The MIT Press

Research

The self-organized learning of noisy environmental stimuli requires distinct phases of plasticity

Steffen Krüppel[1,2] and Christian Tetzlaff[1,2]

[1]Department of Computational Neuroscience, Third Institute of Physics - Biophysics, Georg-August-University, Göttingen, Germany
[2]Bernstein Center for Computational Neuroscience, Georg-August-University, Göttingen, Germany

ABSTRACT

Along sensory pathways, representations of environmental stimuli become increasingly sparse and expanded. If additionally the feed-forward synaptic weights are structured according to the inherent organization of stimuli, the increase in sparseness and expansion leads to a reduction of sensory noise. However, it is unknown how the synapses in the brain form the required structure, especially given the omnipresent noise of environmental stimuli. Here, we employ a combination of synaptic plasticity and intrinsic plasticity—adapting the excitability of each neuron individually—and present stimuli with an inherent organization to a feed-forward network. We observe that intrinsic plasticity maintains the sparseness of the neural code and thereby allows synaptic plasticity to learn the organization of stimuli in low-noise environments. Nevertheless, even high levels of noise can be handled after a subsequent phase of readaptation of the neuronal excitabilities by intrinsic plasticity. Interestingly, during this phase the synaptic structure has to be maintained. These results demonstrate that learning and recalling in the presence of noise requires the coordinated interplay between plasticity mechanisms adapting different properties of the neuronal circuit.

AUTHOR SUMMARY

Everyday life requires living beings to continuously recognize and categorize perceived stimuli from the environment. To master this task, the representations of these stimuli become increasingly sparse and expanded along the sensory pathways of the brain. In addition, the underlying neuronal network has to be structured according to the inherent organization of the environmental stimuli. However, how the neuronal network learns the required structure even in the presence of noise remains unknown. In this theoretical study, we show that the interplay between synaptic plasticity—controlling the synaptic efficacies—and intrinsic plasticity—adapting the neuronal excitabilities—enables the network to encode the organization of environmental stimuli. It thereby structures the network to correctly categorize stimuli even in the presence of noise. After having encoded the stimuli's organization, consolidating the synaptic structure while keeping the neuronal excitabilities dynamic enables the neuronal system to readapt to arbitrary levels of noise resulting in a near-optimal classification performance for all noise levels. These results provide new insights into the interplay between different plasticity mechanisms and how this interplay enables sensory systems to reliably learn and categorize stimuli from the surrounding environment.

INTRODUCTION

Learning to distinguish between different stimuli despite high levels of noise is an important ability of living beings to ensure survival. However, the underlying neuronal and synaptic processes of this ability are largely unknown.

The brain is responsible for controlling movements of an agent's body in response to the perceived stimulus. For instance, the agent should run away from a predator or run after the prey. To do so, the agent needs to be able to reliably classify the perceived stimulus despite its natural variability (e.g., different individuals of the same predator species) or noise (e.g., impaired vision by obstacles). In general, the sensory processing systems of the brain map the stimulus representation onto subsequent brain areas yielding successive representations which are increasingly sparse in activity and expansive in the number of neurons. If the feed-forward synaptic weights realizing this mapping are structured according to the inherent organization of the stimuli (e.g., lion versus pig), the increased sparseness and expansion lead to a significant reduction of noise and therefore to a reliable classification (Babadi & Sompolinsky, 2014). However, it remains unclear how the synapses form the required structure despite noise during learning. Furthermore, how can the system reliably adapt to varying levels of noise (e.g., being in a silent forest compared with near a loud stream)?

In the mouse olfactory system, for instance, 1,800 glomeruli receiving signals from olfactory sensory neurons project to millions of pyramidal neurons in the piriform cortex yielding an expansion of the stimulus representation (Franks & Isaacson, 2006; Mombaerts et al., 1996). Activity of the glomeruli is relatively dense with 10%–30% of glomeruli responding to a given natural odor (Vincis, Gschwend, Bhaukaurally, Beroud, & Carleton, 2012), while in the piriform cortex activity drops to 3%–15% indicating an increase in sparseness (Poo & Isaacson, 2009; Stettler & Axel, 2009). A similar picture can be observed in the *Drosophila* olfactory system. Here, 50 glomeruli project to about 2,500 Kenyon cells in the mushroom body (Balling, Technau, & Heisenberg, 1987; Jefferis et al., 2007). While about 59% of projection neurons respond to a given odor, only 6% of Kenyon cells do (Turner, Bazhenov, & Laurent, 2008). Similar ratios have been observed in the locust olfactory system (Perez-Orive et al., 2002). In the cat visual system, the primary visual cortex has 25 times as many outputs than it receives inputs from the LGN (Olshausen, 2003). In addition, V1-responses to natural visual stimuli are significantly sparser than in the LGN (Dan, Atick, & Reid, 1996; Vinje & Gallant, 2000). Both principles of increased expansion and sparseness of stimulus representations apply to other sensory processing systems as well (Brecht & Sakmann, 2002; Chacron, Longtin, & Maler, 2011; Deweese & Zador, 2003).

The functional roles of increased sparseness as well as expansion have already been proposed in the Marr-Albus theory of the cerebellum (Albus, 1971; Marr, 1969). Here, different representations are thought to evoke different movement responses even though the activity patterns overlap. The Marr-Albus theory demonstrates that through expansion and the sparse activity of granule cells, the overlapping patterns are mapped onto nonoverlapping patterns that can easily be classified. A recent theoretical study has focused on sparse and expansive feed-forward networks in sensory processing systems (Babadi & Sompolinsky, 2014). Here, small variations in activity patterns are caused by internal neuronal noise, input noise, or changes in insignificant properties of the stimuli. For reliable stimulus classification, these slightly varying activity patterns belonging to the same underlying stimulus should evoke the same response in a second layer (or brain area) of a sparse and expansive feed-forward network. Surprisingly, although the network is sparse and expansive, random synaptic weights increase both noise and overlap of activity patterns in the second layer. On the other hand, the same network with

Synaptic weights:
The average transmission efficacy of a synapse quantified as a single number being adapted by diverse plasticity processes.

Olfactory system:
The olfactory system includes all brain areas processing sensory information related to the sense of smell.

Piriform cortex:
One brain area in the cerebrum processing sensory information of the sense of smell.

Mushroom body:
A brain area in insects that is important for odor-related learning and memory.

Cerebellum:
A brain area responsible for the control of movement-related functions such as coordination, timing, or precision.

synaptic weights structured according to the organization of stimuli reduces the noise and overlap of activity patterns, simplifying subsequent classification. How a network is able to learn the organization of stimuli, shape its synaptic structure according to this organization, and do so even in the presence of noise is so far unknown.

Synaptic plasticity:
General term for different kinds of biological mechanisms adapting the weights of synapses depending on neuronal activities.

Homeostatic synaptic plasticity:
Synaptic plasticity mechanism adapting the synaptic weights such that the neuronal dynamics remain in a desired "healthy" regime.

Intrinsic plasticity:
General term for different kinds of biological mechanisms adapting the firing threshold or excitability of a neuron.

The generally accepted hypothesis of learning is that it is realized by changes of synaptic weights by the process of (long-term) synaptic plasticity (Hebb, 1949; Martin, Grimwood, & Morris, 2000). Synaptic weights are strengthened or weakened depending on the activity of the pre- and postsynaptic neurons (Bi & Poo, 1998; Bliss & Lømo, 1973; Markram, Lübke, Frotscher, & Sakmann, 1997). Hebbian plasticity describes the process of increasing a synaptic weight if the activity of the two connected neurons is correlated (Hebb, 1949). Several theoretical studies indicate that Hebbian plasticity alone would lead to divergent synaptic and neuronal dynamics, thus requiring homeostatic synaptic plasticity (Triesch, Vo, & Hafner, 2018; G. G. Turrigiano, Leslie, Desai, Rutherford, & Nelson, 1998) to counterbalance and stabilize the dynamics (Miller & MacKay, 1994; Tetzlaff, Kolodziejski, Timme, & Wörgötter, 2011; Yger & Gilson, 2015; Zenke & Gerstner, 2017; Zenke, Hennequin, & Gerstner, 2013). In addition, neurons adapt their excitability by the process of intrinsic plasticity (Triesch, 2007; Zhang & Linden, 2003). Intrinsic plasticity regulates the excitability of a given neuron so as to maintain a desired average activity (Benda & Herz, 2003; Desai, Rutherford, & Turrigiano, 1999; LeMasson, Marder, & Abbott, 1993; G. Turrigiano, Abbott, & Marder, 1994), which leads to, for instance, the optimization of the input-output relation of a neuron (Triesch, 2007) or the encoding of information in firing rates (Stemmler & Koch, 1999). Several theoretical studies indicate that the interplay of intrinsic plasticity with synaptic plasticity allows neuronal systems to infer the stimulus intensity (Monk, Savin, & Lücke, 2018; Monk, Savin, & Lücke, 2016), to perform independent component analysis (Savin, Joshi, & Triesch, 2010), or to increase their computational capacity (Hartmann, Lazar, Nessler, & Triesch, 2015; Lazar, Pipa, & Triesch, 2009). However, it remains unclear whether this interplay allows sensory systems, on the one hand, to learn the organization of stimuli despite noise and, on the other hand, to adapt to variations of the noise level.

In the present study, we show that in an expansive network intrinsic plasticity regulates the neuronal activities such that the synaptic weights can learn the organization of stimuli even in the presence of low levels of noise. Interestingly, after learning, the system is able to adapt itself according to changes in the level of noise it is exposed to—even if these levels are high. To do so, intrinsic plasticity has to readapt the excitability of the neurons while the synaptic weights have to be maintained, indicating the need of a two-phase learning protocol.

In the following, first, we present the basics of our theoretical model and methods and demonstrate the ability of a feed-forward network with static random or static structured synaptic weights, respectively, to distinguish between noisy versions of $1,000$ different stimuli (similar to Babadi & Sompolinsky, 2014). Then, we introduce the synaptic and intrinsic plasticity rules considered in this study. We train the plastic feed-forward network during an *encoding phase* without noise and test its performance afterwards by presenting stimuli of different noise levels. Intriguingly, the self-organized dynamics of synaptic and intrinsic plasticity yield a performance and network structure similar to the static network initialized with structured synaptic weights. Further analyses indicate that the performance of the plastic network to classify noisy stimuli greatly depends on the neuronal excitability, especially for high levels of noise. Hence, after learning without noise, we changed the noise level in order to test the performance but let intrinsic plasticity readapt the excitability of the neurons. This *readaptation phase* significantly increases the performance of the network. Note, however, if synaptic plasticity is present

during this second phase, the increase in performance is impeded by a prolonged and severe performance decrease. In the next step, we show that in the encoding phase with both intrinsic and synaptic plasticity the network can also learn from noisy stimuli if the level of noise is low. Again, high levels of noise impede learning and classification performance. Interestingly, after the subsequent readaptation phase the network initially trained with low-noise stimuli performs just as well as a network trained with noise-free stimuli, demonstrating the robustness of this learning mechanism to noise.

RESULTS

Model Setup and Classification Performance

The main question of this study concerns how sparse and expansive neural systems, such as sensory processing areas, learn the inherent organization of stimuli enabling a reduction of noise. To tackle this question, similar to a previous study (Babadi & Sompolinsky, 2014), we consider a neural network that consists of two layers of rate-based neurons, with the first layer being linked to the second layer via all-to-all feed-forward synaptic connections. The first layer, called stimulus layer, is significantly smaller ($N_S = 1,000$ neurons) than the second one, called cortical layer ($N_C = 10,000$ neurons). The activity patterns of the stimulus layer serve as stimuli or inputs to the cortical layer. These stimulus patterns are constructed of firing rates $S_i \in \{0, 1\}$ of the stimulus neurons $i \in \{1, ..., N_S\}$ with 0 representing a silent neuron and 1 a maximally active one. Neurons belonging to the cortical layer posses a membrane potential u_j ($j \in \{1, ..., N_C\}$) modeled by a leaky integrator receiving the inputs from the stimulus layer. The membrane potential of a cortical neuron is transformed into a firing rate C_j using a sigmoidal transfer function. Similar to the stimulus neurons, we consider the minimal and maximal firing rates $\mathcal{F}^{\min} = 0$ and $\mathcal{F}^{\max} = 1$. Note that the point of inflection of the sigmoidal transfer function ε_j, also called cortical firing threshold, is neuron-specific.

