Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Structural synaptic plasticity can explain graded amnesia, catastrophic forgetting, and the spacing effect

Andreas Knoblauch, Edgar Körner, Ursula Körner, Friedrich Sommer

2014

Preprint:

This is an accepted article published in PLOS-ONE. The final authenticated version is available online at: https://doi.org/[DOI not available]

OPEN ACCESS Freely available online

PLOS ONE

Structural Synaptic Plasticity Has High Memory Capacity and Can Explain Graded Amnesia, Catastrophic Forgetting, and the Spacing Effect

Andreas Knoblauch[1,2]*, Edgar Körner[2], Ursula Körner[2], Friedrich T. Sommer[3]

1 Engineering Faculty, Albstadt-Sigmaringen University, Albstadt, Germany, **2** Honda Research Institute Europe, Offenbach am Main, Germany, **3** Redwood Center for Theoretical Neuroscience, University of California, Berkeley, California, United States of America

Abstract

Although already William James and, more explicitly, Donald Hebb's theory of cell assemblies have suggested that activity-dependent rewiring of neuronal networks is the substrate of learning and memory, over the last six decades most theoretical work on memory has focused on plasticity of existing synapses in prewired networks. Research in the last decade has emphasized that structural modification of synaptic connectivity is common in the adult brain and tightly correlated with learning and memory. Here we present a parsimonious computational model for learning by structural plasticity. The basic modeling units are "potential synapses" defined as locations in the network where synapses can potentially grow to connect two neurons. This model generalizes well-known previous models for associative learning based on weight plasticity. Therefore, existing theory can be applied to analyze how many memories and how much information structural plasticity can store in a synapse. Surprisingly, we find that structural plasticity largely outperforms weight plasticity and can achieve a much higher storage capacity per synapse. The effect of structural plasticity on the structure of sparsely connected networks is quite intuitive: Structural plasticity increases the "effectual network connectivity", that is, the network wiring that specifically supports storage and recall of the memories. Further, this model of structural plasticity produces gradients of effectual connectivity in the course of learning, thereby explaining various cognitive phenomena including graded amnesia, catastrophic forgetting, and the spacing effect.

Citation: Knoblauch A, Körner E, Körner U, Sommer FT (2014) Structural Synaptic Plasticity Has High Memory Capacity and Can Explain Graded Amnesia, Catastrophic Forgetting, and the Spacing Effect. PLoS ONE 9(5): e96485. doi:10.1371/journal.pone.0096485

Editor: Eleni Vasilaki, University of Sheffield, United Kingdom

Received November 9, 2013; **Accepted** April 8, 2014; **Published** May 23, 2014

Copyright: © 2014 Knoblauch et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: FTS received support from the National Science Foundation through the grants NSF-0855272 and NSF-1219212. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: knoblauch@hs-albsig.de

Introduction

Traditionally, learning and memory are attributed to *weight plasticity*, that is, the modification of the strength of existing synapses according to variants of the Hebb rule [1–5]. Although the theory of weight plasticity has been crucially important in neuroscience and applications of artificial neural networks, it could not easily explain various fundamental memory-related effects in cognitive psychology such as graded amnesia, the prevention of catastrophic forgetting, and the spacing effect.

Another form of synaptic plasticity is *structural plasticity*, that is, the creation and erasure of synapses [6–13]. Originally thought of setting up connectivity during development [14–16] or after injuries [17,18], it has recently been shown to correlate with memory formation and learning in the healthy adult brain [19–23].

Here we introduce and analyze a simple computational model of structural plasticity which exhibits surprisingly high memory capacity and is able to explain the mentioned cognitive effects. A key to understanding the role of structural plasticity in memory has to do with the observation that the brain, even its most densely connected local circuits, is far from being fully connected [24,25]. Thus, for any given network computation, the existing synapses may or may not provide the optimal structure of the network. To assess the match between existing synapses and the synapses required by a computation, we define *effectual connectivity* as the fraction of required synapses that are present in the network. By erasure and creation of synapses, structural plasticity can "migrate" synapses and thereby increase the effectual connectivity for a given network function. By integrating our model with well-known Hopfield- or Willshaw-type neural network models of memory storage and retrieval [16,26,27] we can quantitatively asses the benefits of structural plasticity compared to weight plasticity. In section 0.6 we show that ongoing structural plasticity can strongly increase storage capacity for sparsely connected networks, which is in line with related approaches counting possible synaptic network configurations [28–30] or analyzing storage capacity for structural plasticity during development [15,16]. Moreover, our theory of structural plasticity suggests immediate explanations for various memory phenomena [31–33]. In particular, in section 7 we analyze the role of structural synaptic plasticity in cortico-hippocampal memory replay and consolidation [34,35], preventing catastrophic forgetting in brains [36,37], graded retrograde amnesia following brain lesions [38–40], and the pedagogically relevant spacing effect of learning [41–43].

Concepts and Models

1 Synapse Ensembles and Effectual Connectivity

Common memory theories based on neural associative network models consider only Hebbian-type weight plasticity in networks with fixed structure, thus, neglecting processes involving structural plasticity. Such models predict that the maximal information that can be stored in a given neural network increases in proportion to the number of synaptic connections rather than number of neurons. Therefore, *storage capacity* C is often expressed in terms of stored information per synapse. For example, $C=0.69$ bit per synapse (bps) for networks of binary synapses [26,44], or $C=0.72$ bps for real-valued synaptic weights [45,46]. To judge how many memories can be stored in a network W connecting two neuron populations u and v each comprising n neurons, it is therefore important to know the *anatomical network connectivity*

$$P := \frac{\#\,\text{synaptic connections}}{n^2} \qquad (1)$$

defined as the chance that there is a synaptic connection between two randomly chosen neurons (Fig. 1A).

For memory theories including structural plasticity the situation is different because we can assume that processes including generation of new synapses, consolidation of useful synapses, elimination of useless synapses, and maintenance of anatomical connectivity at a given level P will effectively "migrate" synapses to locations that are most appropriate for storing a particular set of memories. Evidently, anatomical connectivity will then be a bad predictor of storage capacity. Rather storage capacity will depend crucially on the number of locations where a synapse could potentially be generated. Such locations have been called potential synapses [29], where *potential network connectivity*

$$P_{\text{pot}} := \frac{\#\,\text{potential synaptic connections}}{n^2} \qquad (2)$$

is the chance that there is a potential synapse between two neurons.

It is now tempting to apply the old memory theories for weight plasticity as well to structurally plastic networks by simply replacing P by P_{pot}. The underlying argument is that the structurally plastic network with potential connectivity P_{pot} would be functionally equivalent to a structurally static network with anatomical connectivity at the same level P_{pot} because real synapses could "migrate" to any one of the $P_{\text{pot}}n^2$ potential locations. Such an approach would be valid only if the number of required synapses does not exceed the number of actual synapses, Pn^2. However, the question which or how many synapses are actually necessary for storing a particular memory set is usually neglected by theories for fixed networks without structural plasticity. Moreover, from such theories it is impossible to infer any temporal dynamics of structural modifications during memory formation.

We therefore have to introduce another type of connectivity measure that specifies how many synapses have actually been formed at time t between neurons that belong to a particular memory representation. More generally, we can specify the *synapse ensemble* requested to support storage of a memory set $\mathfrak{M}$ by a $n\times n$ matrix $\mathbf{S}$. In the simplest case $\mathbf{S}$ is binary where non-zero matrix entries with $S_{ij}=1$ "tag" potential synapses from neuron i to j that need to be realized or consolidated for storing the memories $\mathfrak{M}$ (Fig. 1B). Then with $\mathbf{W}$ being the $n\times n$ matrix of actual synaptic weights (with $W_{ij}=0$ if there is no real synapse from i to j), we define the *effectual connectivity of memories* $\mathfrak{M}$ as the "overlap" of actual and requested synaptic weights, for example,

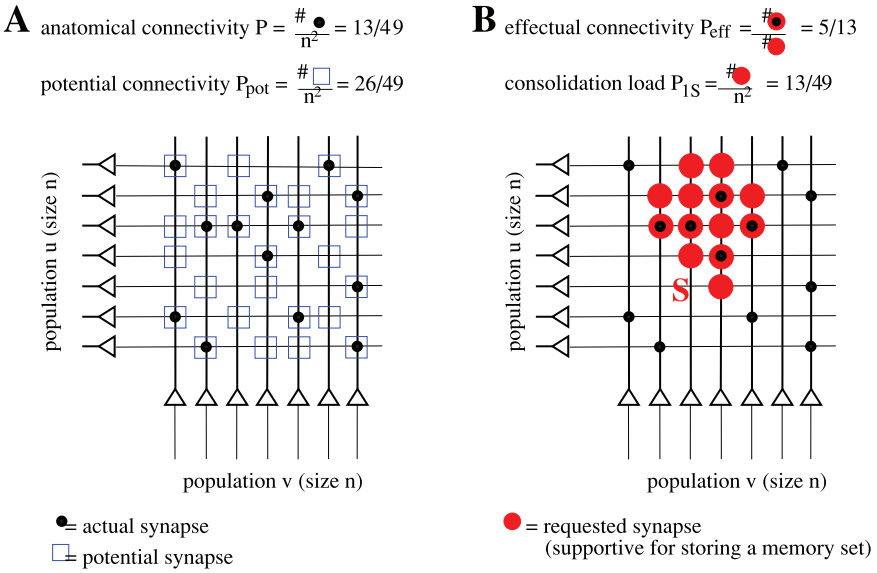

Figure 1. Definitions of network connectivity. Illustration of different connectivity measures for a synaptic network W connecting neuron populations u to v (which may be identical for recurrent networks). **A,** *Anatomical connectivity* P and *potential connectivity* P_{pot} are fractions of neuron pairs (u_i,v_j) connected by an actual (black circles) and potential synapse (blue rectangles), respectively. **B,** The *consolidation signal* S_{ij} specifies the ensemble of neuron pairs that request a synapse ($S_{ij}=1$, red circles) to support storage of a given memory set. The corresponding *effectual connectivity* P_{eff} is then the fraction of neuron pairs requesting a synapse that are already connected by an actual synapse. The *consolidation load* P_{1S} is the fraction of neuron pairs that request a synapse.
doi:10.1371/journal.pone.0096485.g001

$$P_{\text{eff}} := \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}S_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}} \qquad (3)$$

for binary synaptic weights with $W_{ij} \in \{0,1\}$ (Fig. 1B). For real-valued weights one could generalize this definition (e.g., $P_{\text{eff}} := \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \min(|W_{ij}|,|S_{ij}|)}{\sum_{i=1}^{n}\sum_{j=1}^{n} |S_{ij}|}$ where S_{ij} may be either binary or real-valued, specifying the "desired" synaptic weight). It is obviously $0 \le P_{\text{eff}} \le 1$ and, for eq. 3, effectual connectivity P_{eff} corresponds simply to the probability that a requested synapse is actually realized and potentiated ($W_{ij}=1$). We call the matrix $\mathbf{S}$ also *learning signal* or *consolidation signal* because it specifies which synapses should be potentiated or stabilized during memory consolidation. For example, simple Hebbian consolidation signals can be based on the correlations between presynaptic and postsynaptic spike activity (see next section). Such $\mathbf{S}$ could be provided either by repeated bottom-up stimulus presentation or, in the case of episodic memory, by replay from a hippocampus-like short-term memory buffer (Fig. 2B–D). The fraction of non-zero entries in $\mathbf{S}$ is called the *consolidation load* P_{1S}. In larger networks it is typically $P \le P_{\text{eff}} \le P_{\text{pot}}$ if locations of requested synapses $\mathbf{S}$ are uncorrelated to the (initial) locations of potential and actual synapses. Our main hypothesis is that the primary function of structural plasticity is to adapt network structure to the particular memories to be stored. This process corresponds to an increase in effectual connectivity P_{eff} from the level of anatomical connectivity P towards the level of potential connectivity P_{pot} which increases storage capacity per synapse as well as space and energy efficiency of the network [47–49].

2 Model of Structural Plasticity and Consolidation

Figure 2A illustrates a minimal state model for a "potential" synapse. Here a potential synapse ij^{ν} is the possible location of a real synapse connecting neuron i to neuron j, for example, a cortical location where axonal and dendritic branches of neurons i and j are close enough to allow the formation of a novel connection by spine growth and synaptogenesis [29]. As dendrites and axons may closely overlap at multiple locations, in general, there may be multiple potential synapses ($\nu = 1,2,\ldots$) between a neuron pair ij. Our minimal model has only three states: A synapse can be either potential but not yet realized (state π), realized but silent (state and weight 0), or realized and consolidated (state and weight 1). For real synapses, state transitions are modulated by the consolidation signal $s = S_{ij}$.

Then *structural plasticity* means the transition processes between states π and 0 described by transition probabilities $p_g := \text{pr}[\text{state}(t+1)=0 | \text{state}(t)=\pi]$ and $p_{e|s} := \text{pr}[\text{state}(t+1) = \pi | \text{state}(t)=0, S_{ij}=s]$. Similarly, *weight plasticity* means the transitions between states 0 and 1 described by $p_{c|s} := \text{pr}[\text{state}(t+1)=1 | \text{state}(t)=0, S_{ij}=s]$ and $p_{d|s} := \text{pr}[\text{state}(t+1) = 0 | \text{state}(t)=1, S_{ij}=s]$. In accordance with the diagram of Fig. 2A, the evolution of synaptic states can then be described by probabilities $p_{\text{state}}(t)$ that a given potential synapse is in a certain $\text{state} \in \{\pi,0,1\}$ at time step $t = 0,1,2,\ldots$,

$$p_1(t) = (1 - p_{d|s(t)})p_1(t-1) + p_{c|s(t)}p_0(t-1)$$

$$p_0(t) = (1 - p_{c|s(t)} - p_{e|s(t)})p_0(t-1) + p_{d|s(t)}p_1(t-1) + p_g p_\pi(t-1)$$

$$p_\pi(t) = (1-p_g)p_\pi(t-1) + p_{e|s(t)}p_0(t-1) = 1 - p_1(t) - p_0(t)\,, \qquad (4)$$

where the (Hebbian) consolidation signal $s(t) = S_{ij}(t)$ may depend on time. Note that we assume p_g to be independent of s because it is unclear how to provide S_{ij} with high spatial precision ij to not yet realized potential synapses. Instead, p_g may rather be under the control of homeostatic mechanisms to keep the number of synapses or the resulting mean firing rates of a neuron at a desired level [50]. The model could easily be extended towards more biological realism by additional state transitions (e.g., from 1 to π [51]), a cascade of further synaptic states [52], or graded synaptic weights [53,54], but here the focus is on the essential properties of the interplay between structural and weight plasticity.

