. Di Cristo, G. et al. Subcellular domain-restricted GABAergic innervation in primary visual cortex in the absence of sensory and thalamic inputs. *Nat. Neurosci.* **7**, 1184–1186 (2004).
142. Hennequin, G., Agnes, E. J. & Vogels, T. P. Inhibitory plasticity: balance, control, and codependence. *Annu. Rev. Neurosci.* **40**, 557–579 (2017).
143. Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C. & Gerstner, W. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* **334**, 1569–1573 (2011).
144. Clopath, C., Büsing, L., Vasilaki, E. & Gerstner, W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nat. Neurosci.* **13**, 344–352 (2010).
145. Christie, I. K., Miller, P. & Van Hooser, S. D. Cortical amplification models of experience-dependent development of selective columns and response sparsification. *J. Neurophysiol.* **118**, 874–893 (2017).
146. Hart, E. & Huk, A. C. Recurrent circuit dynamics underlie persistent activity in the macaque frontoparietal network. *eLife* **9**, e52460 (2020).
147. Hooks, B. M. et al. Laminar analysis of excitatory local circuits in vibrissal motor and sensory cortical areas. *PLoS Biol.* **9**, e1000572 (2011).
148. Mao, T. et al. Long-range neuronal circuits underlying the interaction between sensory and motor cortex. *Neuron* **72**, 111–123 (2011).
149. Briggs, F. & Usrey, W. M. Emerging views of corticothalamic function. *Curr. Opin. Neurobiol.* **18**, 403–407 (2008).
150. Guo, Z. V. et al. Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* **545**, 181–186 (2017).
151. Lin, I.-C., Okun, M., Carandini, M. & Harris, K. D. Equations governing dynamics of excitation and inhibition in the mouse corticothalamic network. *bioRxiv* https://doi.org/10.1101/2020.06.03.132688 (2020).
152. Wolf, F., Engelken, R., Puelma-Touzel, M., Weidinger, J. D. F. & Neef, A. Dynamical models of cortical circuits. *Curr. Opin. Neurobiol.* **25**, 228–236 (2014).
153. Ahmadian, Y. & Miller, K. D. What is the dynamical regime of cerebral cortex? *arXiv* 1908.10101v2 (2019).
154. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
155. Chaisangmongkon, W., Swaminathan, S. K., Freedman, D. J. & Wang, X.-J. Computing by robust transience: how the fronto-parietal network performs sequential, category-based decisions. *Neuron* **93**, 1504–1517.e4 (2017).
156. Amaral, D. G. & Witter, M. P. The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience* **31**, 571–591 (1989).
157. Le Duigou, C., Simonnet, J., Teleñczuk, M. T., Fricker, D. & Miles, R. Recurrent synapses and circuits in the CA3 region of the hippocampus: an associative network. *Front. Cell. Neurosci.* **7**, 262 (2014).
158. Dimsdale-Zucker, H. R., Ritchey, M., Ekstrom, A. D., Yonelinas, A. P. & Ranganath, C. CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields. *Nat. Commun.* **9**, 294 (2018).
159. Guzowski, J. F., Knierim, J. J. & Moser, E. I. Ensemble dynamics of hippocampal regions CA3 and CA1. *Neuron* **44**, 581–584 (2004).
160. Lee, I., Yoganarasimha, D., Rao, G. & Knierim, J. J. Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3. *Nature* **430**, 456–459 (2004).
161. Leutgeb, S. Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science* **305**, 1295–1298 (2004).
162. Vazdarjanova, A. & Guzowski, J. F. Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. *J. Neurosci.* **24**, 6489–6496 (2004).
163. Mizuseki, K., Royer, S., Diba, K. & Buzsáki, G. Activity dynamics and behavioral correlates of CA3 and CA1 hippocampal pyramidal neurons. *Hippocampus* **22**, 1659–1680 (2012).
