bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Associations between memory traces emerge in a generic neural circuit model through STDP

Christoph Pokorny[1], Matias J. Ison[2], Arjun Rao[1], Robert Legenstein[1], Christos Papadimitriou[3], and Wolfgang Maass[1*]

[1]Institute for Theoretical Computer Science, Graz University of Technology, Graz, Austria
[2]School of Psychology, University of Nottingham, Nottingham, UK
[3]EECS, University of California, Berkeley, CA, USA

[*]Corresponding author: maass@igi.tugraz.at

September 14, 2017

Abstract

The formation of associations between memory items, that enables recall of one memory component by activating another, is a fundamental operation of higher brain function. Recent neural recordings provide insight into the way how such associations are encoded on the level of neurons in the human medial temporal lobe (MTL). We show that important features of these experimental data can be reproduced by a generic neural circuit model consisting of excitatory and inhibitory spiking neurons with data-based short- and long-term synaptic plasticity. A key result of the experimental data and the model is that the association process causes the emergence of overlaps between the assemblies of neurons that encode the memory components. These overlaps appear in the experiments and the model at the same time when the association becomes computationally functional. Hence our model elucidates computational and plasticity processes that are likely to shape memory systems in the brain.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Significance statement

One commonly assumes that memory items are encoded by sparsely distributed groups of neurons, often referred to as assemblies, that fire whenever a memory item is activated. An important question is how combinations of several memory items are encoded. Recent experimental data suggest that the assemblies for memory items expand during an association process, so that overlaps of the assemblies emerge. This result is surprising from the perspective of neural network models, where one commonly assumed that assemblies for memory items remain largely invariant. We show that a simple model for recurrent neural circuits with data-based forms of synaptic plasticity reproduces the new experimental data, and thereby provides the basis for more data-based neural network models for memory associations.

1 Introduction

It has long been known that the hippocampus and the medial temporal lobe (MTL) are fundamental for the formation of new episodic memories (Eichenbaum 2004, Moscovitch 1994, Squire et al. 2004). In the past few years, numerous experiments have started to elucidate how individual memories are stored in the brain. Advances in recording techniques, population analyses, transgenics, and pharmacogenetic and optogenetic tools have led, among others, to the discovery of neural map-like representations of the environment (McNaughton et al. 2006, Moser et al. 2008), the involvement of neural assemblies in the MTL representing non-spatial information (Wood et al. 1999, Aronov et al. 2017), abstract concepts (Quian Quiroga et al. 2005) and the identification of memory "engram" cells (Liu et al. 2012, Tonegawa et al. 2015, Josselyn et al. 2015). There is also consensus that episodic memory relies on the rapid formation of associations (Wirth et al. 2003, Morris et al. 2003, Quian Quiroga 2016). In humans, individual cells are involved in the representation of item-location associations (Miller et al. 2013), long-term item-item associations (De Falco et al. 2016), and new associations between faces and places (Ison et al. 2015). Connectionist models of memory function have successfully given account of numerous experimental observations, including memory deficits following MTL damage and recognition memory (Rolls et al. 1998, McClelland et al. 1995, Norman and O'Reilly 2003). Many of these models started from adapting Marr's ideas (Marr 1971), where a given event is assumed to be encoded by a sparse pattern of activity of hippocampal neurons that represent specific combinations of elements of the event. Information is then carried between the hippocampus and the neocortex, which eventually represents the event with a widely distributed pattern of neural activity (McClelland et al. 1995). Modelling the learning of associations has proved more challenging (McCloskey and Cohen 1989). Theoretical studies of neural network models (Valiant 2000a;b; 2005) proposed mechanisms for the formation of associations based

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

on the hypothesis that the composite memory item is represented by a new assembly of neurons that had not participated in the encoding of its components, while the assemblies for the memory components remain invariant. In contrast, recent neural recordings from the human MTL (Ison et al. 2015, De Falco et al. 2016) show that the formation of associations between memory items creates overlaps between the memory traces for these memory items. These modifications of memory traces are not consistent with standard neural network models for memory that assume that assemblies for memory items remain largely invariant during the formation of associations between them.

We asked under what conditions a neural network model could reproduce the experimental data about modifications of memory traces in the human MTL during the formation of an association. In order to support comparisons of neural responses in the model with recordings from neurons in the human MTL we considered models in the form of generic recurrent networks of excitatory and inhibitory spiking neurons. We included data-based short-term plasticity of synapses, i.e., connection-type-specific mixtures of paired-pulse depression and facilitation, in the model in order to achieve a biologically more realistic network dynamics. Long-term synaptic plasticity was modeled through standard forms of spike-timing-dependent plasticity (STDP).

This simple neural network model reproduced not only the finding of (Ison et al. 2015) that associations between memory items causes overlaps between the corresponding assemblies. In addition it reproduced the finding that very few additional neurons become selective for the combined memory in the MTL during the formation of an association. Furthermore, overlaps between assemblies for memory components appeared in the data and in our model roughly at the same trial when the association becomes computationally functional, in the sense that an associated memory component can be recalled by giving another component of the combined memory as a cue. Altogether our model proposes a new foundation for modelling plasticity and computational processes that underlie the astounding capabilities of the human memory system.