For the microscopic simulations of individual synapses as displayed in Figs. 4 and 6 we have used the Felix++ simulation

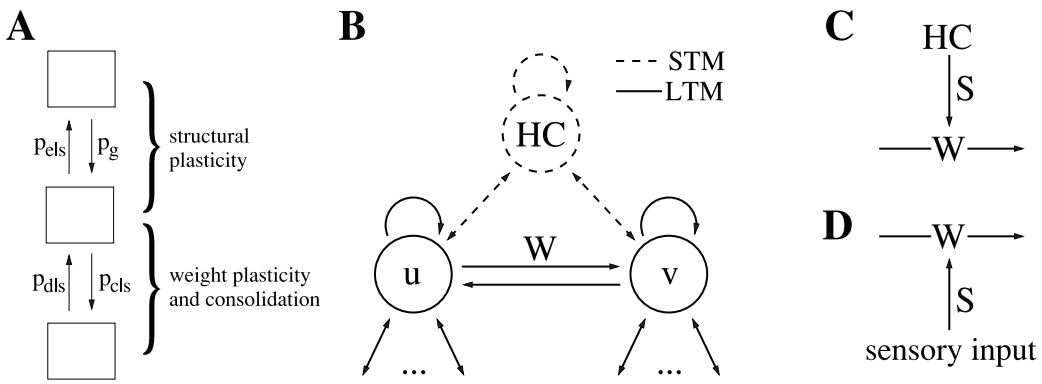

Figure 2. Model of structural plasticity and consolidation. A, State/transition model of a single potential synapse (see text for details). **B**, In the following we consider potential synapses in a network W, for example, connecting two cortical neuron populations u and v. Memories correspond to associations between activity patterns u^{μ} and v^{μ}. We will specifically analyze how well noisy activity patterns $\tilde{u}^{\mu}$ can reactivate the corresponding memories v^{μ} in order to estimate storage capacity. **C, D**: LTM storage (solid) by structural plasticity requires repetitive reactivation of activity patterns in cortical populations u and v to provide an appropriate consolidation signal S to the synapses. This may happen by repeated bottom-up stimulation (**D**) or, for episodic memories, by top-down replay (**C**) from a HC-type STM buffer (dashed). LTM = long-term memory; STM = short-term memory; HC = hippocampus.
doi:10.1371/journal.pone.0096485.g002

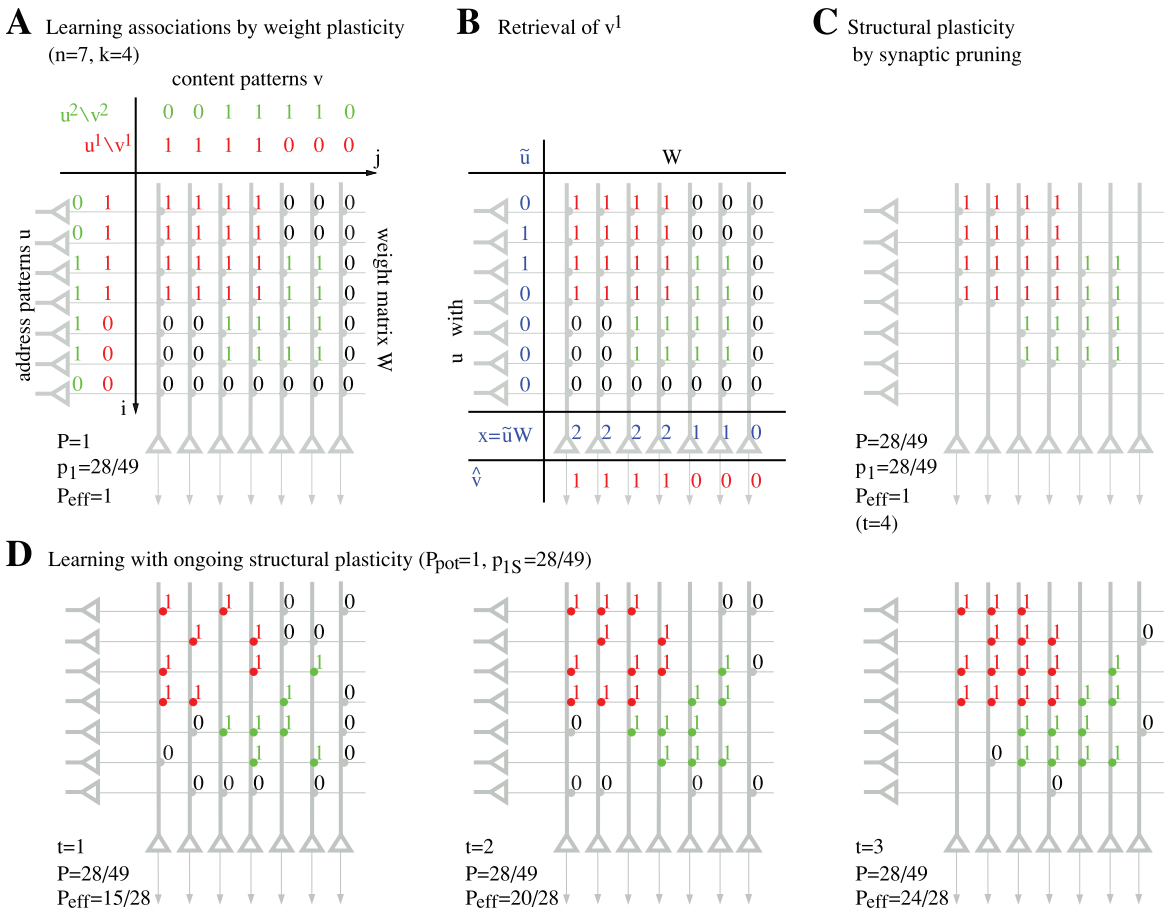

Figure 3. Learning in Willshaw-type associative networks. A, Memory storage by Hebbian weight plasticity (Eq. 5) in a fully connected network ($P=1$). Address patterns u^μ are associated to content patterns v^μ where $\mu=1,\ldots,M$ (here $M=2$). Each memory is represented by a binary activity vector of length $n=7$ having $k=4$ active units (which define the corresponding cell assembly). **B**, One-step retrieval of the first memory from a noisy query pattern $\tilde{u}\approx u^1$ having two of the four active units in u^1 ($\lambda=0.5$). Here $\tilde{u}\approx u^1$ can perfectly reactivate the corresponding memory pattern in population v ($\hat{v}=v^1$) applying a firing threshold $\Theta=\sum_i \tilde{u}_i=2$ on dendritic potentials $x_j=\sum_{i=1}^m W_{ij}\tilde{u}_i$. **C**, As a simple form of structural plasticity, silent synapses can be pruned *after* learning. The resulting network has only 28 (instead of 49) synapses corresponding to a lower anatomical connectivity $P\approx 0.57$, whereas the effectual connectivity is still $P_{\text{eff}}=1$. Thus, pruning does not change network function, but increases stored information per synapse. **D**, Ongoing structural plasticity can similarly increase storage capacity during more realistic learning in networks with low anatomical connectivity (here $P=28/49\approx 0.57$). During each time step $t=1,2,3,4$, Hebbian weight plasticity potentiates and consolidates synapses ij with non-zero consolidation signal $S_{ij}>0$ (which equals W_{ij} of panel A), whereas the remaining silent synapses are eliminated and replaced by new synapses at random locations. Note that the resulting network at $t=4$ is the same as in panel C.
doi:10.1371/journal.pone.0096485.g003

tool [55] to implement large networks with many potential synapses and to simulate network evolution by random sampling of synaptic state variables in discrete time steps. A simple match of the simulation time scale to physiological data can be obtained from the mean lifetime of unconsolidated unrequested synapses: For $p_{e|0}>0$ the mean lifetime is $\sum_{t=0}^{\infty} t p_{e|0}(1-p_{e|0})^{t-1}=\frac{p_{e|0}}{1-p_{e|0}}\sum_{t=1}^{\infty} t(1-p_{e|0})^t=\frac{1}{p_{e|0}}$ simulation steps. This may be compared, for example, to the few days lifetime reported for unstable spines in adult animals [10].

On the network level we use corresponding *macroscopic* variables $P_1^{(s)}$, $P_0^{(s)}$, and $P_\pi^{(s)}$ defined as the fraction of neuron pairs that have a potential synapse in a certain state and receive a certain consolidation signal s. From this we can derive the connectivity variables defined in the previous section, in particular, $P:=\sum_s P_1^{(s)}+P_0^{(s)}$ and $P_{\text{eff}}=P_1^{(1)}/P_{1S}$ for binary s (see Sect. Mathematical Analysis I for details). In most simulations of (adult) memory processes (Figs. 4,3D,6), we have assumed that the rates of synapse generation and elimination are in homeostatic balance to maintain either a constant anatomical network connectivity P or a constant number Pn^2 of actual synapses.

The relation between synapse and network variables is non-trivial in general because there may be multiple potential synapses $v=1,2,\ldots$ per neuron pair ij (see Sect. Mathematical Analysis I.1), for example around 5–10 between two connected neighboring cortical neurons [56–60]. Nevertheless, we argue that even our simple binary model with only a *single synapse* per connected neuron pair bears significant biological relevance because it has been reported that the number of actual synapses per connected neuron pair and also the total synaptic weight is surprisingly similar across neurons (see discussion section; cf. [59,61]). Therefore, we have analyzed this simple model to obtain the results presented below and in Section 6 (see Figs. 4–5). To improve biological realism of our simulation experiments in Section 7 (Fig. 6), we have tested our ideas also with a second model variant that allows *multiple synapses* per neuron pair, where

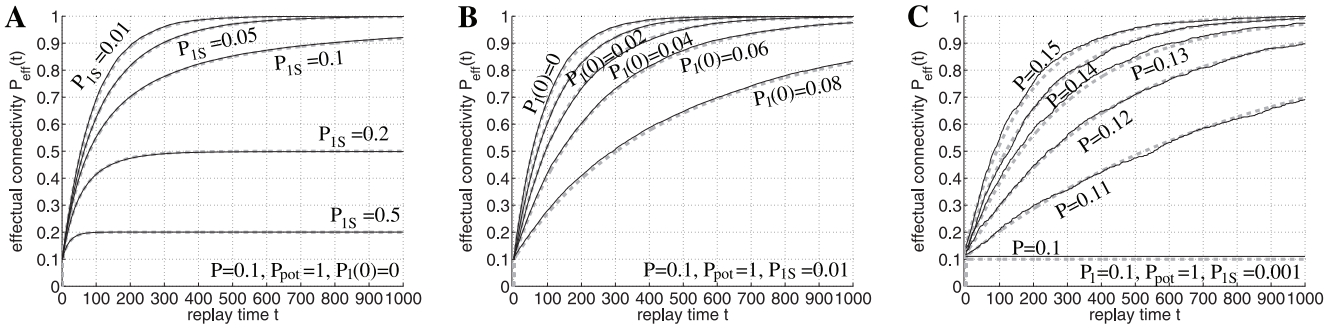

Figure 4. Increase of effectual connectivity during memory consolidation with ongoing structural plasticity. Each curve shows the evolution of effectual connectivity P_{eff} as a function of time t for different parameters P (anatomical connectivity), P_{pot} (potential connectivity), P_{1S} (consolidation load), and $P_1(0)$ (fraction of initially consolidated synapses). Data are from single microscopic network simulations (solid black; cf. Eq. 4; network size $n=1000$) and macroscopic theory (dashed gray; Eq. 11). See Table 1 for further simulation parameters. **A**: $P_{\text{eff}}(t)$ for different consolidation loads P_{1S} and constant $P=0.1$, $P_{\text{pot}}=1$, $P_1(0)=0$. **B**: $P_{\text{eff}}(t)$ for different fractions of initially consolidated synapses $P_1(0)$ and constant $P=0.1$, $P_{\text{pot}}=1$, $P_{1S}=0.01$. **C**: $P_{\text{eff}}(t)$ for different anatomical connectivities P and constant $P_1(0)=0.1$, $P_{\text{pot}}=1$, $P_{1S}=0.001$.
doi:10.1371/journal.pone.0096485.g004

each of the Pn^2 actual synapses of the network can be allocated to one of the $P_{\text{pot}}n^2$ potential locations independently of other synapses. Additional simulations (not shown) have indicated that both model variants yield qualitatively very similar results unless the replay time for a given consolidation signal was very long. Then the second model variant tended to accumulate all available synapses at the locations specified by the consolidation signal such that neuron pairs were connected by a large number of synapses.

3 Models for Memory Storage and Retrieval

The model presented so far is of general relevance for any neural theory of memory, because it is independent of any specific mechanisms for memory storage and retrieval: Any learning and storing mechanisms are only implicitly conveyed by the learning signal S that "tags" potential synapses for later consolidation. Similarly, memory recall is not directly described in the model so far. Rather, our theory describes effectual connectivity P_{eff} which is closely linked to retrieval performance for a given memory set. To explain this link and to allow a more quantitative performance evaluation, the next section instantiates and analyzes our model within a common neural network framework of memory storage and recall.

A particularly simple memory model based on Hebbian learning of binary synapses is the *Steinbuch* or *Willshaw model* [26,44,62]. In the general *hetero-associative* setup (Fig. 3A), memories correspond to binary spike activity vectors u^μ and v^μ stored in a synaptic connection W linking two neuron populations u and v. By choosing the *auto-associative* setup with identical u and v, the Willshaw model can be applied as well to model memory processes in local recurrent connections (cf. Fig. 2B). The average number k of one-entries in an activity vector is called *pattern activity* and corresponds to the mean size of local Hebbian cell assemblies in populations u and v. After *storing* a set of M memory associations in a network without structural plasticity, the weight of an actual synapse connecting neuron u_i to neuron v_j is

$$W_{ij}=\min\left(1,\sum_{\mu=1}^{M}u_i^\mu\cdot v_j^\mu\right)\in\{0,1\}. \qquad (5)$$

Note that a synapse in the Willshaw model is actually a special case of our model of a potential synapse because Eq. 5 instantiates Eq. 4 for $S_{ij}=W_{ij}$, $P=P_{\text{pot}}=P_{\text{eff}}$, $p_{c|1}=1$, and $p_g=p_{e|s}=p_{c|0}=p_{d|s}=0$.

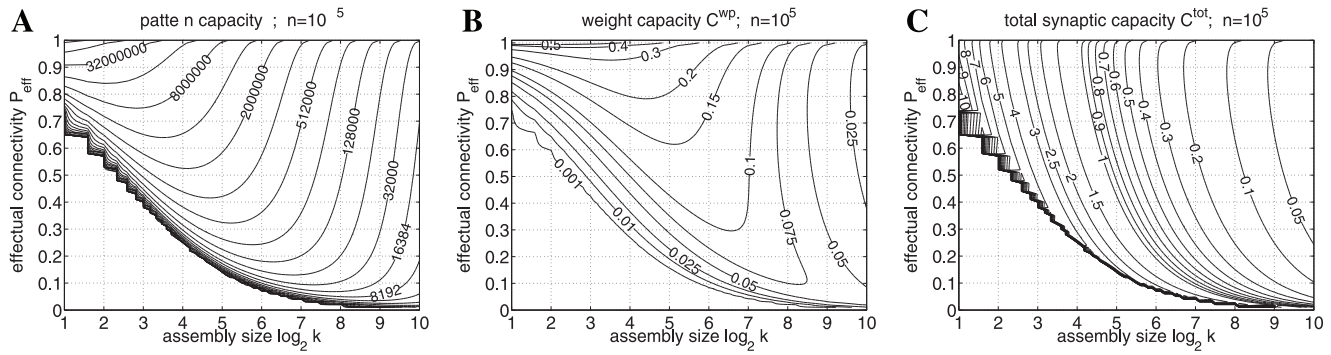

Figure 5. Storage capacities for a finite Willshaw network having the size of a cortical macrocolumn ($n=100000$). A, Contour plot of pattern capacity M_ϵ (number of stored memories) as a function of assembly size k (number of active units in a memory vector) and effectual network connectivity P_{eff} assuming output noise level $\epsilon=0.01$ and noise-free input patterns ($\lambda=1$, $\kappa=0$). **B,** Weight capacity C_ϵ^{wp} for the same setting as in panel **A**. **C,** Total storage capacity C_ϵ^{tot} including structural plasticity for the same setting as in **A**. Note that even modest increases of P_{eff} can strongly increase storage capacity, in particular for sparse neural activity (small k) [82]. All data computed from Gaussian approximation of dendritic potential distributions (see appendix II. 2).
doi:10.1371/journal.pone.0096485.g005