164. Markram, H. et al. Reconstruction and simulation of neocortical microcircuitry. *Cell* **163**, 456–492 (2015).
165. Wang, X.-J. Macroscopic gradients of synaptic excitation and inhibition in the neocortex. *Nat. Rev. Neurosci.* **21**, 169–178 (2020).
166. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
167. Ben-Yishai, R., Bar-Or, R. L. & Sompolinsky, H. Theory of orientation tuning in visual cortex. *Proc. Natl Acad. Sci. USA* **92**, 3844–3848 (1995).
168. Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186–191 (2015).
169. Kim, S. S., Rouault, H., Druckmann, S. & Jayaraman, V. Ring attractor dynamics in the *Drosophila* central brain. *Science* **356**, 849–853 (2017).
170. Daie, K., Svoboda, K. & Druckmann, S. Targeted photostimulation uncovers circuit motifs supporting short-term memory. *bioRxiv* https://doi.org/10.1101/623785 (2019).
171. Deutsch, D. et al. The neural basis for a persistent internal state in *Drosophila* females. *bioRxiv* https://doi.org/10.1101/2020.02.13.947952 (2020).
172. Runyan, C. A., Piasini, E., Panzeri, S. & Harvey, C. D. Distinct timescales of population coding across cortex. *Nature* **548**, 92–96 (2017).
173. Rolls, E. T. An attractor network in the hippocampus: theory and neurophysiology. *Learn. Mem.* **14**, 714–731 (2007).
174. Nakazawa, K. Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science* **297**, 211–218 (2002).
175. Carandini, M. & Heeger, J. D. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2012).
176. Chettih, S. N. & Harvey, C. D. Single-neuron perturbations reveal feature-specific competition in V1. *Nature* **567**, 334–340 (2019).
177. Busse, L., Wade, A. R. & Carandini, M. Representation of concurrent stimuli by population activity in visual cortex. *Neuron* **64**, 931–942 (2009).
178. Abeles, M. *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge University Press, 1991).
179. Diesmann, M., Gewaltig, M.-O. & Aertsen, A. Stable propagation of synchronous spiking in cortical neural networks. *Nature* **402**, 529–533 (1999).
180. Kumar, A., Rotter, S. & Aertsen, A. Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nat. Rev. Neurosci.* **11**, 615–627 (2010).
181. Kumar, A., Rotter, S. & Aertsen, A. Conditions for propagating synchronous spiking and asynchronous firing rates in a cortical network model. *J. Neurosci.* **28**, 5268–5280 (2008).
182. Murphy, B. K. & Miller, K. D. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron* **61**, 635–648 (2009).
183. Hennequin, G., Vogels, T. P. & Gerstner, W. Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron* **82**, 1394–1406 (2014).
184. Khan, A. G. et al. Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nat. Neurosci.* **21**, 851–859 (2018).
185. Najafi, F. et al. Excitatory and inhibitory subnetworks are equally selective during decision-making and emerge simultaneously during learning. *Neuron* **105**, 165–179.e8 (2020).
186. Vyas, S., Golub, M. D., Sussillo, D. & Shenoy, K. V. Computation through neural population dynamics. *Annu. Rev. Neurosci.* **43**, 249–275 (2020).
187. Gallego, J. A., Perich, M. G., Miller, L. E. & Solla, S. A. Neural manifolds for the control of movement. *Neuron* **94**, 978–984 (2017).
188. Emiliani, V., Cohen, A. E., Deisseroth, K. & Häusser, M. All-optical interrogation of neural circuits. *J. Neurosci.* **35**, 13917–13926 (2015).
189. Zhang, Z., Russell, L. E., Packer, A. M., Gauld, O. M. & Häusser, M. Closed-loop all-optical interrogation of neural circuits in vivo. *Nat. Methods* **15**, 1037–1040 (2018).
190. Carrillo-Reid, L., Han, S., Yang, W., Akrouh, A. & Yuste, R. Controlling visually guided behavior by holographic recalling of cortical ensembles. *Cell* **178**, 447–457.e5 (2019).