The different activity patterns of the stimulus layer are organized into $P = 1,000$ stimulus clusters (Figure 1A). Each stimulus cluster $\nu \in \{1, ..., P\}$ consists of one characteristic activity pattern, called central stimulus pattern $\bar{S}^\nu$, which represents the underlying stimulus (e.g., a lion; black dots in the stimulus layer's phase space in Figure 1A). To construct these central stimulus patterns, for each cluster ν and each stimulus neuron i we randomly choose a firing rate $\bar{S}_i^\nu \in \{0, 1\}$ with equal probability, thus resulting in random patterns of ones and zeros (see Figure 1B for schematic examples). In addition, a stimulus cluster contains all noisy versions S^ν of the underlying stimulus (e.g., a lion behind a tree or a rock; indicated by blue halos in Figure 1A) generated by randomly flipping firing rates $\bar{S}_i^\nu$ of the cluster's central stimulus pattern from 1 to 0 or vice versa with probability $\Delta S/2$ (Figure 1B); ΔS thus reflects the average noise level of all noisy stimulus patterns as well as the stimulus cluster's size in the stimulus layer's phase space. If $\Delta S = 0$, the cluster is only a single point in the stimulus layer's phase space (the central stimulus pattern $\bar{S}^\nu$) and is thus noise-free. The maximum value of the stimulus cluster size $\Delta S = 1$ represents a cluster that is distributed evenly across the entire phase space. Here, the noise is so strong that no information remains. The stimulus cluster size ΔS can be retrieved by the normalized Hamming distance between patterns of the same cluster:

$$\Delta S = \left\langle \frac{\sum_{i=1}^{N_S} |S_i^\nu - \bar{S}_i^\nu|}{N_S \cdot 1/2} \right\rangle_{S^\nu, \nu}, \tag{1}$$

with the brackets denoting the average over all noisy stimulus patterns S^ν of all stimulus clusters ν.
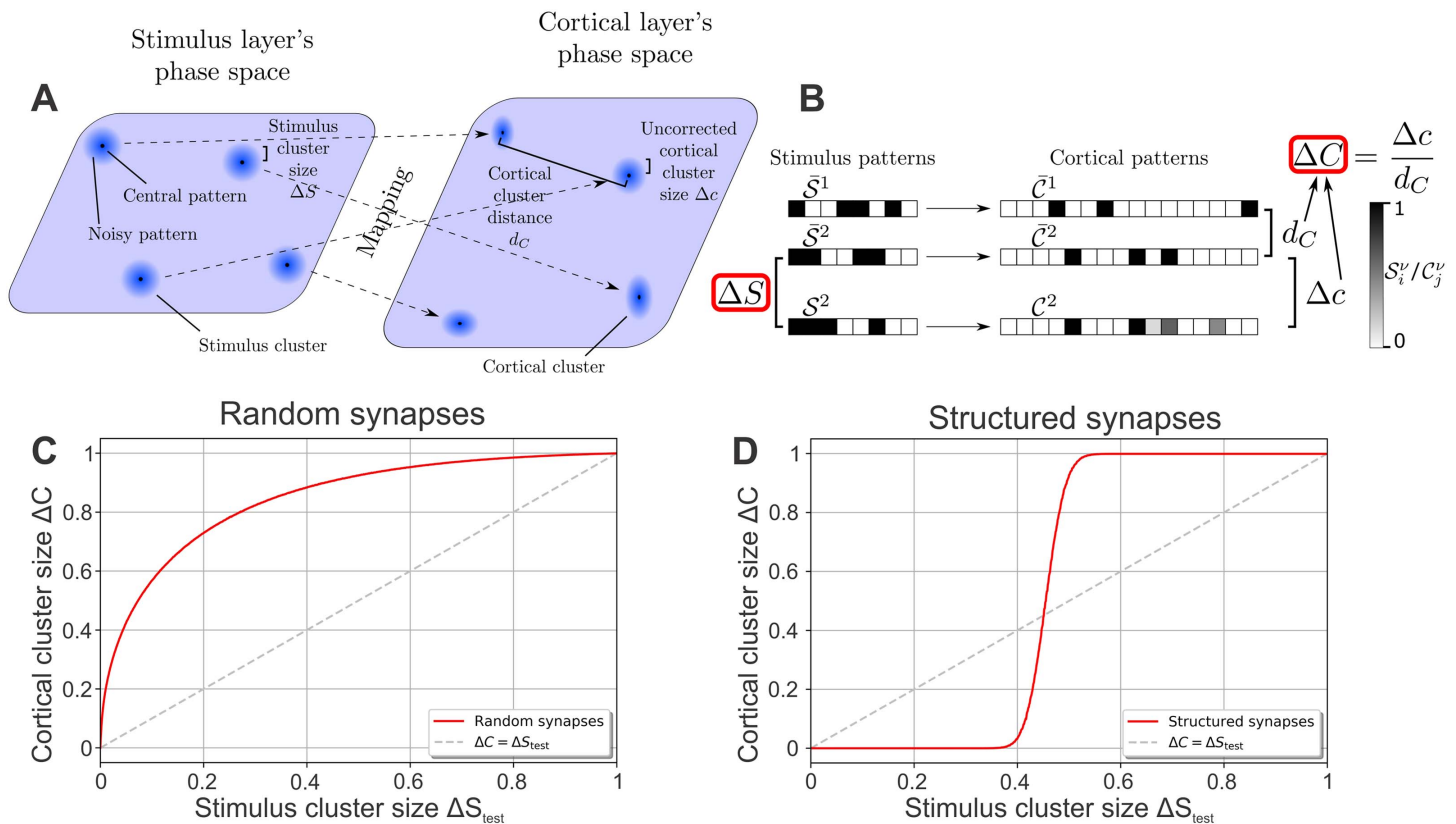
Figure 1. Network model and mathematical approach to quantify the ability of an expansive, sparse network to reduce noise. (A) The feed-forward network consists of two layers of rate-based neurons with the stimulus layer projecting stimuli onto the cortical layer via all-to-all feed-forward synaptic connections. Stimuli are organized in $P = 1,000$ clusters, with each cluster ν consisting of a characteristic central pattern $\bar{\mathcal{S}}^\nu$ (black dots) and noisy patterns $\mathcal{S}^\nu$ (blue halos around dots). The size of the stimulus clusters ΔS corresponds to the level of noise and is indicated schematically by the size of the blue halos. Stimulus clusters are mapped by the synaptic connections to cortical clusters containing central cortical patterns $\bar{\mathcal{C}}^\nu$ and noisy cortical patterns $\mathcal{C}^\nu$. (B) Illustration of different patterns and measures used in this study. The activity of each neuron (box) is indicated by its gray scale (left: stimulus layer; right: cortical layer). The central pattern $\bar{\mathcal{S}}^\nu$ of each stimulus cluster (underlying stimulus) evokes a specific central pattern $\bar{\mathcal{C}}^\nu$ in the cortical layer. Noisy versions of a central stimulus pattern (here $\mathcal{S}^2$) activate different cortical patterns with their average distance Δc from the original pattern depending on the structure of the feed-forward synaptic weights. (C) Random synaptic weights increase the cluster size for all stimulus cluster sizes, that is, $\Delta C > \Delta S_{\text{test}}$; the noise in the stimuli is thus amplified by the network. (D) Synapses that are structured in relation to the organization of the underlying stimuli (stimulus central patterns $\bar{\mathcal{S}}^\nu$) decrease the size of clusters, that is, the noise, up to a medium noise level ($\Delta S_{\text{test}} \approx 0.45$). (C, D) Dashed line indicates $\Delta C = \Delta S_{\text{test}}$.

Every activity pattern of the stimulus layer elicits an activity pattern in the cortical layer, such that stimulus clusters are mapped to cortical clusters (dashed arrows in Figure 1A). Similar to the stimulus cluster, each cortical cluster consists of one central pattern $\bar{\mathcal{C}}^\nu$ (evoked by the noise-free stimulus $\bar{\mathcal{S}}^\nu$) and noisy patterns $\mathcal{C}^\nu$ (evoked by the noisy stimuli $\mathcal{S}^\nu$). Because of the complex mapping of the stimulus patterns onto the cortical layer via the feed-forward synaptic weights, it is not clear how the level of noise is affected by this mapping. Therefore, we estimate the noise in the cortical layer in analogy to Equation 1:

$$\Delta c = \left\langle \frac{\sum_{j=1}^{N_C} |\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{N_C \cdot Z(\mathcal{C}^\nu, \bar{\mathcal{C}}^\nu)} \right\rangle_{\mathcal{C}^\nu, \nu}, \tag{2}$$

where $Z(\mathcal{C}^\nu, \bar{\mathcal{C}}^\nu)$ is a normalization factor (see Methods for more details). As different stimulus clusters are mapped by the same feed-forward weights onto cortical clusters, random

correlations between the cortical clusters could be induced. To account for these correlations we calculate the average distance between clusters by

$$d_C = \left\langle \frac{\sum_{j=1}^{N_C} |\bar{\mathcal{C}}_j^\kappa - \bar{\mathcal{C}}_j^\lambda|}{N_C \cdot Z(\bar{\mathcal{C}}^\kappa, \bar{\mathcal{C}}^\lambda)} \right\rangle_{\kappa,\lambda} , \qquad (3)$$

and correct Equation 2 using this cortical cluster distance (Equation 3) analogous to a signal-to-noise/noise-to-signal ratio to obtain the cortical cluster size

$$\Delta C = \frac{\Delta c}{d_C} . \qquad (4)$$

Therefore, if each pattern $\mathcal{S}^\nu$ of a stimulus cluster ν is mapped onto a different (random) pattern $\mathcal{C}^\nu$ in the cortical layer, $\Delta C = 1$ and the cluster is distributed evenly over the entire cortical layer's phase space. If each pattern of a stimulus cluster is mapped onto the same pattern of the cortical cluster (the central pattern $\bar{\mathcal{C}}^\nu$), $\Delta C = 0$.

In summary, both the stimulus cluster size ΔS as well as the cortical cluster size ΔC are measures for the amount of random fluctuations of different activity patterns belonging to the same underlying stimulus. As such, a network tasked with reducing these random fluctuations should decrease the cluster size, that is, $\Delta C < \Delta S$.

Static Networks

Central to the performance in reducing the cluster size or noise are the feed-forward synaptic weights ω_{ji} between neurons. In the following, we predefine the synaptic weights and test the performance of the network for different levels of noise ΔS_{test} while keeping the synaptic weights fixed. For each noise level ΔS_{test}, we create noisy stimulus patterns for all clusters and use them to evaluate the average noise ΔC in the cortical layer. By doing so, we obtain a performance curve $\Delta C(\Delta S_{\text{test}})$ of the network. If the weights are initialized randomly, here drawn from a Gaussian distribution $\mathcal{N}(0, 2/\sqrt{N_S})$, the cortical cluster size ΔC is always larger than the stimulus cluster size ΔS_{test}, as the performance curve (red line in Figure 1C) is above the identity line ($\Delta C = \Delta S_{\text{test}}$; dashed line) for all values of ΔS_{test}. In other words, the noise of the stimuli (ΔS_{test}) is amplified by the network by increasing the variations between different cortical patterns of the same underlying stimulus ($\Delta C > \Delta S_{\text{test}}$). Note that this amplification of noise is present even though the network is expansive and sparse (Babadi & Sompolinsky, 2014).

This picture changes if the weights are structured according to the organization of the environmental stimuli. To portray such a structure, we initialize the synaptic weights according to Babadi and Sompolinsky (2014) and Tsodyks and Feigelman (1988):

$$\omega_{ji} = \frac{100}{N_S} \sum_{\nu=1}^{P} (\bar{\mathcal{S}}_i^\nu - 1/2)(\mathcal{R}_j^\nu - \mathcal{F}^T) . \qquad (5)$$

Note that the factor 100 ensures that the synaptic weights are in the same order of magnitude as in later analyses. Equation 5 results in a mapping of the central stimulus patterns $\bar{\mathcal{S}}^\nu$ to randomly generated, $\mathcal{F}^T$-sparse cortical patterns $\mathcal{R}^\nu$ ($\mathcal{F}^T = 0.001$). Interestingly, this mapping yields a reduction of noise for up to medium levels ($\Delta S_{\text{test}} \lesssim 0.45$) such that the cortical cluster size ΔC is smaller than the stimulus cluster size ΔS_{test} (Figure 1D). In other words, as already shown in a previous study (Babadi & Sompolinsky, 2014), a structured network reduces

small fluctuations of representations of the same underlying stimulus. Note that in the random as well as the structured network each cortical neuron has an individual firing threshold ε_j. Neuron-specific thresholds are required in order to ensure that every cortical neurons' average response to the central stimulus patterns equals the target activity; that is, $\langle \bar{\mathcal{C}}_j^\nu \rangle_\nu = \mathcal{F}^T$ for all j. We chose $\mathcal{F}^T = 0.001$ as this results in all cortical neurons of the structured network firing in response to exactly one central stimulus pattern, and remaining silent in response to all others (as $\mathcal{F}^T P = 1$), which simplifies the qualitative analysis of the results. In the structured network, the method used for initializing the firing thresholds of each cortical neuron places them at the center of the strongest and second strongest membrane potentials evoked by the central stimulus patterns.

These results show that expansive and sparse networks reduce the noise of stimuli if the synaptic weights from the stimulus to the cortical layer are structured according to the underlying organization of stimuli (here according to the central stimulus patterns $\bar{\mathcal{S}}^\nu$). So far, we have used Equation 5 to artificially set the synaptic weights to the correct values. The question remains how a network can learn these values from the environmental stimuli.

Plastic Network

As demonstrated above, a network with random synaptic weights increases the level of noise, while a structured network decreases it (Figure 1C, D). How can a network develop this structure in a self-organized manner given only the environmental stimuli? To investigate this question, we initialized a network with the same random synaptic weights as above, that is, Gaussian distributed ω_{ji}, and let the system evolve over time using plasticity mechanisms that adapt the synaptic weights and neuronal excitabilities. These plasticity mechanisms are assumed to depend on local quantities only and thus on the directly accessible neuronal activities and synaptic weights (Gerstner & Kistler, 2002; Tetzlaff et al., 2011). Given this assumption, the environmental stimuli influence the dynamics of the plasticity mechanisms as the stimulus patterns determine the activities of the neurons. We consider two plasticity processes: Synaptic weights are controlled by Hebbian correlation learning and an exponential decay term (for weight stabilization),

$$\dot{\omega}_{ji} = \mu \mathcal{S}_i \mathcal{C}_j - \eta \omega_{ji} \, , \tag{6}$$

while a faster intrinsic plasticity mechanism regulates the firing thresholds ε_j of the cortical neurons so as to achieve the target firing rate $\mathcal{F}^T = 0.001$:

$$\dot{\varepsilon}_j = \kappa(\mathcal{C}_j - \mathcal{F}^T), \tag{7}$$

with the parameters μ, η, κ determining the timescales of the mechanisms. Similar to previous studies (Lazar et al., 2009; Miner & Triesch, 2016; Triesch, 2007), we consider that the process of intrinsic plasticity is faster than synaptic plasticity.

Training is carried out in repeated learning steps or trials. In each learning step L, we present all central stimulus patterns $\bar{\mathcal{S}}^\nu$ ($\nu \in \{1, ..., P\}$) to the network once, ensuring there is no chronological information (see Methods for details). This corresponds to a stimulus cluster size $\Delta S_{\text{learn}} = 0$ or noise-free learning. At different stages of learning (that is, after different numbers of learning steps), we test the performance of the network for different levels of noise ΔS_{test} as has been done for the static networks.