2 Methods

2.1 Details of the network model

Our network model was a randomly connected recurrent neural network consisting of 432 excitatory and 108 inhibitory stochastic spike response model neurons (Jolivet et al. 2006). In this model, the response kernel defines the shape of postsynaptic potentials (PSPs). We defined the response kernel of both excitatory and inhibitory neurons as a double-exponential function

$$\epsilon(t) = K(e^{-t/\tau_f} - e^{-t/\tau_r}) \tag{1}$$

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with a rise time constant of $\tau_r = 2\,\text{ms}$, a fall time constant of $\tau_f = 20\,\text{ms}$, and a cut-off after $t = 100\,\text{ms}$. The scaling factor K was computed to obtain a peak value of 1. The instantaneous firing rate r of a neuron depends on the current membrane potential $u_m(t)$ and was given in our simulations by the transfer function

$$r(t) = r_0 e^{0.25 u_m(t)}, \tag{2}$$

with $r_0 = 1.238\,\text{Hz}$. The membrane potential of an excitatory neuron was given by

$$u_m(t) = \sum_i w_i^{\text{Inp}} y_i(t) + \sum_j w_j^{\text{EE}} z_j(t) - \sum_k w_k^{\text{IE}} h_k(t) + \alpha_{\text{E,offset}} + \alpha_{\text{E}}, \tag{3}$$

which was the weighted sum of PSPs $y_i(t)$ caused by external inputs, PSPs $z_j(t)$ caused by excitatory neurons in the recurrent circuit, PSPs $h_k(t)$ caused by inhibitory neurons, and its excitability. The neuronal excitability consisted of a variable component α_{E} drawn from a log-normal distribution with $\mu = 2.64 \cdot 10^{-3}$ and $\sigma = 0.23 \cdot 10^{-3}$ (of the underlying normal distribution) which was shifted by a constant offset of $\alpha_{\text{E,offset}} = -250 \cdot 10^{-3}$ towards negative values. The membrane potential for inhibitory neurons was defined in an analog way but without external inputs and a different value for $\alpha_{\text{I,offset}} = -150 \cdot 10^{-3}$. After generating a spike, neurons entered a refractory period of random length drawn from gamma distributions (shape parameter $k = 2$) with a mean value of 10 ms for excitatory and 3 ms for inhibitory neurons.

Figure 1A shows a schematic representation of the network architecture. Excitatory and inhibitory pools of neurons were reciprocally connected, while each pool was also recurrently connected. Additionally, a pool of 200 external neurons (termed input neurons) was connected to the excitatory population.

All 540 network neurons were arranged in a 3D grid of $6 \times 6 \times 15$ neurons. The recurrent connectivity among excitatory neurons was uniform, in accordance with experimental data on the anatomy of CA3 which show a rather uniform recurrent connectivity among excitatory neurons within this area (Guzman et al. 2016). We chose a connection probability of $p_{\text{EE}} = 50\,\%$ to have a sufficient number of connections between excitatory neurons, despite the small network size. All other connection probabilities between excitatory and inhibitory neurons were exponentially distance-dependent, resulting in very strong and local inhibition. The distance-dependent connection probabilities between pairs of neurons were defined as

$$p(d) = ce^{-d/\lambda}, \tag{4}$$

where d denotes the distance between the neurons in grid units, $\lambda = 0.25$ denotes the length scale, and c is a scaling parameter with values that depended on the connection type as shown in Table 1. This resulted in average connection probabilities of around 4 % for E→I and I→I, and around 5 % for I→E connections (100 % in close vicinity). The pool of input neurons was randomly connected to the excitatory population with a uniform connection probability of 50 %.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