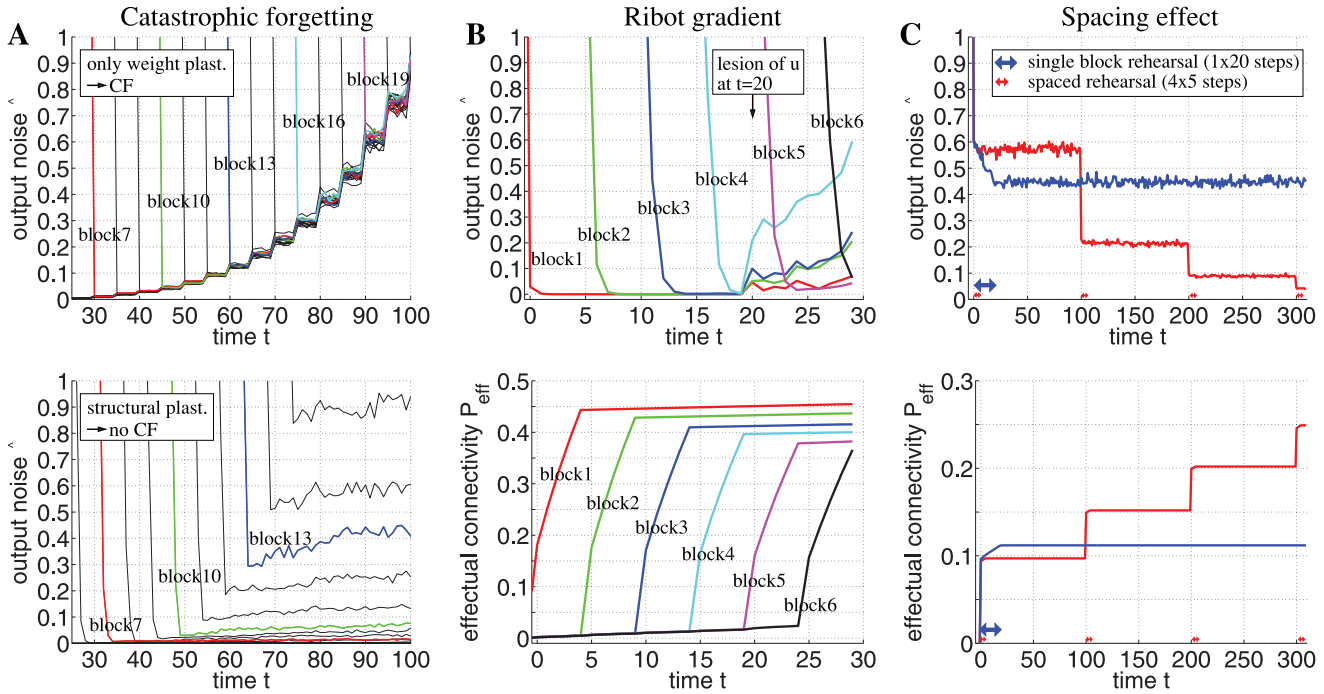

Figure 6. Simulation of catastrophic forgetting, Ribot gradients, and the spacing effect. A, Networks without structural plasticity suffer from catastrophic forgetting (top), but networks with structural plasticity do not (bottom). Plots show output noise $\hat{\epsilon}$ over time t simulating networks of size $n=1000$ and activity $k=50$ storing 25 memory blocks one after the other (only the interesting part between storage of blocks 6 and 21 are visible). Each curve (with a distinct color) corresponds to $\hat{\epsilon}$ for noisy test patterns of a particular memory block with $c=45$ correct and $f=5$ false active units. The steep descent of each curve corresponds to the time when the Hippocampus started to replay the corresponding memory block for 5 time steps. **B**, Networks employing structural plasticity show Ribot gradients after a cortical lesion (top) due to gradients in effectual connectivity (bottom). The lesion was simulated by deactivating half of the neurons in population u at time $t=20$. **C**, Networks employing structural plasticity reproduce the spacing effect of learning. In the first simulation (blue) novel memories were rehearsed once for 20 time steps (blue arrow at $t=0-19$). In a second simulation (red) the same total rehearsal time was "spaced" or distributed to four brief intervals of five steps each (red arrows at $t=0-4$, $t=100-104$, $t=200-204$, and $t=300-304$). Here the network achieves a higher effectual connectivity P_{eff} (bottom) and less retrieval noise ϵ (top). See Sections 2, 3 and Table 1 for further details and simulation parameters.
doi:10.1371/journal.pone.0096485.g006

Memory retrieval means the re-activation of a previously stored content pattern $\hat{\boldsymbol{v}}^{\mu}$ in neuron population v following the activation of a (noisy) address pattern $\tilde{\boldsymbol{u}}^{\mu}$ in population u. The simplest retrieval procedure is "one-step retrieval" with adaptive threshold control [63]. Specifically, an input pattern $\tilde{\boldsymbol{u}}$ is propagated synchronously from population u to population v as illustrated in Fig. 3B. Then dendritic potentials of the neurons in population v are given by simple vector-matrix-multiplication, $x := \tilde{\boldsymbol{u}}^T W$, and the retrieval output $\hat{\boldsymbol{v}}$ is obtained from x by applying a vector of spike thresholds $\boldsymbol{\Theta}$,

$$\hat{v}_j = \begin{cases} 1, & x_j = \left(\sum_{i=1}^{m} \tilde{u}_i W_{ij}\right) \geq \Theta_j \\ 0, & \text{otherwise} \end{cases}, \qquad (6)$$

where $\boldsymbol{\Theta}$ is chosen to obtain close to k active units in $\hat{\boldsymbol{v}}_j$. We can then evaluate retrieval quality by estimating the *output noise* level

$$\hat{\epsilon} := \frac{(n-k)q_{01} + kq_{10}}{k} \qquad (7)$$

defined as the mean Hamming distance $d_H(\hat{\mathbf{v}}, \mathbf{v}^{\mu}) := \sum_{j=1}^{n} |\hat{v}_j - \hat{v}_j^{\mu}|$ between retrieval output $\hat{\mathbf{v}}$ and the original memory $\mathbf{v}^{\mu}$ normalized to the cell assembly size k. Here $q_{01} := \text{pr}[\hat{v}_j = 1 | v_j^{\mu} = 0]$ and $q_{10} := \text{pr}[\hat{v}_j = 0 | v_j^{\mu} = 1]$ are component error probabilities. Similarly, we can define input noise $\tilde{\epsilon} := ((n-k)p_{01} + kp_{10})/k$ as the normalized Hamming distance between input pattern $\tilde{\boldsymbol{u}}$ and the original address memory $\boldsymbol{u}^{\mu}$. We will also express input noise in terms of parameters $\lambda := 1 - p_{10}$ (completeness) and $\kappa := p_{01}(n-k)/k$ (add noise).

We have used one-step retrieval for some of our experiments (Fig. 5) because it is most easy to analyze, for example, for estimating the memory capacity of a single network (see below). However, for the investigation of memory phenomena, there exist more realistic retrieval methods that are based on spiking neurons and iterative (gamma range) oscillatory activity propagation [64,65]. As such models are computationally very demanding, in particular, when simulating longer time intervals in the range of months to years, it is more favorable to use simple iterative extensions of one-step retrieval [27,63,66,67]) that can still mimic many relevant properties of the realistic models.

In particular, iterative retrieval avoids the most serious limitation of one-step retrieval, that is, the lack of a sufficient attractor behavior: High output noise after one-step retrieval does not exclude perfect retrieval after iterated retrieval steps. In fact, as long as the output noise level after the first step is smaller than the input noise level, the iterative retrieval procedure is likely to reduce output noise to zero in subsequent retrieval steps. As a consequence, for *individual* memories, the relation between input and output noise will be much steeper if using the iterative models:

A memory pattern can be retrieved either perfectly or the number of component errors is very high. Still, one-step retrieval is useful by providing lower bounds (because of its suboptimality) and upper bounds (assuming zero input noise) of the true storage capacity.

For our long-term simulations of memory phenomena (Fig. 6) we have therefore extended the Willshaw model in two ways: First, similar as illustrated by Fig. 2B, we have included also Willshaw-type auto-associative connections in addition to the hetero-associative link from u to v in order to account for the rich recurrent connectivity of cortex and to enable iterative refinement of retrieval outputs. Second, we have implemented an *iterative retrieval* procedure as follows (cf. [63]): In an initial step, the input pattern $\tilde{\boldsymbol{u}}$ is propagated through the hetero-associative connections from u to population v, in which the k neurons with the largest dendritic potentials become active, resulting in a preliminary retrieval result $\hat{\boldsymbol{v}}^{(0)}$. In similar further steps, this preliminary result was then iteratively propagated through the auto-associative network of population v yielding refined retrieval outputs $\hat{\boldsymbol{v}}^{(i)}$ for $i=1,2,\ldots$ (where all recurrent connections to u were inactivated). Typically, a small number of iterations was sufficient to obtain stable outputs. For evaluation of output noise $\hat{\epsilon}$ we used the activity pattern $\hat{\boldsymbol{v}}^{(3)}$ after 3 iterations and compared it to the original memory pattern v^{μ} to estimate component error probabilities q_{01} and q_{10} (see Eq. 7).

For the simulations involving structurally plastic networks and long-term consolidation (Fig. 6) we have divided the overall memory set into multiple blocks $\beta=1,2,\ldots$ each containing several individual memory patterns. Each memory block defines a consolidation signal S^{β} that is identical to the Willshaw matrix (Eq.5) obtained from the corresponding subset of memories. Thus, memory blocks are consolidated one after the other, each for a certain number of simulation steps, by reactivating the corresponding activity patterns in populations u and v to mimic either hippocampal short-term storage and top-down replay (Fig. 2B,C) or repeated bottom-up rehearsal of the corresponding memories (Fig. 2B,D). Fig. 6 shows simulations with structural plasticity in the connection W linking u to v. By contrast, the recurrent connections within u and v were prewired without any structural plasticity and auto-associatively stored the individual patterns u^{μ} and v^{μ} with a fixed connectivity ($P=1$ for Fig. 6A, upper panel; $P=0.2$ for Fig. 6A, lower panel; $P=0.1$ for Fig. 6B,C). Table 1 summarizes the remaining simulation parameters.

4 Definitions of Storage Capacity

The *storage capacity* is the amount of information (in bits) that a neural network can store (and retrieve) per synapse. There are two contributions to the total capacity C^{tot} of a synapse,

$$C^{\mathrm{tot}} \leq C^{\mathrm{wp}} + C^{\mathrm{sp}}. \tag{8}$$

First, the *weight capacity* C^{wp} is the information stored by modification of the synaptic weight for a fixed network structure. (a more general definition could as well include any other modifications of synaptic *state* variables such as synaptic transmission delay). Second, the *structural capacity* C^{sp} is the information stored by selecting an appropriate target location for a synapse with fixed weight. We would like to evaluate storage capacity at a limited small output noise level ϵ (see Eq. 7): The "stored information" can then be computed from the *pattern capacity* M_{ϵ} defined as the maximum number of memories that can be stored at noise level ϵ, whereas the *weight capacity* $C^{\mathrm{wp}}_{\epsilon}$ is the stored information normalized to the number of synapses in a static network (no structural plasticity) with connectivity P,

$$M_{\epsilon} := \max\{M : \hat{\epsilon} \leq \epsilon\} \tag{9}$$

$$C^{\mathrm{wp}}_{\epsilon} := \frac{M_{\epsilon} T(k/n, q_{01}, q_{10})}{Pn} \tag{10}$$

where $T(q; q_{01}, q_{10})$ is the transinformation (or mutual information) when transmitting independent memory components v^{μ}_j (with $q := \mathrm{pr}[v^{\mu}_j = 1] = k/n$) over a binary channel (with transition probabilities q_{01} and q_{10} as in Eq. 7) and receiving $\hat{v}_j$ (for details see appendix A in [16]). In general, it is difficult to disentangle the two contributions C^{wp} and C^{sp}. Thus, in the results section we will compute the total capacity C^{tot} for some special cases.

Results

5 Structural Plasticity Increases Effectual Connectivity

In the previous section we have introduced effectual connectivity P_{eff} as a measure of how well a given set of memories is stored in a synaptic network. Without any structural changes of the network, P_{eff} will obviously remain constant, for example, at the level of anatomical connectivity P for novel memories that do not correlate with the current network structure. It is therefore more interesting to investigate the dynamics of P_{eff} during phases of ongoing structural plasticity. For consistency with experimental observations it seems most reasonable to focus on a parameter range where structural plasticity operates on a slower time scale than Hebbian-type weight plasticity ($p_{e|0} \ll p_{c|1}$), but on a faster time scale than the lifetime of stable consolidated synapses ($p_{e|0} \gg p_{d|1}$).

It is indeed possible to analyze our model in such a parameter regime: In Sect. Mathematical Analysis I.2 we compute the temporal evolution of effectual connectivity during consolidation of a novel memory set under the following simplifying assumptions: 1) Large networks with $n \gg 1$ such that all macroscopic variables $P^{(s)}_{\mathrm{state}}$ are close to their means; 2) at most a single synapse per neuron pair; 3) binary consolidation signal $s \in \{0,1\}$; 4) new memories specified by S are independent of initial network structure and any old memories; 5) immediate consolidation with $p_{c|s} = s$; 6) $p_{d|1} = p_{e|1} = 0$; 7) p_g and $p_{e|0}$ in homeostatic balance such that $P(t)$ is constant. Then effectual connectivity for a new set of memories increases from $P_{\mathrm{eff}}(0) = P_1(0)$ before any learning starts to

$$\begin{aligned} P_{\mathrm{eff}}(t) &\approx P_{\mathrm{pot}} - (P_{\mathrm{pot}} - P)/\Pi_{t_1=1}^{t-1} \\ &\quad \left(1 + p_{e|0} \frac{P - (1-P_{1S})(1-p_{d|0})^{t_1} P_1(0) - P_{1S} P_{\mathrm{eff}}(t_1)}{P_{\mathrm{pot}} - P}\right) \\ &\approx P_{\mathrm{pot}} - (P_{\mathrm{pot}} - P) e^{-(t-1)\ln\left(1 + p_{e|0}\frac{P - P_1(0)}{P_{\mathrm{pot}} - P}\right)} \end{aligned} \tag{11}$$

assuming that S is provided at each time step $t=1,2,\ldots$ (e.g., by memory replay) and $P_1(0) := P^{(0)}_1 + P^{(1)}_1$ is the fraction of initially consolidated synapses (corresponding to old memories). The second approximation additionally presumes $P_{1S} \ll 1$ and

Table 1. Simulation parameters.

Figure No.	synapse model	P	n	k	λ	κ	#blocks	M/block	$p_{e\|0}$
4A	single	0.1	1000	10	-	-	1	100–6931	0.1
4B	single	0.1	1000	10	-	-	1	100	0.1
4C	single	0.1−0.5	1000	10	-	-	1	100	0.1
6A, upper	multi	1	1000	50	0.9	0.1	25	12	0
6A, lower	multi	0.2	1000	50	0.9	0.1	25	4	1
6B	multi	0.1	1000	50	0.9	0.1	6	4	1
6C	multi	0.1	1000	50	0.9	0.1	1	20	0.01

Model parameters for the simulations shown in Figs. 4 and 6. All simulations used $P_{\text{pot}}=1$, $p_{c|s}=s$ and $p_{e|1}=p_{d|s}=0$ for $s\in\{0,1\}$ and a homeostatic setting maintaining a constant anatomical connectivity P where the number of eliminated synapses per time step was equal to the number of generated synapses, i.e., $P_0(i)p_{d|0}=(P_{\text{pot}}-P)p_g$, where P_{pot} is the potential connectivity and P_0 is the fraction of unconsolidated silent synapses at time t. Synapse model: "single" = at most a single synapse per neuron pair; "multi" = multiple synapses per neuron pair are possible. P is anatomical connectivity of the synaptic projection from population u to v. n is population size of u and v. k is cell assembly size (i.e., number of one-entries in memory patterns). λ and κ denote completeness and add noise in input patterns $\tilde{u}$ (i.e., $\tilde{u}$ has $c:=\lambda k$ correct and $f:=\kappa k$ false one-entries). # blocks is number of memory blocks that are subsequentially stored. M/block is the number of memories per memory block. $p_{e|s}$, $p_{c|s}$, $p_{d|s}$ are the state transition probabilities of the (potential) synapses.

doi:10.1371/journal.pone.0096485.t001

$p_{d|0}\ll 1$. Thus, convergence of P_{eff} towards P_{pot} requires $P_{1S}\le P/P_{\text{pot}}$ (for $p_{d|0}>0$) or $P_{1S}\le(P-P_1(0))/(P_{\text{pot}}-P_1(0))$ (for $p_{d|0}=0$). Also note that during the first consolidation step there is a quick increase from $P_{\text{eff}}(0)=P_1(0)$ to $P_{\text{eff}}(1)=P$ followed by a much slower increase towards P_{pot} in the subsequent steps. Section 7.1 relates this behavior to the spacing effect as a possible explanation why several brief learning sessions are generally more effective than a single long session.