As learning progresses (Figure 2A), the performance curve develops from the random network's (red line), which amplifies stimulus noise, into one similar to the structured network's performance curve (blue compared with magenta line). The plasticity mechanisms (Equations 6 and 7) enable the network to encode the organization of the stimuli (existence of different
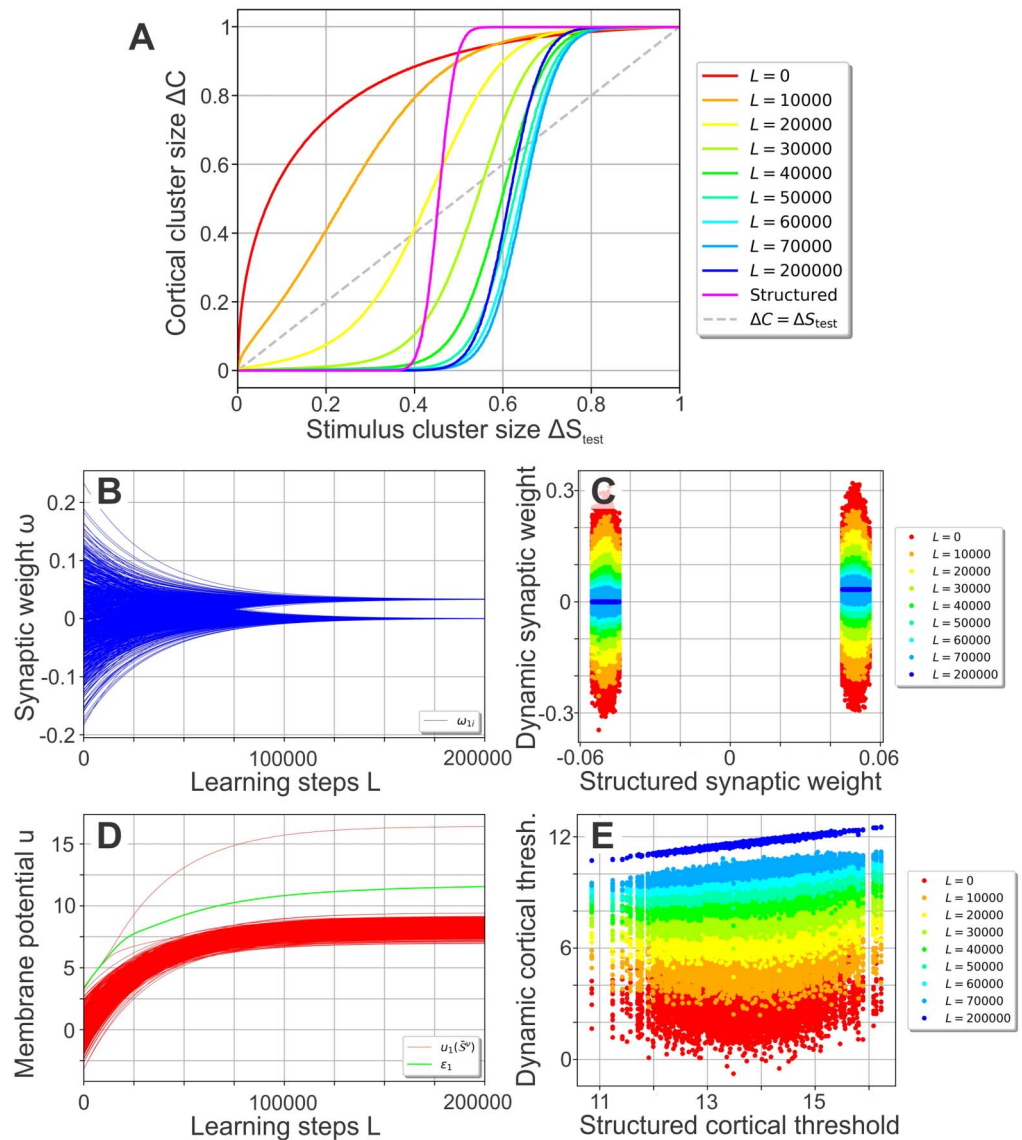

Figure 2. Self-organization of the synaptic and neuronal structure via synaptic and intrinsic plasticity in a noise-free environment. (A) By repeatedly presenting one stimulus pattern $\mathcal{S}^\nu$ per cluster per learning step L using a stimulus cluster size $\Delta S_{\text{learn}} = 0$ (i.e., presenting the central stimulus patterns $\bar{\mathcal{S}}^\nu$), the network's performance develops from the noise-amplification of a random network (red, equal to Figure 1C) to a performance significantly decreasing the level of noise for ΔS_{test} up to about 0.6 (blue). (B, C) During learning, the synaptic weights develop into a bimodal distribution (B; only the weights connecting to neuron 1 are shown) that is correlated to the distribution of the static, structured network (C). (D) For each cortical neuron (here shown for neuron 1), the firing threshold (green) increases such that only one central stimulus pattern can evoke a membrane potential larger than the threshold (red lines depict membrane potentials). (E) Similar to the synaptic weights (C), the firing thresholds tend to become correlated to the ones of the static, structured network.

clusters) in a self-organized manner, with most of the performance gained in the first $L = 60,000$ learning steps: During learning, the synaptic weights evolve from the initial Gaussian distribution into a bimodal distribution with peaks at about 0.033 and 0 (see Figure 2B for an example). The emergence of the bimodal weight distribution and its link to the network performance can be explained as follows: Because of the random initialization of the synaptic weights, each central stimulus pattern leads to a different membrane potential in a given cortical neuron such that all P stimuli together yield a random distribution of evoked membrane potentials (see, e.g., Figure 2D; red lines depict membrane potentials). As the target firing rate is chosen such that each neuron ideally responds to only one central stimulus pattern (as $\mathcal{F}^T P = 1$), intrinsic plasticity adapts the firing threshold ε_j of a neuron such that one of the evoked membrane potentials leads to a distinctly above-average firing rate. Consequently, synapses connecting stimulus neurons being active at the corresponding stimulus pattern with the considered cortical neuron are generally strengthened the most by Hebbian synaptic plasticity. These synapses will likely form the upper peak of the final synaptic weight distribution (Figure 2B). Meanwhile, all other synaptic weights are dominated by the synaptic weight decay (second term in Equation 6) and will later form the lower peak of the distribution at zero. As the continued differentiation of the synaptic weights increases the evoked membrane potential of the most influential central stimulus pattern, these two processes of synaptic and neuronal adaptation drive each other. Interestingly, the resulting synaptic weights are correlated to the structured synapses (Figure 2C) initialized using Equation 5 (here the cortical patterns $\mathcal{R}^\nu$ of Equation 5 were generated using the central cortical patterns $\mathcal{C}^\nu$ of the plastic network at the corresponding learning step L; see Methods for further details). Note that the cortical firing thresholds ε_j of the plastic network become correlated to the values of the static, structured one as well (Figure 2E).

In summary, synaptic and intrinsic plasticity interact and adapt the neuronal network such that, in a noise-free environment, it learns to encode the organization of the stimuli in a way comparable to a static, prestructured network. The trained network is then able to reduce the noise of environmental stimuli even for noise levels up to about 0.6.

The Functional Role of the Cortical Firing Thresholds

While being structurally similar, the performance of the trained, plastic network (Figure 2A, blue) appears significantly better than the performance of the static, structured network (magenta). This fact is not self-explanatory, since both the synaptic weights as well as the cortical firing thresholds are strongly correlated between both networks (Figure 2C, E). However, a closer look at the cortical firing thresholds and their link to the performance of the network reveals the cause of this difference:

In the trained network ($L = 200,000$), as mentioned before, each cortical neuron should fire in response to the central stimulus pattern $\bar{\mathcal{S}}^\nu$ of exactly one cluster and stay silent otherwise. As an example, we will focus on cortical neuron $j = 1$, which fires in response to the central stimulus pattern $\bar{\mathcal{S}}^{842}$ of cluster $\nu = 842$ and remains silent in response to all other central stimulus patterns. In general, two types of errors can occur.

False negatives (a stimulus of cluster 842 is presented and cortical neuron 1 falsely does not fire): Noisy patterns of cluster 842 elicit a distribution of membrane potentials in cortical neuron 1 (Figure 3A), which depends on the stimulus cluster size ΔS_{test}, that is, the level of noise. All noisy stimulus patterns $\mathcal{S}^{842}$ that evoke a membrane potential in neuron 1 that is higher than the neuron's firing threshold ε_1 result in a strong activation of neuron 1. The neuron
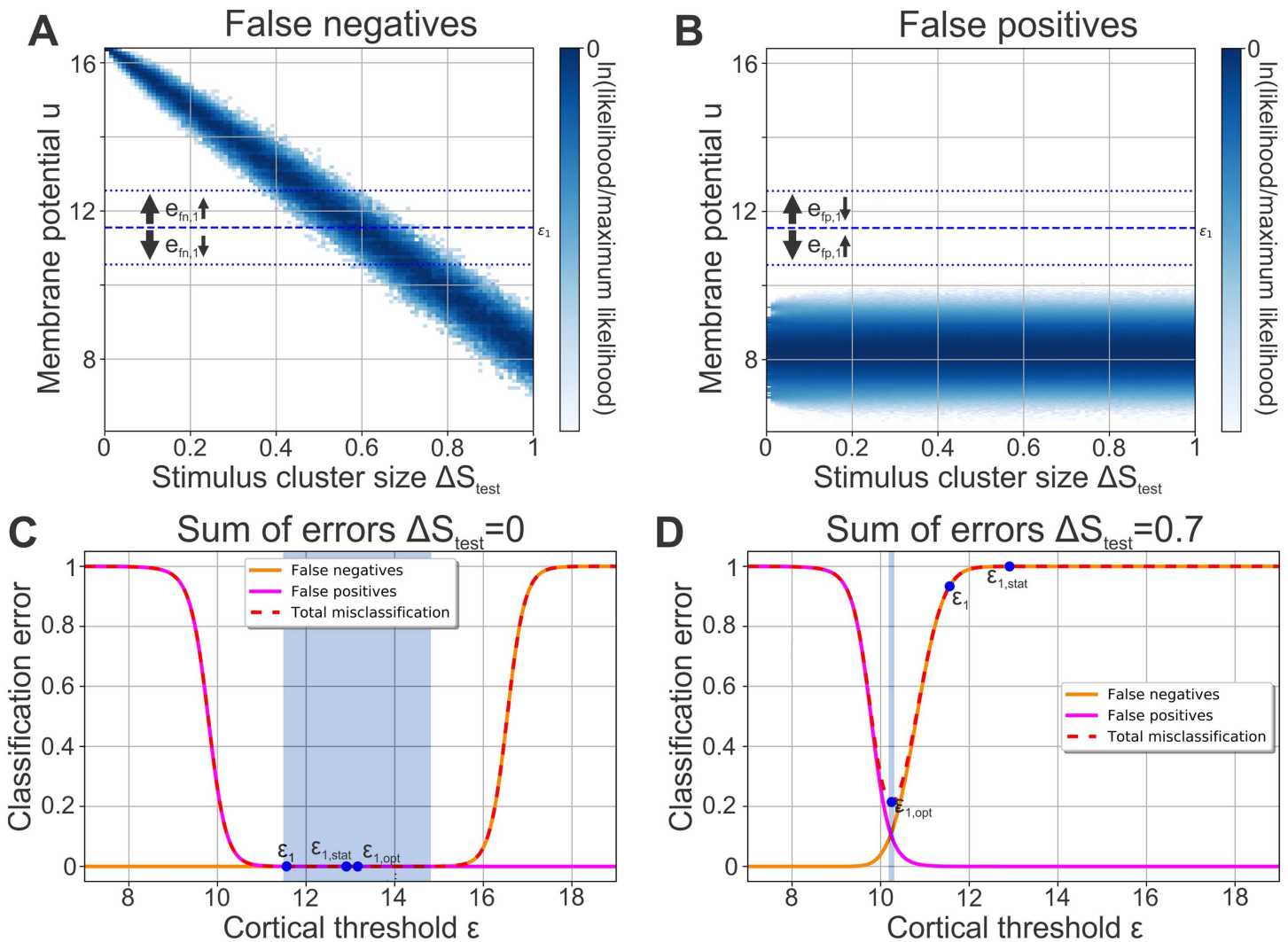
Figure 3. The Classification performance of each neuron depends on its firing threshold. In a single cortical neuron (here neuron $j = 1$), multiple noisy stimulus patterns of the same stimulus cluster elicit a distribution of membrane potentials. Two distinct distributions can be identified: (A) The distribution of membrane potentials evoked by noisy stimulus patterns belonging to the cluster whose central pattern elicits firing in the given cortical neuron (blue; here cluster $\nu = 842$). For any ΔS_{test}, all stimuli yielding a membrane potential that is below the neuron's firing threshold (dashed line; ε_1) do not elicit a strong neuronal response representing false negatives. The distribution significantly depends on the level of noise ΔS_{test}. (B) The membrane potential distribution in response to noisy stimulus patterns of the clusters the neuron is not tuned to ($\nu \neq 842$). Here, all stimuli yielding a membrane potential above the firing threshold are false positives. (C) $\Delta S_{\text{test}} = 0$: A higher firing threshold ε leads to more false negatives (orange) but fewer false positives (magenta) and vice versa for a lower threshold. The sum of errors (dashed red) is negligible in a large regime (blue area: gradient is less than 0.001). (D) $\Delta S_{\text{test}} = 0.7$: With higher levels of stimulus noise, the total error and the classification performance depend critically on the firing threshold. (C, D) $\varepsilon_{1,\text{opt}}$: optimal value of the firing threshold for the given level of noise ΔS_{test} yielding the lowest total error; ε_1: value of the firing threshold after learning with noise-free stimuli (ΔS_{test}; Figure 2); $\varepsilon_{1,\text{stat}}$: firing threshold in the static network (Figure 1D).

therefore classifies these $\mathcal{S}^{842}$ correctly as belonging to cluster 842. However, noisy patterns $\mathcal{S}^{842}$ evoking a lower membrane potential than ε_1 do not elicit strong activation of cortical neuron 1. These noisy patterns are falsely classified as not belonging to cluster 842 and correspond to false negatives.

False positives (a stimulus of a cluster $\nu \neq 842$ is presented and cortical neuron 1 falsely fires): Similar to the analysis of false negatives, the analysis of false positives can be done with

clusters whose central patterns should not elicit activity in cortical neuron 1. The distribution of membrane potentials evoked by noisy patterns of these clusters does not significantly depend on the stimulus cluster size ΔS_{test} (Figure 3B). Noisy stimulus patterns $\mathcal{S}^\nu$ ($\nu \neq 842$) are classified correctly as not part of cluster 842 if neuron 1's membrane potential is lower than its firing threshold ε_1. All noisy patterns evoking a higher membrane potential falsely lead to a firing of cortical neuron 1. They correspond to false positives.