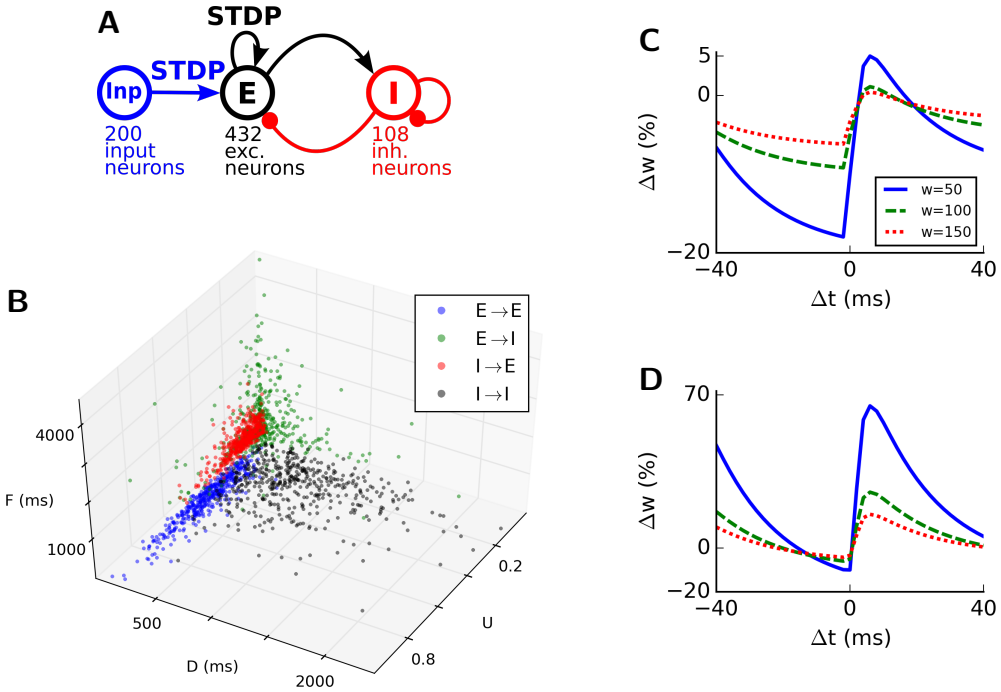

Figure 1: (A) Network architecture, consisting of 200 input neurons (Inp), 432 excitatory neurons (E), and 108 inhibitory neurons (I). Inp→E as well as recurrent E→E synapses were subject to STDP. (B) Illustration of the distribution of UDF parameter values for short-term synapse dynamics in three-dimensional parameter space for different types of connections (blue: E→E; green: E→I; red: I→E; black: I→I). UDF values were generated from bounded gamma distributions with means and SDs given in Table 2. (C, D) STDP rules for (C) synapses from external inputs and (D) recurrent excitatory synapses that were used in our model, showing the relative weight change after two spike pairings at a pairing frequency of 20 Hz for different initial weight values.

Synaptic delays for input connections d_{Inp} and recurrent excitatory connections d_{EE}, as well as inhibitory delays $d_{\mathrm{I}} := d_{\mathrm{EI}} = d_{\mathrm{II}} = d_{\mathrm{IE}}$ were drawn from normal distributions with mean values as given in Table 1 and a coefficient of variation $CV = 0.5$.

Initial synaptic weights were drawn from gamma distributions with mean values as given in Table 1 and $CV = 0.7$. After initialization, all weights between excitatory and inhibitory neurons in the circuit were re-scaled as described in Section 2.1.1 to account for effects of short-term synaptic plasticity.

As can be seen in Table 1, weights to and from inhibitory neurons were rather large as compared to those involving only excitatory neurons. This was an artifact of the small network size. Large weights were required to compensate with a small number of inhibitory connections – due to local connectivity in the small network – for the high recurrent excitatory connectivity. In order to test whether the network still operated in a biologically reasonable regime, we computed the I-to-E ratio of the sums of ingoing weights from input and network neurons to an excitatory neuron. Since synapses within

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Connection	**Conn. prob.**	**Dist.-dep.**	**Init. weights**	**Delays**
Inp→E	50 %	*uniform*	$\mu_{\mathrm{IE}} = 15$	$\mu = 5\,\mathrm{ms}$
			$CV = 0.7$	$CV = 0.5$
E→E	50 %	*uniform*	$\mu_{\mathrm{EE}} = 2.5$	$\mu = 5\,\mathrm{ms}$
			$CV = 0.7$	$CV = 0.5$
E→I	4 %	$c = 2 \cdot 10^5$	$\mu_{\mathrm{EI}} = 1000$	$\mu = 2\,\mathrm{ms}$
		$\lambda = 0.25$	$CV = 0.7$	$CV = 0.5$
I→E	5 %	$c = 4 \cdot 10^5$	$\mu_{\mathrm{IE}} = 1375$	$\mu = 2\,\mathrm{ms}$
		$\lambda = 0.25$	$CV = 0.7$	$CV = 0.5$
I→I	4 %	$c = 1 \cdot 10^5$	$\mu_{\mathrm{II}} = 6000$	$\mu = 2\,\mathrm{ms}$
		$\lambda = 0.25$	$CV = 0.7$	$CV = 0.5$

Table 1: Connection parameters between input (Inp), excitatory (E), and inhibitory (I) pools of neurons. The connection probabilities were either constant (uniform) or exponentially distance-dependent with parameters c and λ. Initial weights were drawn from gamma distributions, synaptic delays from normal distributions with given means μ and CVs.

the network were subject to short-term plasticity (see Section 2.1.1), synaptic efficacies are changing on a fast time scale in the model. We therefore estimated their efficacy values in two ways: (i) for the first presynaptic spike after a (theoretically infinitely) long silent period of the presynaptic neuron and (ii) at equilibrium with an assumed constant presynaptic firing rate of $f_0 = 5\,\mathrm{Hz}$. We found that inhibition dominates recurrent excitation by a factor of 2 (i) to 3 (ii), which is in good agreement with experimental findings in neocortex (Haider et al. 2013). Experimental data suggest that inhibition dominates excitation also in area CA3 of the hippocampus (Calfa et al. 2015, Atallah and Scanziani 2009).

