The interplay of synaptic plasticity and scaling enables self-organized formation and 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

***For correspondence:** johannes.auth@weizmann.ac.il (JMA)

[†]These authors contributed equally to this work

Present address: [§]Department of Neurobiology, The Weizmann Institute of Science, Rehovot, Israel

Abstract It is commonly assumed that memories about experienced stimuli are represented by groups of highly interconnected neurons called 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 synaptic plasticity and homeostatic synaptic scaling organizes synaptic weight adaptations such that a new stimulus forms a new memory and where different stimuli are assigned to distinct cell assemblies. The resulting dynamics can reproduce experimental in-vivo data, focusing on how diverse factors 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.

Introduction

Learning and memorizing information about the environment over long time scales is a vital function of neural circuits of living beings. For this, different elements of a neural circuit - the neurons and synapses - have to coordinate themselves to accurately form, organize, and allocate internal long-term representations of the different pieces of information received. While many processes and dynamics of the single elements are well documented, their coordination remains obscure. How do neurons and synapses self-organize to form functional, stable, and distinguishable memory representations? Moreover, what mechanisms underlie the self-organized coordination yielding such representations?

Current hypotheses center on long-term synaptic plasticity as the core mechanism for robust coordination of the circuits' elements into memory representations (***Martin et al., 2000***; ***Barbieri and Brunel, 2008***; ***Palm et al., 2014***; ***Takeuchi et al., 2014***). Different from this, in this theoretical study, we identify that formation of functional memory representations requires the interaction of synaptic plasticity with (slower) synaptic scaling, which is a mechanisms that enforces network homeostasis (***Turrigiano et al., 1998***; ***Hengen et al., 2013***; ***Tetzlaff et al., 2011***). This interaction ensures the reliable, self-organized coordination of synaptic and neuronal elements into memories.

As discussed in the following, the organization of memory representations is subdivided into two processes: memory formation and allocation.

The reliable *formation* of memory, hence of internal stimulus representations in neural circuits, is explained by the Hebbian hypothesis. In brief, the Hebbian hypothesis (**James, 1890**; **Konorski, 1948**; **Hebb, 1949**; **Palm et al., 2014**; **Holtmaat and Caroni, 2016**) states that, when a neural circuit receives a new piece of information, the corresponding stimulus activates a group of neurons and, via activity-dependent long-term synaptic plasticity (**Bliss and Lomo, 1973**; **Levy and Steward, 1983**; **Martin et al., 2000**; **Malenka and Bear, 2004**), adapts the weights or efficacies of synapses between these neurons. This adaptation remodels the activated group of neurons to a strongly interconnected group of neurons called Hebbian cell assembly (CA). This newly formed CA serves as an internal long-term representation (long-term memory) of the corresponding stimulus (**Hebb, 1949**; **Palm et al., 2014**; **Holtmaat and Caroni, 2016**). Recall of this 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 by the strong recurrent interconnections between CA-neurons resulting in pattern completion (**Hunsaker and Kesner, 2013**; **Palm et al., 2014**). Several experimental and theoretical studies have investigated the formation and recall of CAs given synaptic plasticity (**Hopfield, 1982**, **1984**; **Amit et al., 1985**; **Tsodyks and Feigelman, 1988**; **Amit et al., 1994**; **Buzsaki, 2010**; **Brunel, 2016**; **Holtmaat and Caroni, 2016**). Recent theoretical studies already indicate that, in addition to synaptic plasticity, homeostatic mechanisms, such as synaptic scaling, are required to keep the neural circuit in a functional regime (**Tetzlaff et al., 2013**; **Litwin-Kumar and Doiron, 2014**; **Zenke et al., 2015**; **Tetzlaff et al., 2015**). However, all the above-mentioned studies investigate the coordination of synaptic and neuronal dynamics within the group of neurons (memory formation), but they do not consider the dynamics determining which neurons are recruited to form this group, or rather why is the stimulus allocated to this specific group of neurons and not to others (memory allocation; (**Rogerson et al., 2014**)).

For a proper *allocation* of stimuli to their internal representations, the synapses from the neurons encoding the stimulus to the neurons encoding the internal representation have to be adjusted accordingly. Considering only these types of synapses ("feed-forward synapses"), several studies show that long-term synaptic plasticity yields the required adjustments of synaptic weights (**Willshaw et al., 1969**; **Adelsberger-Mangan and Levy, 1992**; **Knoblauch et al., 2010**; **Babadi and Sompolinsky, 2014**; **Kastellakis et al., 2016**; **Choi et al., 2018**) such that each stimulus is mapped to its corresponding group of neurons. However, theoretical studies (**Sullivan and de Sa, 2006**; **Stevens et al., 2013**), which investigate the formation of input-maps in the visual cortex (**Kohonen, 1982**; **Obermayer et al., 1990**), show that the stable mapping or allocation of stimuli onto a neural circuit requires, in addition to synaptic plasticity, homeostatic mechanisms such as synaptic scaling. Note that all these studies focus on the allocation of stimuli to certain groups of neurons; however, they do not consider the dynamics of the synapses within these groups ("recurrent synapses").

Thus, up to now, it remains unclear how a neural circuit coordinates in a self-organized manner the synaptic and neuronal dynamics underlying the reliable allocation of stimuli to neurons with the simultaneous dynamics underlying the proper formation of memory representations. If these two memory processes are not tightly coordinated, the neural system could show awkward, undesired dynamics: On the one hand, memory allocation could map a stimulus to a group of unconnected neurons, which impedes the formation of a proper CA. On the other hand, the formation of a CA could bias the dynamics of allocation such that multiple stimuli are mapped onto the same CA disrupting the ability to discriminate between these stimuli.

In this theoretical study, we show in a network model that long-term synaptic plasticity (**Hebb, 1949**; **Abbott and Nelson, 2000**; **Gerstner and Kistler, 2002**) together with the slower, homeostatic processes of synaptic scaling (**Turrigiano et al., 1998**; **Turrigiano and Nelson, 2004**; **Tetzlaff et al., 2011**; **Hengen et al., 2013**) leads to the self-organized coordination of synaptic weight changes at feed-forward and recurrent synapses. Remarkably, the synaptic changes of the recurrent synapses yields the reliable formation of CAs. In parallel, the synaptic changes of the feed-forward synapses links the newly formed CA with the corresponding stimulus-transmitting neurons without interrupting already learned ones assuring the proper allocation of CAs. The model reproduces *in-*

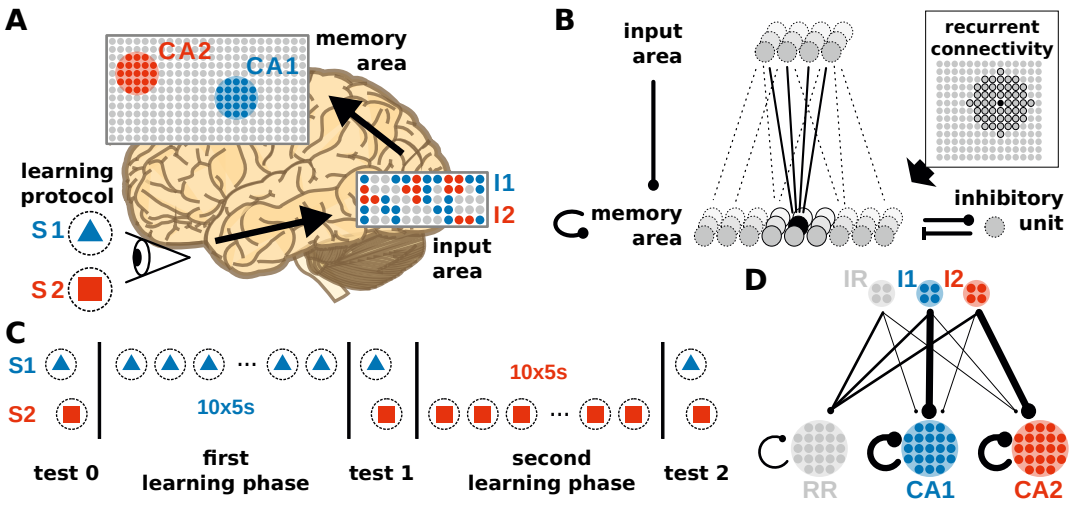

Figure 1. The neural system to investigate the coordination of synaptic and neuronal dynamics underlying the allocation and formation of memories consists of two areas receiving external stimuli. (A): The neural system receives different external stimuli (e.g., a blue triangle, S1, or a red square, S2) yielding the activation of subsets of neurons (colored dots) in the input (I1 and I2) and memory area (CA1 and CA2). (brain image by Hugh Guiney under license CC BY-SA 3.0) (B): Each excitatory neuron in the memory area receives inputs from a random subset of excitatory neurons being in the input area, from its neighboring neurons of the memory area (indicated by the dark gray units in the inset) and from a global inhibitory unit. All synapses between excitatory neurons are plastic regarding the interplay of long-term synaptic plasticity and synaptic scaling. (C): Throughout this study, we consider two learning phases during each a specific stimulus is repetitively presented. In addition, test phases are considered with stopped synaptic dynamics for analyses. (D): Schematic illustration of the average synaptic structure ensuring a proper function of the neural system. This structure should result from the neuronal and synaptic dynamics in conjunction with the stimulation protocol. IR and RR represent populations of remaining neurons being not directly related to the dynamics of memory formation and allocation. Details see main text.

vivo experimental data and provides testable predictions. Furthermore, the analysis of a population model, capturing the main features of the network dynamics, allows us to determine three generic properties of the interaction between synaptic plasticity and scaling, which enable the formation and allocation of memory representations in a reliable manner. These properties of synaptic adaptation are that (i) synaptic weights between two neurons with highly-correlated activities are strengthened (homosynaptic potentiation), (ii) synaptic weights between two neurons with weakly-correlated activities are lowered (heterosynaptic depression), and (iii) the time scale of synaptic weight changes are regulated by the postsynaptic activity level.

Results

Throughout this study, we consider a neural system receiving environmental stimuli (e.g., geometrical shapes as a blue triangle, S1, or a red square, S2), which evoke certain activity patterns within the input area (e.g., blue pattern I1 evoked by S1; ***Figure 1*** A). Each activity pattern, in turn, triggers via several random feed-forward synapses the activation of a subset of neurons in the recurrently connected memory area. For simplicity, all neurons of the system are considered to be excitatory; only a single all-to-all connected inhibitory unit (resembling an inhibitory population) in the memory area regulates its global activity level (***Figure 1*** B). Furthermore, as we investigate here neuronal and synaptic dynamics happening on long time scales, we neglect the influence of single spikes and, thus, directly consider the dynamics of the neuronal firing rates (see *Materials and Methods*). All synapses between the excitatory neurons (feed-forward and recurrent) are adapted by activity-dependent long-term plasticity. Due to the recurrent connections, the memory area should robustly form internal representations of the environmental stimuli, while simultaneously

the feed-forward synapses should provide a proper allocation of the stimuli (activity patterns in input area) onto the corresponding representations. In the following, we will show in a step-by-step approach that the interplay of long-term synaptic plasticity (here conventional Hebbian synaptic plasticity) with homeostatic synaptic scaling (see Methods; ***Tetzlaff et al.*** (***2011***, 2013); ***Nachstedt and Tetzlaff*** (***2017***)) coordinates synaptic changes such that proper formation and allocation of memory is ensured.

First, stimulus S1 is repetitively presented ten times (first learning phase; ***Figure 1*** C). Given this stimulation, the resulting synaptic adaptations of feed-forward and recurrent synapses should yield the proper formation of an internal representation indicating that the dynamics underlying memory allocation (changes of feed-forward synapses) does not impede the recurrent dynamics of CA-formation. After the formation of this representation, next, we repetitively present a different stimulus S2 (second learning phase). Due to this stimulation, the neural system should form another CA representing S2, which is independent of the first one. The proper formation of a second CA indicates that memory allocation is not biased by recurrent dynamics enabling a reliable discrimination between stimuli. Please note that we consider three test phases (***Figure 1*** C), during which synaptic dynamics are fixed, to enable the investigation of the resulting response dynamics of the circuit according to the different stimuli. Otherwise, the system is always plastic. In general, we expect that the neural system should form strongly interconnected groups of neurons according to each learning stimulus (memory formation; CA1 and CA2 in ***Figure 1*** D indicated by thicker lines) while remaining neurons in the memory area last weakly interconnected (RR). In addition, the synapses from the neurons in the input area, which are activated by a specific stimulus (I1 and I2), to the corresponding CAs should have larger weights while all other feed-forward connections remain rather weak (memory allocation; I1 to CA1 and I2 to CA2). After showing that the interplay of synaptic plasticity and scaling yields such a synaptic structure, in the next step, we derive a population model of the neural system and analyze the underlying synaptic and neuronal dynamics to identify the required generic properties determining the synaptic adaptations. Finally, we demonstrate that our theoretical model matches and provides potential explanations for a series of experiments revealing the relation between neuronal dynamics and the allocation of memory (***Yiu et al., 2014***). In addition, we compile some experimentally verifiable predictions, which would confirm the here-proposed hypothesis that the interplay between synaptic plasticity and scaling is required for the proper formation and allocation of memories.