Both false positives and false negatives depend on the firing threshold ε_j of a neuron j. For all values of ΔS_{test}, a lower firing threshold would generally lead to less false negatives ($e_{\text{fn},j}$; Figure 3A) but simultaneously to more false positives ($e_{\text{fp},j}$; Figure 3B) and vice versa for a higher firing threshold. Consequently, there is a trade-off between false negatives and false positives with their sum being related to the network's performance or cortical cluster size (see Methods for derivation):

$$\Delta C \approx e_{\text{tot},j} = e_{\text{fn},j} + e_{\text{fp},j} \qquad \forall j \, . \tag{8}$$

The performance of the network or the total error $e_{\text{tot},j}$ thus depends on a cortical neuron's firing threshold in a nonlinear manner. Given noise-free stimuli ($\Delta S_{\text{test}} = 0$), in a large regime of different values for the firing threshold cortical neuron 1 makes almost no classification error (dashed red line in Figure 3C; gradient in shaded blue area is less than 0.001). For a higher noise level (e.g., $\Delta S_{\text{test}} = 0.7$, Figure 3D), there is no such extended regime of low-error threshold values. Instead, small variations of the firing threshold can drastically change the classification performance, since the membrane potential response distributions overlap at these noise levels (Figure 3A, B).

During training without noise ($\Delta S_{\text{learn}} = 0$), the neuronal firing threshold ε_1 rose to the lower bound of the low-error regime of $\Delta S_{\text{test}} = 0$ (blue area; Figure 3C). In the static network, however, firing thresholds $\varepsilon_{1,\text{stat}}$ were placed at the center of the highest and second highest membrane potentials in response to central stimulus patterns, leading to much higher values. Therefore, if the network performance is tested for small stimulus clusters (low noise ΔS_{test}; Figure 3C), the static and the plastic network have a similar total error and classification performance. For larger stimulus clusters (high noise levels ΔS_{test}; Figure 3D), on the other hand, the higher firing thresholds of the static network lead to considerably more misclassification and consequently to a higher cortical cluster size ΔC. Consequently, the fact that the relation between the threshold and its classification error $e_{j,\text{tot}}$ depends on the noise ΔS_{test} provides an explanation for the large performance differences between the static structured and the plastic network (Figure 2A).

This example (Figure 3) demonstrates that the value of the neuron-specific threshold $\varepsilon_{j,\text{opt}}$ optimizing a neuron's classification performance depends on the stimulus cluster size ΔS_{test} or current level of noise (dotted lines in Figure 4A for neuron 1 in blue and neuron 2 in green). The firing thresholds after training (solid lines in Figure 4A), however, are independent of ΔS_{test}, as they are determined by the noise present during training ($\Delta S_{\text{learn}} = 0$). For $\Delta S_{\text{test}} \lesssim 0.5$ these thresholds are within the regime of low total error (shaded areas indicate the low-error regime for each neuron marked by blue area in Figures 3C and 3D) yielding a high classification performance of the network. However, for $\Delta S_{\text{test}} \gtrsim 0.5$ the thresholds ε_j resulting from training without noise ($\Delta S_{\text{learn}} = 0$) start to deviate significantly from the optimal thresholds $\varepsilon_{j,\text{opt}}$, leading to a decreasing classification performance (Figure 2A and Figure 4C, solid lines for total error of individual neurons). Interestingly, the deviation from the optimal threshold is accompanied by a decrease of the average activity level (solid lines; Figure 4B), while the
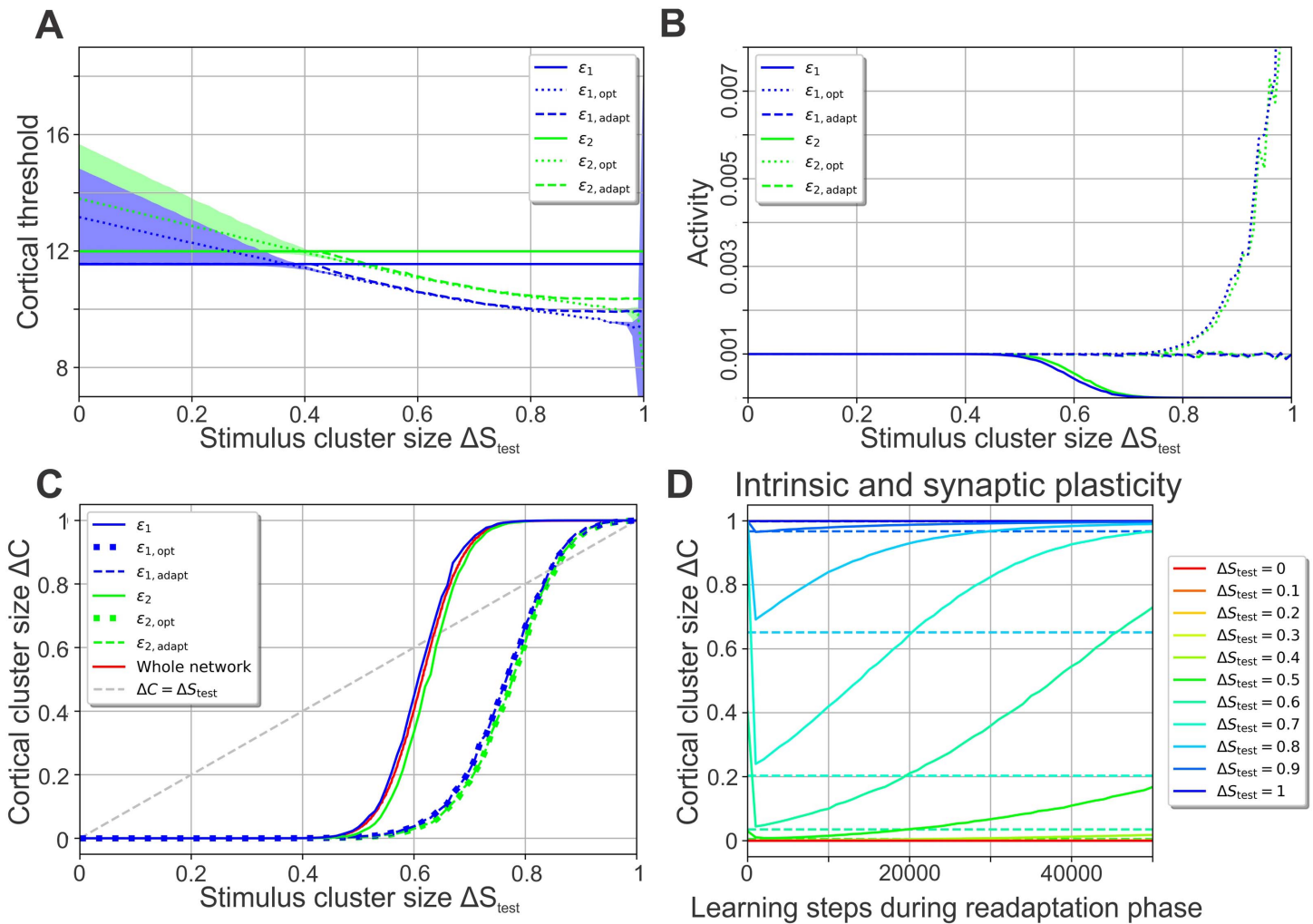
Figure 4. A second learning phase—the readaptation phase—enables the neuronal system to readapt to arbitrary noise levels using intrinsic plasticity. (A–C) After learning without noise, a second learning phase with the noise level ΔS_{test} and only intrinsic plasticity active enables the thresholds to readapt from the values after the first learning phase ε_j (solid lines) to adapted values $\varepsilon_{j,adapt}$ (dashed lines), close to the optimal threshold values $\varepsilon_{j,opt}$ (dotted lines) increasing performance. Blue: neuron 1; green: neuron 2. (A) ΔS_{test}-dependency of cortical thresholds; shaded areas indicate regimes of low error gradient (Figure 3C); (B) ΔS_{test}-dependency of average activities; (C) ΔS_{test}-dependency of total error (dashed lines lie on top of dotted lines). Solid red line shows performance of whole network (from Figure 2A), confirming Equation 8. (D) If synaptic plasticity is present during the second learning phase as well, ΔC initially drops because of intrinsic plasticity and then increases with ongoing presentation of noisy stimuli, indicating a disintegration of the synaptic structure (solid lines; different colors represent different noise levels). Dashed lines indicate ΔC-values for a second learning phase with intrinsic plasticity alone.

optimal thresholds would keep the cortical activity close to the target activity $\mathcal{F}^T = 0.001$ (dotted lines; for $\Delta S_{test} \gtrsim 0.85$ the total error is high and nearly independent of the threshold; see Supplementary Figure 1). We thus expect that after initial learning, intrinsic plasticity could readapt the neuronal firing thresholds according to the present level of noise such that the target activity is maintained and the thresholds $\varepsilon_{j,adapt}$ approximate the optimal threshold values $\varepsilon_{j,opt}$.

We therefore considered a second learning phase, the *readaptation phase*, which is conducted after the initial training or encoding phase is completed. In the readaptation phase, the stimulus cluster size will be the same that the performance is tested for, that is, ΔS_{test}. For now, synaptic plasticity is deactivated as we will only focus on intrinsic plasticity adapting the

cortical firing thresholds $\varepsilon_{j,\text{adapt}}$. To implement this readaptation phase, after the first learning phase is completed, we repeatedly presented one noisy pattern $\mathcal{S}^\nu$ per cluster using a stimulus cluster size ΔS_{test}. Threshold adaptation was stopped when the mean of all cortical thresholds changed by less than 0.0001% in one step, which resulted in less than 7,000 steps for each ΔS_{test}. As expected, intrinsic plasticity adjusts the firing thresholds during this second phase so as to achieve the target firing rate $\mathcal{F}^T$ for all ΔS_{test} (dashed lines; Figure 4B). Furthermore, the adapted thresholds $\varepsilon_{j,\text{adapt}}$ (dashed lines; Figure 4A) are similar to the optimal thresholds $\varepsilon_{j,\text{opt}}$ (dotted lines). This leads to a near-optimal classification performance, which is considerably better than without a readaptation phase (Figure 4C, dashed lines lie on top of dotted line).

Importantly, if synaptic plasticity is also present during this second learning phase, ΔC increases dramatically with ongoing readaptation (solid lines in Figure 4D; different colors represent different noise levels). The initial drop of ΔC is due to intrinsic plasticity (dashed lines show final ΔC-values for intrinsic plasticity alone), while synaptic plasticity leads to a prolonged deterioration of the previously learned synaptic structure if stimuli are too noisy. We therefore conclude that the network has to maintain the synaptic weight structure during the readaptation phase, which we recreate by turning synaptic plasticity off. By doing so, the neuronal system can reliably adjust to stimuli of various noise levels using intrinsic plasticity for adapting the excitability of neurons.

Plastic Networks in Noisy Environments

Up to now, we have shown that a sparse, expansive network can learn the underlying organization of noise-free stimuli ($\Delta S_{\text{learn}} = 0$) by means of synaptic and intrinsic plasticity. Afterwards, a readaptation phase with intrinsic plasticity alone enables the network to readapt to any arbitrary level of noise ΔS_{test} (Figures 4A–C). However, if synaptic plasticity is active during the readaptation phase, the noise of stimuli leads to a disintegration of the synaptic structure (Figure 4D). Therefore, it is unclear whether the network can also learn the organization of stimuli from noisy—instead of noise-free—stimuli by using synaptic plasticity.

To test this, we now investigate the effect of noisy stimuli during training in the encoding phase (i.e., $\Delta S_{\text{learn}} > 0$). To do so, we present one noisy stimulus pattern $\mathcal{S}^\nu$ per cluster in each learning step L using a stimulus cluster size ΔS_{learn}. In noisy environments with up to $\Delta S_{\text{learn}} = 0.2$, cortical neurons show neuronal and synaptic dynamics (Figure 5A, B) similar to noise-free learning (Figure 2B, D). Synaptic weights and firing thresholds become correlated to the static, structured network (Figure 5E, F) to a comparable degree (Figure 2C, E). Nevertheless, because of the noise of the stimuli, some cortical neurons do not manage to separate one stimulus cluster from all others (Figure 5D, $\sim 24\%$ of all neurons for $\Delta S_{\text{learn}} = 0.2$). Consequently, multiple clusters trigger the Hebbian term of synaptic plasticity (Equation 6) such that all synaptic weights approach a medium value (Figure 5C). These synaptic weights diminish the correlation to the static, structured synaptic weights as the final distribution is slightly broader (Figure 5E) than the one from learning without noise (Figure 2C). Furthermore, the cortical neurons without structured incoming synaptic weights (unimodal weight distribution) on average have a lower final firing threshold (blue outliers in Figure 5F).