Figure 4 shows that the approximations accurately predict microscopic model simulations. Consolidation becomes slower for larger consolidation loads P_{1S} which limits maximal storage capacity (panel A; see Section 6). Similarly, consolidation becomes slower for increasing fractions $P_1(0)$ of initially consolidated synapses (panel B). As P_1 will correlate with the number of previously consolidated memories and, thus, with age, this implies that memory consolidation should be faster in young compared to old subjects, even if the anatomical connectivity P would be constant over lifetime. Moreover, the corresponding gradients in P_{eff} resulting after a fixed number of consolidation steps can be related to gradients in memory performance in graded retrograde amnesia (Section 7.2) and the absence of catastrophic forgetting (Section 7.1). Finally, panel C shows that even slight increases in anatomical connectivity (as reported after learning new concepts or tasks [68]; cf. Fig. 7) can strongly speed-up memory consolidation if a large proportion of synapses are in the consolidated state (as expected for adult networks after synaptic pruning [14,15]).

Our analysis and further simulations (data not shown) reveal that the described increase of P_{eff} is very stable and occurs for virtual any plausible configuration of model parameters. Before we discuss the mentioned memory phenomena in more detail, the following shows that, by increasing P_{eff}, structural plasticity can store much more information per synapse than Hebbian-type weight plasticity.

6 How Much Information can a Synapse Store?

It is a well-known result of information theory [69] that optimally coding an entity taken at random from a set of $\mathfrak{n}$ different entities takes $\mathrm{ld}\,\mathfrak{n}$ bits of information [69] (where $\mathrm{ld}:=\log_2$). From this we can derive simple upper bounds for the maximal information that a synapse can store by counting the number of possible synaptic states, i.e. the number of possible weights and locations, that can be realized by weight plasticity and structural plasticity, respectively. The resulting upper bounds for weight capacity C^{wp} and structural capacity C^{sp} are

$$C^{\text{wp}}\le \mathrm{ld}\,N \qquad \text{and} \qquad C^{\text{sp}}\le \mathrm{ld}\,n, \tag{12}$$

assuming that weight plasticity can choose one out of N possible discrete weights for an individual synapse, and structural plasticity can choose between n targets where to grow a novel synapse. These bounds could trivially be reached by an ideal observer that has direct access to synaptic attributes (i.e., weights and locations). However, here we are rather interested in how much information a synaptic network can store *and* safely retrieve employing biologically plausible mechanisms. In particular, we have to measure the amount of retrieved information from plausible neural output variables such as spikes or mean firing rates. For this it is necessary to link our theory to concrete neural network models of memory storage and retrieval, such as Willshaw and Hopfield-type models ([26,27,45,70,71]; see section 3).

Our theory yields the surprising result that the weight capacity C^{wp} in the brain might actually be negligible compared to structural capacity C^{wp}. First, it is well understood that weight capacity of biologically plausible memory models is limited by hard theoretical bounds suggesting $C^{wp}<0.72$ bit per synapse even for an infinite computing precision with $N\to\infty$ [27,45,46,72,73]. Second, due to noisy transmission characteristics and various adaptation mechanisms, real synapses are likely to have a rather small number of functionally distinctive states, perhaps N being on the order of ten or even binary [74–76]. Third, unlike N, the number of potential targets n may actually be very large in the brain: For example, for a cortical neuron n is on the order of 10^5 corresponding to the number of neighboring cells within the same macrocolumn [24], and the number of targets n may be even much larger because each neuron may have a large number of functionally distinct dendritic compartments [28]. Fourth, it has been recently shown that the upper bound of structural capacity can be tightly reached for synaptic pruning following learning in completely connected networks [16,53].

Before generalizing these results to ongoing structural plasticity in sparsely connected networks, let us first re-analyze the classical Willshaw model (without structural plasticity) as illustrated in Fig. 3A,B. There, synaptic weight plasticity follows a simple binary Hebbian rule (Eq. 5). Due to $p_{d|s}=0$ (cf. Eq. 4) the fraction of consolidated synapses p_1 increases monotonically with M until it reaches a maximal value $p_{1\epsilon}$ beyond which the output noise $\hat{\epsilon}$ exceeds the tolerable level ϵ. Some theory presented in Sect. Mathematical Analysis II.1 shows that the corresponding pattern capacity M_ϵ crucially depends on $p_{1\epsilon}$: For networks of size n, randomly generated cell assemblies of size k, and input noise with $\lambda\in(0,1]$ and $\kappa=0$, it is (see text below Eq. 28 in Sect. Mathematical Analysis II.1)

$$p_{1\epsilon}(P_{\text{eff}})\approx \max\left(0,\frac{\left(\frac{\epsilon k}{n-k}\right)^{\frac{1}{\lambda k}}-(1-P_{\text{eff}})}{P_{\text{eff}}}\right)\in(0,1)$$

$$M_\epsilon(P_{\text{eff}})\approx \lambda\frac{n^2}{k\ln\frac{n-k}{\epsilon k}}\ln(1-P_{\text{eff}}(1-p_{1\epsilon}))\ln(1-p_{1\epsilon})\eta \qquad (13)$$

where factor $\eta\approx(1+(\ln\epsilon)/\ln(k/n))^{-1}$ comes close to one for large networks. Multiplication by the stored information per memory and dividing by the number of synapses gives the well known weight capacity of the Willshaw model (see Sect. Mathematical Analysis II. 1),

$$C^{wp}_\epsilon(P_{\text{eff}})\approx\lambda\frac{\text{ld}(1-P_{\text{eff}}(1-p_{1\epsilon}))\ln(1-p_{1\epsilon})}{P_{\text{eff}}}\eta\le 0.69 \qquad (14)$$

where the upper bound $C^{wp}=0.69$ bps can be reached for large networks, $P_{\text{eff}}=1$, $p_{1\epsilon}=0.5$, sparse activity $k\sim\log n$, and zero input noise with $\lambda=1$.

In previous works on structural plasticity we have focused on *synaptic pruning* of silent synapses after learning all memories in a *fully connected* network (Fig. 3C). Here we extend these results to networks with incomplete ("diluted") connectivity and ongoing structural plasticity. Let us first consider synaptic pruning which has been described as one of three phases during brain development (e.g., in humans, synaptic density increases until age of 2–3 years, then remains stable until 5 y, then decreases until puberty and remains relatively stable during adulthood; cf. [14,51,77]; see also Fig. 7):

1. Synaptic overgrowth: The synaptic generation rate is much larger than the elimination rate, $p_g\gg p_{e|s}$, such that anatomical connectivity P can come close to potential connectivity P_{pot}.
2. Critical consolidation phase: Weight plasticity potentiates and consolidates useful synapses that support memory contents specified by the consolidation signal S, e.g., $p_{c|s}=s$, $p_{d|s}=1-s$.
3. Synaptic pruning: Useless synapses are eliminated, e.g., $p_{e|0}\gg p_g$ (cf. Fig. 3C).

Because only a fraction p_1 of the synapses survives phase three, the total storage capacity at maximal M_ϵ (where $p_1=p_{1\epsilon}$) is obtained from renormalizing Eq. 14,

$$C^{\text{tot}}_\epsilon := C^{wp}_\epsilon/p_{1\epsilon} \qquad (15)$$

Using $p_{1\epsilon}$ from Eq. 13 reveals that $C^{\text{tot}}\sim\text{ld}\,n$ for sufficiently small cell assembly sizes k (see Sect. Mathematical Analysis II. 1). Thus, the Willshaw model with structural plasticity comes close to

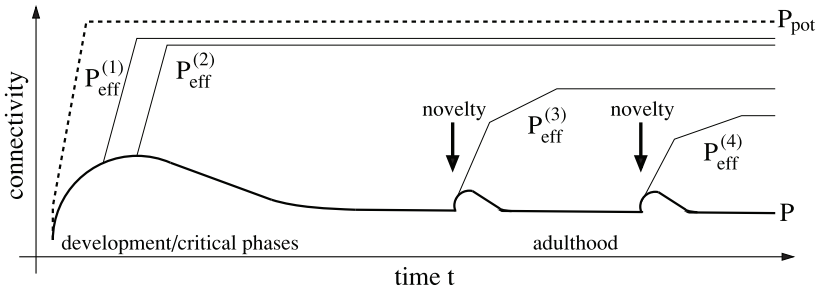

Figure 7. Sketch of network connectivity reflecting lifelong structural plasticity. During development anatomical connectivity P (thick solid) quickly increases reaching a peak level (around 2–3y in humans), where the initial increase is followed by a short period of stable connectivity (until age 5y in humans), a phase of significant decrease of connectivity until puberty, and finally a phase of stable connectivity during adulthood [14,51,77]. Recent experiments suggest a temporary novelty-driven (thick arrows) increase of connectivity during adulthood [23,68,116]. Our model of structural plasticity predicts that learning is fastest for high levels of anatomical connectivity and structural plasticity. Thus, memories acquired during early phases can reach higher levels of effectual connectivity ($P^{(1)}_{\text{eff}}$,$P^{(2)}_{\text{eff}}$; thin solid lines) compared to memories acquired during later phases ($P^{(3)}_{\text{eff}}$,$P^{(4)}_{\text{eff}}$). The resulting gradients in effectual connectivity can explain various memory phenomena (see Section 7 for details).
doi:10.1371/journal.pone.0096485.g007

the information-theoretic capacity bound (Eq. 12). We have shown elsewhere that $C^{\mathrm{tot}} = \mathrm{ld}\ n$ can be reached tightly with much weaker assumptions on cell assembly sizes and effectual connectivity by inhibitory implementations of the Willshaw model [46,78] and both excitatory and inhibitory implementations of Bayesian networks with discrete synaptic weights [53,54,79].

Unlike in development, during adulthood anatomical connectivity is stable. This means that ongoing generation and elimination of synapses must be in homeostatic balance such that the total number of synaptic connections remains approximately constant over time [14,80,81]. In the following we show that ongoing structural plasticity during adulthood can reach the same high storage capacity as during development, although this process may require significantly more time. The basic idea is that the three developmental processing phases (synaptic generation, consolidation, and elimination) run in parallel during each time step t. For example, by choosing the synapse parameters.

$$p_{c|s} = s, \quad p_{d|1} = 0, \quad p_{d|0} > 0, \quad \text{and} \quad P_\pi(t)p_g(t) = P_0(t)p_{e|s}(t) \quad (16)$$

the anatomical connectivity P remains constant and, in essence, all actual synapses "migrate" to the locations ij specified by the consolidation signal S_{ij} (cf. Fig. 3D). *IF* S specifies *all* memories to be stored, S is applied during each time step, and the consolidation load P_{1S} is sufficiently large such that $P \le P_{1S}P_{\mathrm{pot}}$, *THEN* memories will be stored at effectual connectivity $P_{\mathrm{eff}} = P/P_{1S} \le P_{\mathrm{pot}}$, there will be no silent synapses left, and the resulting total capacity C^{tot} is given by Eq. 15. In particular, for $P = P_{1S}P_{\mathrm{pot}}$ the resulting network will be identical as for developmental learning described before (see Fig. 3D and compare to Fig. 3C). This shows that also adult learning in structurally plastic networks with constant low anatomical connectivity can reach the information theoretic bound $C^{\mathrm{tot}} = \mathrm{ld}\ n$ (see Eq. 12).

In the following we apply our theory to networks with biologically relevant parameters. For example, a typical network size may correspond to a cortical macrocolumn of size 1 mm^3 containing about $n = 10^5$ neurons and relatively dense recurrent connections with an anatomical connectivity of about $P = 0.1$ [24,25]. Then we can estimate potential connectivity P_{pot} from experimental measurements of the *filling fraction* P/P_{pot} defined as the fraction of potential synapses that is actually realized (i.e., in state 0 or state 1). For typical $P/P_{\mathrm{pot}} \approx 0.2$ [29], structural plasticity of dendritic spines alone may account already for $P_{\mathrm{pot}} \approx 0.5$ within a neocortical macrocolumn. The corresponding storage capacities are depicted in Figure 5. Note that without structural plasticity ($P_{\mathrm{eff}} = P = 0.1$) the storage capacity remains tiny, e.g., $C^{\mathrm{wp}} \ll 0.1$ for $P = 0.1$. In particular, sparse activity patterns [82] cannot be stored at a low connectivity, e.g., $k \le 64$ requires $P > 0.1$ to stabilize even a single memory pattern.

By contrast, networks employing structural plasticity with potential connectivity $P_{\mathrm{pot}} > 0.1$ can have a large total capacity $C^{\mathrm{tot}} \gg 1$. Interestingly, C^{tot} increases with decreasing connectivity. Thus, even slight increases of effectual connectivity towards $P_{\mathrm{pot}} \approx 0.5$ can strongly increase number of stored memories (M) and even maximize stored information per synapse (C^{tot}). Note that an increase in P_{eff} during consolidation would also allow a simultaneous decrease of activity k to maximize capacity. This means that consolidation involving structural plasticity and sparsification will move the "working point" from the lower right towards the upper left in the contour plots of Fig. 5. Thus, by emulating high effectual connectivity, structural plasticity may also support the sparsification of memory representations [82–85] and stabilize small cell assemblies that would appear unstable for a fixed low connectivity [86,87].

The following sections show that structural plasticity, in addition to increasing storage capacity, can explain several well known memory phenomena in the brain much better than previous theories.

7 Relevance of Structural Plasticity for Memory Phenomena

7.1 Absence of Catastrophic Forgetting. Artificial neural networks such as multi-layer-perceptrons are well known to suffer from what was called catastrophic forgetting (CF) or the stability-plasticity dilemma [36,88–91]. It is the problem that optimizing synaptic weights to store a set of new memories will deteriorate or even destroy previous memories. Freezing synaptic weights can avoid CF, but it also hampers the ability to learn new memories.

Another form of CF has been described for Hopfield-type network models of associative memory [92]. Here CF means that a neural network with fixed structure can almost perfectly store and retrieve memories until the maximal pattern capacity M_ϵ is reached. However, exceeding M_ϵ even by a few additional patterns can destroy the ability to retrieve any of the memories. The same problem occurs when increasing the number of stored memory patterns in the Willshaw-type binary learning models (Fig. 3A, B), even before the point where *all* synapses are uniformly potentiated and therefore have lost specific information about the memory patterns.