Formation and allocation of one memory representation

Before learning, feed-forward as well as recurrent synapses, on average, do not show any structural bias (***Figure 2*** A, test 0) such that the presentation of an environmental stimulus (e.g., S1 or S2) triggers via the activation of a stimulus-specific pattern within the input area (I1 or rather I2) the activation of a random pattern of active neurons in the memory area. To evaluate whether these activated neurons are directly connected with each other, which serves as the basis for the formation of a CA (***Hebb, 1949***; ***Palm et al., 2014***), we consider the average shortest path length (ASPL; ***Appendix 1***) between these neurons. In general, the ASPL is a graph theoretical measure to assess the number of units (here neurons) along the shortest path to be taken in the network to reach one specific unit starting from another unit averaged across all pairs of units (here we consider only the pairs of activated neurons). Thus, as the ASPL-value is high, we conclude that the activated neurons in the memory area are not directly connected with each other (***Figure 2*** B, test 0); thus, the activity pattern is mainly determined by the random feed-forward connections.

By contrast, if a stimulus (here stimulus S1) is repeatedly presented in a given time interval, the neuronal and synaptic dynamics of the network reshapes the pattern of activated neurons in the memory area such that the final pattern consists of a group of interconnected neurons (decrease in ASPL; ***Figure 2*** B, test 1). As shown in our previous studies (***Tetzlaff et al., 2013***; ***Nachstedt and Tetzlaff, 2017***), 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

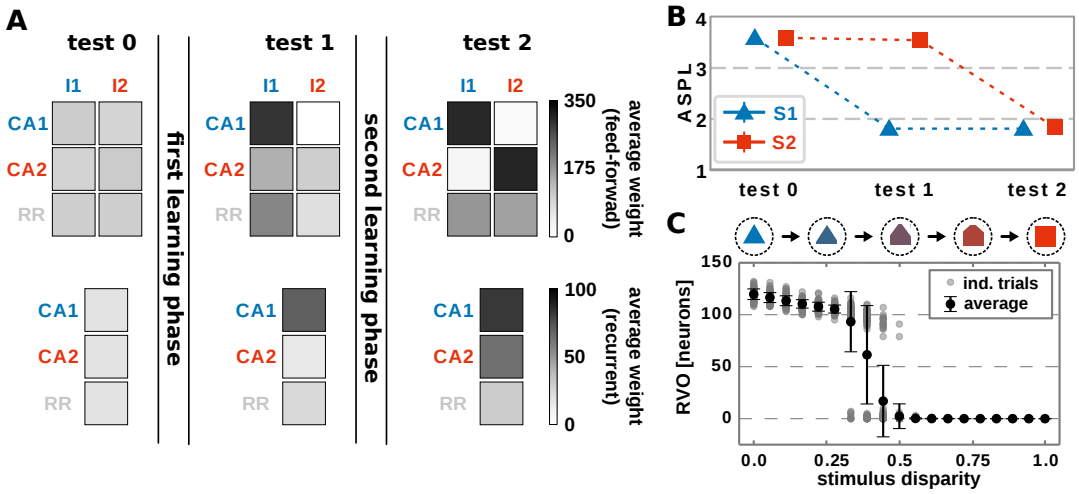

Figure 2. The interaction of conventional Hebbian synaptic plasticity and synaptic scaling enables the stimulus-dependent formation and allocation of memory representations in a neuronal network model. (A): During the test phases, the resulting network structure is evaluated. Top row: average synaptic weight of feed-forward synapses from input populations I1 and I2 activated by the corresponding stimuli to the groups of neurons which become a CA (CA1 and CA2) and others (RR). Please note that we first train the network, then determine the resulting CAs with corresponding neurons, and retroactively sort the neurons into the CA-groups. Bottom row: average synaptic weight of recurrent synapses within the corresponding groups of neurons (CA1, CA2, and other neurons RR). Before learning (test 0), no specific synaptic structures are present. After the first learning phase (test 1), the first group of neurons becomes strongly interconnected, thus, a CA (CA1), which becomes also strongly connected to the active input population I1. (test 2) The second learning phase yields the formation of a second CA (CA2), which is linked to the second input population I2. (B): The formation of CAs is also indicated by the reduction of the average shortest path length (ASPL) between stimulus-activated neurons in the memory area. (Error bars are small and overlapped by symbols). (C): After both learning phases, the response vector overlap (RVO) between neurons activated by S1 and activated by S2 depends non-linearly on the disparity between the stimulus patterns. (A-C): Data presented are mean values over 100 repetitions. Explicit values of mean and standard deviation are given in ***Appendix 2***.

Figure 2–Figure supplement 1. The robustness of CA formation according to parameter variations.

Figure 2–Figure supplement 2. All synaptic weights of feed-forward and recurrent connections.

recurrent synapses without significantly altering other synapses (***Figure 2*** A, test 1, bottom; neurons are sorted into groups retroactively; see ***Figure 2–Figure Supplement 2*** for exemplary, complete weight matrices). Taken together with the decreased ASPL this indicates 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? A repeated presentation of stimulus S1 yields an on average strengthening of synapses projecting from the stimulus-neurons I1 to the whole memory area (***Figure 2*** A, test 1, top). Essentially, the synapses from stimulus-I1-neurons to CA1-neurons have a significantly stronger increase in synaptic weights than the controls (CA2 and RR). In more detail, presenting stimulus S1 initially activates an increasing number of mainly unconnected neurons in the memory area (***Figure 3***; first presentation, fifth row, dark red dots; each dot represents the firing rate of one neuron in the memory area. Note that nearby neurons are connected with each other as indicated in the inset of ***Figure 1*** B and ***Figure 3–Figure Supplement 1***. This does not indicate the physical distance between neurons.). However, the ongoing neuronal activation during following stimulus presentations increases the recurrent synaptic weights (***Figure 3***, fourth row; each dot represents the average incoming weight of recurrent synapses a corresponding neuron receives) and also feed-forward weights from stimulus I1 to the neurons (***Figure 3***, second row; each dot represents the average incoming weight a neuron in the memory area receives from all feed-forward synapses from I1-neurons of the input area), which, in turn, increases the activity. This positive feedback loop between synaptic weights and neuronal activity leads to the emergence

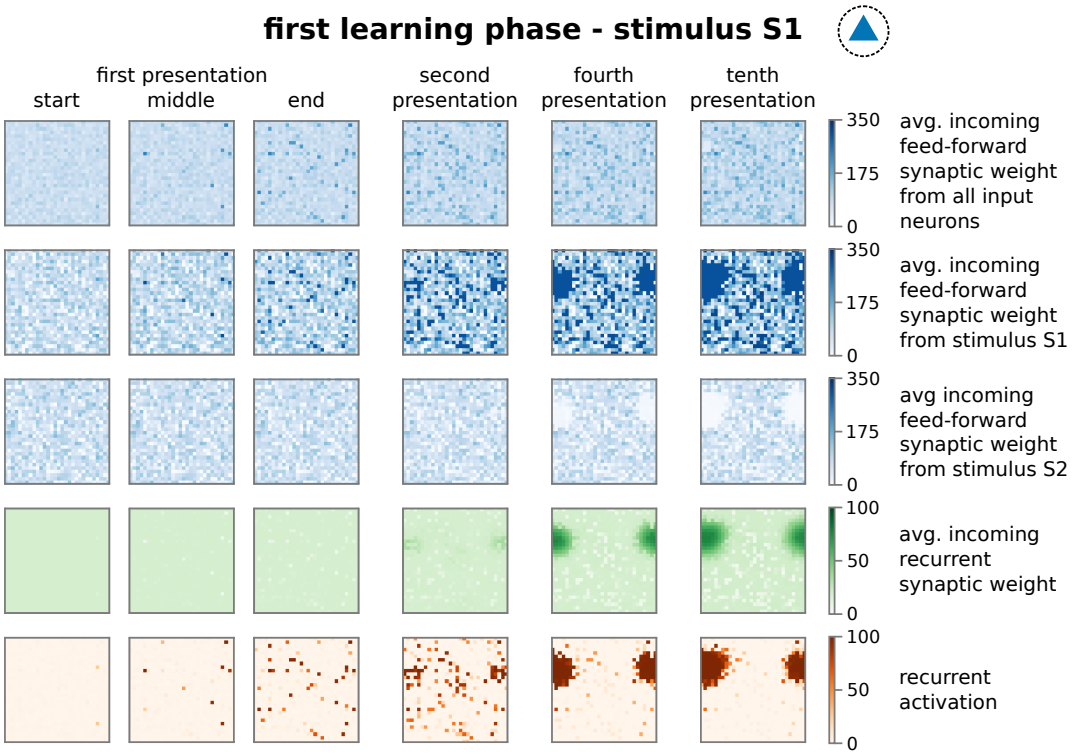

Figure 3. The repetitive presentation of a stimulus (here S1) triggers changes in feed-forward and recurrent synaptic weights as well as neural activities resulting to the proper formation and allocation of a CA. Each panel represents properties of the recurrent network in the memory area with neurons ordered on a $30x30$ grid as indicated in ***Figure 2–Figure Supplement 1*** A at different points of the protocol during the first learning phase. Thus, each dot in a panel represents a property of a neuron in the memory area, which is connected to the neighboring dots as shown in ***Figure 2–Figure Supplement 1*** B. These properties are; first row: average feed-forward synaptic weights from all input neurons; second row: average feed-forward synaptic weights from the subset of I1-input neurons; third row: average feed-forward synaptic weights from the subset of I2-input neurons; forth row: average incoming synaptic weight from neurons of the memory area (recurrent synapses); fifth row: firing rate of the corresponding neuron. Please note that we consider torus-like periodic boundary conditions.

Figure 3–Figure supplement 1. Indexing of neurons in subplots and topology.

of groups of activated neurons, which are directly connected with each other (second presentation, fifth row). Such an interconnected group grows out, incorporating more directly connected neurons, until inhibition limits its growth (see below) and, furthermore, suppresses sparse activation in the remaining neurons by competition (fourth to tenth presentation). The recurrent weights among CA-neurons are increased until an equilibrium between Hebbian synaptic plasticity and synaptic scaling is reached (fourth row). Please note that, as our previous studies show (***Tetzlaff et al., 2011***, ***2013***), the interplay between these two mechanisms yields the existence of this equilibrium; otherwise synaptic weights would grow unbounded even if the neural activity is limited (see also detailed analysis below). Interestingly, the weights of the feed-forward synapses show a different dynamics as of the recurrent synapses. The average over all synaptic weights linking from the input area to each neuron in the memory area (first row) indicates that synapses connected to CA-neurons have a similar average weight compared to synapses connected to other neurons. This implies that the synaptic weight changes of the feed-forward connections to the CA-neurons are on average not significantly different than controls. However, if the feed-forward synapses are sorted according to the stimulus-affiliation of the presynaptic neuron, we see that only the weights of synapses from the S1-stimulus-neurons to the emerging CA-neurons are strengthened (second row, dark blue spot; see also ***Figure 2*** A, test 1, I1 to CA1), while weights of synapses from other stimulus-neurons to the

CA-neurons are on average decreased (third row, white spot; see also ***Figure 2*** A, test 1, I2 to CA1). 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 same stimulus is presented later again. These results reveal that the interaction of synaptic plasticity and scaling self-organizes for a wide parameter regime (see ***Figure 2–Figure Supplement 1***) synaptic changes at recurrent and feed-forward connections to form and allocate a memory representation in a previously random neuronal network.

