The interplay of synaptic plasticity and scaling enables the self-organized allocation of multiple memory representations

Johannes Maria Auth[1,2◕¤], Timo Nachstedt[1,2◕], Christian Tetzlaff[1,2*]

1 Third Institute of Physics, Georg-August-Universität Göttingen, Germany
2 Bernstein Center for Computational Neuroscience, Georg-August-Universität Göttingen, Germany

◕These authors contributed equally to this work.
¤Current Address: Department of Neurobiology, The Weizmann Institute of Science, Rehovot, Israel
* tetzlaff@phys.uni-goettingen.de

Abstract

It is commonly assumed that memories about experienced stimuli are represented in the brain by groups of highly interconnected neurons called Hebbian cell assemblies. This requires allocating and storing information in the neural circuitry, which happens through synaptic weight adaptation. It remains, however, largely unknown how memory allocation and storage can be achieved and coordinated to allow for a faithful representation of multiple memories without disruptive interference between them. In this theoretical study, we show that the interplay between conventional Hebbian synaptic plasticity and homeostatic synaptic scaling organizes synaptic weight adaptations such that, on the one hand, a new stimulus forms a new memory while, on the other hand, different stimuli are assigned to distinct Hebbian cell assemblies. We show that the resulting dynamics can reproduce experimental in-vivo data focusing on how other neuronal and synaptic factors, such as neuronal excitability and network connectivity, influence memory formation. Thus, the here presented model suggests that a few fundamental synaptic mechanisms may suffice to implement memory allocation and storage in neural circuitry.

Author summary

The survival in a changing environment requires the reliable learning, storage, and organization of relevant stimuli in the neuronal circuit. This theoretical work addresses the important issue of how a neuronal circuit coordinates the learning-related synaptic adaptations to properly assign new information to groups of neurons and to form long-lasting memory representations of multiple stimuli. We show that this requires only three synaptic properties – homosynaptic potentiation paired with heterosynaptic depression both driven by the postsynaptic activity level. These properties naturally arise from the generic interplay between conventional Hebbian synaptic plasticity and homeostatic synaptic scaling. Therefore, this study shows that the complex processes of memory allocation and storage can be attributed to ubiquitous synaptic mechanisms in the neural circuitry.

Introduction

Learning and memorizing information about the surrounding environment is a vital function of neural circuits of living beings. For this, a neural circuit has to accurately form and organize internal representations of the different pieces of information received.

The Hebbian hypothesis [1–5] states that, when a neural circuit receives a new piece of information, the corresponding stimulus activates a group of neurons and, via activity-dependent synaptic plasticity [6–9], adapts the weights of related synapses. This process leads to the formation of a strongly interconnected group of neurons or Hebbian cell assembly (CA), which serves as an internal memory representation of the stimulus [3–5]. The recall of the memory translates into the activation of the respective CA. In order to recognize similar pieces of information, similar stimuli also have to be able to activate the corresponding CA. This is enabled, on the one hand, by strong recurrent interconnections between CA-neurons resulting in pattern completion [4, 10] and, on the other hand, by an accurate adjustment of the synapses connecting stimuli with CAs (memory allocation). Up to now, experimental and theoretical studies mainly focussed on revealing either the formation of CAs [5, 11–19] or the allocation of stimuli to neurons [20–28]. By contrast, it is still unknown how a neural circuit coordinates in a self-organized manner the reliable allocation of stimuli to neurons with the proper formation of a memory representation.

In this theoretical study, we show in a network model that conventional Hebbian synaptic plasticity [3, 29, 30] together with the slower, homeostatic processes of synaptic scaling [31–34] leads to the self-organized coordination of weight changes at feed-forward and recurrent synapses enabling the accurate formation and allocation of CAs. The model reproduces *in-vivo* experimental data and provides testable predictions. Furthermore, the analysis of a population model, capturing the main features of the network dynamics, enables us to determine three generic properties of synaptic adaptations required to form and allocate memory representations in a reliable manner. These properties are (i) homosynaptic potentiation, (ii) heterosynaptic depression, and (iii) the down-regulation of synaptic weight changes by the postsynaptic activity level.

Materials and methods

Numerical simulations

The three neuronal populations in the numerical model consist of 936 excitatory neurons (36 in input area, 900 in memory area) and a single inhibitory unit describing a population of inhibitory neurons which are connected to the excitatory neurons in an all-to-all manner. All neurons are described by a rate-coded leaky integrator model. The memory area is arranged as a quadratic neural grid of 30x30 units. Each neuron within the grid receives excitatory connections from four randomly chosen input neurons. In addition, it is recurrently connected to its circular neighborhood of radius four (measured in neuronal units; for visualization see SI text) and to the global inhibitory unit.

Neuron model

The three neuronal populations considered, i.e. the input area, the memory area, as well as a global inhibitory unit are described by differential equations of a rate-coded leaky integrate neuron model. The membrane potential $u_{\mathrm{i}}^{\mathrm{rec}}$, $i \in \{1, ..., N^{\mathrm{M}}\}$ of each neuron i

in the memory area is described as follows:

$$\frac{\mathrm{d}u_\mathrm{i}}{\mathrm{d}t} = -\frac{u_\mathrm{i}}{\tau} + R\left(\sum_{\mathrm{j}}^{N^\mathrm{M}} w_{\mathrm{i,j}}^{\mathrm{rec}} F_\mathrm{j} + w_{\mathrm{i,inh}} F_\mathrm{inh} + \sum_{\mathrm{k}}^{N^\mathrm{I}} w_{\mathrm{i,k}}^{\mathrm{ff}} I_\mathrm{k}\right), \tag{1}$$

with the membrane time constant τ, membrane resistance R, number of neurons in the memory area N^M, number of neurons in the input area N^I, and input rate I_k. The membrane potential is converted into a firing rate F_i by a sigmoidal transfer function with maximal firing rate α, steepness β and inflexion point ϵ:

$$F_\mathrm{i}(u_\mathrm{i}) = \frac{\alpha}{1 + \exp(\beta(\epsilon - u_\mathrm{i}))}. \tag{2}$$

The global inhibitory unit is also modeled as a rate-coded leaky integrator receiving inputs from all neurons of the memory area. Its membrane potential u_inh follows the differential equation 3 with inhibitory membrane time scale τ_inh and resistance R_inh. The potential is converted into a firing rate F_inh by a sigmoidal transfer function (equation 4).

$$\frac{\mathrm{d}u_\mathrm{inh}}{\mathrm{d}t} = -\frac{u_\mathrm{inh}}{\tau_\mathrm{inh}} + R_\mathrm{inh} \cdot \sum_\mathrm{i}^{N^\mathrm{M}} w_\mathrm{inh,i} F_\mathrm{i}, \tag{3}$$

$$F_\mathrm{inh}(u_\mathrm{inh}) = \frac{\alpha}{1+\exp(\beta(\epsilon - u_\mathrm{inh}))}. \tag{4}$$

As the neurons in the input area form the stimuli, their output activation is set manually. Thus no further description is needed.