CF poses problems for technical applications, but also for modeling memory processes because it does not normally occur in our brains. It has been argued that the capacity of the brain might just be too large for running into CF during a normal lifetime. In addition, several alternative solutions have been suggested. For example, many previous approaches suggested to have an additional hidden neural layer (e.g., between populations u and v) in which a new node is allocated for each new input that deviates significantly from previously stored items. The underlying idea is that in a modular organization, separate subnetworks (comprising different subsets of neuron in the intermediate layer) could be trained independently to represent different memories or categories. Such approaches include ART-type architectures [90], emergent category-specific modularity [93], hard-wired modularity [94], and also ideas involving grandmother cells [95] or, in technical terms, look-up-tables [16]. One problem with these approaches is that some high-level mechanism is required for allocating or even generating new neurons in the intermediate layer. However, in most parts of the adult brain, there is little evidence for structural plasticity involving neuron genesis. But without neurogenesis such models also predict catastrophic forgetting at a later time unless plasticity is explicitly switched off after all neurons in the intermediate reservoir have been allocated. Alternative high level mechanisms for preventing CF involve pseudo-rehearsal using self-generated training stimuli from previously learned memories [92]. In the following we are focusing on solutions to CF that can be built at the level of synapses. For example, palimpsests network models [96–98] assume a slow decay of synaptic weights ($p_d > 0$) to prevent approaching the network's capacity limit, however, are not plausible for long-term storage in neocortex. Similarly, synaptic cascade models [52] introduce several consolidated states $1^{(i)}$ with decreasing decay rates $p_d^{(i)} > p_d^{(i+1)}$. However, this cannot prevent exponential decay of memories unless the lowest decay rate is zero causing again CF.

A novel role in preventing CF can be attributed to structural synaptic plasticity: Fig. 6A illustrates simulation experiments

investigating consolidation of multiple memory blocks each consisting of several novel memories. Each memory block is stored in the hippocampus and replayed to neocortical cell populations u and v for a certain time as described before (Fig. 2B, C). As expected, without any structural plasticity ($p_g = p_e = 0$) the network exhibits CF when approaching the capacity limit (upper panel). In contrast, CF is absent in networks with structural plasticity (lower panel). In this case, early stored memories remain stable all the time whereas the ability to store novel memories fades gradually when approaching the capacity limit. This behavior is more consistent with aging effects of human memory [99] and results from the fraction of consolidated synapses steadily increasing with age and the number of stored memories. Correspondingly, the fraction of unconsolidated synapses participating in structural plasticity gradually decreases with age as observed in neurophysiological experiments [21].

More precisely, for memories stored with a certain effectual connectivity P_{eff}, structural plasticity can prevent CF only if the filling fraction is below the maximal fraction of consolidated synapses at the capacity limit, $P/P_{\text{pot}} < p_{1\epsilon}(P_{\text{eff}})$ (see Eq. 13). This condition ensures that the total number of synapses, Pn^2, is smaller than the maximally allowed number of consolidated synapses, $p_{1\epsilon}(P_{\text{eff}})P_{\text{pot}}n^2$, at the network's capacity limit. If fulfilled, the network can never exceed its capacity limit which effectively prevents catastrophic forgetting. Brain networks could satisfy this condition by maintaining a constant (or slowly decreasing; cf, Fig. 7) anatomical connectivity P and by adapting cell assembly size k appropriately in relation to network size n and some target effectual connectivity P_{eff}. Thus, early memories can be consolidated up to some target connectivity P_{eff} which depends on the replay time per memory block. However, at least if replay time per memory remains constant over lifetime, then for later memories P_{eff} and $p_{1\epsilon}(P_{\text{eff}})$ will decrease gradually with the decreasing fraction of available structurally plastic synapses, $P - P_1$ (see Fig. 4B). Therefore, the ability to learn new memories will begin to fade when $p_{1\epsilon}(P_{\text{eff}})$ approaches P/P_{pot}.

7.2 Ribot gradients in retrograde amnesia. Patients with lesions of the hippocampus or neighboring neocortex in the medial temporal lobe often suffer from graded retrograde amnesia [38,40,100,101]. This form of memory loss shows characteristic "Ribot gradients" describing the tendency that recently stored memories are more likely to be lost than remote memories acquired at an earlier time. Simple palimpsests-type memory models (with $p_d > 0$) cannot account for these findings, in fact they predict the reverse effect [96–98].

A body of previous work has proposed that the lesions may disrupt cortico-hippocampal memory replay and, as a result, recent memories disappear because they are not sufficiently consolidated in intact neocortex [34,35,38,39,102–104]. According to such models, the cause of Ribot gradients is a gradient in accumulated replay and consolidation time [102,104].

In one of the models [102], for example, replay is controlled by a random walk over the attractor-landscape in Hopfield-type networks where each stored memory v^μ corresponds to one of the attractors. After acquiring the μth memory, each memory obtains an $1/\mu$ share of replay time. It is concluded that Ribot gradients occur because early memories (smaller μ) can accumulate a larger total consolidation time of about $\sum_{\mu_1=\mu}^{M} 1/\mu_1$ than recent memories, resulting in a larger strength of the memory trace.

Such models predict either that memories would be replayed and consolidated for an unlimited time [102] or that Ribot gradients would occur only for memories acquired during a limited time interval before the lesion occurred [104]. Although there are not yet final experimental answers [34,105], both predictions may be in conflict with evidence that novel memories are buffered and replayed by the hippocampus for a limited time only [34,38,39] and that, depending on the lesion size, graded amnesia can reach back to early childhood [38].

Synaptic learning based on structural plasticity offers an alternative explanation for Ribot gradients without relying on unlimited memory replay (Fig. 6B). According to our model, the substrate of Ribot gradients are gradients in effectual connectivity P_{eff} instead of (or in addition to) gradients in accumulated consolidation time. Even with constant replay time per memory, remote memories are stored with a larger P_{eff} than recent memories, for the very same reasons that explained the absence of catastrophic forgetting. Correspondingly, output noise $\hat{\epsilon}$ will be largest for most recent memories. During normal operation $\hat{\epsilon}$ is sufficiently low to accurately retrieve both remote and recent memories. However, cortical or hippocampal lesions will increase noise-levels such that memories get lost for which P_{eff} is below some critical value, or equivalently, that have been stored after some critical time point.

7.3 Spacing effect. Another interesting feature of memory is that learning new items is more effective if rehearsal is spaced over time compared to single block rehearsal [41–43,106]. For example, learning a list of vocabularies in two sessions each lasting 10 minutes turns out to be more effective than learning in a single session lasting 20 minutes. This so-called spacing effect is remarkably robust and occurs in many explicit and implicit memory tasks in humans and many animals being effective over many time scales from single days to months.

Previous cognitive models attributed the spacing effect either to deficient processing of repeated items during single block rehearsal [107] or to improved storage by exploiting context variability between spaced rehearsal sessions [108]. Typically, these explanations presumed specific high-level structures and mechanisms of memory systems including attention, novelty, and context processing. Although detailed modeling of memory systems may be required to explain specific properties in particular memory tasks, the ubiquity of the spacing effect suggests a common underlying mechanism at the cellular level. We propose that structural plasticity in sparsely connected neural networks is such a mechanism.

Figure 6C shows that structurally plastic networks reproduce the spacing effect naturally when learning a new set of memories in a similar protocol as described for the previous simulations (only here the memory replay should be interpreted more generally as rehearsal, not necessarily generated by the hippocampus). In the first simulation (blue) the memories are rehearsed in a single long time block, while in the second simulation (red) rehearsal is spaced over several shorter blocks such that total rehearsal time is equal for both simulations. For spaced rehearsal the resulting effectual connectivity P_{eff} of the memories turns out to be much higher and, correspondingly, the output noise $\hat{\epsilon}$ much lower than for single block rehearsal.

Further simulation experiments (not shown) have indicated that the spacing effect induced by structural plasticity is very stable. Similar to the psychological experiments, it is remarkably difficult to find conditions without spacing effect. In essence, the spacing effect occurs if weight plasticity is faster than structural plasticity and if consolidated synapses are more stable than silent synapses ($p_{e|0} > p_{e|1}$). Both properties are strongly supported by experiments [4,10,21,109]. In this case, our theory predicts that even in brief rehearsal sessions Hebbian plasticity can quickly consolidate all available synapses useful to store a set of memories. Thus, instead of continuing a rehearsal session, it is better to wait until structural

plasticity has grown additional useful synapses that can then be consolidated in a brief second rehearsal session. As a consequence, spacing effects will necessarily occur whenever learning in the brain depends on structural plasticity. Interestingly, our model with structural plasticity can also quantitatively reproduce long-term spacing effects as recently observed in psychological experiments that investigated optimal spacing intervals to maximize memory retention [110,111].

Discussion

One important limitation in the brain seems to be the number or density of functional (non-silent) synapses, both for anatomical and metabolic reasons. For example, the number of synapses per cortical volume is remarkably similar across different species [112], and theoretical considerations suggest that the energy consumption of the brain is dominated by the number of postsynaptic potentials or, equivalently, the number of functional non-silent synapses [47–49]. In face of these limitation, it might be beneficial that learning in brain circuits "moves" synapses to computationally useful locations [16,31,53,113].

To get a quantitative grip of these ideas we have introduced the concept of effectual connectivity, a macroscopic measure for how useful network structure is for memory storage. Structural plasticity can increase effectual connectivity while keeping the anatomical connectivity (P) at a low constant level. This has been analyzed for a simple model of structural plasticity assuming the following three basic mechanisms: (1) blind synaptogenesis, (2) consolidation of useful synapses, and (3) elimination of irrelevant synapses. Further, we have focused on the most plausible parameter range where structural plasticity (1,3) operates on a slower time scale than weight plasticity and consolidation (2), but the lifetime of consolidated synapses is long compared to the turnover of unstable synapses (see Section 2 and Section 5 for details; cf. [4,10,21]). In our current model implementation we identify strong synapses with stable synapses (weight and state 1) as well as weak synapses with unstable synapses (weight and state 0). This contrasts with some experimental results suggesting that silent synapses could be quite stable [114] whereas even strong synapses could be eliminated, for example, during development [51]. Such findings may be explained by the probabilistic nature of state transitions in our synapse model or a dissociation between synaptic strength and stability, perhaps including a cascade of several different stable and unstable states [52].

Our model is applicable to learning during development, as well as during adulthood (Fig. 7). During development the three mechanisms appear to dominate different phases separated on a large time scale of years [14–16,51,77,115]. Still, on a smaller time scale of days or months [20,21,23], ongoing structural plasticity, involving the three mechanisms simultaneously, could control the anatomical connectivity to be approximately constant (see Eq. 16). Such homeostatic regulation of generation and elimination of synapses is even more evident during adulthood where the anatomical connectivity appears almost stable over several decades [14,51,77]. However, recent experiments demonstrate that there can be novelty-driven excursions from homeostatic balance on the time scale of several days in specific cortical areas of the adult brain, for example, during learning of motor memories [23,68,116]. This phenomenon can be understood within our modeling framework as a different control strategy of the anatomical connectivity, one which is driven by learning load. Specifically, in instances of high learning load, up-regulating the anatomical network connectivity is the means to achieve faster learning by increasing the number of unstable silent synapses that may be recruited into new memories by structural plasticity and consolidation. Taken together, the model can explain the major differences of structural plasticity during development and adulthood by shifts in how metabolic constraints and learning speed are leveraged.

To simulate structural and weight plasticity we have used a simple three state Markov model of a potential synapse where state transition probabilities (with exception of p_g) depend on a Hebbian-type consolidation signal S_{ij} (see Fig. 2A, Eq. 4). Our plasticity model generalizes the binary Willshaw model [26,44] and strongly simplifies realistic weight plasticity models, for example, those based on spike-timing dependent synaptic plasticity (STDP) where potentiation depends on the precise temporal order of presynaptic and postsynaptic spikes [117–119]. In fact, it has been discussed controversially whether STDP-type learning rules would at all be consistent with the Hebbian idea that "what fires together wires together" because, unlike the Willshaw model, simple STDP models predict decoupling of neurons firing at the same time [120–123]. However, we have recently shown that more realistic STDP models (including dendritic propagation delays and parameters fitted to physiological data) are generally consistent with Hebbian learning and local cell assemblies [124].

Similarly, we argue that our model is also consistent with more realistic models of structural plasticity based on homeostatic mechanisms for maintaining mean neuronal firing rates at a constant level [20,50]. In such models, generation and elimination of synapses is induced by firing rates being below and above the homeostatic level, respectively. This is similar to our model with a homeostatic constraint for maintaining a constant anatomical connectivity P (see Section 2), because the mean firing rate of a neuron (e.g., during phases of ongoing activity [125]) will strongly correlate with the number of synapses on its dendrite (cf. [53,126]). Thus, keeping firing rates in homeostasis is essentially equivalent to maintaining the number of synapses per neuron and, thus, P, at a constant level. In our simulations, we have explicitly adjusted the generation rate p_g in each step in order to keep P constant, but in a more realistic setting, p_g could as well be driven by factors representing each neuron's mean firing rate.

Thus, we argue that both Hebbian and homeostatic structural plasticity are necessary to optimize information storage: Hebbian structural plasticity (via $p_{e|s}$) is necessary to eliminate those synapses that are not useful for storing a memory set. But homeostatic structural plasticity (via p_g) is also necessary: First, to balance the requirements of fast learning (large P) and space and energy efficiency (low P). Second, homeostatic structural plasticity may also contribute to *uniformly* sample new memory representations ν^μ from the space of all possible activity patterns (with unit usages $\#\{\mu : \nu_j^\mu = 1\}$ being equal for all neurons ν_j), which is known to be optimal for minimizing output noise and maximizing storage capacity in multi-layer networks (see Fig. 7 in [127]; cf. [126,128,129]): For example, a neuron representing only a few memories will have few state-1 synapses and, correspondingly, low firing rates. This may increase p_g to generate new state-0 synapses, rendering this neuron more plastic and receptive for being used to represent new memories, thereby increasing state-1 synapse number and firing rates until the desired homeostatic level is reached. Some previous works have actually argued that non-Hebbian homeostatic structural plasticity could be sufficient to explain memory formation [18,130]. Although this may hold true if cell assemblies representing different memories would be spatially separated with only little overlap, our results emphasize also the need of Hebbian-type structural plasticity with a specific elimination of unconsolidated synapses. Without Hebbian structural plasticity it seems impossible to stabilize a larger number of

overlapping cell assemblies and to come close to the high memory capacity of our model [16].

By introducing the concepts of effectual connectivity P_{eff} and consolidation signal S, our theory remains largely independent of a specific underlying neural network model of memory. In fact, the performance of the specific model in terms of output noise $\hat{\epsilon}$ is generally a non-linear monotonic function f of effectual connectivity, e.g., $\hat{\epsilon}=f(P_{\text{eff}})$, where f depends on the network model, network size, number of active units per memory vector, number of stored memories, and other factors. Here we have investigated Willshaw-type networks with binary synapses [16,26,44] because they give a simple and intuitive answer to the question which synapses are irrelevant and thus eligible for pruning. However, the efficiency of structural plasticity generalizes to learning employing graded synaptic states [53,54,79]. Previous approaches to memory formation by structural plasticity have also discussed that memories could be encoded in the number of synapses rather than by changing weights of individual synapses [28].

There are several lines of evidence suggesting that the binary weight model (corresponding to states 0 and 1) is already quite useful, in particular, if one would add suitable noise terms to account for distributed synaptic strength: First, experiments indicate that real synapses may have only a small number of functionally distinctive states or may even be binary [74–76,131]. Second, real synapses tend to scale their strengths such that in the soma (where spikes are generated) the resulting postsynaptic potentials have a relatively constant amplitude [61]. Third, anatomical experiments have shown that the number of real synapses per connected neuron pair is relatively constant in cortical areas [59] which indicates active regulation, for example, based on spike correlations [132,133]. Together, these findings support the hypothesis that the number of synapses per neuron pair and the strength of synapses at different dendritic locations might be co-regulated in order to keep the effect of a neuron onto a *connected* neighbor close to a desired constant magnitude. From a functional viewpoint, this perfectly makes sense at least for some functions such as memory storage (or the storage of "random" memory indices [134]) where binary synapses are optimal for storing sparse neural activity patterns [46,53,73].