We refer to the above described values of the parameters as the *standard values*. In Section 3.5, we analyzed the impact of parameter variations on the emergence of memory associations. In this analysis, we considered two factors: First, the global scale w_{scale} of all weights between excitatory and inhibitory neurons. For the standard values described here, we have $w_{\mathrm{scale}} = 1$. By changing the global scale, all initial network weights are re-scaled by this factor. Second, we considered the fraction $g_{\mathrm{E}} = \frac{\mu_{\mathrm{EE}}}{\mu_{\mathrm{IE}}}$ between the mean initial weight of recurrent synapses from excitatory to excitatory neurons and the mean initial weight of synapses form inhibitory to excitatory neurons. By changing this parameter (while leaving μ_{EE} fixed), we changed the fraction between inhibition and recurrent excitation.

All network simulations and data analyses were done in Python 2.7.13 using the NEST Simulator 2.6.0 (Gewaltig and Diesmann 2007) with a time resolution of 1 ms together with the PyNEST Interface (Eppler et al. 2009).

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2.1.1 Short- and long-term synaptic plasticity

Our model for synaptic connections included data-based short-term plasticity, i.e., a mixture of paired-pulse depression and facilitation that depended on the type of the pre- and postsynaptic neuron. This can be described by three parameters (Markram et al. 1998): U (release probability), D (time constant for depression), and F (time constant for facilitation). Values of UDF parameters for synapses between different cell types and layers in the somatosensory cortex of rats have been reported in (Gupta et al. 2000) and more recently in (Markram et al. 2015). Additionally, in the adult human brain, no facilitation but frequency-dependent depression only was found in synaptic connections between layer 2/3 pyramidal neurons (Testa-Silva et al. 2014). In our model, values for UDF parameters were drawn from bounded gamma distributions with mean values and standard deviations (SD) for E→E synapses taken from human experimental data (Testa-Silva et al. 2014), for E→I and I→I synapses taken from the range of values among the most frequent connection types of the recent experimental results reported in (Markram et al. 2015), and for I→E synapses taken from (Gupta et al. 2000). The UDF parameters were bound between [0.001, 0.999] (parameter U) and [0.1 ms, 5000 ms] (parameters D, F) respectively. The used mean and SD values are summarized in Table 2 and the corresponding distributions are illustrated in the three-dimensional UDF parameter space in Figure 1B.

After random initialization of the weights and UDF parameters according to the distributions specified above, all weights were re-scaled based on their steady-state values in the following way (Markram et al. 1998, Sussillo et al. 2007). For a given constant presynaptic firing rate f_0, steady-state values of the synaptic weights w_{dyn} of the dynamic synapses can be computed as

$$w^*_{\mathrm{dyn}}(f_0) = A \cdot R^*(f_0) \cdot u^*(f_0), \tag{5}$$

where A is the absolute synaptic efficacy. The steady-state values for the synaptic availability R and the synaptic utilization u at a given rate r are given by

$$R^*(r) = \frac{1 - \exp(-1/rD)}{1 - (1 - u^*(r))\exp(-1/rD)} \tag{6}$$

and

$$u^*(r) = \frac{U}{1 - (1 - U)\exp(-1/rF)}. \tag{7}$$

The initial weight w_{init} (absolute synaptic efficacy) of each synapse in the network was re-scaled such that it corresponded to its actual dynamic weight at an assumed constant presynaptic firing rate $f_0 = 5\,\mathrm{Hz}$, by

$$w = \frac{w_{\mathrm{init}}}{R^*(f_0) \cdot u^*(f_0)}. \tag{8}$$

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Connection	U	D (ms)	F (ms)	Type
E→E	0.45 ± 0.17	144 ± 67	0 ± 0	depressing
E→I	0.09 ± 0.12	138 ± 211	670 ± 830	facilitating
I→E	0.16 ± 0.10	45 ± 21	376 ± 253	facilitating
I→I	0.25 ± 0.13	706 ± 405	21 ± 9	depressing

Table 2: Parameters for short-term plasticity in the model: U (release probability), D (time constant for depression), and F (time constant for facilitation). Mean ± SD values were taken from experimental results given in (Testa-Silva et al. 2014; E→E connections), (Gupta et al. 2000; I→E connections) and (Markram et al. 2015; remaining connections) respectively.

This re-scaling of weights was beneficial to reduce initial transients and to decrease the burn-in time of the network needed to reach steady-state activity after starting the simulation.

In addition, synaptic connections from input to excitatory as well as between excitatory neurons in our model were subject to a specific type of STDP. We used a theoretically tractable variant of the standard STDP learning rule (Nessler et al. 2013) which reproduces for a medium pairing frequency of around 20 Hz experimentally measured STDP curves (see Figure 1B in (Pecevski and Maass 2016) and (Sjöström et al. 2001)). Weight updates for each postsynaptic spike were given by

$$\Delta w_{\text{Inp}} = \eta(y(t)f(w_{\text{Inp}}) - 1) \tag{9}$$

for synapses from input neurons to excitatory neurons in the circuit. Here, $\eta = 5$ scales the size of the weight updates and $y(t)$ is the (unweighted) sum of PSPs (as in Equation (3)) caused by external inputs. The function $f(w)$ implemented a weight dependency of the update and was given by

$$f(w) = \beta \left(1 + \frac{a}{(a \cdot w + b)^2}\right) \tag{10}$$

as derived in (Jonke 2013), with parameters $a = 2 \cdot 10^{-4}$, $b = 1 \cdot 10^{-2}$, and $\beta = 1$. Synaptic connections between excitatory neurons in the network were subject to the same plasticity updates, but with a parameter $\beta = 5$. This scaling of the positive part of the STDP window function resulted in a larger potentiation and shorter depression time window. Both STDP induction rules are illustrated in Figure 1C and D for a pairing frequency of 20 Hz. Total relative weight changes were limited to [0 %, 200 %] for input synapses and to [0 %, 1000 %] for recurrent excitatory synapses respectively.