Formation and allocation of a second memory representation

After showing that the synaptic dynamics of the feed-forward connection does not impede the formation of a CA as internal representation of a stimulus, next, we will demonstrate that the recurrent dynamics (thus, a formed CA) does not obstruct the feed-forward dynamics given new stimuli. Clearly the presence of a memory representation can alter the self-organizing dynamics shown before, which could impede the proper formation and allocation 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, different stimulus S2 (second learning phase; ***Figure 1*** C) after the proper formation of the CA associated to stimulus S1 and analyze whether a second CA is formed which is independent of the first one. Similar to the first learning phase, the repeated presentation of stimulus S2 (here, stimulus-associated activity patterns in the input area I1 and I2 have a *stimulus disparity* equals 1 indicating no overlap between patterns; see ***Appendix 1***) yields via activity pattern I2 in the input area the activation of a group of interconnected neurons in the memory area (decreased ASPL; ***Figure 2*** B, test 2, red). In addition, the stimulation triggers a strengthening of the corresponding recurrent synaptic weights (***Figure 2*** A, bottom row, test 2, CA2; ***Figure 4***). 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 (***Figure 2*** C, disparity equals 1; ***Figure 4***, forth row, tenth presentation; ***Figure 2–Figure Supplement 2***). As indicated by the response vector overlap (RVO; basically the number of neurons activated by both stimuli), this depends on the disparity between stimuli; for quite dissimilar stimuli both CAs are separated (disparity $\gtrsim 0.5$ yields RVO ≈ 0), for more similar stimuli the system undergoes a state transition ($0.3 \lesssim$ disparity $\lesssim 0.5$ yields RVO > 0), and for quite similar stimuli both stimuli activate basically the same group of neurons (disparity $\lesssim 0.3$ yields RVO > 100 given that 120 ± 4 neurons are on average part of a CA; ***Figure 5–Figure Supplement 1***). Note that the latter demonstrates that the network does correctly assign noisy versions of a learned stimulus pattern (*pattern completion* (***Hunsaker and Kesner, 2013***)) instead of forming a new CA, while the first case illustrates that the network performs *pattern separation* (***Hunsaker and Kesner, 2013***) to distinguish different stimuli. 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 input pattern and newly formed CA are adapted accordingly (***Figure 2*** A, test 2, I2 to CA2). 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 (***Figure 2*** A). But, although stimulus S1 is not present, the second learning phase leads to a weakening of synapses projecting from corresponding input neurons I1 to the newly formed CA considerably below control (***Figure 2*** A, test 2, I2 to CA1). Similarly, during the first learning phase the synaptic weights between input neurons I2 and the first cell assembly are also weakened (***Figure 2*** A, test 1, I2 to CA1). Apparently, this weakening of synapses from the other, non-assigned stimulus to a CA reduces the chance of spurious activations. In addition, as we will show in the following, this weakening is also an essential property enabling the proper functioning of the neural system. In summary, these results of our theoretical network model show that the self-organized processes resulting from the interplay

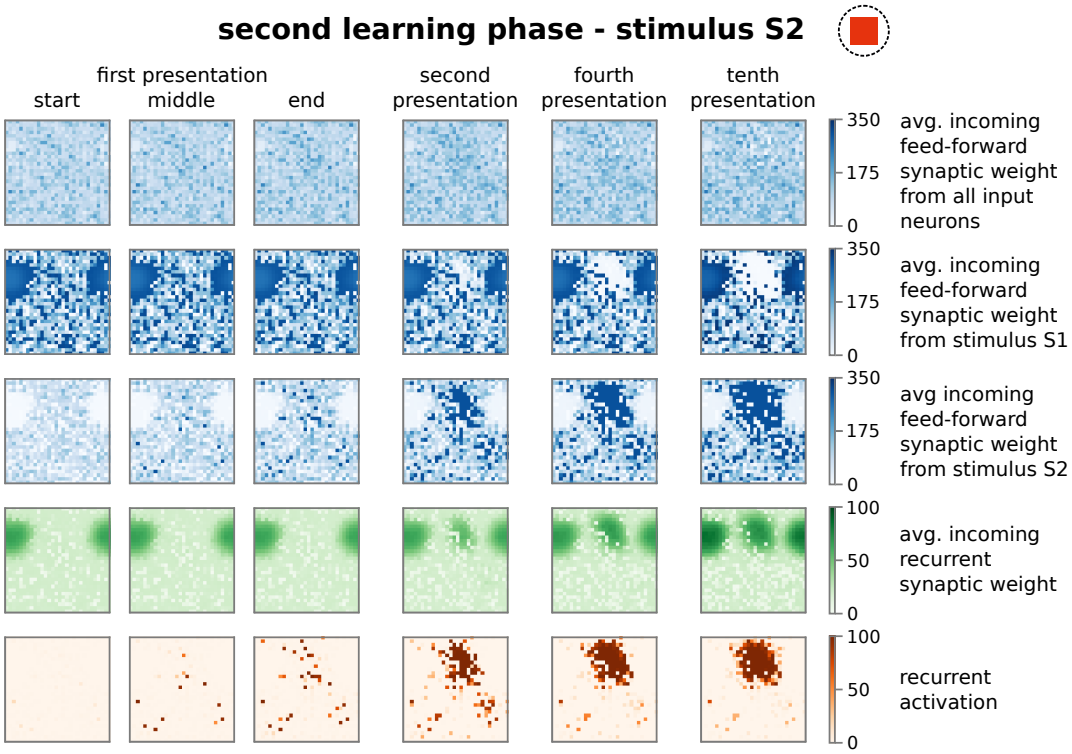

Figure 4. The presentation of a second stimulus (here S2) yields the formation of a second, distinct group of strongly interconnected neurons or CA. The structure of the subplots is the same as in ***Figure 3***. The first learning phase yields the encoding of a highly interconnected subpopulation of neurons in the memory area (***Figure 2*** & ***Figure 3***). However, due to the interplay between Hebbian synaptic plasticity and synaptic scaling this CA cannot be activated by the second stimulus S2. Instead, the process of initially scattered activation (dark red dots in the fifth row; first presentation) and the following neuronal and synaptic processes (fourth to tenth presentation), as described before, are repeated yielding the formation of a second CA representing stimulus S2. Please note that both representations do not overlap (see ***Figure 2–Figure Supplement 2***).

between synaptic plasticity and scaling coordinates the synaptic and neuronal dynamics such that the proper formation and allocation of several memory representations without significant interferences between them is enabled.

Generic properties of synaptic adaptations required for the formation and allocation of memory representations

In order to obtain a more detailed understanding of the self-organizing coordination of synaptic and neuronal dynamics by synaptic plasticity and scaling underlying the reliable formation and allocation of memory representations, we have to reduce the complexity of the model to enable (partially) analytical investigations. As already indicated by the above shown results (***Figure 2*** A), the main features of the self-organizing network dynamics can be described by considering the synaptic weights averaged over the given neuronal populations. Thus, we assume that the different involved neuronal populations in the input (input-pattern I1 and I2 neurons) and memory area (populations CA1 and CA2 becoming CAs in the memory area) are by themselves homogeneous allowing the derivation of a theoretical model describing the average population dynamics (***Figure 5*** A). For this, we combine the neuronal dynamics of all neurons within such a population (groups of I1-, I2-, CA1-, CA2-neurons) and describe them by the average neuronal dynamics of the population such that we obtain four variables each describing the average firing rate of one population or group of neurons (I1: $\bar{I}_1$; I2: $\bar{I}_2$; CA1: $\bar{F}_1$; CA2: $\bar{F}_2$). Similarly, we combine the synaptic dynamics of all synapses to

describe them by the average synaptic dynamics for all connections within a neuronal population (within CA1: $\bar{w}_1^{\text{rec}}$; within CA2: $\bar{w}_2^{\text{rec}}$) and for all connections between populations (from I1 to CA1: $\bar{w}_{11}^{\text{ff}}$; from I1 to CA2: $\bar{w}_{21}^{\text{ff}}$; from I2 to CA1: $\bar{w}_{12}^{\text{ff}}$; from I2 to CA2: $\bar{w}_{22}^{\text{ff}}$). As activities and weights of the remaining neurons and synapses (IR and RR groups of neurons in ***Figure 1*** D) remain small, in the following, we neglect their influence on the system dynamics. The inhibition is considered similarly (see *Materials and Methods*) and the values of some system parameters are taken from full network simulations (***Figure 5–Figure Supplement 1***). Please note, we re-scale the average neuronal activities and the average synaptic weights such that, if the weights equal one, learning is completed (see *Materials and Methods*). By considering the average neuronal and synaptic dynamics, we map the main features of the complex network dynamics with $\approx N^2$ dimensions (N is the number of neurons in the network) to a 8-dimensional theoretical model.

Given such a population model of our adaptive network, first, we investigate the formation of a memory representation in a blank, random neural circuit (thus, $\bar{I}_1 > 0$ and $\bar{I}_2 = 0$). As mentioned before, the advantage of using a population model is the reduced dimensionality enabling, amongst others, the analytical calculation of the nullclines of the system to classify its dynamics. Nullclines show states of the system in which the change in one dimension equals zero (***Glendinning, 1994***; ***Izhikevich, 2007***). Thus, the intersection points of all nullclines of a system are the fixed points of the systems dynamics describing the system states in which no change occurs. Such fixed points can either be stable (attractive), thus the system will tend to converge into this point or state, or unstable (repulsive) meaning that the system will move away from this state (***Appendix 4***). Such dynamics can be summarized in a phase space diagram, in which each point indicates a possible state of the system and we can determine from the relative position of this state to the nullclines and fixed points to which new state the system will go to. The population model of the network model has 8 dimensions, thus, 8 nullclines and a 8-dimensional phase space (***Figure 5*** A); however, by inserting solutions of some nullclines into others (***Appendix 3***), we can project the dynamics onto two dimensions for visualization. In the following, we project the dynamics onto the average recurrent synaptic weights of both populations in the memory area ($\bar{w}_1^{\text{rec}}$, $\bar{w}_2^{\text{rec}}$) to investigate the dynamics underlying the formation of a CA during the first learning phase (***Figure 5*** B, top; please see ***Figure 5*** B, bottom, for corresponding activity levels).

Thus, the projection of the solutions of the nullclines of the system dynamics onto the average recurrent synaptic weights of the two neuronal groups shows that the recurrent dynamics during the first learning phase are dominated by three fixed points: one is unstable (orange, 7; more specifically, it is a saddle point) and two are stable (green, 2 and 3). As the recurrent synaptic weights before learning are in general rather weak (which "sets" the initial state in this diagram), the fixed points 4, 8, 9, 10 cannot be reached by the system and, thus, they do not influence the here discussed dynamics. The two stable fixed points represent that one of the two neuronal populations becomes a strongly interconnected CA, while the other remains in a weakly interconnected state. For instance, in state 2 the first population is a CA as it is strongly interconnected ($\bar{w}_1^{\text{rec}} \gg 0$) and the second population of neurons remains weakly connected ($\bar{w}_2^{\text{rec}}$ has only about 10% of the value of $\bar{w}_1^{\text{rec}}$). The unstable or repulsive fixed point lies on the identity line ($\bar{w}_1^{\text{rec}} = \bar{w}_2^{\text{rec}}$) having the same distance to both stable, attractive states. The resulting mirror symmetry in the phase space implies that the dynamics on the one side of the identity line, reaching the stable fixed point lying on this side, equals the dynamics on the other side. Please note that the form of the nullclines and, thus, the existence and positions of the fixed points of the system dynamics depend on the mechanisms determining the synaptic adaptations. In other words, given a strong stimulus, the interplay between synaptic plasticity and scaling coordinates the recurrent synaptic dynamics such that the system by itself has to form a CA (reach one of both stable states). The question, which of both stable states is reached, translates into the question, to which group of neurons will the stimulus be allocated. As both groups are before learning quite similar, the initial state of the system will be close to the identity line. Only a minor difference between both groups (e.g., slightly different recurrent synaptic weights or a little different number of feed-forward synapses)