Although our definition of effectual connectivity P_{eff} is tailored for the analysis of structural plasticity and memory storage, it shares many features with previous definitions of effective connectivity, e.g., based on "Granger causality" or "transfer entropy" used for analyzing the functional structure of brain networks from measured neural activity [135–137]. For example, transfer entropy $T_{u\to v}$ [137] is a measure of the directional information flow from one brain area u to another area v. In the simplest case the transfer entropy between activities $u(t)$ and $v(t)$ measured in two brain areas u and v is defined as $T_{u\to v} := \sum p(v(t+1),u(t),v(t)) \log \frac{p(v(t+1)|u(t),v(t))}{p(v(t+1)|v(t))}$ where $p(.)$ denotes the distribution of activity patterns, see Eq. 4 in [137] for details. This measure is very similar to the transinformation-based capacity measure $C^{\text{wp}}(P_{\text{eff}})$ (see Eqs. 10,14) which depends monotonically on P_{eff} rendering effectual connectivity an equivalent measure of how well an input activity pattern u^μ in one area can reactivate a corresponding target pattern v^μ in another area. In fact, the equivalence of the two measures, $T_{u\to v} \sim C^{\text{wp}}(P_{\text{eff}})$, can be shown for a simplified model of neural activity propagation in brain areas [138].

Adding to previous results of storage capacity based on counting possible synaptic network configurations [28–30] (cf. Eq. 12), our model proves that simple memory networks of n neurons with structural plasticity can indeed store *and* retrieve up to $C^{\text{tot}} \sim \log n$ bits per synapse. By comparison, even with real-valued synapses that have an infinite number of states, Hebbian-type weight plasticity without structural plasticity achieves less than one bit per synapse [72,73,139,140]. Technical adaptations of our model to applications such as information storage and pattern recognition have exhibited advantages in terms of recognition time and memory requirements compared to methods based on traditional weight plasticity [16,53,127].

Besides increasing storage capacity and energy efficiency of neural networks, our results suggest that structural plasticity is a key element in understanding various memory phenomena. One key prediction of the model under homeostatic maintenance of anatomical connectivity P are time-dependent gradients in effectual connectivity P_{eff}, such that memories from an earlier time have higher P_{eff} than memories from a later time. These gradients occur because consolidation of an increasing number of memories will continuously decrease the number of "migratable" (not yet consolidated) synapses and, thus, learning of new memories becomes slower and slower. We have shown that such gradients in P_{eff} can explain both aging effects and the absence of catastrophic forgetting because learning may stop just before the number of stored memories reaches the critical capacity limit [31,36,99]. The same gradients in P_{eff} can also explain Ribot gradients in amnesic patients suffering from lesions of the medio-temporal lobe [38–40]. Ribot gradients can also be explained by gradients in accumulated consolidation time, assuming unlimited cortico-hippocampal consolidation [102,104]. However, our model is unique in producing Ribot gradients even for finite consolidation times, in accordance with findings of a time-limited role of the hippocampal system in consolidation [34,38,39].

Last, our model is able to bridge different models, describing the spacing effect [43] on psychological [41,42,106] and molecular levels [141] by identifying structural synaptic plasticity as the potential cellular mechanism for spacing effects. The presence of structural plasticity in the adult brain is not only strongly supported by recent experimental evidence. As our results show, it is necessary to achieve high storage capacity and energy efficiency, and inevitably causes spacing effects. Structural plasticity is consistent with psychological theories that explained the spacing effect by encoding variability [106,108] but attributes the increased variability for spaced rehearsal to the changing pattern of synaptic connections rather than a changing learning context. While previous models based on delayed synaptic consolidation induced by molecular signaling cascades [52,141] may account for short-term spacing effects on the time-scale of minutes, structural plasticity can also explain long-term spacing effects on the time scale of months to years [110,111]. As the temporal profile of optimal learning depends on parameters of structural plasticity, predictions from theories of structural plasticity will be testable by future experiments that can link memory performance (behavioral data) and structural plasticity (physiological data) in cortical areas where these memories are stored.

Mathematical Analysis

I Temporal Dynamics of Effectual Connectivity P_{eff}

I.1 Relation between synapse and network states. As will be shown, effectual connectivity P_{eff} is a macroscopic network state that can be computed from the (microscopic) states of individual potential synapses. For this we first have to describe the relation between microscopic synaptic state variables $p_{\text{state}}(t)$ (Eq. 4) and the corresponding macroscopic connectivity variables $P_{\text{state}}(t)$. As indicated in the main text this relation is non-trivial

(see text below Eq. 4), because there may be multiple actual and/or potential synapses between each neuron pair ij, whereas connectivity of a neuron pair ij has to be defined in terms of the presence of *at least* one synapse or the absence of *all* synapses. For example, we could define neuron pair ij to be in state 1 if there is at least one potential synapse ij^v that is in state 1. Similarly, we define that $\text{state}(ij)=0$ iff $\text{state}(ij)\neq 1$ and there is at least one real synapse with $\text{state}(ij^v)=0$. Finally, $\text{state}(ij)=\pi$ iff $\text{state}(ij)\notin\{0,1\}$ and there is at least one potential synapse with $\text{state}(ij^v)=\pi$.

Next we divide neuron pairs into distinct groups, where two neuron pairs are in the same group if they receive identical consolidation signals $s(t)$. Then, in analogy to Eq. 4 we can define the (macroscopic) fractions $P^{(s)}_{\text{state}}$ of neuron pairs ij belonging to group $s=S_{ij}$ and being in a certain $\text{state}\in\{\pi,0,1\}$,

$$P^{(s)}_1 = P^{(s)}_{\text{pot}} \sum_{\mathfrak{n}=1}^{\infty} \mathfrak{p}(\mathfrak{n}) \left(1-(1-p^{(s)}_1)^{\mathfrak{n}}\right) \tag{17}$$

$$P^{(s)}_\pi = P^{(s)}_{\text{pot}} \sum_{\mathfrak{n}=1}^{\infty} \mathfrak{p}(\mathfrak{n}) \left(p^{(s)}_1\right)^{\mathfrak{n}} \tag{18}$$

$$P^{(s)}_0 = P^{(s)}_{\text{pot}} - P^{(s)}_1 - P^{(s)}_\pi \tag{19}$$

where $P^{(s)}_{\text{pot}}$ is the fraction of neuron pairs that have a potential synapse and receive consolidation signal $S_{ij}=s$ (typically $P^{(s)}_{\text{pot}}=P_{\text{pot}}P^{(s)}_{1S}$ if the matrix of potential connections is independent of the stored memories), and $\mathfrak{p}(\mathfrak{n})$ is the probability that there are exactly $\mathfrak{n}$ potential synapses given that there is at least one potential synapse for neuron pair ij. See ref. [59] for neuroanatomical estimates of $\mathfrak{p}(\mathfrak{n})$ in various cortical areas.

From this we can compute the macroscopic state variables P_{state} defined as the fractions of neuron pairs ij that are in a particular $\text{state}\in\{\emptyset,\pi,0,1\}$ (where state $\emptyset$ denotes neuron pairs without any potential synapses) and the various connectivity measures defined in Section 1,

$$P_{\text{state}}(t) = \sum_s P^{(s)}_{\text{state}}(t) \quad \text{for state} \in\{\emptyset,\pi,0,1\} \tag{20}$$

$$P(t) = P_0(t) + P_1(t) \tag{21}$$

$$P_{\text{pot}}(t) = P_\pi(t) + P_0(t) + P_1(t) \tag{22}$$

$$P_{1S} = \sum_{s\neq 0} \sum_{\text{state}\in\{\emptyset,\pi,0,1\}} P^{(s)}_{\text{state}}(t) \tag{23}$$

$$P_{\text{eff}}(t) = \frac{\sum_{s\neq 0} P^{(s)}_1(t)}{P_{1S}} \ . \tag{24}$$

By these definitions we are in the position to do microscopic simulations of networks of potential synapses and compute the corresponding connectivity measures (e.g., as we have done for Fig. 6; see also Section 1).

While we have worked out a general theoretical framework of structural plasticity [142], the following analyses will be limited to the much simpler case where a neuron pair has at most one synapse, $\mathfrak{p}(\mathfrak{n})=1$. Such a setting is justified by experimental findings that there is an active regulation of the total connection strength of the synapses connecting two neurons towards a constant value (see discussion section).

I.2 Increase of P_{eff} towards P_{pot}. To prove Eq. 11 let us now analyze the temporal dynamics of effectual connectivity P_{eff} under simplified conditions. Specifically, we analyze the increase of P_{eff} towards P_{pot} during consolidation in a *large* network with *constant* anatomical connectivity P having at most a *single* potential synapse per neuron pair. For this we will assume a simple *constant* consolidation signal, i.e., ongoing rehearsal or replay with $s(t)=S_{ij}\in\{0,1\}$ for $t>0$. Constant P requires a *homeostatic constraint* where generation and elimination of synapses are in approximate balance,

$$p_g(t) = \frac{p_{e|0}P^{(0)}_0(t-1) + p_{e|1}P^{(1)}_0(t-1)}{P_\pi}, \tag{25}$$

where $P^{(s)}_0$ is as defined in Sect. Mathematical Analysis I.1. Furthermore, we assume $p_{c|s}=s\in\{0,1\}$ and sufficiently large neuron populations u and v with sizes $n\gg 1$ (cf. Fig. 3) such that $P_{\text{eff}}(t)$ and $P_0(t)$ (and $P_1(t)=P-P_0(t)$) are always close to their expectations. Thus, at any point in time, there exist Pn^2 synapses distributed over $P_{\text{pot}}n^2$ possible locations. Before learning starts, the network has already $P_1(0)n^2$ consolidated synapses (e.g., due to earlier learned memories) that are unrelated to the novel memories specified by S_{ij}. Thus, initially $P_{\text{eff}}(0)=P_1(0)$ (Eq. 24). After the first learning step at $t=1$ all available synapses get potentiated and consolidated, $P_{\text{eff}}(1)=P$. For $t>1$ it is

$$P_{\text{eff}}(t) = P_{\text{pot}}\left(1-\prod_{t_1=1}^{t}(1-\frac{G(t_1)}{L(t_1)})\right)$$

where $G(t)$ is the number of new synapses generated at time t (which equals the number of eliminated synapses), $L(t)$ is the number of potential locations to put them, and $\prod\ (1-G/L)$ is the probability that a given potential synapse ij with $S_{ij}=1$ is not yet realized and consolidated until time t. For $t=1$ we can assume $G(1)=Pn^2$ and $L(1)=P_{\text{pot}}n^2$. For $t>1$ it is $G(t)=p_{e|0}P_0(t-1)n^2$ and $L(t)=P_{\text{pot}}n^2-Pn^2+G(t)$, where the number of unconsolidated synapses, P_0n^2, computes from

$$P_0(t) \approx P-(1-P_{1S})P_1(0)(1-p_{d|0})^t - P_{1S}P_{\text{eff}}(t)\ .$$

i.e., all real synapses minus initially consolidated (and not yet deconsolidated) synapses minus the newly consolidated synapses marked by S. Thus, the factors in the product become $1-G(t_1)/L(t_1)\approx(P_{\text{pot}}-P)/(P_{\text{pot}}-P+p_{e|0}P_0(t_1-1)$. Therefore

$$P_{\text{eff}}(t) \approx P_{\text{pot}}\left(1-(1-P/P_{\text{pot}})\prod_{t_1=2}^{t}\frac{1}{1+p_{e|0}P_0(t_1-1)/(P_{\text{pot}}-P)}\right)$$

proving Eq. 11. The second approximation in Eq. 11 becomes valid if all product terms are approximately equal, i.e., if $P_{1S} \ll 1$ (set of novel memories is small) and $tp_{d|0} \ll 1$ (deconsolidation during the time interval of rehearsal or replay is negligible). Note that here the increase of $P_{\text{eff}}(t)$ does not depend on $p_{d|1}$ since synapses with $S_{ij}=1$ that get deconsolidated are immediately ($p_{c|1}=1$) reconsolidated.

II Evaluation of Memory Capacity

II.1 Asymptotic analysis for one-step retrieval. As argued in Section 6, the storage capacity of structurally plastic networks where memories are stored with effectual connectivity P_{eff} is equivalent to the capacity of a structurally static network with increased anatomical connectivity $P=P_{\text{eff}}$ (cf. Fig. 3). Therefore the following computes the storage capacity for one-step retrieval in the Willshaw network without any structural plasticity ($p_{e|s}=p_g=p_{d|s}=0$, $p_{c|s}=s$; see Section 3 and Fig. 3A) where synaptic weights are given by Eq. 5.

For the following approximate asymptotic analysis we use several simplifications. First, Address and content memory patterns u^μ, v^μ are binary random vectors of size n each having k active units (i.e., k is the size of a Hebbian cell assembly representing the memory in population u or v). Second, the The query pattern $\tilde{u}$ has $c := \lambda k$ randomly chosen "correct" one-entries of an address pattern u^μ (where $0<\lambda\leq 1$) but no additional "false" one-entries ($f := \kappa k = 0$). Third, as previously suggested [78,143–145], we assume that each neuron j can optimize its firing threshold $\Theta_j := c'(j)$ according to the number of connected active "correct" query neurons, that is, $c'(j) := \#\{i : \tilde{u}_i = 1 \text{ and state}(ij) \in \{0,1\}\}$.

Let us first estimate error probabilities after storing M associations. We have $q_{10}=0$ due to the assumptions of optimal threshold control and zero add noise ($\kappa=0$). To see this note that $W_{ij}=1$ for any actual synapse ij with $\tilde{u}_i=1$ (which implies $u_i^\mu=1$ due to the zero add noise assumption) and $v^\mu=1$. Therefore the dendritic potential x_j will equal $\Theta_j=c'(j)$ and thus $\hat{v}_j=1$ if $v_j^\mu=1$. By contrast, q_{01} depends on the probability p_1 that a given synapse is potentiated (see Eqs. 4, 5). After storing M memory associations we have

$$p_1 = 1-\left(1-\frac{k^2}{n^2}\right)^M . \qquad (=P_{1S}) \tag{26}$$

This follows from the fact that a synapse is potentiated with probability k/n during presentation of a single memory. After presentation of all M memories, the synapse will therefore still be in state 0 (unpotentiated) with probability $p_0=(1-k/n^2)^M$. The state probability p_1 has been called "memory load" or "matrix load" in previous works [16] because, for fully connected networks, p_1 corresponds to the fraction of one-entries in the weight matrix. From Eq. 26 we obtain that a "low neuron" j with $v_j^\mu=0$ may fire with error probability

$$q_{01} \approx \sum_{c'=0}^{c} p_B(c';c,P)p_1^{c'} = (1-P(1-p_1))^c . \tag{27}$$

where $p_B(x;N,P) := \binom{N}{x}P^x(1-P)^{N-x}$ is the binomial probability. Note that $c'(j)$ follows a binomial distribution such that $\text{pr}[c'(j)=c_1]=p_B(c_1;c,P)$. Thus, the sum in Eq. 27 averages over all possible values of c' where the error probability given c' is $\text{pr}[\hat{v}_j=1|v_j^\mu=0]=p_1^{c'}$. This is because an error requires that all c' relevant synapses of neuron j are potentiated, where the probability of one synapse being potentiated is p_1. An exact analysis shows that this binomial approximation of q_{01} becomes exact in the limit of large networks and sufficiently small cell assemblies with $k=O(n/\log^2 n)$ (see [129]; see also Section II.2).