2.1.2 Input patterns and trial-to-trial variability

In each simulation, we repeatedly presented three input patterns to the network (see Section 2.2.1). In total, we generated 20 such triples of input patterns. Each input

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

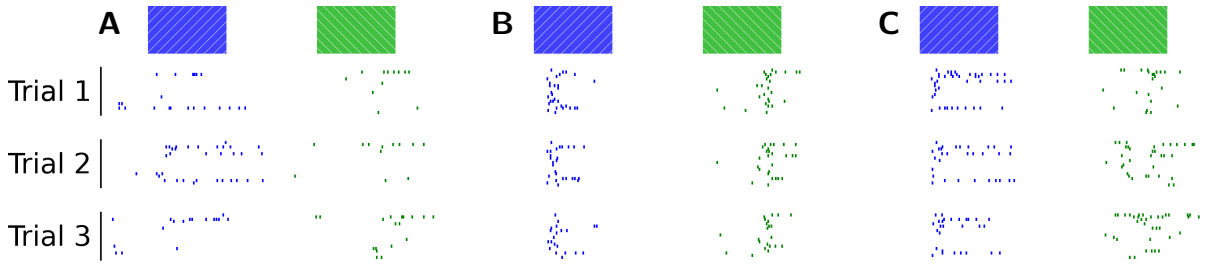

Figure 2: Trial-to-trial variability of 20 arbitrarily selected assembly neurons of the blue and green assemblies in response to their preferred input patterns (A) before any STDP took place, (B) after plasticity phases 1 and 2 (i.e., after the emergence of assemblies), and (C) after plasticity phase 3 (i.e., after the formation of associations). Different rows represent examples from 3 different trials. The widths of the blue and green symbols at the top of the figure indicate the input pattern presentation intervals of 100 ms. Spikes were extracted from time windows of -50 ms to 150 ms relative to pattern onsets.

pattern was composed of 200 independently drawn Poisson spike trains over 100 ms with a firing rate of 4 Hz. The exact timing of spikes was then kept constant for each pattern.

From each pattern triple, we chose an arbitrary pattern and termed it the *blue* pattern, another one the *green* pattern, and the third one the *red* pattern for easy reference in the text. As detailed below, a neural assembly emerged for each of these patterns due to synaptic plasticity, and these were termed the *blue*, the *green*, and the *red* assembly respectively.

One specific set of three patterns was used for most simulations throughout this article, and we refer to this triple as the *default patterns*. For some simulations, 13 pattern triples were chosen out of the 20 pattern sets according to a criterion detailed in *Results*.

At each presentation of such a pattern, each spike train was superimposed with a freshly generated Poisson spike train with a rate of 3 Hz. In this way, a biologically realistic trial-to-trial variability of network responses to the same input could be observed. Figure 2 illustrates this inherent trial-to-trial variability of 20 arbitrarily selected assembly neurons in response to their preferred input patterns before any STDP took place, after the first two plasticity phases (i.e., after the emergence of assemblies, see below), and after the third plasticity phase (i.e., after the formation of associations, see below).

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2.2 Details to emergent memory traces

2.2.1 Plasticity phases

In total, we considered three plasticity phases. In the simulations for Section 3.2, only the first two phases were applied. The third phase was added in the simulations for Sections 3.3–3.5. In the first two phases – which lasted for around 250 s each – input patterns were presented at random time points to the network as input. After a pattern presentation, a time period was randomly chosen from the uniform distribution between 0.5 s and 3 s during which input neurons emitted freshly generated Poisson spike trains at 5 Hz. After this noise period, the next pattern was randomly chosen and presented and so on. The first pattern at the beginning of a phase was chosen randomly with a uniform probability over the three patterns (but not presented to the network; see below). Subsequent patterns were drawn based on a switching probability of 75 %, meaning that with probability 0.75, a switch to another pattern occurred (which was drawn uniformly over the other patterns), and with 25 % probability the same pattern was repeated.

When generating such a sequence of input patterns and noise periods of a given length, the exact duration of the resulting plasticity phase was determined by the following rules: The first pattern was always omitted so that each sequence started with a noise period at the beginning. In case there was a pattern exactly at the end of a sequence, also the last pattern was omitted so that each sequence ended with a noise period. In case there was already a noise period at the end of a sequence, this noise period was not truncated. So when referring to a plasticity phase of 250 s, the exact duration could be (with a pattern duration of 100 ms) in a range between [249.8 s, 253.0 s] and was 250.8 s on average. The same rules apply to all other simulation phases accordingly.