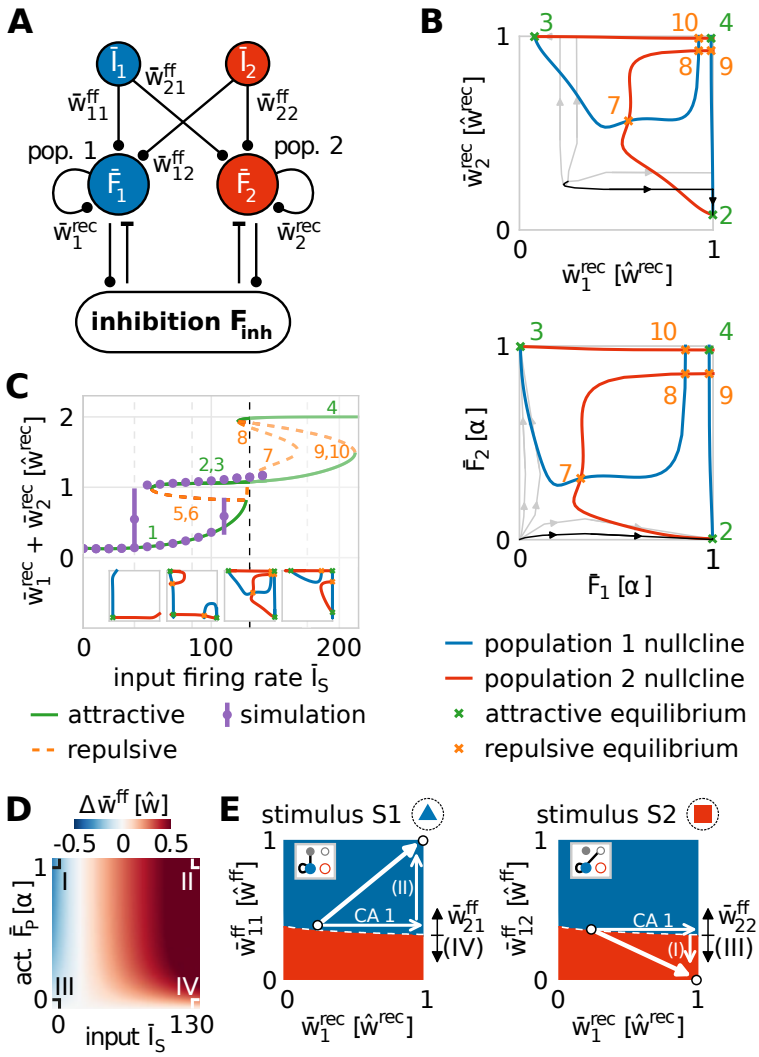

Figure 5. Population model of network dynamics enables analytical derivation of the underlying neuronal and synaptic dynamics. (A): Schema of the population model for averaged network dynamics (bars above variables indicate the average over all neurons in the population. $\bar{I}_s$: firing rate of input population $s \in \{1,2\}$; $\bar{F}_p$: neural activity of population $p \in \{1,2\}$; $\bar{w}^{ff}_{ps}$: weight of feed-forward synapses from population s to p; $\bar{w}^{rec}_p$: weight of recurrent synapses within population p. (B): The intersections of the population nullclines projected into weight (top) and activity (bottom) space reveal several fixed points (attractive: green; repulsive: orange) indicating the formation of a CA (green markers 2 and 3), if the system deviates from the identity line. Numbers correspond to labels of fixed points. (C): The bifurcation diagram (labels as in (B) and insets) 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). Solutions of the full network model (purple dots) and population model match. (D): The dynamics of feed-forward synaptic weights depends on the firing rate of the input population and of the population in the memory area. There are four different cases (I-IV) determining the system dynamics. (E): These cases (indicated by arrows with Roman numbers: I-IV) together with the potentiation of recurrent synapses (arrow labeled 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 (red) will represent the presented stimulus (left: stimulus S1; right: stimulus S2). Details see main text.

Figure 5–Figure supplement 1. Parameter estimation for population model.

results to a small variation in the initial condition of the first learning phase such that the system is slightly off the identity line (see traces for examples). Given the difference, the system will be on one side of the identity line and converge to the corresponding fixed point implying that the

corresponding group of neurons will become the internal representation (e.g., black trace). Note that this symmetry-dependent formation of a CA is quite robust as long as the input firing rate is above a certain threshold ($\bar{I}_A \gtrsim 120$), which agrees with results from the more detailed network model discussed before (purple dots; ***Figure 5*** C). Below this threshold, the system remains in a state both groups of neurons are not becoming CAs (state 1 in ***Figure 5*** C; see also first two insets). Thus, the existence of the threshold predicts that a new, to-be-learned stimulus has to be able to evoke sufficient activity in the input area (above the threshold) to trigger the processes of memory formation; otherwise, the system will not learn.

In parallel to the development of the recurrent synaptic weights, the synaptic weights of the feed-forward connections change to assure proper memory allocation. Thus, we derive analytically the activity-dependency of the interaction of synaptic plasticity and scaling and obtain the change of the feed-forward synaptic weights ($\Delta\bar{w}^{\text{ff}}$), expected during the first learning phase, given different activity conditions of the input ($\bar{I}_s$) and CA-populations ($\bar{F}_p$, $s, p \in \{1, 2\}$; ***Figure 5*** D and ***Appendix 5***). As expected, the combination of both activity levels for a certain duration determines whether the weights of the feed-forward synapses are potentiated (red), depressed (blue), or not significantly adapted (white). In general, if both activities are on a quite high level, synapses are potentiated (case II; so-called homosynaptic potentiation; ***Miller*** (***1996***)). If the presynaptic activity (input population) is on a low level and the postsynaptic activity (CA-population) is on a high level, on average, feed-forward synapses are depressed (case I; so-called heterosynaptic depression; ***Miller*** (***1996***)). 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. For this, given the presentation of a to-be-learned stimulus, we have to consider the basins of attraction of the system in the phase space (***Figure 5*** E) projected onto different types of connections (insets; gray indicates the active input population). If the system is in a certain state, we marked by the color of this state which group of neurons will become a CA and will be assigned to the stimulus presented (left: S1 is presented; right: S2 is presented). The mirror symmetry described before (***Figure 5*** B) maps to the boundary (white dashed line in ***Figure 5*** E) between both basins of attraction (blue: population 1 becomes the internal representation; red: population 2 becomes the CA). Thus, during the first learning phase (stimulus S1; ***Figure 5*** E, left), a small variation in initial conditions breaks the symmetry such that the system is, in the example highlighted in ***Figure 5*** B (black trace), in the basin of attraction of population 1 becoming the internal representation (dot nearby symmetry line in blue area). This leads to the strengthening of the recurrent synapses within population 1 forming a CA (increase of $\bar{w}_1^{\text{rec}}$; ***Figure 5*** B, E). In parallel, the synaptic strengthening induces an increase of the activity level of the population ($\bar{F}_1$; black trace in ***Figure 5*** B, bottom) yielding, together with the high activity level of input population I1 ($\bar{I}_1 \gg 0$), an average increase of the corresponding feed-forward synapses ($\bar{w}_{11}^{\text{ff}}$; case II in ***Figure 5*** D). These synaptic changes push the system further away from the symmetry condition (white arrows; ***Figure 5*** E, left) implying a more stable memory representation. Note that changing the strength of synapses connecting input population I1 with population 2 ($\bar{w}_{21}^{\text{ff}}$) could result in 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 in ***Figure 5*** D). Thus, during the first learning phase, the formation and allocation of an internal representation is dictated by the subdivision of the system phase space into different basins of attraction of the stable fixed points such that small variations in the before-learning state of the network predetermines the dynamics during learning. This subdivision, in turn, emerges from the interplay of synaptic plasticity and scaling.

How do these synaptic and neuronal dynamics of the allocation and formation of the first CA influence the dynamics of the second learning phase? In general, the formation of a CA acts as a variation or perturbation of the initial condition breaking the symmetry for the second learning

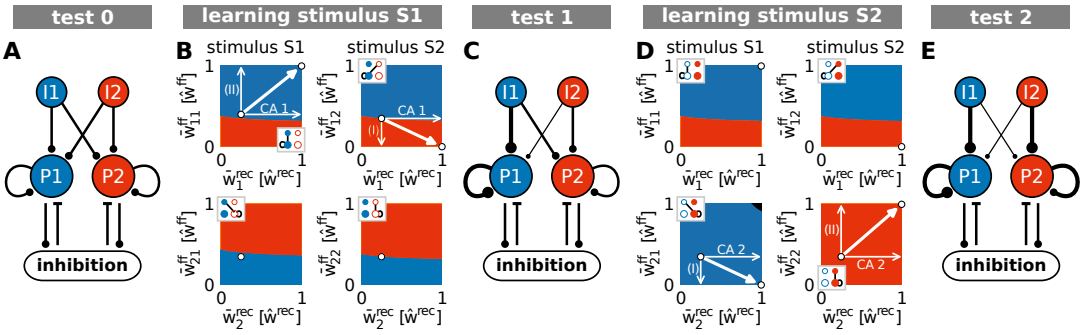

Figure 6. Summary of the synaptic changes and their implication on the formation and allocation of memory representations. The interaction of synaptic plasticity and scaling brings a blank network (A; test 0) during the first learning phase (B) to an intermediate state (C; test 1). From this intermediate state, the second learning phase (D) yields the system into the desired end state (E; test 2), in which each stimulus is allocated to one CA ($S1/I1$ to pop. 1/CA1 and $S2/I2$ to pop. 2/CA2). (A), (C), (E): Thickness of lines is proportional to average synaptic weight. (B), (D): Similar to panels in ***Figure 5*** E. Black area indicates regimes in which both populations would be assigned to the corresponding stimulus.

phase (stimulus S2; ***Figure 5*** E, right). The formation of the first CA pushes the system into the blue area (CA1-arrow). This indicates that, if stimulus S2 is presented, the feed-forward synapses would be adapted such that population 1 would also represent stimulus S2. This would impede the discrimination ability of the network between stimulus S1 and S2. However, during the first learning phase, as the input population I2 of stimulus S2 is inactive, synapses projecting from I2-input neurons to population 1 neurons are depressed (case I in ***Figure 5*** D; downward arrow in ***Figure 5*** E, right) and the system switches into the red area. This area indicates that, if stimulus S2 is presented as during the second learning phase, population 2 would form a CA representing stimulus S2 and not population 1. Please note that this switch can be impeded by adapting the connections from the I2-input population to population 2 during the first learning phase ($\bar{w}^{\mathrm{ff}}_{22}$) shifting the symmetry condition (black arrows in ***Figure 5*** E, right). But, similar to before, this effect is circumvented by the system, as population 2 is basically inactive resulting to case III (***Figure 5*** D). 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 form a new CA during the second learning phase (see ***Figure 6*** for further phase space projections and dynamics during second learning phase). These results indicate that this intermediate state can only be reached if synaptic adaptations comprise three properties implied by the four cases I-IV (***Figure 5*** D): (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).

Modeling experimental findings of competition-based memory allocation

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 (***Yiu et al., 2014***; ***Frankland and Josselyn, 2015***) 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 *Materials and 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

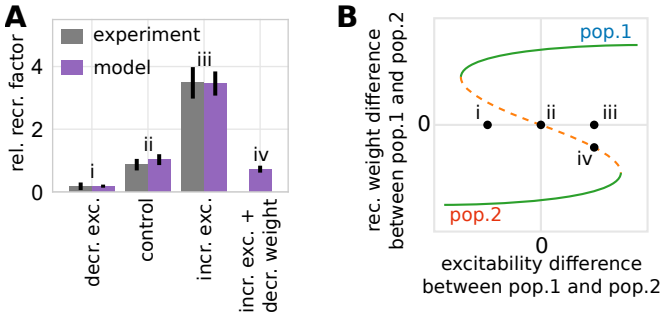

Figure 7. The model of synaptic plasticity and scaling matches experimental *in-vivo* data and provides experimentally verifiable predictions. (A): The artificial modification of the excitability of a subset of neurons alters the probability of these neurons to become part of a memory representation (normalized to control). Experimental data are taken from (***Yiu et al., 2014***). Data presented are mean values with standard error of the mean. Labels correspond to instances shown in (B). (B): The alteration of the excitability of one group of neurons (here pop. 1) compared to others yields a shift of the positions of the system's fixed points and basins of attractions (here shown schematically; for details see ***Figure 7–Figure Supplement 1***) inducing a bias towards one population (e.g., from instance ii to iii by increasing the excitability of population 1). (A), (B): The model analysis yields the prediction that this excitability-induced bias can be counterbalanced by, for instance, additionally decreasing the average synaptic weight of the manipulated population before learning (here by a factor of 0.1). This additional manipulation shifts the initial state of the network back to the symmetry condition (orange; instance iv).