191. Marshel, J. H. et al. Cortical layer-specific critical dynamics triggering perception. *Science* **365**, eaaw5202 (2019).
192. Yang, W., Carrillo-Reid, L., Bando, Y., Peterka, D. S. & Yuste, R. Simultaneous two-photon imaging and two-photon optogenetics of cortical circuits in three dimensions. *eLife* **7**, e32671 (2018).
193. Balasubramanian, K. et al. Propagating motor cortical dynamics facilitate movement initiation. *Neuron* **106**, 526–536.e4 (2020).
194. Lytton, W. W. Computer modelling of epilepsy. *Nat. Rev. Neurosci.* **9**, 626–637 (2008).
195. Liou, J.-Y. et al. A model for focal seizure onset, propagation, evolution, and progression. *eLife* **9**, e50927 (2020).
196. Cossart, R. et al. Dendritic but not somatic GABAergic inhibition is decreased in experimental epilepsy. *Nat. Neurosci.* **4**, 52–62 (2001).
197. Renart, A. et al. The asynchronous state in cortical circuits. *Science* **327**, 587–590 (2010).
198. Tetzlaff, T., Helias, M., Einevoll, G. T. & Diesmann, M. Decorrelation of neural-network activity by inhibitory feedback. *PLoS Comput. Biol.* **8**, e1002596 (2012).
199. Sippy, T. & Yuste, R. Decorrelating action of inhibition in neocortical networks. *J. Neurosci.* **33**, 9813–9830 (2013).
200. Zohary, E., Shadlen, M. N. & Newsome, W. T. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370**, 140–143 (1994).
201. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**, 358–366 (2006).
202. Busche, M. A. & Konnerth, A. Impairments of neural circuit function in Alzheimer's disease. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150429 (2016).
203. Zott, B., Busche, M. A., Sperling, R. A. & Konnerth, A. What happens with the circuit in Alzheimer's disease in mice and humans? *Annu. Rev. Neurosci.* **41**, 277–297 (2018).
204. Palop, J. J. et al. Aberrant excitatory neuronal activity and compensatory remodeling of inhibitory hippocampal circuits in mouse models of Alzheimer's disease. *Neuron* **55**, 697–711 (2007).
205. Schmolesky, M. T., Wang, Y., Pu, M. & Leventhal, A. G. Degradation of stimulus selectivity of visual cortical cells in senescent rhesus monkeys. *Nat. Neurosci.* **3**, 384–390 (2000).
206. Yu, S., Wang, Y., Li, X., Zhou, Y. & Leventhal, A. G. Functional degradation of extrastriate visual cortex in senescent rhesus monkeys. *Neuroscience* **140**, 1023–1029 (2006).
207. Vossel, K. A., Tartaglia, M. C., Nygaard, H. B., Zeman, A. Z. & Miller, B. L. Epileptic activity in Alzheimer's disease: causes and clinical relevance. *Lancet Neurol.* **16**, 311–322 (2017).
208. Vossel, K. A. et al. Seizures and epileptiform activity in the early stages of Alzheimer disease. *JAMA Neurol.* **70**, 1158–1166 (2013).
209. Sanchez, P. E. et al. Levetiracetam suppresses neuronal network dysfunction and reverses synaptic and cognitive deficits in an Alzheimer's disease model. *Proc. Natl Acad. Sci. USA* **109**, E2895–E2903 (2012).
210. Palop, J. J. & Mucke, L. Network abnormalities and interneuron dysfunction in Alzheimer disease. *Nat. Rev. Neurosci.* **17**, 777–792 (2016).
211. Verret, L. et al. Inhibitory interneuron deficit links altered network activity and cognitive dysfunction in Alzheimer model. *Cell* **149**, 708–721 (2012).