During the first plasticity phase of 250 s, only synaptic connections from input neurons to excitatory neurons in the network were subject to STDP. During the second plasticity phase of 250 s, only recurrent synaptic connections between excitatory neurons in the network were subject to STDP, see *Results* for a discussion. The third plasticity phase is discussed in Section 2.3.1.

2.2.2 Definition of assembly neurons

The experiments of (Ison et al. 2015) studied the formation of associations between memory traces for a specific set of images. Neurons were classified as belonging to the memory trace (assembly) for a specific image if they significantly responded to this image based on a Wilcoxon rank-sum test ($p < 0.05$) between baseline and response intervals. As additional criterion, a median firing rate in response intervals across trials

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of at least 2 Hz was required. Neurons that satisfied these criteria for any of the images were termed visually responsive units (VRUs).

In an analogous manner we defined that a neuron in our recurrent network belonged to the assembly for a particular input pattern if it satisfied the same firing rate criterion and significantly responded to this input pattern based on the same Wilcoxon rank-sum test. For this purpose, we ran an additional simulation phase without any plasticity (termed test phase 1) for around 325 s (see exact rules in Section 2.2.1) after the first two plasticity phases. Within this test phase, not only the blue, green, and red patterns, but also a combined pattern (blue and green) as used in the third plasticity phase (see Section 2.3.1) were presented to the network. We extracted the average firing rates for each excitatory neuron within the baseline interval [-100 ms, 0 ms] and response interval [10 ms, 110 ms] relative to the onsets of each blue, green, and red pattern presentation (termed trial; around 44 trials per pattern). Based on this, we computed the median firing rate in the response interval across all trials. Using a Wilcoxon rank-sum test ($p < 0.05$) we tested if the firing rate of a neuron across all trials was significantly higher in the response than in the baseline interval for the blue, green, and red input patterns. A neuron that showed a significant response to an input pattern and had a median firing rate in the response interval of at least 2 Hz was defined as assembly neuron for a particular (preferred) input pattern.

We refer to the neurons that belonged to the assembly for at least one of the three input patterns as pattern responsive units (PRUs). Neurons belonging to more than one assembly are referred to as multi-responsive units (MRUs).

The same statistical test was also applied to the trials with combined patterns during test phases 1 and 2 (see Section 2.3.2) in an analogous manner, in order to estimate the number of originally non-responsive units (i.e., units that did not significantly respond to any separate or the combined pattern) that became responsive to a separate or combined pattern after the third plasticity phase (i.e., after the formation of associations; see Section 3.3).

2.3 Details to emergent associations

2.3.1 Third plasticity phase for emergence of associations

During the third plasticity phase, a combined pattern was repeatedly presented to the network. The sequence of input patterns and noise periods was randomly generated according to the same rules as described in Section 2.2.1 for a duration of around 36 s, but using only the combined input pattern. However, in contrast to the previous phases, we assured that exactly 20 patterns were presented. Since the whole sequence of input

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

patterns and noise periods was precomputed beforehand at the beginning of the simulation, this could be done by repeating the random generation process until the number of input patterns was 20, resulting in an average duration of 36.8 s.

Here, we always combined the blue and the green pattern (the identity of which was arbitrarily assigned before as discussed above). The combined pattern was constructed by superimposing the frozen Poisson firing patterns of the 200 external input neurons for the two input patterns, again superimposed with fresh Poisson spike trains at 3 Hz. In other words, during a presentation of the combined pattern, an input neuron did spike at all spike times of the blue pattern, as well as at all spike times of the green pattern, and at additional spike times defined by a freshly and for this neuron individually generated 3 Hz Poisson spike train. STDP was continuously active for synapses between input and excitatory neurons as well as for recurrent excitatory synapses during this third phase of plasticity.

2.3.2 Definition of pair-coding units (PCUs)

PCUs were identified in the same way as described in the experimental procedures in (Ison et al. 2015), but with a baseline and response interval as defined in Section 2.2.2. Specifically, we ran an additional simulation phase without any plasticity (termed test phase 2) for around 325 s (see exact rules in Section 2.2.1) after the third plasticity phase. As in test phase 1, the blue, green, red, as well as the combined pattern (blue and green) were presented to the network, and the average firing rates for each excitatory neuron within the baseline and response intervals relative to the onsets of the blue, green, and red pattern presentations were extracted (around 44 trials per pattern). Again, we used a Wilcoxon rank-sum test ($p < 0.05$) to test if the firing rate of a neuron across all trials was significantly higher in the response than in the baseline interval for the blue, green, and red input patterns.

We only considered PRUs that had exactly one of the components of the combined pattern as their preferred (P) stimulus, i.e., the blue (green) input pattern for neurons belonging to the blue (green) assembly. Accordingly, the other component was defined as their non-preferred (NP) stimulus, i.e., the green (blue) input pattern for neurons belonging to the blue (green) assembly. The red input pattern (which was not part of the combined pattern) was defined as the non-associated (NA) stimulus for this subset of PRUs.