Figure 7–Figure supplement 1. Excitability bifurcation diagram as obtained from the population model

compared the results to experimental findings (***Figure 7*** A).

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 (***Figure 5*** B), this phenomenon can be explained as follows: the manipulation of the excitability in one population of neurons changes the distance between the repulsive state (orange; ***Figure 7*** 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 (e.g., instance iii for increased excitability of population 1). 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 that this group of neurons becomes 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; (***Erlenhardt et al., 2016***)), the network's initial condition is shifted such that the CREB-induced influence on the relative recruitment factor is counterbalanced (***Figure 7*** A, instance iv). Thus, the match between model and *in-vivo* experimental data supports the here-proposed hypothesis that the combination of Hebbian synaptic plasticity and synaptic scaling coordinates the synaptic and neuronal dynamics enabling the self-organized allocation and formation of memory representations.

Discussion

Our 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 (***Tetzlaff et al., 2013***; ***Litwin-Kumar and Doiron, 2014***; ***Zenke et al., 2015***). 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 (***Kohonen, 1982***; ***Sullivan and de Sa, 2006***; ***Stevens et al., 2013***), 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 (***Sjöström et al., 2001***; ***Graupner and Brunel, 2010***).

Note that 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 (***Abbott and Nelson, 2000***; ***Tetzlaff et al., 2011***; ***Turrigiano and Nelson, 2004***; ***Turrigiano, 2017***; ***Zenke and Gerstner, 2017***).

Similar to previous studies (***Tetzlaff et al., 2013***, ***2015***; ***Nachstedt and Tetzlaff, 2017***), 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 (***Kim et al., 2013a***; ***Kastellakis et al., 2016***); 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 (***Kim et al., 2013a***,b; ***Kastellakis et al., 2016***) also imply the three generic properties (i-iii) supporting our findings. Further investigations are required to assess possible differences between different realizations of the three generic properties. For this, the here used theoretical methods from the field of nonlinear dynamics (***Glendinning, 1994***; ***Izhikevich, 2007***) seem to be promising given their ability to derive and classify fundamental system dynamics, which can be verified by experiments.

Our results indicate that the combined dynamics of Hebbian synaptic plasticity and synaptic scaling are sufficient to coordinate the synaptic and neuronal dynamics underlying the allocation and formation of memory representations. Interestingly, for several different stimuli the 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 (neurons encoding several stimuli). However, amongst others, experimental results indicate that memory representations can overlap (***Holtmaat and Caroni, 2016***; ***Cai et al., 2016***; ***Yokose et al., 2017***) and, in addition, theoretical studies show that overlaps increase the storage capacity of a neuronal network (***Tsodyks and Feigelman, 1988***) and can support memory recall (***Recanatesi et al., 2015***). 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 (***Figure 7***; ***Kim et al.*** (***2013a***); ***Yiu et al.*** (***2014***)). Furthermore, the dynamics of CREB seem to be time-dependent (***Yiu et al., 2014***; ***Frankland and Josselyn, 2015***; ***Kastellakis et al., 2016***). 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. Thus, our study shows that the interplay between synaptic plasticity and scaling is required to include all three generic properties of synaptic adaptation enabling a proper formation and allocation of memories. In addition, given the here-derived theoretical model, other mechanisms can be included to investigate systematically their functional implication on the self-organized, complex system dynamics underlying the multitude of memory processes.

Materials and Methods

Numerical simulations

The neuronal system in the numerical model consist of 936 excitatory neurons (36 in input area, 900 in memory area) and a single inhibitory unit. The inhibitory unit describes a population of inhibitory neurons which are connected to the excitatory neurons in an all-to-all manner. Given the long time scales considered in this study, 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 ***Figure 1*** B and ***Figure 3–Figure Supplement 1***) and to the global inhibitory unit. Initially, recurrent synaptic weights of existing connections equal to $0.25 \cdot \hat{w}^{\text{rec}}$ and feed-forward synaptic weights are drawn from a uniform distribution $\{0, 0.7 \cdot \hat{w}^{\text{ff}}\}$ ($\hat{w}^{\text{rec}} = \sqrt{\frac{\kappa^{\text{rec}}\alpha^2}{\alpha - F^{\text{T}}}} \approx 77.5$ and $\hat{w}^{\text{ff}} = \sqrt{\frac{\kappa^{\text{ff}}\alpha \cdot 130}{\alpha - F^{\text{T}}}} \approx 306.1$). Connections to and from the inhibitory neuron are at fixed and homogeneous weight (for detailed values see ***Table 1***).

Neuron model

The membrane potential $u_{\text{i}}^{\text{rec}}$, $i \in \{1, ..., N^{\text{M}}\}$ of each excitatory neuron i in the memory area is described as follows:

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

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 firing rate I_{k} of input neuron 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:

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

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 ***Equation 4***):

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

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

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 weight changes of the excitatory feed-forward (equation ***Equation 5***) and recurrent synapses (equation ***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 $w^{\mathsf{ff}}_{\mathsf{i,k}}$ from input neuron $k \in \{0, \cdots, N^{\mathsf{I}}\}$ to memory neuron $i \in \{0, \cdots, N^{\mathsf{M}}\}$ is:

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

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

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

$c^{\mathsf{ff}}_{\mathsf{i,k}}$ and $c^{\mathsf{ff}}_{\mathsf{i,k}}$ are the entries in the feed-forward and recurrent connectivity matrices of value 1, if the connection exists, or otherwise of value 0.

All other connections are considered to be non-plastic.

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
$w_{\mathsf{inh,i}}$	synaptic weight to inhibitory unit	0.6
$w_{\mathsf{i,inh}}$	synaptic weight from inhibitory unit	1200

Coding framework

The differential equations have been solved numerically with the Euler method with a time step of 5 ms using Python 3.5.

Simulation of self-org. formation of two CAs

The system undergoes two learning phases presenting two completely dissimilar input patterns I1 (first phase) and I2 (second phase). For input pattern I1 half of the input neurons are set to be active at 130 Hz whereas the other half remains inactive at 0 Hz and vice versa for input pattern I2. During a 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 (***Yiu et al., 2014***), 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 $\bar{w}_i^{\text{rec}}$ between neurons of the population, and the mean weight $\bar{w}_i^{\text{ff}}$ 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 (***Figure 5*** E and ***Figure 6***), 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 *appendices*.

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 5–Figure Supplement 1*** C; $n^{\text{ff}} = 2.3$, ***Figure 5–Figure Supplement 1*** 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 2–Figure Supplement 2*** 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 ($i \in \{1, 2\}$):

$$\frac{\mathrm{d}\bar{w}_{ik}^{\text{ff}}}{dt} = \mu\left(\bar{F}_i\bar{I}_k + (\kappa^{\text{ff}})^{-1}(F^{\text{T}} - \bar{F}_i)(\bar{w}_{ik}^{\text{ff}})^2\right)., \tag{10}$$

$$\frac{\mathrm{d}\bar{w}_i^{\text{rec}}}{\mathrm{d}t} = \mu\left(\bar{F}_i^2 + (\kappa^{\text{rec}})^{-1}(F^{\text{T}} - \bar{F}_i)(\bar{w}_i^{\text{rec}})^2\right) \tag{11}$$

Data Display

The population model applies the same combined learning rule as the numerical simulation. We thus consider the memorization process completed when the dynamic fixed point of synaptic plasticity is reached, i.e. Hebbian plasticity and scaling compensate each other.

In order to focus the reader's attention onto the populations dynamics as well as for illustrative reasons, we scale the explicit values of synaptic weights and activation in their depiction (***Figure 5*** B-E, ***Figure 6*** B,D, ***Figure 7–Figure Supplement 1***) with the respective values of their dynamic fixed points (weights; $\hat{w}^{\text{ff}}$, $\hat{w}^{\text{rec}}$) or the maximum value (activation; α). Hence, their values simply range from 0 to 1.

Recruitment Basins

For determining the recruitment basins (***Figure 5*** E and ***Figure 6***), we exploit the symmetry of the system and that, in general, only one of the two stimuli (S1 or S2) is active. Accordingly, we approximate the second, inactive input to zero and neglect the respective feed-forward synapses. The population model is integrated with the given initial values of the feed-forward and recurrent weights and $\bar{u}_1 = \bar{u}_2 = \bar{u}_{\text{inh}} = 0$ for $100\,\text{s}$. At $t = 100\,\text{s}$, we evaluate which of the two populations is active. ***Table 2*** provides the exact used initial values.

Table 2. Initial values for Recruitment Basin Plots

Figure	$\bar{u}_1$	$\bar{u}_2$	$\bar{u}_{\text{inh}}$	$\bar{w}^{\text{ff}}_{1A}$	$\bar{w}^{\text{ff}}_{2A}$	$\bar{w}^{\text{rec}}_{1}$	$\bar{w}^{\text{rec}}_{2}$
2E left; 2E right†; 3B top left, 3B top right†, 3B bottom right*†, 3D top left, 3D top right†	0	0	0	$0, \Delta w^{\text{ff}}_{1A}, \ldots, \hat{w}^{\text{ff}}_{1A}$	$0.35\,\hat{w}^{\text{ff}}_{2A}$	$0, \Delta w^{\text{rec}}_{1}, \ldots, \hat{w}^{\text{rec}}_{1}$	$0.25\,\hat{w}^{\text{rec}}_{2}$
3B bottom left*	0	0	0	$0, \Delta w^{\text{ff}}_{1A}, \ldots, \hat{w}^{\text{ff}}_{1A}$	$0.40\,\hat{w}^{\text{ff}}_{2A}$	$0, \Delta w^{\text{rec}}_{1}, \ldots, \hat{w}^{\text{rec}}_{1}$	$0.25\,\hat{w}^{\text{rec}}_{2}$
3D bottom left*	0	0	0	$0, \Delta w^{\text{ff}}_{1A}, \ldots, \hat{w}^{\text{ff}}_{1A}$	$1.00\,\hat{w}^{\text{ff}}_{2A}$	$0, \Delta w^{\text{rec}}_{1}, \ldots, \hat{w}^{\text{rec}}_{1}$	$1.00\,\hat{w}^{\text{rec}}_{2}$
3D bottom right*†	0	0	0	$0, \Delta w^{\text{ff}}_{1A}, \ldots, \hat{w}^{\text{ff}}_{1A}$	$0.00\,\hat{w}^{\text{ff}}_{2A}$	$0, \Delta w^{\text{rec}}_{1}, \ldots, \hat{w}^{\text{rec}}_{1}$	$1.00\,\hat{w}^{\text{rec}}_{2}$

$$\Delta w^{\text{ff}}_{1A} = 0.001\hat{w}^{\text{ff}}_{1A},\ \Delta w^{\text{rec}}_{1} = 0.001\hat{w}^{\text{rec}}_{1}$$

*) Using symmetry by commutating populations. †) Using symmetry by commutating inputs.

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

Abbott LF, Nelson SB. Synaptic plasticity: taming the beast. Nat Neurosci. 2000; 3(11):1178–1183. doi: https://doi.org/10.1038/81453.

Adelsberger-Mangan DM, Levy WB. Information maintenance and statistical dependence reduction in simple neural networks. Biol Cybern. 1992; 67(5):469–477. doi: https://doi.org/10.1007/BF00200991.

Amit DJ, Brunel N, Tsodyks M. Correlations of cortical Hebbian reverberations: theory versus experiment. J Neurosci. 1994; 14(11):6435–6445. doi: https://doi.org/10.1523/JNEUROSCI.14-11-06435.1994.

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

Babadi B, Sompolinsky H. Sparseness and Expansion in Sensory Representations. Neuron. 2014; 83(5):1213–1226. doi: https://doi.org/10.1016/j.neuron.2014.07.035.

Barbieri F, Brunel N. Can attractor network models account for the statistics of firing during persistent activity in prefrontal cortex? Front Neurosci. 2008; 2(1):114–122. doi: https://doi.org/10.3389/neuro.01.003.2008.

Bliss TVP, Lomo T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J Physiol. 1973; 232:331–356. doi: https://doi.org/10.1113/jphysiol.1973.sp010273.

Brunel N. Is cortical connectivity optimized for storing information? Nat Neurosci. 2016; 19:749–755. doi: https://doi.org/10.1038/nn.4286.