In general, low levels of noise ($\Delta S_{\text{learn}} \lesssim 0.25$) are tolerated by the network without large losses in performance (Figure 6A). The failed-learning cortical neurons (Figure 5C, D), which become more with higher noise levels (see Supplementary Figure 3), have a negative effect on the performance of the network. At $\Delta S_{\text{learn}} \gtrsim 0.25$, the noise is so strong that the system is not able to recognize and learn the underlying organization of stimuli (that is, the existence of
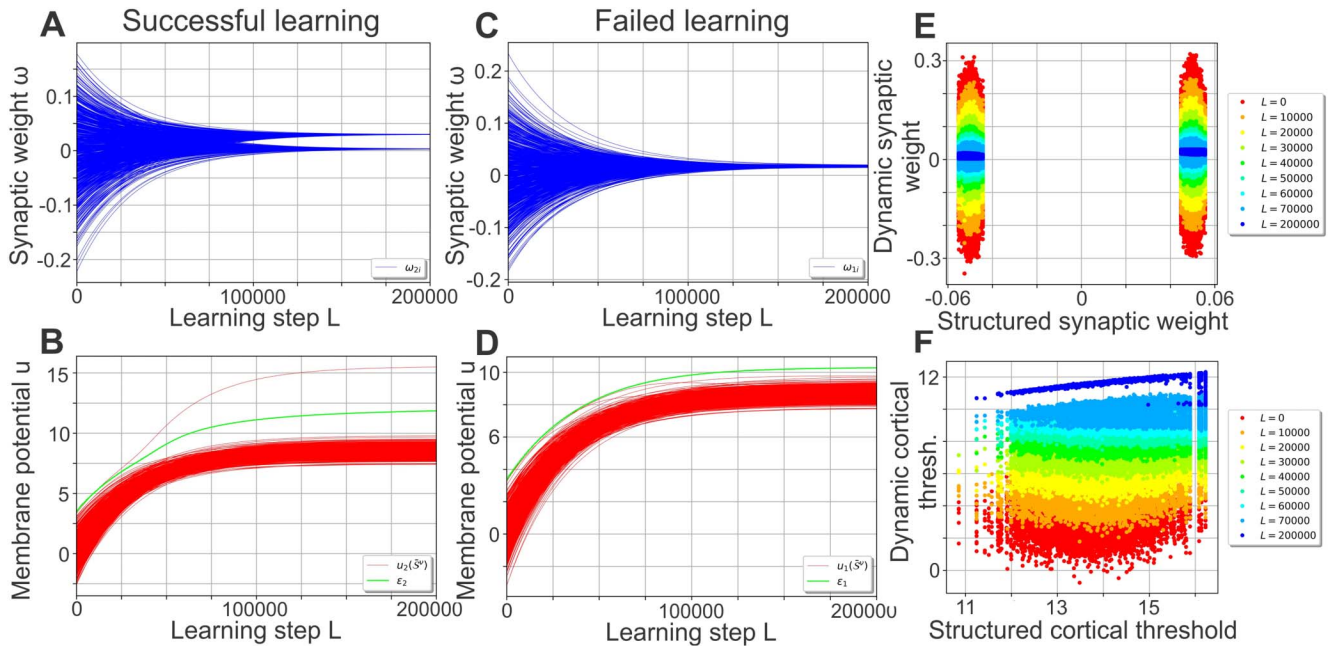
Figure 5. Self-organization of the synaptic and neuronal structure in a noisy environment. The dynamics of synaptic and intrinsic plasticity enable the sparse, expansive network to learn the underlying organization of stimuli from noisy stimulus patterns (here $\Delta S_{learn} = 0.2$). (A, B) The majority of cortical neurons develop a distribution of incoming synaptic weights (A, blue lines) and membrane potential responses (B, red lines) similar to the ones learning without noise (Figure 2B, D). Here shown for neuron 2. Green line in (B) denotes the threshold. (C, D) However, the noise prevents some neurons ($\sim 24\%$) to form a proper synaptic structure (C), yielding a firing threshold (D) that does not separate the membrane potential evoked by one cluster from the others. Therefore, these neurons are not tuned to one specific cluster. Here shown for neuron 1. (E, F) Overall, the network trained by noisy stimuli develops synaptic weights (E) and firing thresholds (F) similarly correlated to the static, structured network than the network trained without noise (Figure 2C, E). The few neurons that failed learning lead to a minor broadening of the distributions.
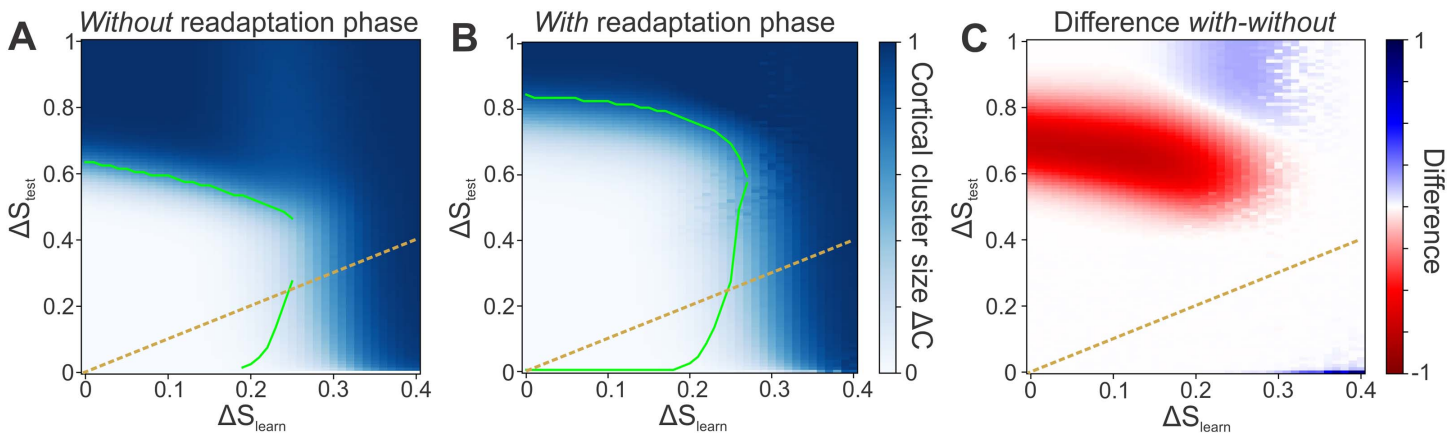

Figure 6. The network can reliably learn from noisy stimuli with and without a readaptation phase. (A) Despite the presence of noise ΔS_{learn} during learning, the network can learn the organization of stimuli and, after encoding, classify stimuli of even higher noise levels ΔS_{test}. However, higher levels of ΔS_{learn} decrease the performance. Color code depicts ΔC, green line marks $\Delta C = \Delta S_{test}$. (B) If the learning phase is followed by a readaptation phase using only intrinsic plasticity and the level of noise ΔS_{test} with which the system is tested, the overall classification performance increases drastically. Now, stimuli with a noise level of up to $\Delta S_{test} \approx 0.8$ can be classified. (C) The readaptation phase leads to a large performance gain for medium and high noise levels ΔS_{test}. Color code depicts the difference between the network without and with a readaptation phase. Red area represents a benefit by using the readaptation phase. (A–C) Orange dashed line: identity line $\Delta S_{learn} = \Delta S_{test}$

different clusters). However, if there is little or even no noise during learning, the network can subsequently not only classify stimuli of that same level of noise, but also classify significantly noisier stimuli (white area above orange dashed identity line). This result indicates that the network does not adapt specifically to only the noise level ΔS_{learn} it is learning from, but that the network generalizes across a broad variety of different noise levels ΔS_{test}. For instance, although the network may learn from stimulus patterns with an average noise level of $\Delta S_{\text{learn}} = 0.1$, it can reliably classify stimuli of noise levels ΔS_{test} from 0 to about 0.6 afterwards.

Furthermore, the performance of a network that was successfully trained in a noisy environment can be drastically improved by a subsequent readaptation phase. Using this second
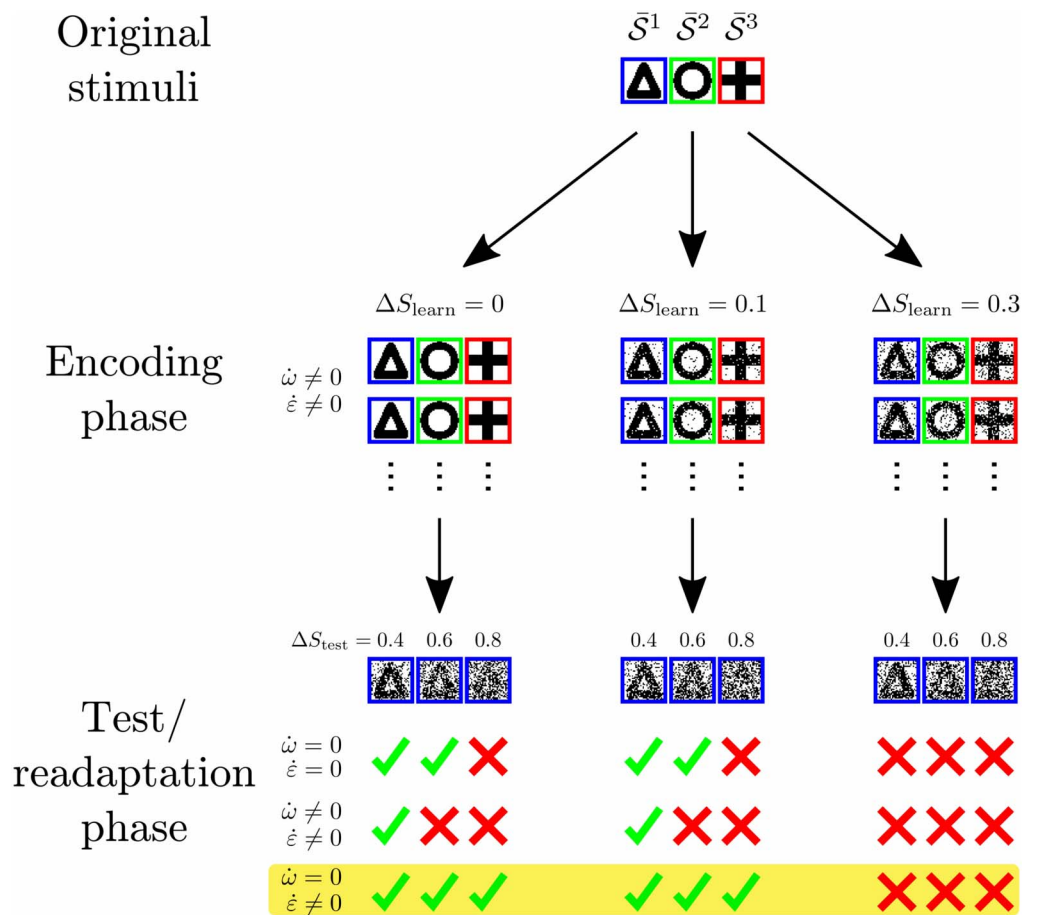

Figure 7. Schematic summary of results. Noisy patterns $\mathcal{S}^\nu$ are repeatedly generated from original stimuli $\bar{\mathcal{S}}^\nu$ (e.g., a triangle, a circle, and a cross) and imprinted on the stimulus layer (encoding phase). If the noise ΔS_{learn} is sufficiently small, synaptic and intrinsic plasticity lead to the formation of structure encoding the organization of stimuli (existence of different geometrical forms). After this initial learning phase, a second learning or readaptation phase enables the network to classify stimuli even in the presence of very high levels of noise ΔS_{test}. Here, only intrinsic plasticity should be present ($\dot{w} = 0$; $\dot{\varepsilon} \neq 0$). This suggests that learning is carried out in two phases: In the first phase, the *encoding phase*, synaptic weights develop to represent the basic organization of the environmental stimuli. This structuring of synaptic weights is most efficient if the noise ΔS_{learn} is low. In the second phase, the *readaptation phase*, learning is dominated by intrinsic plasticity while synaptic weights have to be maintained. The cortical firing thresholds are then able to quickly adapt to the current level of noise ΔS_{test}. Thereby, intrinsic plasticity approximates the optimal thresholds for a given value of ΔS_{test} maximizing performance.

phase in order to (re)adapt the neuronal excitabilities to the level of noise ΔS_{test} that will subsequently be tested for enables the network to classify stimuli up to even higher noise levels of $\Delta S_{\text{test}} \approx 0.8$ (Figure 6B). Consequently, the readaptation phase provides a significant advantage for a large regime of stimulus cluster sizes (red area in Figure 6C). Even more so, stimulus clusters with sizes $\Delta S_{\text{test}} \in (0.6, 0.8)$ can only be classified by using the readaptation phase. The decrease in performance for noise levels between $\Delta S_{\text{learn}} \in (0.2, 0.3)$ and $\Delta S_{\text{test}} \in (0.8, 1.0)$ (blue area) is not crucial given the low level of performance (Figure 6A).

In summary, sparse, expansive networks can learn the clustered organization of noisy stimuli (underlying stimuli might be triangle, circle, and cross like in Figure 7) by the interplay of synaptic and intrinsic plasticity in a self-organized manner. During the initial encoding phase, low levels of noise ΔS_{learn} can be tolerated by the system, while higher levels of noise obstruct the network's ability to learn the organization of stimuli. After the encoding phase, the network can reliably classify noisy patterns of up to $\Delta S_{\text{test}} \approx 0.6$ if synaptic weights and neuronal firing thresholds are fixed ($\dot{w} = 0; \dot{\varepsilon} = 0$). On the other hand, the performance decreases significantly if both synaptic and intrinsic plasticity are allowed to modify the network's structure during the presentation of these noisy stimuli ($\dot{w} \neq 0; \dot{\varepsilon} \neq 0$). Interestingly, if the synaptic structure is maintained while the excitability of the cortical neurons can adapt ($\dot{w} = 0; \dot{\varepsilon} \neq 0$), the network can successfully classify stimuli even in the presence of very high levels of noise (see Figure 7 bottom for examples). These results suggest that learning in the presence of noise requires two distinct phases of plasticity: initial learning of the organization of environmental stimuli via synaptic and intrinsic plasticity in the encoding phase followed by the readaptation phase using only intrinsic plasticity in order to readapt to the current level of noise.

DISCUSSION

How do neuronal systems learn the underlying organization of the surrounding environment in realistic, noisy conditions? In this study, we have shown that sparse and expansive networks can reliably form the required neuronal and synaptic structures via the interplay of synaptic and intrinsic plasticity. Among others, our results indicate that after learning the classification of diverse environmental stimuli in the presence of high levels of noise works best if the synaptic structure is more rigid than the neuronal structure, namely the excitabilities of the neurons. Thereby, our model predicts that higher levels of noise lead to lower firing thresholds or (on average) increased neuronal excitabilities (Figure 4A).

Furthermore, our model predicts that classification performance is highest if the system is adapted to the perceived level of noise. We propose the following psychophysical experiment related to pattern recognition in order to test this prediction: First, subjects have to learn a set of previously unknown patterns, such as visual or auditory patterns. Second, they have to identify noisy versions of these patterns. We propose that the classification performance of a given noisy pattern depends on the history of patterns the subject perceived beforehand. Specifically, our model predicts that a given noisy pattern is classified most reliably if the previously perceived patterns had the same level of noise. By transferring this protocol to an animal model, the predicted course of the adaptation of the firing thresholds could be verified, too.