PCUs were defined as PRUs within this subset that had a non-significant response before (i.e., in test phase 1) and a significant response after (i.e., in test phase 2) the third plasticity phase to their NP stimulus. Additionally, single-trial increases after the third plasticity phase in the response intervals of the NP stimulus were required to be significantly larger (Wilcoxon rank-sum test; $p < 0.05$) than the ones of the NA stimulus.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Single-trial increases were computed as the firing rates in the response interval of a given stimulus during tests phase 2 minus the mean firing rates in the response interval of this stimulus over all trials during test phase 1. As an additional constraint, neurons which showed no longer a significant response to the P stimulus after the third plasticity phase were excluded (1 out of 23 PCUs identified otherwise).

2.4 Details to neuronal and functional learning curves

Behavioral performance of learned associations in (Ison et al. 2015) was tested during learning by showing an image and asking the participant to select the corresponding associated image from a list of images (Task 3). Similarly, in our network model, we investigated after which combined pattern presentation in the third plasticity phase the formed associations became functionally useful, in the sense that a downstream network could infer the associated pattern for a given input pattern. For this purpose, we ran stepwise simulations where increasing numbers of the combined pattern (from 0-20) were presented during the third plasticity phase, and observed the resulting numbers of PCUs, weight changes between the associated assemblies, and readout performance using two different kinds of readout strategies.

2.4.1 Readout A: Readout based-on spike counts

Spike counts from all neurons belonging to an assembly were extracted from the baseline interval of [-100 ms, 0 ms] and the response interval of [10 ms, 110 ms] relative to stimulus onsets of the blue, green, and red patterns during test phase 2 (i.e., after the third plasticity phase; see Section 2.3.2; around 44 trials per pattern). For each assembly, the mean spike count differences between the response and the baseline interval over all neurons belonging to the corresponding assembly were computed. When presenting a blue (green) input pattern, the spike count differences for its associated (green (blue)) and non-associated (red) assembly were compared, and the assembly with a higher spike count difference was selected as output of the readout. This readout can be seen as mimicking the task to select the associated image from a list of images in a multiple choice test in the experiments of (Ison et al. 2015). The total readout performance was estimated by computing the fraction of times (of all blue (green) trials) the correct (i.e., the associated) assembly was selected.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2.4.2 Readout B: Linear readout

For each of the three assemblies (red, blue, and green), a linear readout was trained with the standard algorithm for a support vector machine with parameter $C = 1$ to detect its activation in response to its preferred (P) input pattern. We used the LinearSVC implementation from the scikit-learn 0.18 library (Pedregosa et al. 2011). 432-dimensional feature vectors for this readout were extracted by taking the non-weighted PSPs (similar to $z_j(t)$ in Equation (3), but with $\tau_r = 4\,\text{ms}$, $\tau_f = 40\,\text{ms}$, $T_c = 200\,\text{ms}$) summed over time at time point 100 ms relative to pattern onsets from all excitatory neurons (i.e., without any prior knowledge about the actual assembly neurons). Each feature vector component was scaled to a range of [0, 1] (MinMaxScaler in the scikit-learn 0.18 library). Each classifier was then trained on 50 % of the data to distinguish target patterns (i.e., assembly activation at its preferred pattern presentation; around 22 presentations) from non-target patterns (around 44 presentations) during test phase 1 (i.e., after the first two phases of plasticity).

We then tested whether this readout was able to detect during test phase 2 (i.e., after the third phase of plasticity) an indirect activation of its corresponding assembly via its associated non-preferred input pattern (again, based on around 22 target and 44 non-target pattern presentations). To take unbalanced classes into account (there were always twice as many non-target as target patterns), the class weight of the smaller class was increased by a factor of 2, using the `class_weight` argument of LinearSVC. We used the *balanced accuracy* as performance measure which is defined as $bacc = \frac{TPR+TNR}{2}$, with TPR being the true positive rate and TNR the true negative rate. The TPR (TNR) is defined as the fraction between the number of samples that were correctly classified as positive (negative) examples and the total number of positive (negative) examples in the test set. Moreover, to increase the sensitivity of the readouts to detecting associations, the separating hyperplane was shifted towards the non-target class by decreasing the corresponding class weight by a factor of $1 \cdot 10^{-3}$.

2.5 Details to the functional impact of network parameters and choices of input patterns

2.5.1 Impact of network parameters

In Section 2.1, we described the network parameters and the chosen standard values we used throughout our simulations. One instantiation of this network setup is referred to as the *standard model*. We next investigated the impact of certain network parameters and the robustness of the results against variations of network initialization (i.e., specific network connectivity and initial parameter values) as well as the specific realization of

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

input patterns. To this end, we ran extensive computer simulations and varied one or two parameters at a time while keeping the standard values for all other parameters. To have comparable ranges of values despite the fact that different parameters have different scales and units, we used relative parameter changes with respect to the corresponding standard value. To cover a reasonably wide range of values, we used logarithmically spaced steps of 25 %, 50 %, 100 %, 200 %, and 400 % of the standard parameter values. In Figure 10A and B, 800 % was included too. Only for the connection probability between excitatory neurons we used absolute probability values in linearly spaced steps from 10 % to 90 %.