Synaptic plasticity

The synaptic weight changes in the excitatory feed-forward (equation 5) and recurrent connections (equation 6) are determined by the combined learning rule of conventional Hebbian synaptic plasticity (first term) and synaptic scaling (second term) with time constants μ, κ^ff, κ^rec, and target firing rate F^T. The differential equation for the synaptic weight of a feed-forward connection $w_{\mathrm{i,k}}^{\mathrm{ff}}$ from input neuron $k \in \{0, \cdots, N^\mathrm{I}\}$ to memory neuron $i \in \{0, \cdots, N^\mathrm{M}\}$ is:

$$\frac{\mathrm{d}w_{\mathrm{i,k}}^{\mathrm{ff}}}{\mathrm{d}t} = \mu\left(F_\mathrm{i} I_\mathrm{k} + (\kappa^\mathrm{ff})^{-1}(F^\mathrm{T} - F_\mathrm{i})(w_{\mathrm{i,k}}^{\mathrm{ff}})^2\right) \cdot c_{\mathrm{i,k}}^{\mathrm{ff}}. \tag{5}$$

The dynamics of the synaptic weight of a recurrent connection $w_{\mathrm{i,j}}^{\mathrm{rec}}$ from memory neuron $j \in \{0, \cdots, N^\mathrm{M}\}$ to memory neuron $i \in \{0, \cdots, N^\mathrm{M}\}$ is determined by:

$$\frac{\mathrm{d}w_{\mathrm{i,j}}^{\mathrm{rec}}}{\mathrm{d}t} = \mu\left(F_\mathrm{i} F_\mathrm{j} + (\kappa^\mathrm{rec})^{-1}(F^\mathrm{T} - F_\mathrm{i})(w_{\mathrm{i,j}}^{\mathrm{rec}})^2\right) \cdot c_{\mathrm{i,j}}^{\mathrm{rec}}. \tag{6}$$

$c_{\mathrm{i,k}}^{\mathrm{ff}}$ and $c_{\mathrm{i,k}}^{\mathrm{ff}}$ are the entries in the feed-forward or recurrent connectivity matrices, respectively. Being either of value 1 (existing connection) or 0 (non-existing connection), they maintain the system's connectivity structure, i.e. prevent the generation of connections.

All other connections are considered to be non-plastic.

Coding framework

The simulation codes were written in Python 3.5. The model system was updated by using the Euler method at time steps of 5 ms.

Table 1. Model Parameters: Variable, Descriptions and used Values

Variable	Description	Value
τ	membrane time constant (memory area)	0.01
R	membrane resistance (memory area)	1/11
N^{M}	number of neurons in memory area	900
N^{I}	number of neurons in input area	36
I_{k}	input rate	{0,130}
α	maximum firing rate	100
β	sigmoid steepness	0.05
ϵ	sigmoid inflexion point	130
μ	plasticity time constant	1/15
F^{T}	target firing rate	0.1
κ^{rec}	scaling time constant (recurrent)	60
κ^{ff}	scaling time constant (feed-forward)	720
τ_{inh}	membrane time constant (inhibitory unit)	0.02
R_{inh}	membrane resistance (inhibitory unit)	1
$w_{\mathrm{inh,i}}$	synaptic weight to inhibitory unit	0.6
$w_{\mathrm{i,inh}}$	synaptic weight from inhibitory unit	1200

Simulation of self-org. formation of two CAs

The system undergoes two learning phases being presented two completely dissimilar stimulus patterns A (first phase) and B (second phase). For stimulus pattern A half of the input neurons are set to be active at 130 Hz whereas the other half remains inactive at 0 Hz. Vice versa for stimulus pattern B. During the respective learning phase the respective pattern is presented 10 times for 5 sec with a 1 sec pause in between. Both learning phases are embraced by test phases in which plasticity is shut off and both patterns are presented for 0.5 sec each to apply measures on the existing memory structures.

Comparison to experimental data

The manipulation of the neuronal excitability has been done by adapting the value for ϵ in the transfer function of the neuron model, i.e. shifting its inflexion point to lower (increased excitability) or higher (decreased excitability) values. Similar to the methods used in experiments [36], we manipulated a subpopulation of neurons within a randomly chosen circular area in the memory area (about 10% of the network). The relative recruitment factor is the relation of recruitment probabilities for manipulated and control neurons averaged over 100 repetitions.

Population model

We consider two non-overlapping populations of N excitatory neurons and one inhibitory unit.

The state of every population $i \in \{1, 2\}$ is determined by its mean membrane potential $\bar{u}_i$, its mean recurrent synaptic weight $\bar{w}_i^{\mathrm{rec}}$ between neurons of the population, and the mean weight $\bar{w}_i^{\mathrm{ff}}$ of feed-forward synapses projecting signals from the currently active input onto the population. We assume that the two populations interact solely through the inhibitory unit whose state is given by its membrane potential u_{inh}. Thus, the dynamics of the model is described by a set of seven differential equations (see following section). To obtain its equilibria, we analytically derive the nullclines $\bar{u}_1^*(\bar{u}_2^*)$

and $\bar{u}_2^*(\bar{u}_1^*)$ and numerically determine their intersections. The stability of an equilibrium is obtained from the sign of the eigenvalues of the system's Jacobi Matrix.

For analyzing in which regimes population 1 and population 2 are assigned to an input stimulus as a function of the initial synaptic weights (Fig 2 E and Fig 3), we initialize the system with the given combination of feed-forward and recurrent average synaptic weights and $\bar{u}_1 = \bar{u}_2 = \bar{u}_{\text{inh}} = 0$, simulate it for 100 sec, and assess which of the two populations is active. For further details and parameter values see SI Text.