Buzsaki G. Neural syntax, cell assemblies, synapsembles, and readers. Neuron. 2010; 68:362–385. doi: https://doi.org/10.1016/j.neuron.2010.09.023.

Cai DJ, Aharoni D, Shuman T, Shobe J, Biane J, Song W, Wei B, Veshkini M, La-Vu M, Lou J, Flores SE, Kim I, Sano Y, Zhou M, Baumgaertel K, Lavi A, Kamata M, Tuszynski M, Mayford M, Golshani P, et al. A shared neural ensemble links distinct contextual memories encoded close in time. Nature. 2016; 534(7605):115–118. doi: https://doi.org/10.1038/nature17955.

Choi JH, Sim SE, Kim JI, Choi DI, Oh J, Ye S, Lee J, Kim T, Ko HG, Lim CS, Kaang BK. Interregional synaptic maps among engram cells underlie memory formation. Science. 2018; 360:430–435. doi: http://dx.doi.org/10.1126/science.aas9204.

Erlenhardt N, Yu H, Abiraman K, Yamasaki T, Wadiche JI, Tomita S, Bredt DS. Porcupine Controls Hippocampal AMPAR Levels, Composition, and Synaptic Transmission. Cell Rep. 2016; 14:782–794. doi: http://dx.doi.org/10.1016/j.celrep.2015.12.078.

Frankland PW, Josselyn SA. Memory allocation. Neuropsychopharmacology. 2015; 40(40):243–243. doi: http://dx.doi.org/10.1038/npp.2014.234.

Gerstner W, Kistler WM. Mathematical formulations of Hebbian learning. Biol Cybern. 2002; 87:404–415. doi: https://doi.org/10.1007/s00422-002-0353-y.

Glendinning P. Stability, instability and chaos: An introduction to the theory of nonlinear differential equations. Cambridge University Press; 1994.

Graupner M, Brunel N. Mechanisms of induction and maintenance of spike-timing dependent plasticity in biophysical synapse models. Front Comput Neurosci. 2010; 4:136. doi: https://doi.org/10.3389/fncom.2010.00136.

Hebb DO. The Organization of Behaviour. Wiley, New York; 1949.

Hengen KB, Lambo ME, Van Hooser SD, Katz DB, Turrigiano GG. Firing rate homeostasis in visual cortex of freely behaving rodents. Neuron. 2013; 80:335–342. doi: https://doi.org/10.1016/j.neuron.2013.08.038.

Holtmaat A, Caroni P. Functional and structural underpinnings of neuronal assembly formation in learning. Nat Neurosci. 2016; 19:1553–1562. doi: https://doi.org/10.1038/nn.4418.

Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci. 1982; 79(8):2554–2558. doi: https://doi.org/10.1073/pnas.79.8.2554.

Hopfield JJ. Neurons with graded response have collective computational properties like those of two-state neurons. Proc Natl Acad Sci. 1984; 81:3088–3092. doi: https://doi.org/10.1073/pnas.81.10.3088.

Hunsaker MR, Kesner RP. The operation of pattern separation and pattern completion processes associated with different attributes or domains of memory. Neurosci Biobehav Rev. 2013; 37:36–58. doi: https://doi.org/10.1016/j.neubiorev.2012.09.014.

Izhikevich E. Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting. MIT Press; 2007.

James W. The principles of psychology. New York: Henry Holt and Company; 1890.

Kastellakis G, Silva AJ, Poirazi P. Linking Memories across Time via Neuronal and Dendritic Overlaps in Model Neurons with Active Dendrites. Cell Reports. 2016; 17(6):1491–1504. doi: https://doi.org/10.1016/j.celrep.2016.10.015.

Kim D, Paré D, Nair SS. Assignment of model amygdala neurons to the fear memory trace depends on competitive synaptic interactions. J Neurosci. 2013; 33:14354–14358. doi: https://doi.org/10.1523/JNEUROSCI.2430-13.2013.

Kim D, Paré D, Nair SS. Mechanisms contributing to the induction and storage of Pavlovian fear memories in the lateral amygdala. Learn Mem. 2013; 20:421–430. doi: 10.1101/lm.030262.113.

Knoblauch A, Palm G, Sommer FT. Memory capacities for synaptic and structural plasticity. Neural Comput. 2010; 22(2):289–341. doi: https://doi.org/10.1162/neco.2009.08-07-588.

Kohonen T. Self-organized formation of topologically correct feature maps. Biol Cybern. 1982; 43(1):59–69. doi: https://doi.org/10.1007/BF00337288.

Konorski J. Conditioned reflexes and neuron organization. Cambridge: Cambridge University Press; 1948.

Levy WB, Steward O. Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. Neuroscience. 1983; 8(4):791–797. doi: https://doi.org/10.1016/0306-4522(83)90010-6.

Litwin-Kumar A, Doiron B. Formation and maintaince of neuronal assemblies through synaptic plasticity. Nat Commun. 2014; 5:5319. doi: https://doi.org/10.1038/ncomms6319.

Malenka RC, Bear MF. LTP and LTD: an embarrassment of riches. Neuron. 2004; 44:5–21. doi: https://doi.org/10.1016/j.neuron.2004.09.012.

Martin SJ, Grimwood PD, Morris RGM. Synaptic plasticity and memory: an evaluation of the hypothesis. Ann Rev Neurosci. 2000; 23(1):649–711. doi: https://doi.org/10.1146/annurev.neuro.23.1.649.

Miller KD. Synaptic economics: competition and cooperation in synaptic plasticity. Neuron. 1996; 17:371–374. doi: https://doi.org/10.1016/S0896-6273(00)80169-5.

Nachstedt T, Tetzlaff C. Working Memory Requires a Combination of Transient and Attractor-Dominated Dynamics to Process Unreliably Timed Inputs. Sci Rep. 2017; 7:2473. doi: https://doi.org/10.1038/s41598-017-02471-z.

Obermayer K, Ritter H, Schulten K. A principle for the formation of the spatial structure of cortical feature maps. Proc Natl Acad Sci. 1990; 87:8345–8349. doi: https://doi.org/10.1073/pnas.87.21.8345.

Palm G, Knoblauch A, Hauser F, Schüz A. Cell assemblies in the cerebral cortex. Biol Cybern. 2014; 108:559–572. doi: https://doi.org/10.1007/s00422-014-0596-4.

Recanatesi S, Katkov M, Romani S, Tsodyks M. Neural network model of memory retrieval. Front Comp Neurosci. 2015; 9:149. doi: https://doi.org/10.3389/fncom.2015.00149.

Rogerson T, Cai DJ, Frank A, Sano Y, Shobe J, Lopez-Aranda MF, Silva AJ. Synaptic tagging during memory allocation. Nat Rev Neurosci. 2014; 15(3):157–169. doi: https://doi.org/10.1038/nrn3667.

Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. Neuroimage. 2010; 52:1059–1069. doi: https://doi.org/10.1016/j.neuroimage.2009.10.003.

Sjöström PJ, Turrigiano GG, Nelson SB. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron. 2001; 32:1149–1164. doi: https://doi.org/10.1016/S0896-6273(01)00542-6.

Stevens JLR, Law JS, Antolik J, Bednar JA. Mechanisms for stable, robust, and adaptive development of orientation maps in the primary visual cortex. J Neurosci. 2013; 33(40):15747–15766. doi: https://doi.org/10.1523/JNEUROSCI.1037-13.2013.

Sullivan TJ, de Sa VR. Homeostatic synaptic scaling in self-organizing maps. Neural Networks. 2006; 19(6-7):734–743. doi: 10.1016/j.neunet.2006.05.006.

Takeuchi T, Duszkiewicz AJ, Morris RGM. The synaptic plasticity and memory hypothesis: encoding, storage and persistence. Philos Trans R Soc Lond B Biol Sci. 2014; 369:20130288. doi: 10.1098/rstb.2013.0288.

Tetzlaff C, Dasgupta S, Kulvicius T, Wörgötter F. The use of Hebbian cell assemblies for nonlinear computation. Sci Rep. 2015; 5:12866. doi: https://doi.org/10.1038/srep12866.

Tetzlaff C, Kolodziejski C, Timme M, Tsodyks M, Wörgötter F. Synaptic scaling enables dynamically distinct short- and long-term memory formation. PLoS Comput Biol. 2013; 9(10):e1003307. doi: https://doi.org/10.1371/journal.pcbi.1003307.

Tetzlaff C, Kolodziejski C, Timme M, Wörgötter F. Synaptic scaling in combination with many generic plasticity mechanisms stabilizes circuit connectivity. Front Comput Neurosci. 2011; 5:47. doi: https://doi.org/10.3389/fncom.2011.00047.

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

Turrigiano GG. The dialectic of Hebb and homeostasis. Philos Trans R Soc Lond B Biol Sci. 2017; 372. doi: 10.1098/rstb.2016.0258.

Turrigiano GG, Leslie KR, Desai NS, Rutherford LC, Nelson SB. Activity-dependent scaling of quantal amplitude in neocortical neurons. Nature. 1998; 391:892–896. doi: https://doi.org/10.1038/36103.

Turrigiano GG, Nelson SB. Homeostatic plasticity in the developing nervous system. Nat Rev Neurosci. 2004; 5:97–107. doi: https://doi.org/10.1038/nrn1327.

Willshaw DJ, Buneman OP, Longuet-Higgins HC. Non-holographic associative memory. Nature. 1969; 222:960–962. doi: http://dx.doi.org/10.1038/222960a0.

Yiu AP, Mercaldo V, Yan C, Richards B, Rashid AJ, Hsiang HL, Pressey J, Mahadevan V, Tran MM, Kushner SA, Woodin MA, Frankland PW, Josselyn SA. Neurons Are Recruited to a Memory Trace Based on Relative Neuronal Excitability Immediately before Training. Neuron. 2014; 83(3):722–735. doi: 10.1016/j.neuron.2014.07.017.

Yokose J, Okubo-Suzuki R, Nomoto M, Ohkawa N, Nishizono H, Suzuki A, Matsuo M, Tsujimura S, Takahashi Y, Nagase M, Watabe AM, Sasahara M, Kato F, Inokuchi K. Overlapping memory trace indispensable for linking, but not recalling, individual memories. Science. 2017; 355:398–403. doi: 10.1126/science.aal2690.

Zenke F, Agnes EJ, Gerstner W. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. Nat Commun. 2015; 6:6922. doi: doi:10.1038/ncomms7922.

Zenke F, Gerstner W. Hebbian plasticity requires compensatory processes on multiple timescales. Philos Trans R Soc Lond B Biol Sci. 2017; 372. doi: 10.1098/rstb.2016.0259.