Now we can compute the storage capacity by limiting output noise $\hat{\epsilon}$ (Eq. 7) by some constant $\epsilon>0$. Thus, we have to solve

$$(n-k)q_{01} \leq \epsilon k \tag{28}$$

for p_1 which gives the maximal matrix load $p_{1\epsilon}$ of Eq. 13 that satisfies $\hat{\epsilon}\leq\epsilon$. With this, solving Eq. 26 for M yields the pattern capacity M_ϵ of Eq. 13. For small ϵ and k/n it is $T\approx -(k/n)\text{ld}(k/n)$ and with Eq. 10 it follows the weight capacity Eq. 14.

For networks with structural plasticity Eq. 13 is still valid but effectual connectivity will be typically larger than anatomical connectivity, $P_{\text{eff}}>P$. As silent synapses are functionally irrelevant and can be pruned (but see the remarks below) we can compute total storage capacity in bits per synapse from renormalizing Eq. 14. Thus, dividing the totally stored information by $P_{\text{eff}}p_{1\epsilon}n^2$ instead of $P_{\text{eff}}n^2$ yields

$$C_\epsilon^{\text{tot}} := \frac{C_\epsilon^{\text{wp}}}{p_{1\epsilon}} \approx \lambda \frac{\text{ld}(1-P_{\text{eff}}(1-p_{1\epsilon}))\ln(1-p_{1\epsilon})}{p_{1\epsilon}P_{\text{eff}}}\eta. \tag{29}$$

For large $P_{\text{eff}}\to 1$ and small $p_{1\epsilon}=(\epsilon l/(n-k))^{\lambda k}\to 0$ the total storage capacity per synapse diverges with network size n,

$$C_\epsilon^{\text{tot}} \approx -\lambda\eta\,\text{ld}\,p_{1\epsilon} \approx \frac{\text{ld}\frac{n}{k}}{k} . \tag{30}$$

Together with Eq. 11 this proves that in networks with structural plasticity, high potential connectivity, and sufficiently small cell assembly size k, it is possible to come close to the information theoretic capacity bound (see Eq. 12).

One limitation of this analysis is the assumption of an optimal threshold control. In fact, an optimal threshold control as presumed above would actually require silent synapses in order to compute spike thresholds $\Theta_j=c'(j)$ in incompletely connected *excitatory* networks with $P_{\text{eff}}<1$ [143,144] (so they should not be pruned). Therefore we will use the resulting expressions for C^{tot} merely for approximating the storage capacity for a more conservative threshold control (see next section). Nevertheless the results are still asymptotically correct for high effectual connectivity $P_{\text{eff}}\to 1$ because then the optimal spike threshold $c'(j)=c$ gets independent of remaining silent synapses [16]. Corresponding results hold true also for *inhibitory* network models where an optimal spike threshold control could easily be realized (including pruning of silent synapses) because it is independent of $c'(j)$ for any P_{eff} [78]. This suggests that structural plasticity could store information in inhibitory networks even more efficiently than in excitatory networks (cf. [13]).

II.2 Numerical evaluation for finite networks. The analysis of the previous section is asymptotically correct for large networks ($n\to\infty$), large connectivity ($P_{\text{eff}}\to 1$), and sparse activity ($k=O(n/\log^2 n)$) [16,129]. It is also useful to get an overview about the qualitative effect of increasing effectual connectivity P_{eff} and its relation to the memory load p_1. To compute storage capacity of finite networks with large activity k and low

connectivity P_{eff} it is possible to do an exact analysis by generalizing the approach of [129]. However, as such an approach would be computationally very expensive, the following develops a Gaussian approximation of dendritic potential distributions, which can reduce reduce computation time by several orders of magnitude. For example, in some preliminary experiments we have evaluated the exact storage capacity $M_\epsilon = 25005$ for $n = 10^5$, $k = 724$, $P_{\text{eff}} = 0.5$ for $\lambda = 1, \kappa = 0, \epsilon = 0.01$ which took about 57 h on a single core of an 2.2 GHz AMD Opteron compute server. By comparison, using the Gaussian approximation developed in this section yields $M_\epsilon = 24851$, quite close to the exact value, but took only 2.5 sec computing time.

Let us first consider the Willshaw-Palm distribution $p_{Ph}(X;k,n,M,P,z)$ defined as the exact probability that a content neuron's dendritic potential x_j equals X given that M random memories are stored in a heteroassociative Willshaw-Palm network with size n, anatomical connectivity P, and (constant) activity k if stimulating with a random pattern (unrelated to the stored memories) with z active units. From Eq. 3.22 in [129] we obtain p_{Ph} for the special case of fully connected networks ($P = 1$),

$$p_{\text{Ph}}(X;k,n,M,1,z) = \binom{z}{x} \sum_{s=0}^{x} (-1)^s \binom{x}{s} (1-\tilde{p}_1)^{s+z-x} \\ (1 - \frac{l}{n}(1 - B(m,k,s+z-x)))^M \tag{31}$$

where $B(a,b,c) := \binom{a-b}{c} / \binom{a}{c} = \Pi_{i=0}^{c-1} (a-b-i)/(a-i) = B(a,c,b)$. In network with general connectivity P each of the z active input units is connected to neuron j with probability P. Therefore the number of connected neurons is binomially distributed and

$$p_{\text{Ph}}(X;k,n,M,P,z) = \sum_{z'}^{z} p_B(z';z,P)\, p_{\text{Ph}}(X;k,n,M,1,z') \tag{32}$$

We can now determine the first two moments of this distribution, The mean $E(x_j)$ can easily be computed from the memory load Eq. 26,

$$\mu_{\text{Ph}} := E(x_j) = zPp_1 \tag{33}$$

and the variance

$$\sigma^2_{\text{Ph}} := E((x_j - \mu_{\text{Ph}})^2) = \sum_{X=0}^{z} (X - \mu_{\text{Ph}})^2 p_{\text{Ph}}(X;k,n,M,P,z) \tag{34}$$

$$= \sum_{z'}^{z} p_B(z';z,P) \sum_{X=0}^{z} (X - \mu_{\text{Ph}})^2 p_{\text{Ph}}(X;k,n,M,1,z') \tag{35}$$

$$= \sum_{z'}^{z} p_B(z';z,P)\, \sigma^*_{\text{Ph}}(z') \tag{36}$$

can be computed from the corresponding variance of a fully connected network which is well approximated by (see Eq. 4.25 in [129])

$$\sigma^*_{\text{Ph}}(z') \approx z'^2 (p_0^{(2)} - p_0^2) + z'(p_0 - p_0^{(2)}) \tag{37}$$

where $p_0 := 1 - p_1$ (cf. Eq. 26) and $p_0^{(2)} := (1 - (k^2/n^2)(2 - k/n))^M$. Therefore the variance of the diluted network is well approximated by

$$\sigma^2_{\text{Ph}} \approx zP(1-p_0) - zP^2(1 - 2p_0 + p_0^{(2)}) + z^2 P^2 (p_0^{(2)} - p_0^2) \tag{38}$$

From these results we can easily compute mean values and variances of the dendritic potential distributions of high and low units. Here high units are neurons v_j with $v_j^\mu = 1$, i.e., neurons that should be activated during retrieval. Similarly, low units are neurons v_j with $v_j^\mu = 0$. Thus, if the query pattern $\tilde{\boldsymbol{u}}$ has exactly c correct units from an address memory u^μ and additionally f randomly chosen false units (not active in u^μ) then the mean and variance of a low unit's dendritic potential will be

$$\mu_{\text{lo}} = (c+f)Pp_1 \tag{39}$$

$$\sigma^2_{\text{lo}} = (c+f)P(1-p_0) - \\ (c+f)P^2(1 - 2p_0 + p_0^{(2)}) + (c+f)^2 P^2 (p_0^{(2)} - p_0^2) \tag{40}$$

and mean and variance of a high unit's dendritic potential will be

$$\mu_{\text{hi}} = cP + fPp_1 \tag{41}$$

$$\sigma^2_{\text{hi}} = cP(1-P) + \\ fP(1-p_0) - fP^2(1 - 2p_0 + p_0^{(2)}) + f^2 P^2 (p_0^{(2)} - p_0^2) \tag{42}$$

Assuming Gaussian distributions we can compute a globally optimal firing threshold Θ that minimizes output noise $\hat{\epsilon}$ by applying some standard methods (e.g., see appendix D in [46]). Then we can determine pattern capacity M_ϵ by doing a binary search to efficiently find the maximal M that satisfies $\hat{\epsilon} \le \epsilon$. Finally, we can determine C_ϵ^{wp} from Eq. 10 and thus also $p_{1\epsilon}$ from Eq. 26 and $C_\epsilon^{\text{tot}} := C^{\text{wp}}/p_{1\epsilon}$. Corresponding data for $n = 10^5$ is shown in Fig. 5.

Acknowledgments

We thank Günther Palm and Marc-Oliver Gewaltig for many fruitful discussions.

Author Contributions

Conceived and designed the experiments: AK. Performed the experiments: AK. Analyzed the data: AK. Contributed reagents/materials/analysis tools: AK. Wrote the paper: AK EK UK FS.

References

1. Hebb D (1949) The organization of behavior. A neuropsychological theory. New York: Wiley.
2. Bliss T, Collingridge G (1993) A synaptic model of memory: long-term potentiation in the hippocampus. Nature 361: 31–39.
3. Frey U, Morris R (1997) Synaptic tagging and long-term potentiation. Nature 385: 533–536.
4. Paulsen O, Sejnowski T (2000) Natural patterns of activity and long-term synaptic plasticity. Current Opinion in Neurobiology 10: 172–179.
5. Song S, Miller K, Abbott L (2000) Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nature Neuroscience 3(9): 919–926.
6. Raisman G (1969) Neuronal plasticity in the septal nuclei of the adult rat. Brain Research 14: 25–48.
7. Engert F, Bonhoeffer T (1999) Dendritic spine changes associated with hippocampal long-term synaptic plasticity. Nature 399: 66–70.
8. Witte S, Stier H, Cline H (1996) In vivo observations of timecourse and distribution of morphological dynamics in Xenopus retinotectal axon arbors. Journal of Neurobiology 31: 219–234.
9. Woolley C (1999) Structural plasticity of dendrites. In: Stuart G, Spruston N, Häusser M, editors, Dendrites., Oxford, UK: Oxford University Press. pp. 339–364.
10. Trachtenberg J, Chen B, Knott G, Feng G, Sanes J, et al. (2002) Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. Nature 420: 788–794.
11. Lamprecht R, LeDoux J (2004) Structural plasticity and memory. Nature Reviews Neuroscience 5: 45–54.
12. DePaola V, Holtmaat A, Knott G, Song S, Wilbrecht L, et al. (2006) Cell type-specific structural plasticity of axonal branches and boutons in the adult neocortex. Neuron 49: 861–875.
13. Schuemann A, Klawiter A, Bonhoeffer T, Wierenga C (2013) Structural plasticity of GABAergic axons is regulated by network activity and GABA-A receptor activation. Frontiers in Neural Circuits 7: 113: 1–16.
14. Huttenlocher P, De Courten C, Garey L, van der Loos H (1982) Synaptogenesis in human visual cortex - evidence for synapse elimination during normal development. Neuroscience Letters 33: 247–252.
15. Chechik G, Meilijson I, Ruppin E (1998) Synaptic pruning in development: A computational account. Neural Computation 10(7): 1759–1777.
16. Knoblauch A, Palm G, Sommer F (2010) Memory capacities for synaptic and structural plasticity. Neural Computation 22(2): 289–341.
17. Keck T, Mrsic-Flogel T, Vaz Afonso M, Eysel U, Bonhoeffer T, et al. (2008) Massive restructuring of neuronal circuits during functional reorganization of adult visual cortex. Nature Neuroscience 11(10): 1162–1167.
18. Butz M, van Ooyen A, Wörgötter F (2009) A model for cortical rewiring following deafferentation and focal stroke. Frontiers in Computational Neuroscience 3: 1–15.
19. Yang G, Pan F, Gan WB (2009) Stably maintained dendritic spines are associated with lifelong memories. Nature 462: 920–924.
20. Butz M, Wörgötter F, van Ooyen A (2009) Activity-dependent structural plasticity. Brain Research Reviews 60(2): 287–305.
21. Holtmaat A, Svoboda K (2009) Experience-dependent structural synaptic plasticity in the mammalian brain. Nature Reviews Neuroscience 10: 647–658.
22. Leuner B, Gould E (2010) Structural plasticity and hippocampal function. Annual Review of Psychology 61: 111–140.
23. Fu M, Zuo Y (2011) Experience-dependent structural plasticity in the cortex. Trends in Neurosciences 34(4): 177–187.
24. Braitenberg V, Schüz A (1991) Anatomy of the cortex. Statistics and geometry. Berlin: Springer-Verlag.
25. Hellwig B (2000) A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. Biological Cybernetics 82: 111–121.
26. Willshaw D, Buneman O, Longuet-Higgins H (1969) Non-holographic associative memory. Nature 222: 960–962.
27. Hopfield J (1982) Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Science, USA 79: 2554–2558.
28. Poirazi P, Mel B (2001) Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. Neuron 29: 779–796.
29. Stepanyants A, Hof P, Chklovskii D (2002) Geometry and structural plasticity of synaptic connectivity. Neuron 34: 275–288.
30. Chklovskii D, Mel B, Svoboda K (2004) Cortical rewiring and information storage. Nature 431: 782–788.
31. Knoblauch A (2009) The role of structural plasticity and synaptic consolidation for memory and amnesia in a model of cortico-hippocampal interplay. In: Mayor J, Ruh N, Plunkett K, editors, Connectionist Models of Behavior and Cognition II: Proceedings of the 11th Neural Computation and Psychology Workshop. Singapore: World Scientific Publishing, pp. 79–90.
32. Knoblauch A (2009) Structural plasticity, cortical memory, and the spacing effect. BMC Neuroscience (Suppl 1): O16.
33. Knoblauch A (2008) On structural plasticity in neural associative networks. HRI-EU Report 08-04, Honda Research Institute Europe GmbH, D-63073 Offenbach/Main, Germany.
34. Ji D, Wilson M (2007) Coordinated memory replay in the visual cortex and hippocampus during sleep. Nature Neuroscience 10(1): 100–107.
35. McClelland J, McNaughton B, O'Reilly R (1995) Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. Psychological Review 102(3): 419–457.
36. French R (1999) Catastrophic forgetting in connectionist networks: causes, consequences and solutions. Trends in Cognitive Sciences 3(4): 128–135.
37. Grossberg S (1987) Competitive Learning: From interactive activation to adaptive resonance. Cognitive Science 11: 23–63.
38. Squire L, Bayley P (2007) The neuroscience of remote memory. Current Opinion in Neurobiology 17: 185–196.
39. Ross R, Eichenbaum H (2006) Dynamics of hippocampal and cortical activation during consolidation of a nonspatial memory. The Journal of Neuroscience 26(18): 4852–4859.
40. Ribot T (1881) Les maladies de la memoire. Paris: Germer Baillare.
41. Crowder R (1976) Principles of learning and memory. Oxford: Lawrence Erlbaum.
42. Greene R (1989) Spacing effects in memory: evidence for a two-process account. Journal of Experimental Psychology: Learning, Memory, and Cognition 15(3): 371–377.
43. Ebbinghaus H (1885) Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Leipzig: Duncker & Humblot.
44. Palm G (1980) On associative memories. Biological Cybernetics 36: 19–31.
45. Palm G (1991) Memory capacities of local rules for synaptic modification. A comparative review. Concepts in Neuroscience 2: 97–128.
46. Knoblauch A (2011) Neural associative memory with optimal bayesian learning. Neural Computation 23(6): 1393–1451.
47. Lennie P (2003) The cost of cortical computation. Current Biology 13: 493–497.
48. Laughlin S, Sejnowski T (2003) Communication in neuronal networks. Science 301: 1870–1874.
49. Attwell D, Laughlin S (2001) An energy budget for signaling in the grey matter of the brain. Journal of Cerebral Blood Flow and Metabolism 21: 1133–1145.
50. Butz M, van Ooyen A (2013) A simple rule for dendritic spine and axonal bouton formation can account for cortical reorganization after focal retinal lesions. PLOS Computational Biology 9(10): e1003259.
51. Elston G, Oga T, Fujita I (2009) Spinogenesis and pruning scales across functional hierarchies. The Journal of Neuroscience 29(10): 3271–3275.
52. Fusi S, Drew P, Abbott L (2005) Cascade models of synaptically stored memories. Neuron 45: 599–611.
53. Knoblauch A (2010) Zip nets: Efficient associative computation with binary synapses. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN). Barcelona, Spain: IEEE World Congress on Computational Intelligence (WCCI), pp. 4271–4278.
54. Knoblauch A (2009) Zip nets: Neural associative networks with non-linear learning. HRI-EU Report 09-03, Honda Research Institute Europe GmbH, D-63073 Offenbach/Main, Germany.
55. Knoblauch A (2003) Synchronization and pattern separation in spiking associative memory and visual cortical areas. PhD thesis, Department of Neural Information Processing, University of Ulm, Germany.
56. Deuchars J, West D, Thomson A (1994) Relationships between morphology and physiology of pyramid-pyramid single axon connections in rat neocortex in vitro. Journal of Physiology 478(3): 423–435.
57. Markram H, Lübke J, Frotscher M, Roth A, Sakmann B (1997) Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. Journal of Physiology 500(Pt 2): 409–440.
58. Angus Silver R, Lübke J, Sakmann B, Feldmeyer D (2003) High-probability uniquantal transmission at excitatory synapses in barrel cortex. Science 302(5652): 1981–1984.
59. Fares T, Stepanyants A (2009) Cooperative synapse formation in the neocortex. Proceedings of the National Academy of Sciences, USA 106(38): 16463–16468.
60. Deger M, Helias M, Rotter S, Diesmann M (2011) Cooperative structural plasticity based on pre- and postsynaptic spike timing. Frontiers in Computational Neuroscience Conference Abstract: BC11: Computational Neuroscience & Neurotechnology Bernstein Conference & Neurex Annual Meeting.
61. London M, Segev I (2001) Synaptic scaling in vitro and in vivo. Nature Neuroscience 4(9): 853–854.
62. Steinbuch K (1961) Die Lernmatrix. Kybernetik 1: 36–45.
63. Schwenker F, Sommer F, Palm G (1996) Iterative retrieval of sparsely coded associative memory patterns. Neural Networks 9: 445–455.
64. Knoblauch A, Palm G (2001) Pattern separation and synchronization in spiking associative memories and visual areas. Neural Networks 14: 763–780.
65. Knoblauch A, Palm G (2002) Scene segmentation by spike synchronization in reciprocally connected visual areas. II. Global assemblies and synchronization on larger space and time scales. Biological Cybernetics 87(3): 168–184.
66. Sommer F, Palm G (1999) Improved bidirectional retrieval of sparse patterns stored by Hebbian learning. Neural Networks 12: 281–297.
67. Kosko B (1988) Bidirectional associative memories. IEEE Transactions on Systems, Man, and Cybernetics 18: 49–60.