Model Definition

The two excitatory populations in the population model are described by their mean membrane potentials $\bar{u}_i$, $i \in \{1, 2\}$, $k \in \{\text{A}, \text{B}\}$:

$$\frac{\mathrm{d}\bar{u}_i}{\mathrm{d}t} = -\frac{\bar{u}_i}{\tau} + R\left(\bar{n}^{\text{rec}}\bar{w}_i^{\text{rec}}\bar{F}_i + w_{i,\text{inh}}F_{\text{inh}} + \sum_k \bar{n}^{\text{ff}}\bar{w}_{ik}^{\text{ff}}\bar{I}_k\right). \tag{7}$$

Here, the time scale τ, the resistance R and the synaptic weight $w_{i,\text{inh}}$ have the same value as in the network simulations. The average number $\bar{n}_i^{\text{rec}}$ of incoming recurrent connections from neurons within the population as well as the number of feed-forward synapses transmitting signals from active inputs to every neuron ($\bar{n}^{\text{ff}}$) are taken from simulations ($\bar{n}^{\text{rec}} = 35$, Figure S5 C; $n^{\text{ff}} = 2.3$, Figure S5 B).

The membrane potential of the inhibitory population is given by

$$\frac{\mathrm{d}u_{\text{inh}}}{\mathrm{d}t} = -\frac{u_{\text{inh}}}{\tau_{\text{inh}}} + R_{\text{inh}}(w_{\text{inh},1}N\bar{F}_1 + w_{\text{inh},2}N\bar{F}_2) \tag{8}$$

with τ_{inh}, R_{inh} and $w_{\text{inh},1} = w_{\text{inh},2}$ corresponding to the respective values in the network simulations. The number N of neurons per population is adjusted to the CAs in the network simulation and chosen as $N = 120$ (Figure S5 A).

The transfer function of the neurons within the population is the same as for individual neurons:

$$\bar{F}_i(\bar{u}_i) = \frac{\alpha}{1 + \exp(\beta(\epsilon - \bar{u}_i))},\ i \in \{1, 2, \text{inh}\}. \tag{9}$$

The synaptic weight changes of recurrent and feed-forward synapses follow the interplay of conventional Hebbian synaptic plasticity and synaptic scaling ($i \in \{1, 2\}$):

$$\frac{\mathrm{d}\bar{w}_i^{\text{rec}}}{\mathrm{d}t} = \mu\left(\bar{F}_i^2 + (\kappa^{\text{rec}})^{-1}(F^{\text{T}} - \bar{F}_i)(\bar{w}_i^{\text{rec}})^2\right), \tag{10}$$

$$\frac{\mathrm{d}\bar{w}_{ik}^{\text{ff}}}{dt} = \mu\left(\bar{F}_i\bar{I}_k + (\kappa^{\text{ff}})^{-1}(F^{\text{T}} - \bar{F}_i)(\bar{w}_{ik}^{\text{ff}})^2\right). \tag{11}$$

Results

Formation and allocation of one memory representation

We consider a recurrently connected excitatory population of neurons (memory area with an all-to-all connected inhibitory population to regulate the global activity level; Fig 1 A) with each neuron receiving inputs from several randomly assigned neurons of a second excitatory population (input area). All synapses between the excitatory neurons (red; feed-forward and recurrent) are adapted by conventional Hebbian synaptic plasticity together with homeostatic synaptic scaling (see Methods; [16, 33, 35]). A stimulus is represented by the activation of a stimulus-specific subset of neurons in the input area forming a stimulus pattern (Fig 1 B).

Throughout this study, we consider initially a learning phase during which stimulus pattern A (magenta; Fig 1 B, C) is repeatedly presented to a blank neural circuit (no encoded CA) to investigate the basic CA formation and allocation dynamics. After the successful formation of a CA (blue; Fig 1 B), in a second learning phase, stimulus pattern B (green) is repeatedly presented to analyze the influence of the encoded CA on the formation and allocation of a further CA (orange). In addition, we consider three test phases (Fig 1 C), during which synaptic dynamics are frozen, to enable the investigation of the resulting response dynamics of the circuit according to the different stimulus patterns.

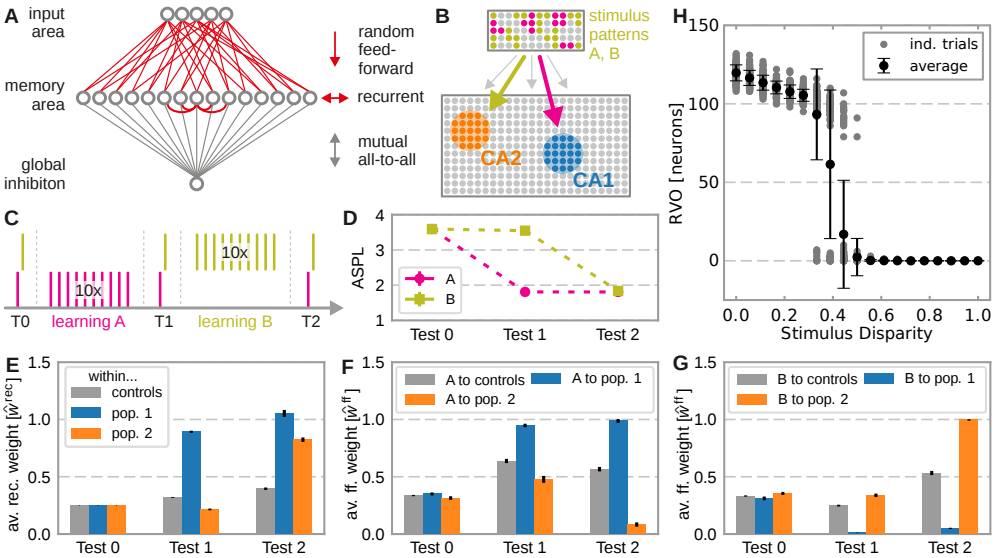

Fig 1. The interaction of conventional Hebbian synaptic plasticity and synaptic scaling enables the stimulus-dependent formation and allocation of memory representations in a neuronal network model. A: The network model consists of an excitatory population of neurons (memory area) receiving inputs from a second excitatory (input area) and one inhibitory population (global inhibition). All synapses in between excitatory neurons (red) are plastic resulting to (B:) the stimulus-specific (A: magenta; B: green) allocation and formation of cell assemblies (CA1: blue; CA2: orange) shown here schematically. C: Stimulation protocol used throughout this study. During the test phases (T0, T1, T2), synaptic dynamics are stopped to analyze the changes induced by the learning phases. D: The two learning phases yield a reduction of the average shortest path length (ASPL) between stimulus-activated neurons in the memory area. E: In addition, the average strength of the intra-population synapses is increased such that pop. 1 and pop. 2 become independent CAs. F, G: Furthermore, the average synaptic strength of the feed-forward connections from the stimulus-neurons to the formed CAs are adapted specifically such that pop. 1 represents stimulus A (F) and pop. 2 represents stimulus B (G). H: The response overlap (RVO) between both CAs depends non-linearly on the disparity between the stimulus patterns. Data presented are mean values with standard deviation over 100 repetitions (H) or rather 1000 repetitions (D-G).