212. Yizhar, O. et al. Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* **477**, 171–178 (2011).
213. Kehrer, C. Altered excitatory-inhibitory balance in the NMDA-hypofunction model of schizophrenia. *Front. Mol. Neurosci.* **1**, 6 (2008).
214. Rubenstein, J. L. R. & Merzenich, M. M. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255–267 (2003).
215. Gogolla, N. et al. Common circuit defect of excitatory-inhibitory balance in mouse models of autism. *J. Neurodev. Disord.* **1**, 172–181 (2009).
216. Nelson, S. B. & Valakh, V. Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders. *Neuron* **87**, 684–698 (2015).
217. Toader, O., von Heimendahl, M., Schuelert, N., Nissen, W. & Rosenbrock, H. Suppression of parvalbumin interneuron activity in the prefrontal cortex recapitulates features of impaired excitatory/inhibitory balance and sensory processing in schizophrenia. *Schizophr. Bull.* **46**, 981–989 (2020).
218. Haider, B. & McCormick, D. A. Rapid neocortical dynamics: cellular and network mechanisms. *Neuron* **62**, 171–189 (2009).
219. Zagha, E. & McCormick, D. A. Neural control of brain state. *Curr. Opin. Neurobiol.* **29**, 178–186 (2014).
220. Ferguson, K. A. & Cardin, J. A. Mechanisms underlying gain modulation in the cortex. *Nat. Rev. Neurosci.* **21**, 80–92 (2020).
221. Zenke, F., Gerstner, W. & Ganguli, S. The temporal paradox of Hebbian learning and homeostatic plasticity. *Curr. Opin. Neurobiol.* **43**, 166–176 (2017).
222. Zenke, F. & Gerstner, W. Hebbian plasticity requires compensatory processes on multiple timescales. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20160259 (2017).
223. Lee, W.-C. A. et al. Anatomy and function of an excitatory network in the visual cortex. *Nature* **532**, 370–374 (2016).
224. Zenke, F., Hennequin, G. & Gerstner, W. Synaptic plasticity in neural networks needs homeostasis with a fast rate detector. *PLoS Comput. Biol.* **9**, e1003330 (2013).
225. Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C. & Nelson, S. B. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature* **391**, 892–896 (1998).
226. Sadeh, S., Clopath, C. & Rotter, S. Emergence of functional specificity in balanced networks with synaptic plasticity. *PLoS Comput. Biol.* **11**, e1004307 (2015).
227. Froemke, R. C. Plasticity of cortical excitatory-inhibitory balance. *Annu. Rev. Neurosci.* **38**, 195–219 (2015).
228. Griffen, T. C. GABAergic synapses: their plasticity and role in sensory cortex. *Front. Cell. Neurosci.* **8**, 91 (2014).
229. Lefort, S., Gray, A. C. & Turrigiano, G. G. Long-term inhibitory plasticity in visual cortical layer 4 switches sign at the opening of the critical period. *Proc. Natl Acad. Sci. USA* **110**, E4540–E4547 (2013).
230. Donato, F., Rompani, S. B. & Caroni, P. Parvalbumin-expressing basket-cell network plasticity induced by experience regulates adult learning. *Nature* **504**, 272–276 (2013).
231. Fino, E. & Yuste, R. Dense inhibitory connectivity in neocortex. *Neuron* **69**, 1188–1203 (2011).
232. Hofer, S. B. et al. Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nat. Neurosci.* **14**, 1045–1052 (2011).
233. Booker, S. A. & Vida, I. Morphological diversity and connectivity of hippocampal interneurons. *Cell Tissue Res.* **373**, 619–641 (2018).
234. Pelkey, K. A. et al. Hippocampal GABAergic inhibitory interneurons. *Physiol. Rev.* **97**, 1619–1747 (2017).

Acknowledgements
The authors thank A. Aertsen, A. Kumar and S. J. Barnes for comments on the manuscript.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Neuroscience thanks T. Tchumatchenko and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.