68. Xu T, Yu X, Perlik A, Tobin W, Zweig J, et al. (2009) Rapid formation and selective stabilization of synapses for enduring motor memories. Nature 462: 915–919.
69. Shannon C, Weaver W (1949) The mathematical theory of communication. Urbana/Chicago: University of Illinois Press.
70. Hertz J, Krogh A, Palmer R (1991) Introduction to the theory of neural computation. Redwood City: Addison-Wesley.
71. Lansner A (2009) Associative memory models: from the cell-assembly theory to biophysically detailed cortex simulations. Trends in Neurosciences 32(3): 178–186.
72. Gardner E, Derrida B (1988) Optimal storage properties of neural network models. JPhys A: Math Gen 21: 271–284.
73. Knoblauch A (2010) Optimal synaptic learning in non-linear associative memory. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN). Barcelona, Spain: IEEE World Congress on Computational Intelligence (WCCI), pp. 3205–3211.
74. Petersen C, Malenka R, Nicoll R, Hopfield J (1998) All-or-none potentiation at CA3-CA1 synapses. Proceedings of the National Academy of Science, USA 95: 4732–4737.
75. Montgomery J, Madison D (2004) Discrete synaptic states define a major mechanism of synapse plasticity. Trends in Neuroscience 27(12): 744–750.
76. O'Connor D, Wittenberg G, Wang SH (2005) Graded bidirectional synaptic plasticity is composed of switch-like unitary events. Proceedings of the National Academy of Sciences, USA 102(27): 9679–9684.
77. Huttenlocher P, Dabholkar A (1997) Regional differences in synaptogenesis in human cerebral cortex. Journal of Comparative Neurology 387: 167–178.
78. Knoblauch A (2007) On the computational benefits of inhibitory neural associative networks. HRI-EU Report 07-05, Honda Research Institute Europe GmbH, D-63073 Offenbach/Main, Germany.
79. Knoblauch A (2013) Efficient associative computation with discrete synapses and structural plasticity in preparation.
80. Huttenlocher P (1979) Synaptic density in human frontal cortex - developmental changes and effects of aging. Brain Research 163(2): 195–205.
81. Bourgeois J, Rakic P (1993) Changes of synaptic density in the primary visual cortex of the macaque monkey from fetal to adult stage. The Journal of Neuroscience 73(7): 2801–2820.
82. Waydo S, Kraskov A, Quiroga R, Fried I, Koch C (2006) Sparse representation in the human medial temporal lobe. Journal of Neuroscience 26(40): 10232–10234.
83. Freedman D, Riesenhuber M, Poggio T, Miller E (2006) Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. Cerebral Cortex 16(11): 1631–1644.
84. McNamara A, Buccino G, Menz M, Gläscher J, Wolbers T, et al. (2008) Neural dynamics of learning sound-action associations. PLoS ONE 3(12): e3845.
85. Weinberger N, Miasnikov A, Chen J (2009) Sensory memory consolidation observed: Increased specificity of detail over days. Neurobiology of Learning and Memory 91: 273–286.
86. Latham P, Nirenberg S (2004) Computing and stability in cortical networks. Neural Computation 16(7): 1385–1412.
87. Aviel Y, Horn D, Abeles M (2005) Memory capacity of balanced networks. Neural Computation 17: 691–713.
88. Grossberg S (1980) How does a brain build a cognitive code. Psychological Review 87: 1–51.
89. McCloskey M, Cohen N (1989) Catastrophic interference in connectionist networks: The sequential learning problem. In: Bower G, editor, Psychology of Learning and Motivation: Vol. 24, New York: Academic Press. 109–164.
90. Carpenter G, Grossberg S (2003) Adaptive resonance theory. In: Arbib M, editor, The Handbook of Brain Theory and Neural Networks, Second Edition, Cambridge, MA: MIT Press. pp. 87–90.
91. Abraham C, Robins A (2005) Memory retention - the synaptic stability versus plasticity dilemma. Trends in Neuroscience 28(2): 73–78.
92. Robins A, McCallum S (1998) Catastrophic forgetting and the pseudorehearsal solution in Hopfield type networks. Connection Science 7: 121–135.
93. French R (1992) Semi-distributed representations and catastrophic forgetting in connectionist networks. Connection Science 4: 365–377.
94. Murre J (1992) Learning and categorization in modular neural networks. LEA, NJ: Hillsdale.
95. Barlow H (1972) Single units and sensation: a neuron doctrine for perceptual psychology. Perception 1: 371–394.
96. Nadal JP, Toulouse G, Changeux J, Dehaene S (1986) Networks of formal neurons and memory palimpsests. Europhysics Letters 1(10): 535–542.
97. van Hemmen J, Keller G, Kühn R (1988) Forgetful memories. Europhysics Letters 5: 663–668.
98. Sandberg A, Lansner A, Petersson K, Ekeberg O (2000) A palimpsest memory based on an incremental Bayesian learning rule. Neurocomputing 32–33: 987–994.
99. Hedden T, Gabrieli J (2004) Insights into the ageing mind: a view from cognitive neuroscience. Nature Reviews Neuroscience 5: 87–96.
100. Zola S (2000) Amnesia I: Neuroanatomic and clinical issues. In: Farah M, Feinberg T, editors, Patient-based approaches to cognitive neuroscience., Cambridge, MA: MIT-Press, chapter 21. 275–290.
101. Baddeley A (1990) Human memory: theory and practice. Hillsdale, NJ: Lawrence Erlbaum.
102. Meeter M, Murre J (2005) TraceLink: A model of consolidation and amnesia. Cognitive Neuropsychology 22(5): 559–587.
103. Buzsaki G (1996) The hippocampo-neocortical dialogue. Cerebral Cortex 6: 81–92.
104. Alvarez R, Squire L (1994) Memory consolidation and the medial temporal lobe: a simple network model. Proceedings of the National Academy of Sciences (USA) 91: 7041–7045.
105. Nadel L, Moscovitch M (1997) Memory consolidation, retrograde amnesia and the hippocampal complex. Current Opinion in Neurobiology 7(2): 217–227.
106. Cepeda N, Pashler H, Vul E, Wixted J, Rohrer D (2006) Distributed practice in verbal recall tasks: A review and quantitative synthesis. Psychological Bulletin 132(3): 354–380. Cepeda/Pashler/Vul/Wixted/Rohrer:2006.
107. Bregman A (1967) Distribution of practice and between-trials interference. Canadian Journal of Psychology 21: 1–14.
108. Glenberg A (1979) Component-levels theory of the effects of spacing of repetitions on recall and recognition. Memory & Cognition 7: 95–112.
109. Martin S, Grimwood P, Morris R (2000) Synaptic plasticity and memory: an evaluation of the hypothesis. Annual Review of Neuroscience 23: 649–711.
110. Knoblauch A (2010) Bimodal structural plasticity can explain the spacing effect in long-term memory tasks. Frontiers in Systems Neuroscience Conference Abstract: Computational and Systems Neuroscience.
111. Cepeda N, Vul E, Rohrer D, Wixted J, Pashler H (2008) Spacing effects in learning: A temporal ridgeline of optimal retention. Psychological Science 19(11): 1095–1102.
112. Abeles M (1991) Corticonics: Neural circuits of the cerebral cortex. Cambridge UK: Cambridge University Press.
113. Knoblauch A (2006) On compressing the memory structures of binary neural associative networks. HRI-EU Report 06-02, Honda Research Institute Europe GmbH, D-63073 Offenbach/Main, Germany.
114. Hofer S (2009) Structural traces of past experience in the cerebral cortex. Journal of Molecular Medicine (Berlin) 88(3): 235–239.
115. Miyoshi S, Okada M (2004) Storage capacity diverges with synaptic efficiency in an associative memory model with synaptic delay and pruning. IEEE Transaction on Neural Networks 15(5): 1215–1227.
116. Yu X, Zuo Y (2011) Spine plasticity in the motor cortex. Current Opinion in Neurobiology 21(1): 169–174.
117. Markram H, Lübke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science 275: 213–215.
118. Bi G, Poo M (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. The Journal of Neuroscience 18(24): 10464–10472.
119. Clopath C, Büsing L, Vasilaki E, Gerstner W (2010) Connectivity reflects coding: a model of voltage-based STDP with homeostasis. Nature Neuroscience 13(3): 344–352.
120. Lubenov E, Siapas A (2008) Decoupling through synchrony in neuronal circuits with propagation delays. Neuron 58: 118–131.
121. Knoblauch A, Sommer F (2003) Synaptic plasticity, conduction delays, and inter-areal phase relations of spike activity in a model of reciprocally connected areas. Neurocomputing 52–54: 301–306.
122. Gerstner W, Kempter R, van Hemmen J, Wagner H (1996) A neuronal learning rule for sub-millisecond temporal coding. Nature 386: 76–78.
123. Fell J, Axmacher N (2011) The role of phase synchronization in memory processes. Nature Reviews Neuroscience 12: 105–118.
124. Knoblauch A, Hauser F, Gewaltig MO, Körner E, Palm G (2012) Does spike-timing-dependent synaptic plasticity couple or decouple neurons firing in synchrony? Frontiers in Computational Neuroscience 6: 1–27.
125. Arieli A, Sterkin A, Grinvald A, Aertsen A (1996) Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. Science 273: 1868–1871.
126. Knoblauch A (2010) Efficient associative computation with binary or low precision synapses and structural plasticity. In: Proceedings of the 14th International Conference on Cognitive and Neural Systems (ICCNS). Boston, MA: Center of Excellence for Learning in Education, Science, and Technology (CELEST), p. 66.
127. Knoblauch A (2008) Best-match hashing with inhibitory associative networks for real-world object recognition. HRI-EU Report 08-05, Honda Research Institute Europe GmbH, D-63073 Offenbach/Main, Germany.
128. Buckingham J, Willshaw D (1992) Performance characteristics of the associative net. Network: Computation in Neural Systems 3: 407–414.
129. Knoblauch A (2008) Neural associative memory and the Willshaw-Palm probability distribution. SIAM Journal on Applied Mathematics 69(1): 169–196.
130. Dammasch I (1989) Structural realization of a hebb-type learning rule. In: Cotterill R, editor, Models of Brain Function., Cambridge University Press. 539–552.
131. Song S, Sjöström P, Reigl M, Nelson S, Chklovskii D (2005) Highly nonrandom features of synaptic connectivity in local cortical circuits. PLOS Biology 3(3): 507–519.
132. Helias M, Rotter S, Gewaltig MO, Diesmann M (2008) Structural plasticity controlled by calcium based correlation detection. Frontiers in Computational Neuroscience 2: 7.
133. Deger M, Helias M, Rotter S, Diesmann M (2012) Spike-timing dependence of structural plasticity explains cooperative synapse formation in the neocortex. PLoS Computational Biology 8(9): e1002689.

134. Teyler T, Rudy J (2007) The hippocampal indexing theory and episodic memory: Updating the index. Hippocampus 17(12): 1158–1169.
135. Sporns O (2007) Brain connectivity. Scholarpedia 2: 4695.
136. Granger C (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37(3): 424–438.
137. Schreiber T (2000) Measuring information transfer. Physical Review Letters 85: 461–464.
138. Knoblauch A (2014) Structural plasticity and effective connectivity in preparation.
139. Tsodyks M, Feigel'man M (1988) The enhanced storage capacity in neural networks with low activity level. Europhysics Letters 6: 101–105.
140. Palm G, Sommer F (1996) Associative data storage and retrieval in neural nets. In: Domany E, van Hemmen J, Schulten K, editors, Models of Neural Networks III, New York: Springer-Verlag. 79–118.
141. Pagani M, Oishi K, Gelb B, Zhong Y (2009) The phosphatase SHP2 regulates the spacing effect for long-term memory induction. Cell 139: 1–13.
142. Knoblauch A (2014) Efficient simulation of structural plasticity in the brain in preparation.
143. Buckingham J, Willshaw D (1993) On setting unit thresholds in an incompletely connected associative net. Network: Computation in Neural Systems 4: 441–459.
144. Graham B, Willshaw D (1995) Improving recall from an associative memory. Biological Cybernetics 72: 337–346.
145. Bosch H, Kurfess F (1998) Information storage capacity of incompletely connected associative memories. Neural Networks 11(5): 869–876.