After the successful learning of the inherent organization of stimuli, in this study we changed the synaptic variability by "turning off" the dynamics of synaptic plasticity (Figure 4). This change of the timescale of synaptic plasticity between the encoding and the readaptation phase could be related to the dynamics during the development of the visual system (Daw, 2003; Daw, Fox, Sato, & Czepita, 1992; Hensch, 2004; Hooks & Chen, 2007). During the critical period, the early visual system is quite susceptible to new sensory experiences and the system

is very plastic. In addition, the visual range during early developmental phases is limited, which could imply lower levels of noise. Thus, the encoding phase in our model could be linked to the critical period. By contrast, the matured visual system is quite rigid, matching the requirements of the readaptation phase, which predicts that the sensory system should be able to adapt to different levels of noise by (only) changing the neuronal excitabilities (Figure 6).

One of the major assumptions of this work, similar to a previous study (Babadi & Sompolinsky, 2014), is that environmental stimuli are organized such that they can be grouped into clusters. Each of these clusters has the same Gaussian noise level ΔS. Natural stimuli, however, have much more structured noise statistics. Nevertheless, the mechanisms considered here that enable the network to compensate for noisy stimuli (i.e., synaptic and intrinsic plasticity) do not specifically rely on the noise being Gaussian. Intrinsic plasticity will still maintain the target firing rate independent of precisely how the membrane potential distributions (Figure 3A, B) are shaped by different types of noise. Given our results (Figure 4), we expect that the neuronal thresholds resulting in the target firing rate will be close to the optimal threshold. Furthermore, the exponential synaptic decay may lead to less reliable presynaptic stimulus neurons having a smaller impact on a cortical neuron's firing. In addition to clusters not being Gaussian shaped, in a natural environment each underlying stimulus may also have a different overall level of noise such that ΔS^ν depends on the cluster ν. However, if the synaptic structure has already been learned during the encoding phase, we expect that cluster-specific $\Delta S^\nu_{\text{test}}$ do not have an impact on the classification performance, as each cortical neuron becomes selective to only one stimulus cluster (Figure 2D). In addition, only the noise level of this selected cluster defines the optimal firing threshold (Figure 3). Therefore, the firing threshold of each neuron can be tuned to its distinct, optimal threshold value, which is independent of the noise levels of other clusters. On the other hand, we expect that different $\Delta S^\nu_{\text{learn}}$ during the encoding phase will lead to over- and underrepresentations of stimulus clusters in the network. Since noise attenuates competition between clusters (Figure 5C, D), clusters with high $\Delta S^\nu_{\text{learn}}$ are less competitive and will subsequently be underrepresented. Nevertheless, the underrepresentation could be an advantage, as stimuli that are too noisy are less informative about the environment than others; consequently, the neuronal system attributes a smaller amount of resources (neurons and synapses) to them. However, the effect of cluster-specific noise on the neuronal and synaptic dynamics have to be investigated further.

Additionally, some stimulus clusters might be perceived more often than others. The corresponding representations would become larger than average, since their relevant synapses are strengthened more often by Hebbian synaptic plasticity, leading to a competitive advantage. Larger representations of more frequently perceived stimulus clusters might provide a behavioral advantage, as these clusters also need to be classified more often. However, the discrepancy between the frequency of such a cluster and the target firing rate of a cortical neuron responding to it might pose a problem. As intrinsic plasticity tries to maintain the target activity, the firing threshold would be placed so high that even slight noise could not be tolerated. One solution might be that neurons could have different target activities (G. G. Turrigiano, 2008) and clusters are selected such that target activity and presentation frequency match. A different mechanism could be global inhibition. A single inhibitory neuron or population of neurons connected to all relevant cortical neurons could homeostatically regulate the activity of the cortical layer by providing inhibitory feedback. Such a mechanism has been identified, for instance, in the *Drosophila* mushroom body (Eichler et al., 2017; Faghihi, Kolodziejski, Fiala, Wörgötter, & Tetzlaff, 2013).

In this study, only one combination of three different plasticity rules was investigated. Of course, many more plasticity mechanisms are conceivable and have been widely studied

(Dayan & Abbott, 2001; Miner & Triesch, 2016; Tetzlaff, Kolodziejski, Markelic, & Wörgötter, 2012; Zenke, Agnes, & Gerstner, 2015). One mechanism could be synaptic scaling regulating the synaptic weights instead of the neuronal excitability such that the neurons reach a certain target firing rate (Desai, Cudmore, Nelson, & Turrigiano, 2002; Hengen, Lambo, Van Hooser, Katz, & Turrigiano, 2013; Keck et al., 2013; Tetzlaff et al., 2011; G. G. Turrigiano et al., 1998). However, the timescale of synaptic scaling is significantly slower than the timescale of intrinsic plasticity, which could increase the duration of the readaptation phase required by the neuronal system to adapt to new levels of noise. On the other hand, faster homeostatic mechanisms (Zenke & Gerstner, 2017) could result in a shorter duration of readaptation. However, the influence of further plasticity mechanisms on the dynamics of sparse, expansive networks has to be analyzed in future studies.

It is usually assumed that homeostatic synaptic plasticity is required for competition (Abbott & Nelson, 2000; Miller, 1996). In the present study, however, competition arises from the interactions of Hebbian synaptic plasticity and homeostatic intrinsic plasticity alone. Homeostatic intrinsic plasticity maintains a certain activity of a given cortical neuron. Stimuli compete for this activity. If one stimulus gains an activity advantage, it will see synapses activated by it strengthened. This leads to less strengthening of other synapses, because the occurrence of Hebbian synaptic plasticity is limited by homeostatic intrinsic plasticity. Synapses will only subsequently be weakened due to homeostatic synaptic plasticity (exponential decay term), which does not interfere in the interaction between Hebbian synaptic and homeostatic intrinsic plasticity generating competition (see Supplementary Figure 2). Consequently, the widely held opinion that homeostatic synaptic plasticity is required for competition might have to be revised.

Even though expansion is a common feature of sensory processing networks, it is not a prerequisite for the results presented here. Nonexpansive networks, too, can learn to distinguish different clusters, although they do not reach the performance of an expansive network (see Supplementary Figure 4). This means that nonexpansive networks as well profit from a two-phase learning protocol as suggested here.

Overall, this study suggests the following answer to how networks learn to classify stimuli in noisy environments: Learning takes place in two distinct phases. The first phase is the *encoding phase*. Hebbian synaptic and homeostatic intrinsic plasticity structure synaptic weights so as to represent the organization of stimuli, with each neuron becoming selectively responsive to a single stimulus cluster. Optimal synaptic structure is achieved if stimuli are noise-free. The second learning phase, called *readaptation phase*, ensues in an arbitrarily noisy environment. Here, synaptic weights have to be maintained in order to preserve the previously learned synaptic structure. Meanwhile, homeostatic intrinsic plasticity regulates the activity of neurons. The firing thresholds are thereby adapted to their optimal values, maximizing classification performance in the current environment (Figure 7).

METHODS

Network and Plasticity Mechanisms

In this study, a two-layered feed-forward network of rate-based neurons is investigated (Figure 1A). The first layer, called stimulus layer, consists of $N_S = 1,000$ neurons, while the second layer, called cortical layer, consists of $N_C = 10,000$ neurons. Feed-forward synaptic connections exist from all stimulus to all cortical neurons. Their synaptic strengths are given

by ω_{ji} where $j \in \{1, ..., N_C\}$ denotes the postsynaptic cortical neuron and $i \in \{1, ..., N_S\}$ the presynaptic stimulus neuron. No recurrent connections are present.

The neurons of the stimulus layer will act as input. As such, the firing rate $\mathcal{S}_i$ of stimulus neuron i will be set to either 0 or 1. Each input therefore is a pattern of firing rates $\mathcal{S}_i \in \{0, 1\}$ on the stimulus layer. These firing rates elicit membrane potentials in the cortical neurons, which follow the leaky integrator equation $\dot{u}_j = -u_j + \sum_{i=1}^{N_S} \omega_{ji} \mathcal{S}_i$. We assume that each input pattern is presented long enough such that the membrane potential mostly resides in the fixed point for the current input. In order to save computation time, we therefore discard the leaky integrator dynamics and simplify the membrane potential to the fixed point of the leaky integrator equation:

$$u_j = \sum_{i=1}^{N_S} \omega_{ji} \mathcal{S}_i . \tag{9}$$

The membrane potential u_j will then be translated into a firing rate $\mathcal{C}_j$ of cortical neuron j via the sigmoidal transfer function

$$\mathcal{C}_j = \frac{\mathcal{F}^{\max}}{1 + \exp(\beta(\varepsilon_j - u_j))} , \tag{10}$$

resulting in cortical firing rates between 0 and $\mathcal{F}^{\max}$. The steepness of the sigmoidal function is given by $\beta = 5$, the maximum firing rate $\mathcal{F}^{\max} = 1$, and the point of inflection ε_j is specific to each cortical neuron j. ε_j corresponds to a neuron-specific firing threshold determining the neuronal excitability.

Intrinsic plasticity regulates this neuron-specific firing threshold ε_j. In order for each cortical neuron j to reach a target firing rate $\mathcal{F}^T = 0.001$, the point of inflection of the sigmoidal transfer curve follows the dynamics

$$\dot{\varepsilon}_j = \kappa(\mathcal{C}_j - \mathcal{F}^T) . \tag{11}$$

The parameter $\kappa = 1 \cdot 10^{-2}$ determines the adaptation speed of intrinsic plasticity. If the firing rate $\mathcal{C}_j$ of cortical neuron j is larger than the target firing rate $\mathcal{F}^T$, the threshold ε_j increases such that $\mathcal{C}_j$ decreases (assuming the input stays constant), and vice versa.

The feed-forward synaptic connections ω_{ji} between the postsynaptic cortical neuron j and the presynaptic stimulus neuron i are controlled by unsupervised synaptic plasticity:

$$\dot{\omega}_{ji} = \mu \mathcal{S}_i \mathcal{C}_j - \eta \omega_{ji} . \tag{12}$$

The parameters $\mu = 1 \cdot 10^{-5}$ and $\eta = 3 \cdot 10^{-8}$ determine the speed of the Hebbian correlation learning term and the exponential decay of synaptic weights, respectively. We assume that synaptic plasticity acts much slower than the presentation time of a single input pattern such that the fixed point of the leaky integrator given by Equation 9 does not significantly change during the presentation of a single input and the simplification thus still holds.

Clustered Stimuli

The structuring of the inputs and the analysis methods are similar to a previous work (Babadi & Sompolinsky, 2014). Here, sensory stimuli are grouped in $P = 1,000$ clusters. Each cluster

comprises different sensory impressions of the same environmental stimulus. Its main component is a characteristic neuronal firing pattern, called the central stimulus pattern $\bar{\mathcal{S}}^\nu$, where $\nu \in \{1, ..., P\}$ denotes the cluster (Figure 1A, B). All central patterns are generated by assigning each stimulus neuron i for each stimulus cluster ν a firing rate $\bar{\mathcal{S}}_i^\nu$ of either 0 or 1 with equal probability. In addition to the central pattern, each cluster also contains noisy variants of the central pattern, called noisy patterns $\mathcal{S}^\nu$. Noisy stimulus patterns are generated by randomly changing the central stimulus pattern's firing rates from 1 to 0 or vice versa with probability $\Delta S/2$. ΔS thereby determines the level of noise and consequently the size of the stimulus clusters, and can range from 0 (no noise) to 1 (no correlation remains). Furthermore, it is the normalized average Hamming distance of noisy stimulus patterns to their central stimulus pattern:

$$\Delta S = \left\langle \frac{\sum_{i=1}^{N_S} |\mathcal{S}_i^\nu - \bar{\mathcal{S}}_i^\nu|}{N_S \cdot 1/2} \right\rangle_{\mathcal{S}^\nu, \nu}, \tag{13}$$

with the angular brackets denoting the average over all noisy stimulus patterns $\mathcal{S}^\nu$ of all clusters ν.

All central and noisy stimulus patterns elicit central and noisy cortical patterns $\bar{\mathcal{C}}^\nu$ and $\mathcal{C}^\nu$, respectively, in the cortical layer of the network. In analogy to Equation (13) the (uncorrected) size of the resulting cortical clusters can be defined as Δc via

$$\Delta c = \left\langle \frac{\sum_{j=1}^{N_C} |\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{N_C \cdot Z(\mathcal{C}^\nu, \bar{\mathcal{C}}^\nu)} \right\rangle_{\mathcal{C}^\nu, \nu}. \tag{14}$$

As the firing rates $\mathcal{C}_j^\nu$ and $\bar{\mathcal{C}}_j^\nu$ can take on values between 0 and 1, a more complex normalization $Z(\mathcal{C}^\kappa, \mathcal{C}^\lambda)$ for the patterns $\mathcal{C}^\kappa$ and $\mathcal{C}^\lambda$ is required:

$$Z(\mathcal{C}^\kappa, \mathcal{C}^\lambda) = \frac{1}{N_C^2} \sum_{l=1}^{N_C} \sum_{m=1}^{N_C} |\mathcal{C}_l^\kappa - \mathcal{C}_m^\lambda|. \tag{15}$$

This normalization quantifies the average overlap two random cortical patterns with the same firing rates would have.

Being generated randomly, the central stimulus patterns are uncorrelated among each other. Because of the propagation of these patterns through the synaptic connections, however, the central cortical patterns might not be uncorrelated. In the context of noise reduction a more appropriate performance measure compensates for the introduced correlation. The cortical cluster size ΔC is therefore defined as

$$\Delta C = \frac{\Delta c}{d_C}, \tag{16}$$

where the cortical cluster distance d_C is a measure of the correlation between central cortical patterns:

$$d_C = \left\langle \frac{\sum_{j=1}^{N_C} |\bar{\mathcal{C}}_j^\kappa - \bar{\mathcal{C}}_j^\lambda|}{N_C \cdot Z(\bar{\mathcal{C}}^\kappa, \bar{\mathcal{C}}^\lambda)} \right\rangle_{\kappa, \lambda}. \tag{17}$$

Classification Errors

In the following, the classification errors $e_{\text{fp},j}$ (false positives) and $e_{\text{fn},j}$ (false negatives) of single neurons in a trained network will be set into relation with the cortical cluster size ΔC. To do so, we will first discuss the cortical cluster distance d_C and the ergodicity of the network. We will then use the results to derive the relation between the cortical cluster size and the classification errors.