The first presentation of a new stimulus to the network triggers the activation of a widely distributed pattern of neurons (see SI Text) being not directly connected with each other (high average shortest path length (ASPL); Fig 1 D, test 0). By contrast, if the stimulus is repeatedly presented in a given time interval (here stimulus A), the dynamics of the network reshapes the pattern of activated neurons such that the final

pattern consists of a group of neighboring, interconnected neurons (decrease in ASPL; magenta, test 1). As shown in our previous studies [16, 35], the combination of synaptic plasticity and scaling together with a repeated activation of an interconnected group of neurons yields an average strengthening of the interconnecting synapses without significantly altering other synapses (blue; Fig 1 E, test 1). The decreased ASPL and the locally strengthened synaptic weights point to the stimulus-dependent formation of a CA during the first learning phase.

However, do the self-organizing dynamics also link specifically the stimulus-neurons with the CA-neurons? The repeated presentation of stimulus pattern A yields an on average strengthening of synapses projecting from the stimulus-neurons to the whole memory area (Fig 1 F, test 1). Essentially, the synapses from the stimulus-neurons to the CA-neurons (blue) have a significantly stronger increase in synaptic weights than the controls (gray). This implies a proper assignment of the stimulus to the newly formed CA during the learning phase resulting in a higher chance of activating the CA-neurons when the stimulus is presented later again.

These results reveal that the interaction of synaptic plasticity and scaling self-organizes for a wide parameter regime (see SI Text) synaptic changes at recurrent and feed-forward connections to form and allocate a memory representation in a blank neuronal network.

Formation and allocation of a second memory representation

Clearly the presence of a memory representation can alter the self-organizing dynamics shown before, which could impede the proper formation of representations of further stimuli. For instance, the existence of a CA in the neuronal network could bias the adaptations of the feed-forward synapses such that a new stimulus is also assigned to this CA. This would imply that the neural circuit is unable to discriminate between the originally CA-associated stimulus and the new stimulus. Thus, to investigate the influence of prior learning, we repeatedly present a second stimulus B (second learning phase) after the proper formation of the CA associated to stimulus A.

Similar to the first learning phase, the repeated presentation of stimulus pattern B (here, patterns A and B have a stimulus disparity of 1 indicating no overlap between stimulus patterns; see SI Text) yields the activation of a group of strongly interconnected neurons (decreased ASPL; green, Fig 1 D, test 2) and a strengthening of the corresponding recurrent synaptic weights (orange; Fig 1 E, test 2). Thus, the stimulus-dependent formation of a new CA is not impeded by the existence of another CA. Furthermore, both CAs are distinguishable as they do not share any neuron in the memory area (Fig 1 H, disparity equals 1). This depends on the disparity between stimuli; for quite dissimilar stimuli (disparity $\gtrsim 0.5$) both CAs are separated, for more similar stimuli ($0.3 \lesssim$ disparity $\lesssim 0.5$) the system undergoes a state transition, and for quite similar stimuli (disparity $\lesssim 0.3$) both stimuli activate basically the same group of neurons. Note that the latter demonstrates that the network does correctly assign noisy versions of a learned stimulus pattern (pattern completion [10]) instead of forming a new CA, while the first case illustrates that the network performs pattern separation [10]. This indicates a correct assignment of the stimuli to the corresponding CAs, such that also in the presence of another CA the weight changes of synapses between stimulus pattern and newly formed CA are adapted accordingly (orange; Fig 1 G, test 2). Thus, the self-organizing dynamics yields the formation and allocation of a new CA during the second learning phase. Note that the synaptic weights of the initially encoded CA are not altered significantly during this phase (compare blue bars at test 1 with at test 2 in Fig 1 E-G). But, although stimulus A is not present, the second learning phase leads to a weakening of synapses projecting from stimulus A neurons to the newly formed CA considerably below control (orange; Fig 1 F, test 2). Similarly, during the first learning

phase the synaptic weights between stimulus B neurons and the first cell assembly are also weakened (blue; Fig 1 G, test 1). Apparently, this weakening of synapses from the other, non-assigned stimulus to a CA reduces the chance of spurious activations. Thus, the self-organized dynamics resulting from the interaction of synaptic plasticity and scaling enables the proper formation and allocation of several memory representations without significant interferences between them.

Generic properties of synaptic adaptations enabling the formation and allocation of memory representations

In order to obtain a more detailed understanding of the self-organizing processes of synaptic plasticity and scaling underlying the reliable formation and allocation of memory representations, we have to reduce the complexity of the model to enable analytical investigations. For this, we assume that the different involved neuronal populations in the input (stimulus A and B neurons) and memory area (populations 1 and 2 becoming CAs in the memory area) are by themselves homogeneous allowing the consideration of average population dynamics (Fig 2 A). This maps the main features of the complex network dynamics to a 7-dimensional model (see Materials and Methods and SI Text).

First, we investigate the formation of a memory representation in a blank neural circuit. The solutions of the nullclines of the dynamics in the space of the average recurrent synaptic weights of the two neuronal groups ($\bar{w}_1^{\mathrm{rec}}$ and $\bar{w}_2^{\mathrm{rec}}$; Fig 2 B, left) show that the dynamics during the first learning phase is dominated by three equilibrium states: one is repulsive (red, 7) and two are attractive (green, 2 and 3; the equilibria 4, 8, 9, 10 do not influence the here discussed dynamics). The two attractive states represent that one of the two neuronal populations becomes a strongly interconnected CA (e.g., pop. 1 in state 2), while the other remains in a weakly interconnected state. The repulsive state lies on the identity line ($\bar{w}_1^{\mathrm{rec}} = \bar{w}_2^{\mathrm{rec}}$) having the same distance to both attractive states. The resulting mirror symmetry in the phase space implies that the dynamics on the one side of the identity line (reaching the attractive state lying on this side) equals the dynamics on the other side. Therefore, given that before learning both populations are similarly weak interconnected, small variations in the initial condition of the network are sufficient to determine which group of neurons will become a CA during learning (see black trace for an example). Note that this symmetry-dependent formation of a CA is quite robust against variations in parameters, as the strength of the stimulus ($\bar{I}_A \gtrsim 120$), which agrees with results from the more detailed network model discussed before (purple dots; Fig 2 C).