Appendix 1

Measures used in the network simulation

Average shortest path length

The *average shortest path length* (ASPL, ***Figure 2*** D) is considered here as a measure to analyze the spatial distribution of activation within the memory area. A high ASPL between neurons indicates that these neurons are spatially broadly distributed across the memory area. By contrast, a low ASPL indicates that the neurons are clustered. In particular, as strongly activated neurons are supposed to become part of a memory representation, we focus on the distribution of highly activated neurons. For this, for each trial, we identified the 10% of neurons with the highest activity level (index set P) and calculated the shortest path length (SPL; (***Rubinov and Sporns, 2010***)); using the *networkX* package for *Python*) between them and averaged over all those paths (denoted by $\langle\cdot\rangle$):

$$\text{ASPL} = \langle \text{SPL}_{\text{i,j}} \rangle_{i,j \in P, \text{i} \neq \text{j}}. \tag{12}$$

Dynamic equilibria of synaptic weights

The *average outgoing recurrent synaptic weight* (***Figure 2*** E) is a measure of the interconnection within a neuronal subpopulation in the memory area (index set Q). We therefore averaged the synaptic weight over all the connections among neurons within the sup-population:

$$\bar{w}^{\text{rec}} = \langle w^{\text{rec}}_{\text{i,j}} \rangle_{i,j \in Q, \text{i} \neq \text{j}}. \tag{13}$$

The *average incoming feed-forward synaptic weight* (***Figure 2*** F,G) is the average synaptic weight of connections between a subpopulation in the memory area (index set Q) and a specific stimulus pattern in the input area (index set H):

$$\bar{w}^{\text{ff}} = \langle w^{\text{ff}}_{\text{i,k}} \rangle_{i \in Q, k \in H}. \tag{14}$$

Response disparity dependent on stimulus similarity

To analyze the response disparity, stimulus S1 is presented 10 times for 5 sec with 1 sec pause in between to form a single CA. After that, plasticity is shut off and we present variations of stimulus S1 with increasing *stimulus disparity* until the stimulus equals stimulus S2 (***Figure 2*** H). Stimulus disparity measures the relative amount of non-overlap between two stimulus patterns, in this case stimulus S1 and its variation (in the following called stimulus S1'). Both stimuli are of identical size $N^{\text{S}} = 0.5 \cdot N^{\text{I}}$, so that the stimulus disparity is calculated as follows:

$$\textit{stimulus disparity}\left(S1,\ S1'\right) \ = \ 1 - \frac{1}{N^{\text{S}}} \cdot \sum_{\text{k}}^{N^{\text{I}}} S_{\text{k}}(S1) \cdot S_{\text{k}}(S1'), \tag{15}$$

with binary stimulus patterns for a given stimulus $X \in S1, S1'$:

$$S_{\text{k}}(X) = \begin{cases} 1, & \text{if } I_{\text{k}}(X) \ = \ 130, \\ 0, & \text{if } I_{\text{k}}(X) \ = \ 0. \end{cases} \tag{16}$$

Thus, a stimulus disparity equal zero describes two identical stimuli, whereas a disparity equal one indicates two non-overlapping stimulus patterns. The input area size of $N^{\mathrm{I}} = 36$ allows for 18 steps in variation of $5.\bar{5}\%$ each. At the end of each presentation, we compare the resulting response in the memory area with the one at the end of the learning phase (i.e. the response to the original stimulus S1). The *response vector overlap* (RVO; ***Figure 2*** H) describes the similarity between the response patterns in the memory area due to the presentation of stimuli S1 and S1′:

$$RVO(S1,\ S1') = \sum_{\mathrm{i}}^{\mathrm{N^M}} R_{\mathrm{i}}(S1) \cdot R_{\mathrm{i}}(S1') \tag{17}$$

with binary response of neuron i to a given stimulus $X \in S1, S1'$:

$$R_{\mathrm{i}}(X) = \begin{cases} 1, & \text{if } F_{\mathrm{i}}(X) \geqslant 0.5 \cdot \alpha, \\ 0, & \text{else.} \end{cases} \tag{18}$$

Appendix 2

Appendix 2 Table 1. Explicit results depicted in ***Figure 2*** A,B: average shortest path length (ASPL) and average synaptic weights.

	test0	**test1**	**test2**
$\bar{w}^{\mathrm{ff}}_{11}$	107 (36)	290 (11)	302 (11)
$\bar{w}^{\mathrm{ff}}_{12}$	96 (22)	145 (19)	25 (15)
$\bar{w}^{\mathrm{ff}}_{21}$	95 (21)	3.03 (0.12)	14.95 (0.13)
$\bar{w}^{\mathrm{ff}}_{22}$	108 (36)	103 (36)	305.004 (0.002)
$\bar{w}^{\mathrm{rec}}_{1}$	19.37 (0)	69 (10)	81 (22)
$\bar{w}^{\mathrm{rec}}_{2}$	19.37 (0)	16.6 (5.2)	63 (20)
$\bar{w}^{\mathrm{rec}}_{\mathrm{RR}}$	19.37 (0)	24.7 (6.2)	30 (11)
ASPL S1	3.592 (0.018)	1.809 (0.014)	1.808 (0.016)
ASPL S2	3.592 (0.017)	3.546 (0.036)	1.837 (0.035)

Appendix 3

Nullclines and Equilibria

The equilibrium values $\bar{w}_i^{\mathrm{ff},*}$ and $\bar{w}_i^{\mathrm{rec},*}$ of the feed-forward and recurrent weights can be obtained as a function of the equilibrium values of the population activities $\bar{F}_i^*$ from ***Equation 10*** and ***Equation 11***:

$$\bar{w}_{ik}^{\mathrm{ff},*}(\bar{F}_i^*) = \sqrt{\frac{\kappa^{\mathrm{ff}}\bar{F}_i^*\bar{I}_k}{\bar{F}_i^* - F^{\mathsf{T}}}}. \tag{19}$$

$$\bar{w}_i^{\mathrm{rec},*}(\bar{F}_i^*) = \sqrt{\frac{\kappa^{\mathrm{rec}}(\bar{F}_i^*)^2}{\bar{F}_i^* - F^{\mathsf{T}}}}, \tag{20}$$

The equilibrium value $\bar{u}_{\mathrm{inh}}^*$ of the membrane potential of the inhibitory population can be formulated as a function of $\bar{F}_1^*$ and $\bar{F}_2^*$ based on ***Equation 8***:

$$\bar{u}_{\mathrm{inh}}^*(\bar{F}_1^*, \bar{F}_2^*) = R_{\mathrm{inh}}\tau_{\mathrm{inh}}(w_{\mathrm{inh},1}N_1\bar{F}_1^* + w_{\mathrm{inh},2}N_2\bar{F}_2^*). \tag{21}$$

By inserting equations ***Equation 20***, ***Equation 19*** and ***Equation 21*** into ***Equation 7*** and using $\bar{F}_i^* = F(\bar{u}_i^*)$, we obtain a system of the two population nullclines that only depends on the equilibrium values $\bar{u}_1^*$ and $\bar{u}_2^*$ ($i \in \{1,2\}$):

$$0 = -\frac{\bar{u}_i^*}{\tau} + R\left(\bar{n}_i^{\mathrm{rec}}\bar{w}_i^{\mathrm{rec},*}(\bar{F}_i^*)\bar{F}_i^* + w_{i,\mathrm{inh}}\bar{F}_{\mathrm{inh}}^*(\bar{F}_1^*, \bar{F}_2^*) + \sum_k \bar{n}^{\mathrm{ff}}\bar{w}_{ik}^{\mathrm{ff},*}(\bar{F}_i^*)\bar{I}_k\right).$$

We solve this system numerically to receive the equilibrium values $\bar{u}_1^*$ and $\bar{u}_2^*$ and, in consequence, by means of equations ***Equation 20***, ***Equation 19*** and ***Equation 21***, also $\bar{w}_1^{\mathrm{rec},*}$, $\bar{w}_2^{\mathrm{rec},*}$, $\bar{w}_{1\mathrm{A}}^{\mathrm{ff},*}$, $\bar{w}_{1\mathrm{B}}^{\mathrm{ff},*}$, $w_{2\mathrm{A}}^{\mathrm{ff},*}$, $\bar{w}_{2\mathrm{B}}^{\mathrm{ff},*}$ and $\bar{u}_{\mathrm{inh}}^*$.

Appendix 4

Stability

The stability of an equilibrium is determined by the sign of the eigenvalue with the largest real part of the system's Jacobi matrix evaluated at the equilibrium. The nonzero terms of the Jacobi matrix are ($i \in \{1,2\}$, $k \in \{\mathrm{S1}, \mathrm{S2}\}$):

$$\frac{\partial \dot{\bar{u}}_i}{\partial \bar{u}_i} = -\frac{1}{\tau} + R\bar{n}_i^{\mathrm{rec}} \bar{w}_i^{\mathrm{lat}} \frac{\partial \bar{F}_i}{\partial \bar{u}_i}, \qquad \frac{\partial \dot{\bar{u}}_i}{\partial \bar{u}_{\mathrm{inh}}} = R w_{i,\mathrm{inh}} \frac{\partial \bar{F}_{\mathrm{inh}}}{\partial \bar{u}_{\mathrm{inh}}},$$

$$\frac{\partial \dot{\bar{u}}_i}{\partial \bar{w}_i^{\mathrm{rec}}} = R\bar{n}_i^{\mathrm{rec}} \bar{F}_i, \qquad \frac{\partial \dot{\bar{u}}_i}{\partial \bar{w}_{ik}^{\mathrm{ff}}} = R\bar{n}_i^{\mathrm{ff}} \bar{I}_k,$$

$$\frac{\partial \dot{\bar{u}}_{\mathrm{inh}}}{\partial \bar{u}_i} = R_{\mathrm{inh}} w_{\mathrm{inh},i} N_i \frac{\partial \bar{F}_i}{\partial \bar{u}_i}, \qquad \frac{\partial \dot{\bar{u}}_{\mathrm{inh}}}{\partial \bar{u}_{\mathrm{inh}}} = -\frac{1}{\tau_{\mathrm{inh}}},$$

$$\frac{\partial \dot{\bar{w}}_i^{\mathrm{rec}}}{\partial \bar{u}_i} = \mu^{\mathrm{rec}} \frac{\partial \bar{F}_i}{\partial \bar{u}_i} \left(2\bar{F}_i - \frac{(\bar{w}_i^{\mathrm{rec}})^2}{\kappa^{\mathrm{rec}}} \right), \qquad \frac{\partial \dot{\bar{w}}_i^{\mathrm{rec}}}{\partial \bar{w}_i^{\mathrm{rec}}} = \frac{2\mu^{\mathrm{rec}}}{\kappa^{\mathrm{rec}}} (F^{\mathsf{T}} - \bar{F}_i) \bar{w}_i^{\mathrm{rec}},$$

$$\frac{\partial \dot{\bar{w}}_{ik}^{\mathrm{ff}}}{\partial \bar{u}_i} = \mu^{\mathrm{ff}} \frac{\partial \bar{F}_i}{\partial \bar{u}_i} \left(\bar{I}_k - \frac{(\bar{w}_{ik}^{\mathrm{ff}})^2}{\kappa^{\mathrm{ff}}} \right), \qquad \frac{\partial \dot{\bar{w}}_{ik}^{\mathrm{ff}}}{\partial \bar{w}_{ik}^{\mathrm{ff}}} = \frac{2\mu^{\mathrm{ff}}}{\kappa^{\mathrm{ff}}} (F^{\mathsf{T}} - \bar{F}_i) \bar{w}_{ik}^{\mathrm{ff}}$$

with

$$\frac{\partial \bar{F}_i}{\partial \bar{u}_i} = \beta \bar{F}_i \left(1 - \frac{\bar{F}_i}{\alpha} \right) \qquad \text{and} \qquad \frac{\partial \bar{F}_{\mathrm{inh}}}{\partial \bar{u}_{\mathrm{inh}}} = \beta \bar{F}_{\mathrm{inh}} \left(1 - \frac{\bar{F}_{\mathrm{inh}}}{\alpha} \right).$$

The eigenvalues of the resulting matrix are determined numerically.