In order to simplify the following derivations for $\Delta C = \frac{\Delta c}{d_C}$, we consider that the cortical cluster distance $d_C = 1$ in a trained network as discussed in the following. In the 'Plastic Network' section (see, e.g., Figure 2) we have demonstrated that a given cortical neuron becomes responsive to a single stimulus cluster during training. The stimulus cluster that a neuron becomes responsive to is usually the one that initially elicits the strongest membrane potential, as the related synapses will experience the greatest strengthening by Hebbian plasticity. This cluster is a random one, though, as the initial membrane potential depends only on the initially random synaptic weights. Consequently, each cortical neuron becomes responsive to a random stimulus cluster. This implies that the cortical clusters are uncorrelated since each cortical neuron's response to a given cluster is random. By definition, the cortical cluster distance d_C is thus equal to 1. We therefore have

$$\Delta C = \Delta c = \left\langle \frac{\sum_{j=1}^{N_C} |\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{N_C \cdot Z(\mathcal{C}^\nu, \bar{\mathcal{C}}^\nu)} \right\rangle_{\mathcal{C}^\nu, \nu} . \tag{18}$$

Next, we will assume that a trained network is ergodic, that is, we can exchange averages over cortical patterns ("time") with averages over cortical neurons ("space"). Specifically, we assume the following relation to hold:

$$\left\langle \frac{\sum_{j=1}^{N_C} |\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{N_C \cdot Z(\mathcal{C}^\nu, \bar{\mathcal{C}}^\nu)} \right\rangle_{\mathcal{C}^\nu, \nu} = \left\langle \frac{\sum_{\nu=1}^{P} |\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{P \cdot Z(\mathcal{C}_j, \bar{\mathcal{C}}_j)} \right\rangle_{\mathcal{C}_j, j} \tag{19}$$

with

$$Z(\mathcal{C}^\nu, \bar{\mathcal{C}}^\nu) = \frac{1}{N_C^2} \sum_{l=1}^{N_C} \sum_{m=1}^{N_C} |\mathcal{C}_l^\nu - \bar{\mathcal{C}}_m^\nu| , \tag{20}$$

$$Z(\mathcal{C}_j, \bar{\mathcal{C}}_j) = \frac{1}{P^2} \sum_{\kappa=1}^{P} \sum_{\lambda=1}^{P} |\mathcal{C}_j^\kappa - \bar{\mathcal{C}}_j^\lambda| . \tag{21}$$

$\mathcal{C}_j$ and $\bar{\mathcal{C}}_j$ are vectors containing the firing rates of cortical neuron j in response to one noisy/central pattern of each cluster.

In the following, we will divide the assumption about the ergodicity of the network in several smaller assumptions and discuss whether they are valid. First, we assume a large system, that is, $P, N_C \to \infty$, which approximates the network studied here consisting of $P = 1,000$ and $N_C = 10,000$ adequately. Second, we need to assume that the set of cortical clusters is homogeneous, that is, ΔC is the same for all clusters. This is a sufficient approximation as ΔS is the same for all clusters and no cluster is preferred by the cortical neurons in a trained network as discussed before. Given these two assumptions, we can drop the average over noisy patterns $\mathcal{C}^\nu$ in Equation 19, because the average over a single noisy pattern of an infinite number of clusters is equal to the average over all noisy patterns of an infinite number of clusters as long as the clusters are homogeneous. Likewise, by assuming that the set of cortical neurons is

homogeneous, we can drop the average over the sets of noisy firing rates $\mathcal{C}_j$ in Equation 19. In a trained network, where all neurons developed a bimodal weight distribution and have the same target firing rate, this is a decent approximation. We can thus write the following:

$$\left\langle \frac{\sum_{j=1}^{N_C} |\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{N_C \cdot Z(\mathcal{C}^\nu, \bar{\mathcal{C}}^\nu)} \right\rangle_{\mathcal{C}^\nu, \nu} = \left\langle \frac{\sum_{\nu=1}^{P} |\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{P \cdot Z(\mathcal{C}_j, \bar{\mathcal{C}}_j)} \right\rangle_{\mathcal{C}_j, j} \tag{22}$$

$$\stackrel{P, N_C \to \infty}{\Longleftrightarrow} \left\langle \frac{\sum_{j=1}^{N_C} |\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{N_C \cdot Z(\mathcal{C}^\nu, \bar{\mathcal{C}}^\nu)} \right\rangle_\nu = \left\langle \frac{\sum_{\nu=1}^{P} |\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{P \cdot Z(\mathcal{C}_j, \bar{\mathcal{C}}_j)} \right\rangle_j \tag{23}$$

$$\Leftrightarrow \frac{1}{P} \sum_{\nu=1}^{P} \frac{\sum_{j=1}^{N_C} |\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{\frac{1}{N_C} \sum_{l=1}^{N_C} \sum_{m=1}^{N_C} |\mathcal{C}_l^\nu - \bar{\mathcal{C}}_m^\nu|} = \frac{1}{N_C} \sum_{j=1}^{N_C} \frac{\sum_{\nu=1}^{P} |\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{\frac{1}{P} \sum_{\lambda=1}^{P} \sum_{\kappa=1}^{P} |\mathcal{C}_j^\lambda - \bar{\mathcal{C}}_j^\kappa|} \tag{24}$$

$$\Leftrightarrow \sum_{\nu=1}^{P} \sum_{j=1}^{N_C} \frac{|\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{\frac{1}{N_C^2} \sum_{l=1}^{N_C} \sum_{m=1}^{N_C} |\mathcal{C}_l^\nu - \bar{\mathcal{C}}_m^\nu|} = \sum_{\nu=1}^{P} \sum_{j=1}^{N_C} \frac{|\mathcal{C}_j^\nu - \bar{\mathcal{C}}_j^\nu|}{\frac{1}{P^2} \sum_{\kappa=1}^{P} \sum_{\lambda=1}^{P} |\mathcal{C}_j^\kappa - \bar{\mathcal{C}}_j^\lambda|} \tag{25}$$

$$\Leftarrow \frac{1}{N_C^2} \sum_{l=1}^{N_C} \sum_{m=1}^{N_C} |\mathcal{C}_l^\nu - \bar{\mathcal{C}}_m^\nu| = \frac{1}{P^2} \sum_{\kappa=1}^{P} \sum_{\lambda=1}^{P} |\mathcal{C}_j^\kappa - \bar{\mathcal{C}}_j^\lambda| \quad \forall j, \nu \tag{26}$$

$$\Leftrightarrow \left\langle \frac{1}{N_C} \sum_{m=1}^{N_C} |\mathcal{C}_l^\nu - \bar{\mathcal{C}}_m^\nu| \right\rangle_l = \left\langle \frac{1}{P} \sum_{\lambda=1}^{P} |\mathcal{C}_j^\kappa - \bar{\mathcal{C}}_j^\lambda| \right\rangle_\kappa \quad \forall j, \nu. \tag{27}$$

Equation 26 is a sufficient, but not a necessary, condition for Equation 25. Therefore, if we can show that Equation 27 is a valid assumption, this suffices (together with the assumptions mentioned above) for the ergodicity of a trained network.

As demonstrated in the Results section and discussed before, every cortical neuron of the trained network is responsive to a single, random cluster. We need to further assume that every central pattern elicits activity in $\mathcal{F}^T N_C = 10$ of all cortical neurons. This is strictly true only on average, but if cortical neurons respond to a random cluster, it is a decent approximation. Consequently, we can do the following transformations:

$$\left\langle \frac{1}{N_C} \sum_{m=1}^{N_C} |\mathcal{C}_l^\nu - \bar{\mathcal{C}}_m^\nu| \right\rangle_l = \left\langle \frac{1}{P} \sum_{\lambda=1}^{P} |\mathcal{C}_j^\kappa - \bar{\mathcal{C}}_j^\lambda| \right\rangle_\kappa \quad \forall j, \nu \tag{28}$$

$$\Leftrightarrow \left\langle \frac{10}{N_C} |\mathcal{C}_l^\nu - 1| + \frac{N_C - 10}{N_C} |\mathcal{C}_l^\nu - 0| \right\rangle_l = \left\langle \frac{1}{P} |\mathcal{C}_j^\kappa - 1| + \frac{P-1}{P} |\mathcal{C}_j^\kappa - 0| \right\rangle_\kappa \quad \forall j, \nu \tag{29}$$

$$\Leftrightarrow \left\langle \mathcal{C}_l^\nu \right\rangle_l = \left\langle \mathcal{C}_j^\kappa \right\rangle_\kappa \quad \forall j, \nu. \tag{30}$$

That is, for each ΔS every cortical pattern ν and every cortical neuron j must have the same average firing rate. This is true given the assumptions we have already discussed: The cortical neurons are homogeneous, that is, they all have a bimodal weight distribution and so forth, and each cortical neuron is responsive to a random cluster.

In total, we have divided the assumption of ergodicity (Equation 19) of a trained network in simpler assumptions that we were able to validate. Using the ergodicity we now have

$$\Delta C = \left\langle \frac{\sum_{\nu=1}^{P} |\mathcal{C}_j^{\nu} - \bar{\mathcal{C}}_j^{\nu}|}{P \cdot Z(\mathcal{C}_j, \bar{\mathcal{C}}_j)} \right\rangle_{\mathcal{C}_j, j} . \tag{31}$$

Similar to the argument made for Equation 23, in an infinitely large network, averaging over an infinite set of noisy firing rates $\mathcal{C}_j$ of a single cortical neuron j is equal to averaging over an infinite set of noisy firing rates $\mathcal{C}_j$ of all cortical neurons as long as the neurons are homogeneous. We can thus drop the average over j:

$$\Delta C = \left\langle \frac{\sum_{\nu=1}^{P} |\mathcal{C}_j^{\nu} - \bar{\mathcal{C}}_j^{\nu}|}{P \cdot Z(\mathcal{C}_j, \bar{\mathcal{C}}_j)} \right\rangle_{\mathcal{C}_j, j} \tag{32}$$

$$\xrightarrow{P, N_C \to \infty} \left\langle \frac{\sum_{\nu=1}^{P} |\mathcal{C}_j^{\nu} - \bar{\mathcal{C}}_j^{\nu}|}{P \cdot Z(\mathcal{C}_j, \bar{\mathcal{C}}_j)} \right\rangle_{\mathcal{C}_j} \tag{33}$$

$$= \left\langle \frac{\sum\limits_{\nu | \bar{\mathcal{C}}_j^{\nu}=1} |\mathcal{C}_j^{\nu} - \bar{\mathcal{C}}_j^{\nu}| + \sum\limits_{\nu | \bar{\mathcal{C}}_j^{\nu}=0} |\mathcal{C}_j^{\nu} - \bar{\mathcal{C}}_j^{\nu}|}{\frac{1}{P} \sum_{\kappa=1}^{P} \left(1 \cdot |\mathcal{C}_j^{\kappa} - 1| + (P-1) \cdot |\mathcal{C}_j^{\kappa} - 0| \right)} \right\rangle_{\mathcal{C}_j} \tag{34}$$

$$= \left\langle \frac{\sum\limits_{\nu | \bar{\mathcal{C}}_j^{\nu}=1} |\mathcal{C}_j^{\nu} - \bar{\mathcal{C}}_j^{\nu}| + \sum\limits_{\nu | \bar{\mathcal{C}}_j^{\nu}=0} |\mathcal{C}_j^{\nu} - \bar{\mathcal{C}}_j^{\nu}|}{1 + \frac{P-2}{P} \sum_{\kappa=1}^{P} \mathcal{C}_j^{\kappa}} \right\rangle_{\mathcal{C}_j} \tag{35}$$

$$= \underbrace{\left\langle \frac{\sum\limits_{\nu | \bar{\mathcal{C}}_j^{\nu}=1} |\mathcal{C}_j^{\nu} - \bar{\mathcal{C}}_j^{\nu}|}{1 + \frac{P-2}{P} \sum_{\kappa=1}^{P} \mathcal{C}_j^{\kappa}} \right\rangle_{\mathcal{C}_j}}_{\text{false negatives } e_{\text{fn},j}} + \underbrace{\left\langle \frac{\sum\limits_{\nu | \bar{\mathcal{C}}_j^{\nu}=0} |\mathcal{C}_j^{\nu} - \bar{\mathcal{C}}_j^{\nu}|}{1 + \frac{P-2}{P} \sum_{\kappa=1}^{P} \mathcal{C}_j^{\kappa}} \right\rangle_{\mathcal{C}_j}}_{\text{false positives } e_{\text{fp},j}} . \tag{36}$$

Initialization

When initialized randomly, the synaptic weights ω_{ji} are drawn from a Gaussian distribution with mean 0 and variance $\frac{2}{\sqrt{N_S}}$. The synaptic weights can also be initialized in a structured manner according to $\omega_{ji} = \frac{100}{N_S} \sum_{\nu=1}^{P} (\bar{\mathcal{S}}_i^{\nu} - \frac{1}{2})(\mathcal{R}_j^{\nu} - \mathcal{F}^T)$ (similar to Babadi & Sompolinsky, 2014; Tsodyks & Feigelman, 1988). The factor 100 scales the synaptic weights such that they are in the same order of magnitude as the synaptic weights in the trained dynamic network ensuring that they are comparable. $\mathcal{R}^{\nu}$ are cortical patterns that are generated using one of the following methods: For Figure 1 $\mathcal{R}^{\nu}$ are random patterns of ones and zeros where each pattern ν and each cortical neuron j has an activity of $\mathcal{F}^T$. For all other results, $\mathcal{R}^{\nu}$ are computed via $\mathcal{R}_j^{\nu} = \Theta(\bar{\mathcal{C}}_j^{\nu} - \mathcal{T}_j)$, where Θ denotes the Heaviside function and the thresholds $\mathcal{T}_j$ are chosen such that each cortical neuron j achieves an activity of $\mathcal{F}^T$. This results in cortical patterns $\mathcal{R}^{\nu}$ that are correlated to the central cortical patterns $\bar{\mathcal{C}}^{\nu}$ of an existing network.

The cortical membrane thresholds ε_j are then initialized such that each cortical neuron j achieves an average firing rate of the target firing rate $\mathcal{F}^T$ at the central cortical patterns. In order to find the corresponding membrane thresholds ε_j, the secant method is used with initial values of 0 and the mean of the highest and second highest (as $\mathcal{F}^T P = 1$) membrane potentials

of cortical neuron j. If structured synaptic weights are used, this leads to ε_j close to the mean of the highest and second highest membrane potentials of neuron j.