In parallel to the development of the recurrent synaptic weights, the synaptic strengths of the feed-forward connections change to assure proper memory allocation. Thus, we analyze the activity-dependency of the interaction of synaptic plasticity and scaling and obtain the change of the feed-forward synaptic weights expected during the learning phase given different activity conditions of the input and memory populations (Fig 2 D). In general, the combination of both activity levels determines whether the weight of a feed-forward synapse is potentiated (red), depressed (blue), or not adapted (white). If both activities are on a quite high level, synapses are potentiated (case II; homosynaptic potentiation). If the presynaptic activity (input) is on a low level and the postsynaptic activity (population) is on a high level, feed-forward synapses are depressed (case I; heterosynaptic depression). However, if the postsynaptic activity is low, synaptic changes are negligible regardless of the level of presynaptic activity (cases III and IV).

The different parts of recurrent and feed-forward synaptic dynamics described before together lead to the formation and allocation of a CA as described in the following. During the first learning phase (stimulus A), a small variation in initial conditions

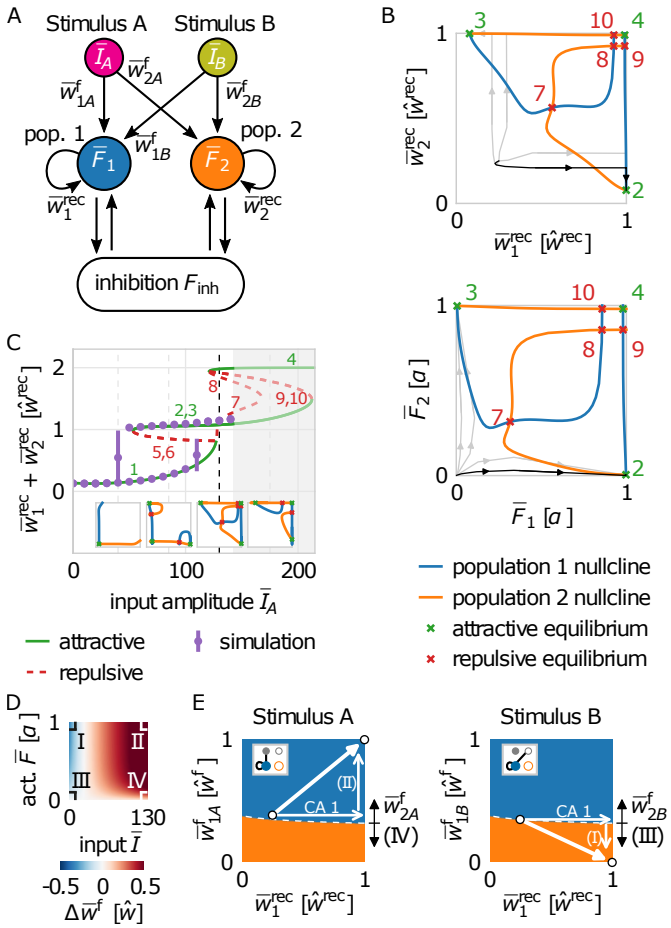

Fig 2. Population model of network dynamics enables analytical derivation of the underlying synaptic dynamics. A: Schema of the population model of the averaged network dynamics (bars indicate the average over all neurons in population $s \in \{A, B, 1, 2\}$). $\bar{I}$: stimulus amplitude; $\bar{F}$: population activity; $\bar{w}^{\mathrm{ff}}_{sp}$: strength of feed-forward synapses between populations s and p; $\bar{w}^{\mathrm{rec}}_{p}$: strength of recurrent synapses within population p. B: The intersections of the population nullclines projected into weight (left) and activity (right) space reveal equilibria (attractive: green; repulsive: red) indicating the formation of a CA (green markers 2 and 3), if the system deviates from the identity line. Numbers correspond to numbers of equilibria in C. α is the maximal neuronal activity. C: The bifurcation diagram of the network indicates that CAs are formed for a wide variety of input amplitudes ($\bar{I}_A \gtrsim 120$). The dashed line illustrates the value used in B (see insets). Solutions of the full network (purple dots) and population model match (deviation for large values of $\bar{I}_A$, gray area; details see SI Text). D: The dynamics of feed-forward synaptic strengths depends on the input amplitude and the population activity subdivided into four different cases (I-IV). E: These cases (I-IV) together with the strengthening of recurrent synapses (CA1) yield the self-organized formation and allocation of CAs. Namely, during the first learning phase, synaptic changes drive the system (white dot) into regimes where either population 1 (blue) or population 2 (orange) will represent the corresponding stimulus (left: stimulus A; right: stimulus B).

breaks the symmetry (white dashed line; Fig 2 E) such that the system is, for instance, in the basin of attraction that population 1 will represent stimulus A (blue area). This

leads to the strengthening of recurrent synapses within population 1 forming a CA (increase of $\bar{w}_1^{\mathrm{rec}}$; Fig 2 B, E). In parallel, the synaptic strengthening induces an increase of the activity level of the population (state 2 in Fig 2 B, right) yielding, together with the active stimulus A, an average increase of the corresponding feed-forward synapses ($\bar{w}_{1A}^{\mathrm{ff}}$; case II). These synaptic changes push the system further away from the symmetry condition (white arrows; Fig 2 E, left) implying a more stable memory representation. Note that changing the strength of synapses connecting stimulus A with population 2 ($\bar{w}_{2A}^{\mathrm{ff}}$) could result to a shift of the symmetry condition (indicated by black arrows) counteracting the stabilization process. However, this effect is circumvented by the system as the second population has a low activity level and therefore corresponding feed-forward synapses are not adapted (case IV).