Appendix 5

Feed-Forward Synaptic Weight Change

For constant pre- and postsynaptic activities (***Figure 5*** D), ***Equation 10*** can be solved analytically by separation of variables. The resulting time-course $w_i^{\mathrm{ff}}(t)$ depends on the given parameters and initial conditions:

$$\bar{w}_i^{\mathrm{ff}}(t) = \begin{cases} \bar{w}_i^{\mathrm{ff},*} \coth\left(\sqrt{\frac{\bar{F}_i \bar{I}_i (\bar{F}_i - F_{\mathsf{T}})}{\kappa^{\mathrm{ff}}}}(t-t_0)\mu^{\mathrm{ff}} + \operatorname{arcoth}\left(\frac{\bar{w}_i^{\mathrm{ff}}(t_0)}{w_i^{\mathrm{ff},*}}\right)\right) & \text{for } \bar{w}_i^{\mathrm{ff}}(t_0) > w_i^{\mathrm{ff},*} \wedge \bar{F}_i^{\mathrm{ff}} > F_{\mathsf{T}} \wedge \bar{I}_i > 0, \\ \bar{w}_i^{\mathrm{ff},*} \tanh\left(\sqrt{\frac{\bar{F}_i \bar{I}_i (\bar{F}_i - F_{\mathsf{T}})}{\kappa^{\mathrm{ff}}}}(t-t_0)\mu^{\mathrm{ff}} + \operatorname{artanh}\left(\frac{\bar{w}_i^{\mathrm{ff}}(t_0)}{w_i^{\mathrm{ff},*}}\right)\right) & \text{for } \bar{w}_i^{\mathrm{ff}}(t_0) < w_i^{\mathrm{ff},*} \wedge \bar{F}_i^{\mathrm{ff}} > F_{\mathsf{T}} \wedge \bar{I}_i > 0, \\ \bar{w}_i^{\mathrm{ff},\dagger} \tan\left(\sqrt{\frac{\bar{F}_i \bar{I}_i (\bar{F}_i - F_{\mathsf{T}})}{\kappa^{\mathrm{ff}}}}(t-t_0)\mu^{\mathrm{ff}} + \arctan\left(\frac{\bar{w}_i^{\mathrm{ff}}(t_0)}{w_i^{\mathrm{ff},*}}\right)\right) & \text{for } \bar{w}_i^{\mathrm{ff}}(t_0) < w_i^{\mathrm{ff},*} \wedge \bar{F}_i^{\mathrm{ff}} < F_{\mathsf{T}} \wedge \bar{I}_i > 0, \\ \left(\frac{1}{\bar{w}_i^{\mathrm{ff}}(t_0)} - \frac{F^{\mathsf{T}} - \bar{F}_i^{\mathrm{ff}}}{\kappa^{\mathrm{ff}}}(t-t_o)\mu^{\mathrm{ff}}\right)^{-1} & \text{for } \bar{F}_i^{\mathrm{ff}} = 0 \vee \bar{I}_i = 0 \end{cases}$$

with

$$\bar{w}_i^{\mathrm{ff},\dagger}(\bar{F}_i^*) = \sqrt{\frac{\kappa^{\mathrm{ff}} \bar{F}_i^* \bar{I}}{F^{\mathsf{T}} - \bar{F}_i^*}}.$$

Appendix 6

Comparison of Bifurcation Curve with Network Simulation

When comparing the equilibrium structure of the population model dependent on the input amplitude (bifurcation parameter) with the equilibria reached in network simulations (***Figure 5*** C), the network simulations are initialized close to the different expected stable configurations. For every input amplitude I, we perform two simulations with different initial conditions:

- $w_{ij}^{\text{rec}} = 0.25\hat{w}^{\text{rec}}$ for all realized recurrent synapses and $w_{ij}^{\text{ff}} = \hat{w}^{\text{ff}}$ for all realized feed-forward synapses.
- $w_{ij}^{\text{rec}} = \hat{w}^{\text{rec}}$ for synapses in between 121 neurons in a circle-shaped population, $w_{ij}^{\text{rec}} = 0.25\hat{w}^{\text{rec}}$ for all other realized recurrent synapses and $w_{ij}^{\text{ff}} = \hat{w}^{\text{ff}}$ for all realized feed-forward synapses.

In each case, the network is simulated for $50,000$ s. Every simulation is repeated 50 times with different random connectivities. To avoid simulation artifacts related to absolute silence of input channels, we assume a small background activity of 0.1α for inactive inputs. In the final state, we either consider all neurons with activity higher than 0.5α or, if there are none, 120 neurons centered around the activity center of the network as population 1. Population 2 is defined as the circular group of 120 neurons with the highest distance (respecting the periodic boundary conditions) to population 1. Within these two population, we evaluate the mean recurrent weight.

Large Input Amplitudes

In the network simulation, the functional role of the inhibitory population is two-fold: On the one hand, inhibition mediates the competition between different populations. This role is also captured by the population model. On the other hand, it prevents an active cell assembly from growing without limit by inhibiting neighboring neurons. This aspect is not reflected in the population model as in the latter the size of the populations is approximated as being fixed. Due to this discrepancy, the population model predicts equilibria also for very large input amplitudes while in the network simulation these input amplitudes lead to full activation of the complete network.

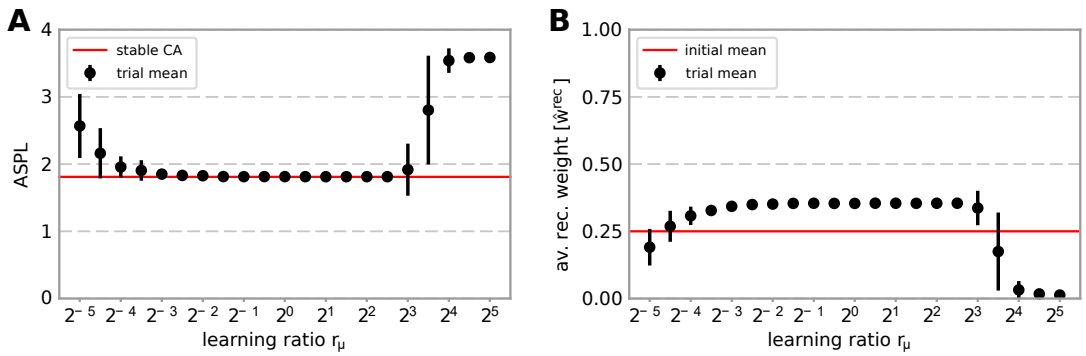

Figure 2–Figure supplement 1. The formation of a cell assembly is robust against changes in the velocity of synaptic adaptations of the feed-forward (μ^{ff}; **Equation 5**) and of the recurrent synapses (μ^{rec}; **Equation 6**). Given one learning phase, (A) ASPL (between the 10 mostly active neurons) as well as (B) average recurrent synaptic weights within the CA indicate that a CA is formed for a wide range of time scale ratios ($r_\mu = \frac{\mu^{rec}}{\mu^{ff}}$). Please note that a low ASPL (≈ 1.8; A, red line) and a significant increase of the average recurrent synaptic weight above the initial mean value (0.25; B, red line) indicate a proper formation of a CA. In the main text, synaptic changes of both types of synapses (feed-forward as well as recurrent) occur with the same time scale $\mu^{ff} = \mu^{rec} = \mu$.

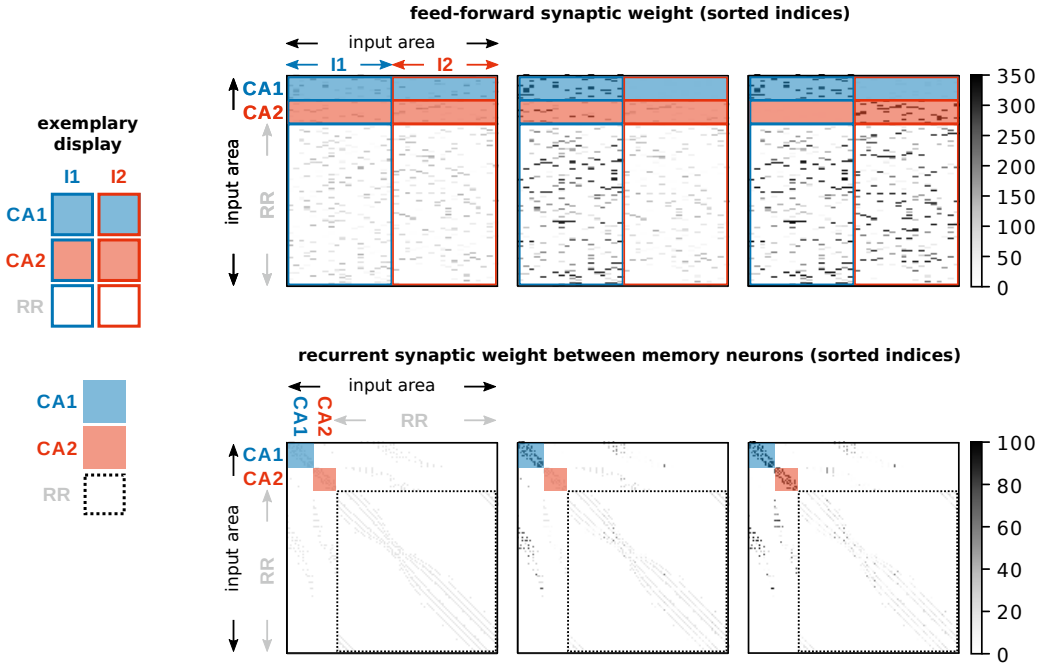

Figure 2–Figure supplement 2. Raw data of the weights of (top) feed-forward and (bottom) recurrent synapses (left) before learning, (middle) after the first learning phase, and (right) after the second learning phase for one system initiation. Indices of neurons in the memory area are sorted according to formed CAs (blue and red shading) while indices of neurons in the input area are sorted according to their input-affilition (blue and red boxes).

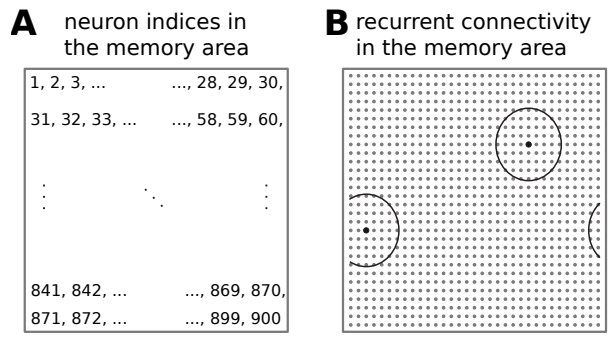

Figure 3–Figure supplement 1. Indexing of neurons and recurrent connectivity in the memory area. (A): The neurons are arranged on a 30x30-grid with indices running from left-top to right-bottom such that each dot in subplots of **Figure 3** and **Figure 4** indicate properties of one neuron. (B): Each neuron is connected to its neighbors, if the position of the neighboring neuron is within a circle of radius 4 (measured in neuronal units; see two circles as examples). Periodic boundary conditions are introduced to avoid boundary and finite-size effects.

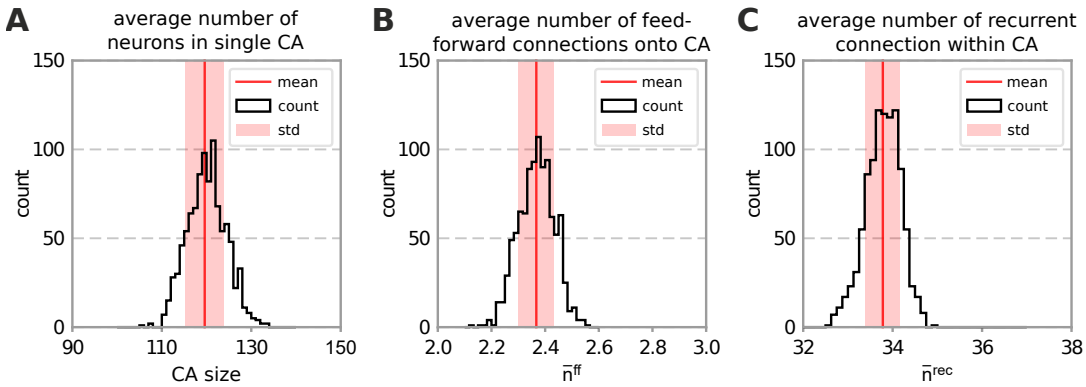

Figure 5–Figure supplement 1. Some parameters of the population model are estimated from the full network model. Each panel shows the histogram of a parameter from 1000 different network initializations. Average values are given as *mean*±*standard deviation*. (A): numbers of neurons within the first formed CA, with average $\bar{N}^{CA} = 120 \pm 4$. (B): average number of feed-forward connections per CA-neuron from the active input population I1 to corresponding CA, with average $\bar{n}^{ff} = 2.37 \pm 0.07$. (C): average number of recurrent connections each neuron within the first formed CA receives from other neurons in the same CA, with average $\bar{n}^{rec} = 33.8 \pm 0.4$).

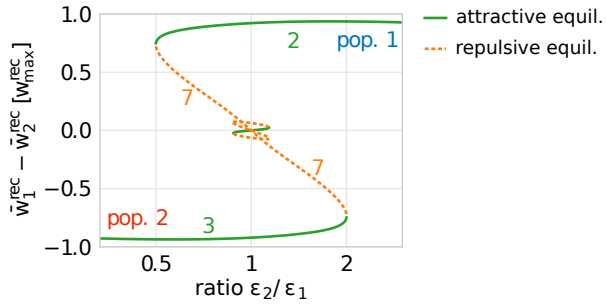

Figure 7–Figure supplement 1. Excitability bifurcation diagram as obtained from the population model. Note that a higher value of ϵ_i means a lower excitability of population i and vice-versa. For a given ratio $\frac{\epsilon_2}{\epsilon_1}$, the used values are $\epsilon_1 = \epsilon_0 \cdot \sqrt{\frac{\epsilon_2}{\epsilon_1}}^{-1}$ and $\epsilon_2 = \epsilon_0 \cdot \sqrt{\frac{\epsilon_2}{\epsilon_1}}$.