Implementation

Training is done by repeatedly presenting stimulus patterns for one time step $\Delta t = 1$ each. The high computational demand of simulating a network with $11,000$ neurons for $200,000$ training steps each containing $1,000$ patterns for just a single parameter set made parallel simulation of patterns necessary and required the usage of a computer cluster. To this end, each training step consists of parallel simulation of one stimulus pattern per cluster. Stimulus patterns are generated using a stimulus cluster size ΔS_{learn} and the current learning step is denoted by L. The synaptic weights ω_{ji} and cortical thresholds ε_j are updated at the end of each training step according to $\Delta\omega_{ji} = \sum_{\nu=1}^{P} \dot{\omega}_{ji}(\mathcal{C}^\nu)$ and $\Delta\varepsilon_j = \sum_{\nu=1}^{P} \dot{\varepsilon}_j(\mathcal{C}^\nu)$, where $\mathcal{C}^\nu$ denotes the cortical pattern of cluster ν that was simulated in this learning step. This is a reasonable approximation if a single learning step changes the network's state only slightly, as is the case in this study.

The cortical cluster size ΔC (cf. Equation 14 and Equation 16) in response to a stimulus cluster size ΔS_{test} is approximated using 10 noisy patterns per cluster.

If intrinsic plasticity is active during the testing phase, repeatedly, one noisy pattern per cluster is simulated using a stimulus cluster size ΔS_{test}. After the mean of all cortical thresholds changed by less than 0.0001%, the cortical cluster size ΔC is calculated for the given stimulus cluster size ΔS_{test}. In order to speed up its computation, we used the central cortical patterns $\bar{\mathcal{C}}^\nu$ and cortical cluster distance d_C from ahead of the adaptation phase and were able to verify that this does not influence the results. The thresholds are reset to their previous values afterwards. The entire procedure is performed for all ΔS_{test}, each requiring fewer than $7,000$ learning steps for the thresholds to converge.

SUPPORTING INFORMATION

Supporting information for this article is available at https://doi.org/10.1162/netn_a_00118.

AUTHOR CONTRIBUTIONS

FUNDING INFORMATION

REFERENCES

Abbott, L. F., & Nelson, S. B. (2000). Synaptic plasticity: Taming the beast. *Nature Neuroscience*, *3*(11s), 1178–1183. https://doi.org/10.1038/81453

Albus, J. S. (1971). A theory of cerebellar function. *Mathematical Biosciences*, *10*(1), 25–61. https://doi.org/10.1016/0025-5564(71)90051-4

Babadi, B., & Sompolinsky, H. (2014). Sparseness and expansion in sensory representations. *Neuron*, *83*(5), 1213–1226. https://doi.org/10.1016/j.neuron.2014.07.035

Balling, A., Technau, G. M., & Heisenberg, M. (1987). Are the structural changes in adult *Drosophila* mushroom bodies memory traces? Studies on biochemical learning mutants.

Journal of Neurogenetics, *4*(1), 65–73. https://doi.org/10.3109/01677068709102334

Benda, J., & Herz, A. V. M. (2003). A universal model for spike-frequency adaptation. *Neural Computation*, *15*(11), 2523–2564. https://doi.org/10.1162/089976603322385063

Bi, G.-q., & Poo, M.-m. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, *18*(24), 10464–10472. https://doi.org/10.1523/JNEUROSCI.18-24-10464.1998

Bliss, T. V. P., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology*, *232*(2), 331–356. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1350458/

Brecht, M., & Sakmann, B. (2002). Dynamic representation of whisker deflection by synaptic potentials in spiny stellate and pyramidal cells in the barrels and septa of layer 4 rat somatosensory cortex. *Journal of Physiology*, *543*(1), 49–70. https://doi.org/10.1113/jphysiol.2002.018465

Chacron, M. J., Longtin, A., & Maler, L. (2011). Efficient computation via sparse coding in electrosensory neural networks. *Current Opinion in Neurobiology*, *21*(5), 752–760. https://doi.org/10.1016/j.conb.2011.05.016

Dan, Y., Atick, J. J., & Reid, R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. *Journal of Neuroscience*, *16*(10), 3351–3362. https://doi.org/10.1523/JNEUROSCI.16-10-03351.1996

Daw, N. W. (2003). Critical periods in the visual system. In B. Hopkins & S. P. Johnson (Eds.), *Neurobiology of infant vision*. Westport, CT: Praeger.

Daw, N. W., Fox, K., Sato, H., & Czepita, D. (1992). Critical period for monocular deprivation in the cat visual cortex. *Journal of Neurophysiology*, *67*, 197–202. https://doi.org/10.1152/jn.1992.67.1.197

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

Desai, N. S., Cudmore, R. H., Nelson, S. B., & Turrigiano, G. G. (2002). Critical periods for experience-dependent synaptic scaling in visual cortex. *Nature Neuroscience*, *5*(8), 783–789. https://doi.org/10.1038/nn878

Desai, N. S., Rutherford, L. C., & Turrigiano, G. G. (1999). Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nature Neuroscience*, *2*(6), 515–520. https://doi.org/10.1038/9165

Deweese, M. R., & Zador, A. M. (2003). Binary Coding in auditory cortex. *Advances in Neural Information Processing Systems*, 117–124.

Eichler, K., Li, F., Litwin-Kumar, A., Park, Y., Andrade, I., Schneider-Mizell, C. M., . . . Cardona, A. (2017). The complete connectome of a learning and memory centre in an insect brain. *Nature*, *548*, 175–182. https://doi.org/10.1038/nature23455

Faghihi, F., Kolodziejski, C., Fiala, A., Wörgötter, F., & Tetzlaff, C. (2013). An information theoretic model of information processing in the *Drosophila* olfactory system: The role of inhibitory neurons for system efficiency. *Frontiers in Computational Neuroscience*, *7*, 183.

Franks, K. M., & Isaacson, J. S. (2006). Strong single-fiber sensory inputs to olfactory cortex: Implications for olfactory coding. *Neuron*, *49*(3), 357–363. https://doi.org/10.1016/j.neuron.2005.12.026

Gerstner, W., & Kistler, W. M. (2002). Mathematical formulations of Hebbian learning. *Biological Cybernetics*, *87*, 404–415.

Hartmann, C., Lazar, A., Nessler, B., & Triesch, J. (2015). Where's the noise? Key features of spontaneous activity and neural variability arise through learning in a deterministic network. *PLoS Computational Biology*, *11*, e1004640. https://doi.org/10.1371/journal.pcbi.1004640

Hebb, D. O. (1949). *The organization of behavior*. New York, NY: Wiley & Sons. Retrieved from http://s-f-walker.org.uk/pubsebooks/pdfs/The_Organization_of_Behavior-Donald_O._Hebb.pdf

Hengen, K. B., Lambo, M. E., Van Hooser, S. D., Katz, D. B., & Turrigiano, G. G. (2013). Firing rate homeostasis in visual cortex of freely behaving rodents. *Neuron*, *80*, 335–342.

Hensch, T. K. (2004). Critical period regulation. *Annual Review of Neuroscience*, *27*, 549–579. https://doi.org/10.1146/annurev.neuro.27.070203.144327

Hooks, B. M., & Chen, C. (2007). Critical periods in the visual system: Changing views for a model of experience-dependent plasticity. *Neuron*, *56*, 312–326. https://doi.org/10.1016/j.neuron.2007.10.003

Jefferis, G. S. X. E., Potter, C. J., Chan, A. M., Marin, E. C., Rohlfing, T., Maurer, C. R., & Luo, L. (2007). Comprehensive maps of *Drosophila* higher olfactory centers: Spatially segregated fruit and pheromone representation. *Cell*, *128*(6), 1187–1203. https://doi.org/10.1016/j.cell.2007.01.040

Keck, T., Keller, G. B., Jacobsen, R. I., Eysel, U. T., Bonhoeffer, T., & Hübener, M. (2013). Synaptic scaling and homeostatic plasticity in the mouse visual cortex in vivo. *Neuron*, *80*, 327–334.

Lazar, A., Pipa, G., & Triesch, J. (2009). SORN: A self-organizing recurrent neural network. *Frontiers in Computational Neuroscience*, *3*, 23.

LeMasson, G., Marder, E., & Abbott, L. F. (1993). Activity-dependent regulation of conductances in model neurons. *Science*, *259*(5103), 1915–1917. https://doi.org/10.1126/science.8456317

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, *275*(5297), 213–215. https://doi.org/10.1126/science.275.5297.213

Marr, D. (1969). A theory of cerebellar cortex. *Journal of Physiology*, *202*(2), 437–470. https://doi.org/10.1113/jphysiol.1969.sp008820

Martin, S. J., Grimwood, P. D., & Morris, R. G. M. (2000). Synaptic plasticity and memory: An evaluation of the hypothesis. *Annual Review of Neuroscience*, *23*(1), 649–711. https://doi.org/10.1146/annurev.neuro.23.1.649

Miller, K. D. (1996). Synaptic economics: Competition and cooperation in synaptic plasticity. *Neuron*, *17*, 371–374.

Miller, K. D., & MacKay, D. J. C. (1994). The role of constraints in Hebbian learning. *Neural Computation*, *6*(1), 100–126. https://doi.org/10.1162/neco.1994.6.1.100

Miner, D., & Triesch, J. (2016). Plasticity-driven self-organization under topological constraints account for nonrandom features

of cortical synaptic wiring. *PLoS Computational Biology*, *12*(2), e1004759.

Mombaerts, P., Wang, F., Dulac, C., Chao, S. K., Nemes, A., Mendelsohn, M., . . . Axel, R. (1996). Visualizing an olfactory sensory map. *Cell*, *87*(4), 675–686. https://doi.org/10.1016/S0092-8674(00)81387-2

Monk, T., Savin, C., & Lücke, J. (2016). Neurons equipped with intrinsic plasticity learn stimulus intensity statistics. In *Advances in neural information processing systems* (pp. 4278–4286).

Monk, T., Savin, C., & Lücke, J. (2018). Optimal neural inference of stimulus intensities. *Scientific Reports*, *8*, 10038. https://doi.org/10.1038/s41598-018-28184-5

Olshausen, B. A. (2003). Principles of image representation in visual cortex. *The Visual Neurosciences*, *2*, 1603–1615.

Perez-Orive, J., Mazor, O., Turner, G. C., Cassenaer, S., Wilson, R. I., & Laurent, G. (2002). Oscillations and sparsening of odor representations in the mushroom body. *Science*, *297*(5580), 359–365. https://doi.org/10.1126/science.1070502

Poo, C., & Isaacson, J. S. (2009). Odor representations in olfactory cortex: "Sparse" coding, global inhibition, and oscillations. *Neuron*, *62*(6), 850–861. https://doi.org/10.1016/j.neuron.2009.05.022

Savin, C., Joshi, P., & Triesch, J. (2010). Independent component analysis in spiking neurons. *PLoS Computational Biology*, *6(4)*, e1000757.

Stemmler, M., & Koch, C. (1999). How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate. *Nature Neuroscience*, *2*(6), 521–527. https://doi.org/10.1038/9173

Stettler, D. D., & Axel, R. (2009). Representations of odor in the piriform cortex. *Neuron*, *63*(6), 854–864. https://doi.org/10.1016/j.neuron.2009.09.005

Tetzlaff, C., Kolodziejski, C., Markelic, I., & Wörgötter, F. (2012). Time scales of memory, learning, and plasticity. *Biological Cybernetics*, *106*(11), 715–726.

Tetzlaff, C., Kolodziejski, C., Timme, M., & Wörgötter, F. (2011). Synaptic scaling in combination with many generic plasticity mechanisms stabilizes circuit connectivity. *Frontiers in Computational Neuroscience*, *5*. https://doi.org/10.3389/fncom.2011.00047

Triesch, J. (2007). Synergies between intrinsic and synaptic plasticity mechanisms. *Neural Computation*, *19*, 885–909.

Triesch, J., Vo, A. D., & Hafner, A.-S. (2018). Competition for synaptic building blocks shapes synaptic plasticity. *eLife*, *7*. https://doi.org/10.7554/eLife.37836

Tsodyks, M. V., & Feigelman, M. V. (1988). The enhanced storage capacity in neural networks with low activity level. *EPL (Europhysics Letters)*, *6*(2), 101. https://doi.org/10.1209/0295-5075/6/2/002

Turner, G. C., Bazhenov, M., & Laurent, G. (2008). Olfactory representations by *Drosophila* mushroom body neurons. *Journal of Neurophysiology*, *99*(2), 734–746. https://doi.org/10.1152/jn.01283.2007

Turrigiano, G., Abbott, L., & Marder, E. (1994). Activity-dependent changes in the intrinsic properties of cultured neurons. *Science*, *264*(5161), 974–977. https://doi.org/10.1126/science.8178157

Turrigiano, G. G. (2008). The self-tuning neuron: Synaptic scaling of excitatory synapses. *Cell*, *135*, 422–435.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., & Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, *391*(6670), 892–896. https://doi.org/10.1038/36103

Vincis, R., Gschwend, O., Bhaukaurally, K., Beroud, J., & Carleton, A. (2012). Dense representation of natural odorants in the mouse olfactory bulb. *Nature Neuroscience*, *15*(4), 537–539. https://doi.org/10.1038/nn.3057

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, *287*(5456), 1273–1276. https://doi.org/10.1126/science.287.5456.1273

Yger, P., & Gilson, M. (2015). Models of metaplasticity: A review of concepts. *Frontiers in Computational Neuroscience*, *9*, 138.

Zenke, F., Agnes, E. J., & Gerstner, W. (2015). Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications*, *6*, 6922.

Zenke, F., & Gerstner, W. (2017). Hebbian plasticity requires compensatory processes on multiple timescales. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *372*. https://doi.org/10.1098/rstb.2016.0259

Zenke, F., Hennequin, G., & Gerstner, W. (2013). Synaptic plasticity in neural networks needs homeostasis with a fast rate detector. *PLoS Computational Biology*, *9*(11), e1003330.

Zhang, W., & Linden, D. J. (2003). The other side of the engram: Experience-driven changes in neuronal intrinsic excitability. *Nature Reviews Neuroscience*, *4*, 886–900.