The formation of a CA acts as a variation or perturbation of the initial condition breaking the symmetry for the second learning phase (stimulus B; Fig 2 E, right). The formation of the first CA pushes the system into the blue area (CA1-arrow). This indicates that, during learning stimulus B, population 1 would also represent stimulus B impeding the discrimination ability of the network between stimulus A and B. However, during the first learning phase, as stimulus B is inactive, synapses projecting from stimulus B neurons to population 1 neurons are depressed (case I) and the system switches into the orange area. This area indicates that, during learning stimulus B, population 2 would represent the stimulus. Please note that this switch can be impeded by adapting the connections from stimulus B to population 2 ($\bar{w}_{2B}^{\mathrm{ff}}$) shifting the symmetry condition. But, similar to before, this effect is circumvented by the system as population 2 is basically inactive (case III). Thus, after the first learning phase, the synaptic dynamics regulated by the combination of Hebbian plasticity and synaptic scaling drives the system into an intermediate state, which implies that the system will definitely reach the desired state during learning the second stimulus (Fig 3). This intermediate state can only be reached if synaptic adaptations comprise three properties implied by the four cases I-IV: (i) homosynaptic potentiation (case I), (ii) heterosynaptic depression (case II), and (iii) the down-regulation of synaptic weight changes by the postsynaptic activity level (cases III and IV).

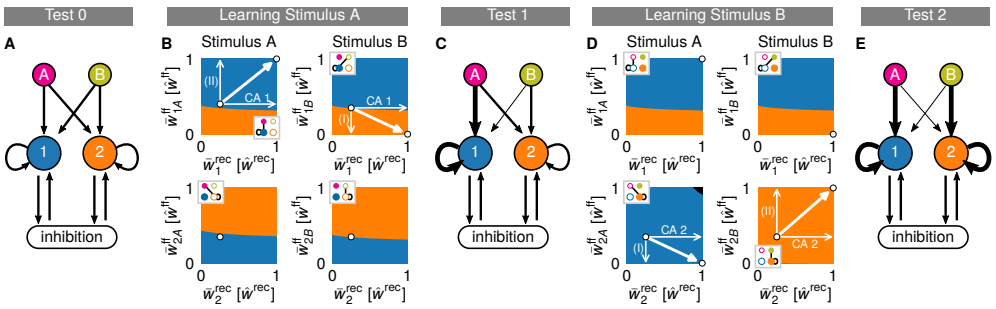

Fig 3. Summary of the synaptic changes underlying the formation and allocation of memory representations. The interaction of synaptic plasticity and scaling brings a blank network (A; test 0) during the first learning phase (B) to an intermediate state (C; test 1). From this intermediate state, the second learning phase (D) yields the system into the end state (E; test 2), in which each stimulus is allocated to one CA (A to pop. 1 and B to pop. 2). A, C, E: Thickness of lines indicates average synaptic strength. B, D: Similar to panels in Fig 2 E. Black indicates regimes in which both populations would represent the corresponding stimulus.

Experimental Data

In addition to the described three properties, the here-proposed model implies the existence of a symmetry condition underlying the formation and allocation of memory representations. Small variations in the initial condition of the system suffice to break this symmetry. These variations could be, aside from noise, enforced experimentally by adapting neuronal parameters in a local group of neurons. Amongst others, several experiments [36, 37] indicate that the probability of a group of neurons to become part of a newly formed memory representation can be influenced by changing their excitability pharmacologically (e.g., by varying the CREB concentration). We reproduced such manipulations in the model by adapting the neuronal excitability of a group of neurons accordingly and analyzing the data similar to experiments (see Methods). Thus, we investigated the probability of a single neuron to become part of a CA averaged over the whole manipulated group of neurons (relative recruitment factor) and compared the results to experimental findings (Fig 4 A).

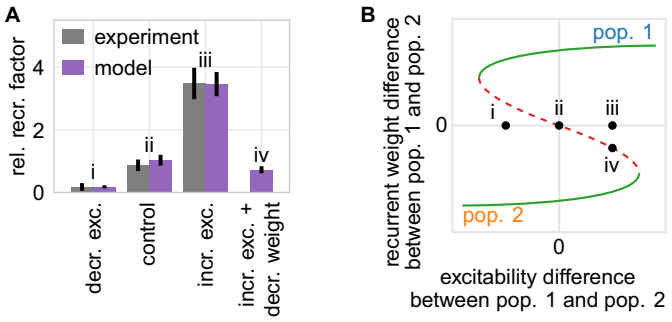

Fig 4. The model of synaptic plasticity and scaling matches experimental *in-vivo* data and provides experimentally verifiable predictions. A: Artificially modifying the excitability of a subset of neurons alters the probability of these neurons to become part of a memory representation (normalized to control). Experimental data are taken from [36]. Data presented are mean values with standard error of the mean. B: The alteration of the excitability of one group of neurons (here pop. 1) compared to others yields a shift of the network's equilibria (here shown schematically; details see SI Text) compared to its initial state (dots) inducing a bias towards one population. A, B: The model analysis yields the prediction (instance iv) that this bias can be counterbalanced by additionally decreasing the average synaptic weight of the manipulated population before learning (here by a factor of 0.1), which shifts the initial state of the network back to the symmetry condition (red).

On the one hand, if the excitability of a group of neurons is artificially increased briefly before learning, the probability of these neurons to become part of the memory representation is significantly enhanced. On the other hand, if the excitability is decreased, the neurons are less likely to become part of the representation. Considering the theoretical results shown before (Fig 2 B), this phenomenon can be explained as follows: the manipulation of the excitability in one population of neurons changes the distance between the repulsive equilibrium state (red; Fig 4 B) to the two attractive states (green). Thus, an increased (decreased) excitability yields a larger (smaller) distance between the repulsive state and the attractive state related to the manipulated population. This larger (smaller) distance implies a changed basin of attraction of the manipulated population enhancing the chance that the initial condition of the network (black dots) lies within this basin. This implies an increase (decrease) of the probability to become a CA as depicted by the variation of the experimentally measured single neuron probability. In addition, the theoretical analysis yields the prediction that the

measured effects will be altered by manipulating other parameters. For instance, if the synaptic weight of the population with increased excitability is on average decreased before stimulus presentation (e.g., by PORCN; [38]), the network's initial condition is shifted such that the CREB-induced influence on the relative recruitment factor is counterbalanced (Fig 4 A, instance iv). Thus, the match between model and *in-vivo* experimental data supports the here-proposed ansatz that the combination of Hebbian synaptic plasticity and synaptic scaling yields the self-organized allocation and formation of memory representations.

Discussion

Our here-presented theoretical study indicates that the formation as well as the allocation of memory representations in neuronal networks depend on the self-organized coordination of synaptic changes at feed-forward and recurrent synapses. We predict that the combined dynamics of conventional Hebbian synaptic plasticity and synaptic scaling could be sufficient for yielding this self-organized coordination as it implies three generic properties: (i) homosynaptic potentiation, (ii) heterosynaptic depression, and (iii) the down-regulation of synaptic weight changes by the postsynaptic activity level.

Previous theoretical studies show that the properties (i) and (ii) are required in recurrent neuronal networks to dynamically form memory representations [16–18]. However, these studies do not consider the feed-forward synaptic dynamics. On the other hand, studies analyzing feed-forward dynamics, such as the self-organization of cortical maps [21, 23, 25], also indicate the importance of homosynaptic potentiation and heterosynaptic depression. However, these studies do not consider the recurrent synaptic dynamics. Only by considering both feed-forward *and* recurrent synaptic dynamics, we revealed the requirement of property (iii) that a low level of postsynaptic activity curtails the synaptic changes which is also supported by experimental evidence [39, 40].

Note that, here, property (iii) is realized by both mechanisms: Hebbian synaptic plasticity as well as synaptic scaling. By contrast, property (i) is implemented by Hebbian synaptic plasticity only and property (ii) is realized by synaptic scaling only. This indicates that synaptic scaling could have an essential role in the allocation and formation of multiple memory representations beyond the widely assumed stabilization of neural network dynamics [29, 32, 33, 41, 42].

Similar to previous studies [16, 35, 43], we consider here an abstract model to describe the neuronal and synaptic dynamics of the network. Despite the abstract level of description, the model matches experimental *in-vivo* data of memory allocation. Other theoretical models match similar experimental data [28, 44]; however, these models are of greater biological detail including more dynamic processes (e.g., short-term plasticity). However, only by considering an abstract model, we have been able to derive analytical expressions such that we could find the requirement of the three generic properties yielding the proper formation and allocation of memories. Remarkably, the synaptic plasticity processes considered in the detailed models [28, 44, 45] also imply the three generic properties (i-iii) supporting our findings.

The here shown results indicate that the combined dynamics of Hebbian synaptic plasticity and synaptic scaling are sufficient to determine the basic dynamics underlying the allocation and formation of memory representations. Interestingly, for several different stimuli the combined dynamics always yields the formation of separated memory representations. In particular the process of heterosynaptic depression impedes the formation of overlaps between several memory representations. However, amongst others, experimental results indicate that memory representations can overlap [5, 46, 47] and, in addition, theoretical studies show that overlaps increase the storage capacity of a neuronal network [14] and can support memory recall [48]. To partially counterbalance

the effect of heterosynaptic depression to enable the formation of overlaps, further time-dependent processes are required. For instance, the CREB-induced enhancement of neuronal excitability biases the neuronal and synaptic dynamics such that the respective subgroup of neurons is more likely to be involved in the formation of a memory representation (Fig 4; [36, 44]). Furthermore, the dynamics of CREB seem to be time-dependent [28, 36, 37]. Therefore, the enhancement of CREB can counterbalance heterosynaptic depression for a given period of time and, by this, could enable the formation of overlaps. We expect that the impact of such time-dependent processes on the dynamics of memories can be integrated into the here-proposed model to analyze the detailed formation of overlaps between memory representations.

Conclusion

In summary, the here-shown theoretical results indicate that the complex dynamics of allocation and formation of multiple memory representations could result from the interplay between Hebbian synaptic plasticity and synaptic scaling. This interplay reliably coordinates in a self-organized manner synaptic changes at feed-forward as well as recurrent synapses as it entails three generic properties. Furthermore, the here-derived theoretical model enables the integration of further neuronal and synaptic processes to assess a general understanding of the dynamics underlying the self-organized allocation and formation of multiple memory representations. For more information, see SI text.

Acknowledgments

We thank Florentin Wörgötter for fruitful comments. The research was funded by the H2020-FETPROACT project Plan4Act (#732266) [JMA, CT], by the Federal Ministry of Education and Research Germany (#01GQ1005A, #01GQ1005B) [TN, CT], and by the International Max Planck Research School for Physics of Biological and Complex Systems by stipends of the country of Lower Saxony with funds from the initiative "Niedersächsisches Vorab" and of the University of Göttingen [TN].

References

1. W. James. *The principles of psychology.* New York: Henry Holt and Company, 1890.

2. J. Konorski. *Conditioned reflexes and neuron organization.* Cambridge: Cambridge University Press, 1948.

3. D. O. Hebb. *The Organization of Behaviour.* Wiley, New York, 1949.

4. G. Palm, A. Knoblauch, F. Hauser, and A. Schüz. Cell assemblies in the cerebral cortex. *Biol. Cybern.*, 108:559–572, 2014.

5. A. Holtmaat and P. Caroni. Functional and structural underpinnings of neuronal assembly formation in learning. *Nat. Neurosci.*, 19:1553–1562, December 2016.

6. T. V. P. Bliss and T. Lomo. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.*, 232:331–356, 1973.

7. W. B. Levy and O. Steward. Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience*, 8(4):791–797, 1983.

8. S. J. Martin, P. D. Grimwood, and R. G. M. Morris. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.*, 23:649–711, 2000.

9. R. C. Malenka and M. F. Bear. LTP and LTD: an embarrassment of riches. *Neuron*, 44:5–21, 2004.

10. M. R. Hunsaker and R. P. Kesner. The operation of pattern separation and pattern completion processes associated with different attributes or domains of memory. *Neurosci. Biobehav. Rev.*, 37:36–58, January 2013.

11. J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, apr 1982.

12. J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA*, 81:3088–3092, 1984.

13. D. J. Amit, H. Gutfreund, and H. Sompolinsky. Storing infinite numbers of patterns in a spin-glas model of neural networks. *Phys. Rev. Lett.*, 55(14):1530–1533, 1985.

14. M. Tsodyks and M. Feigelman. Enhanced storage capacity in neural networks with low level of activity. *Europhys. Lett.*, 6(2):101–105, 1988.

15. D. J. Amit, N. Brunel, and M. Tsodyks. Correlations of cortical Hebbian reverberations: theory versus experiment. *J. Neurosci.*, 14(11):6435–6445, 1994.

16. C. Tetzlaff, C. Kolodziejski, M. Timme, M. Tsodyks, and F. Wörgötter. Synaptic scaling enables dynamically distinct short- and long-term memory formation. *PLoS Comput. Biol.*, 9(10):e1003307, 2013.

17. A. Litwin-Kumar and B. Doiron. Formation and maintaince of neuronal assemblies through synaptic plasticity. *Nat. Commun.*, 5:5319, 2014.

18. F. Zenke, E. J. Agnes, and W. Gerstner. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nat. Commun.*, 6:6922, apr 2015.

19. N. Brunel. Is cortical connectivity optimized for storing information? *Nat. Neurosci.*, 19:749–755, May 2016.

20. D. J. Willshaw, O. P. Buneman, and H. C. Longuet-Higgins. Non-holographic associative memory. *Nature*, 222:960–962, 1969.

21. T. Kohonen. Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, 43(1):59–69, 1982.

22. D. M. Adelsberger-Mangan and W. B. Levy. Information maintenance and statistical dependence reduction in simple neural networks. *Biol. Cybern.*, 67:469–477, 1992.

23. T. J. Sullivan and V. R. de Sa. Homeostatic synaptic scaling in self-organizing maps. *Neural Networks*, 19(6-7):734–743, jul 2006.

24. A. Knoblauch, G. Palm, and F. T. Sommer. Memory capacities for synaptic and structural plasticity. *Neural Comput.*, 22(2):289–341, 2010.

25. J.-L. R. Stevens, J. S. Law, J. Antolik, and J. A. Bednar. Mechanisms for stable, robust, and adaptive development of orientation maps in the primary visual cortex. *J. Neurosci.*, 33(40):15747–15766, 2013.

26. B. Babadi and H. Sompolinsky. Sparseness and expansion in sensory representations. *Neuron*, 83(5):1213–1226, sep 2014.

27. T. Rogerson, D. J. Cai, A. Frank, Y. Sano, J. Shobe, M. F. Lopez-Aranda, and A. J. Silva. Synaptic tagging during memory allocation. *Nat. Rev. Neurosci.*, 15(3):157–169, feb 2014.

28. G. Kastellakis, A. J. Silva, and P. Poirazi. Linking memories across time via neuronal and dendritic overlaps in model neurons with active dendrites. *Cell Reports*, 17(6):1491–1504, nov 2016.

29. L. F. Abbott and S. B. Nelson. Synaptic plasticity: Taming the beast. *Nat. Neurosci. (Suppl.)*, 3:1178–1183, 2000.

30. W. Gerstner and W. M. Kistler. Mathematical formulations of Hebbian learning. *Biol. Cybern.*, 87:404–415, 2002.

31. G. G. Turrigiano, K. R. Leslie, N. S. Desai, L. C. Rutherford, and S. B. Nelson. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391:892–896, 1998.

32. G. G. Turrigiano and S. B. Nelson. Homeostatic plasticity in the developing nervous system. *Nat. Rev. Neurosci.*, 5:97–107, Feb 2004.

33. C. Tetzlaff, C. Kolodziejski, M. Timme, and F. Wörgötter. Synaptic scaling in combination with many generic plasticity mechanisms stabilizes circuit connectivity. *Front. Comput. Neurosci.*, 5:47, 2011.

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

35. T. Nachstedt and C. Tetzlaff. Working memory requires a combination of transient and attractor-dominated dynamics to process unreliably timed inputs. *Sci. Rep.*, 7:2473, 2017.

36. A. P. Yiu, V. Mercaldo, C. Yan, B. Richards, A. J. Rashid, H.-L. Hsiang, J. Pressey, V. Mahadevan, M. M. Tran, S. A. Kushner, M. A. Woodin, P. W. Frankland, and S. A. Josselyn. Neurons are recruited to a memory trace based on relative neuronal excitability immediately before training. *Neuron*, 83(3):722–735, aug 2014.

37. P. W. Frankland and S. A. Josselyn. Memory allocation. *Neuropsychopharmacology*, 40(40):243–243, jan 2015.

38. N. Erlenhardt, H. Yu, K. Abiraman, T. Yamasaki, J. I. Wadiche, S. Tomita, and D. S. Bredt. Porcupine controls hippocampal ampar levels, composition, and synaptic transmission. *Cell Rep.*, 14:782–794, February 2016.

39. P. J. Sjöström, G. G. Turrigiano, and S. B. Nelson. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32:1149–1164, 2001.

40. M. Graupner and N. Brunel. Mechanisms of induction and maintenance of spike-timing dependent plasticity in biophysical synapse models. *Front. Comput. Neurosci.*, 4:23, 2010.

41. G. G. Turrigiano. The dialectic of Hebb and homeostasis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 372, March 2017.

42. F. Zenke and W. Gerstner. Hebbian plasticity requires compensatory processes on multiple timescales. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 372, March 2017.

43. C. Tetzlaff, S. Dasgupta, T. Kulvicius, and F. Wörgötter. The use of Hebbian cell assemblies for nonlinear computation. *Sci. Rep.*, 5:12866, 2015.

44. D. Kim, D. Paré, and S. S. Nair. Assignment of model amygdala neurons to the fear memory trace depends on competitive synaptic interactions. *J. Neurosci.*, 33:14354–14358, September 2013.

45. D. Kim, D. Paré, and S. S. Nair. Mechanisms contributing to the induction and storage of pavlovian fear memories in the lateral amygdala. *Learn. Mem.*, 20:421–430, July 2013.

46. D. J. Cai, D. Aharoni, T. Shuman, J. Shobe, J. Biane, W. Song, B. Wei, M. Veshkini, M. La-Vu, J. Lou, S. E. Flores, I. Kim, Y. Sano, M. Zhou, K. Baumgaertel, A. Lavi, M. Kamata, M. Tuszynski, M. Mayford, P. Golshani, and A. J. Silva. A shared neural ensemble links distinct contextual memories encoded close in time. *Nature*, 534(7605):115–118, may 2016.

47. J. Yokose, R. Okubo-Suzuki, M. Nomoto, N. Ohkawa, H. Nishizono, A. Suzuki, M. Matsuo, S. Tsujimura, Y. Takahashi, M. Nagase, A. M. Watabe, M. Sasahara, F. Kato, and K. Inokuchi. Overlapping memory trace indispensable for linking, but not recalling, individual memories. *Science*, 355:398–403, January 2017.

48. S. Recanatesi, M. Katkov, S. Romani, and M. Tsodyks. Neural network model of memory retrieval. *Front. Comp. Neurosci.*, 9:149, 2015.

49. M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52:1059–1069, 2010.