All simulations were done using 10 different global random seeds, having an effect on both network generation (e.g., connection matrix, initial parameter values) and the specific realization of input patterns. The average results over these 10 simulations are reported. Parameters that were investigated include synaptic delays between excitatory neurons d_{EE} (mean value; the standard value was 5 ms), inhibitory delays $d_{\mathrm{I}} := d_{\mathrm{EI}} = d_{\mathrm{II}} = d_{\mathrm{IE}}$ for E→I, I→I, and I→I connections respectively (mean value; standard value: 2 ms), f_0 (steady-state firing rate used for weight re-scaling; see Section 2.1.1; standard value: 5 Hz), g_{E} (fraction between the mean initial weight of E→E connections and the mean initial weight of I→E connections: $g_{\mathrm{E}} = \frac{\mu_{\mathrm{EE}}}{\mu_{\mathrm{IE}}}$; standard value: 1/550), w_{scale} (global scaling factor for initial weights of all connections between excitatory and inhibitory neurons; standard value: 1.0), $\alpha_{\mathrm{E,offset}}$ (excitability offset of excitatory neurons; standard value: $-250 \cdot 10^{-3}$), and $\alpha_{\mathrm{I,offset}}$ (excitability offset of inhibitory neurons; standard value: $-150 \cdot 10^{-3}$).

2.5.2 Impact of input patterns on assembly sizes and number of PCUs

In order to understand the variability of asssembly sizes and number of PCUs we investigated the impact of the random choice of input patterns on the resulting assembly sizes and numbers of PCUs. For this purpose, we ran extensive simulations with 20 random triples of input patterns (see Section 2.1.2) applied to a fixed random network (standard model).

We presented input patterns to the network as in plasticity phase 1 for 325 s in its initial state (i.e., before any synaptic plasticity has taken place). During this presentation period, STDP was not active in the network. The summed activity of the excitatory population in response to the different input patterns, estimated with a Gaussian kernel with $\sigma = 10$ ms, was extracted and averaged across all repetitions (around 59 repetitions per pattern). Figure 3 shows an example of three frozen input spike patterns applied to the network (panel A), and the resulting summed population activity in response to these three input patterns (panel B). A close similarity between the estimated spike densities of the frozen patterns (Figure 3A bottom) and the population responses can be observed, indicating that the initial network activity is driven by the input. Moreover, we

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

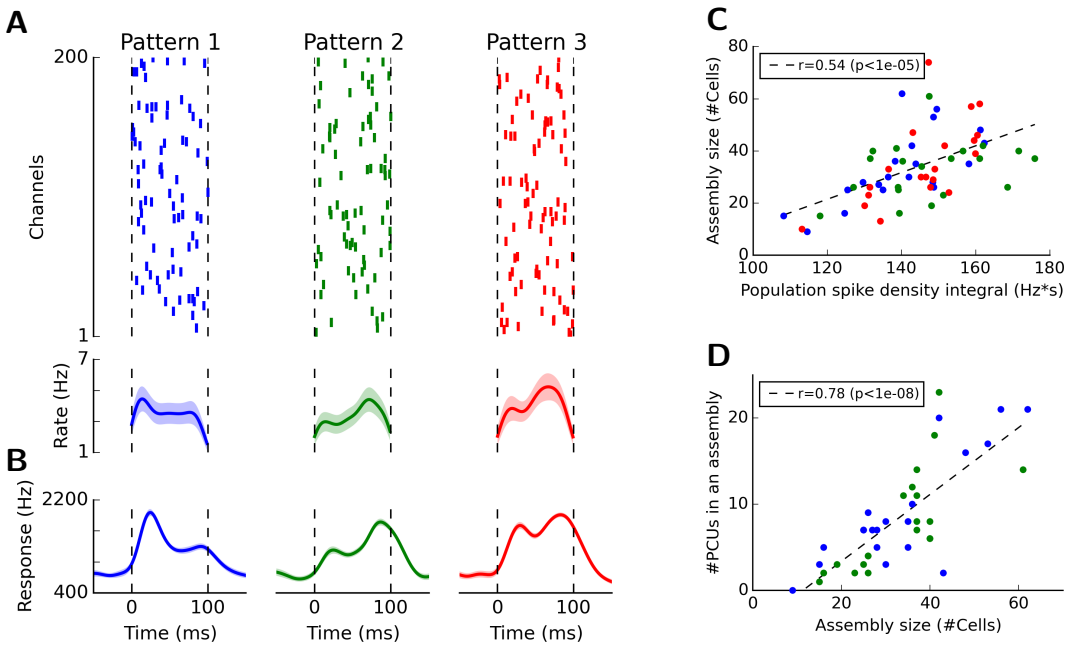

Figure 3: Dependencies between the network response to an input pattern and the resulting assembly size and number of PCUs. (A) Frozen spike patterns applied to the network. Poisson spike trains are drawn for each of the 200 input channels with an average firing rate of 4 Hz per channel. Below: Average spike densities (± SEM) of the frozen spike patterns estimated with a Gaussian kernel with $\sigma = 10$ ms. (B) Summed activity of the excitatory population in response to the three different input patterns shown in (A) in the initial state of the network, averaged over around 59 repetitions per pattern. Shaded areas represent the SEM. (C) Scatter plot showing the integral of the summed population activity from (B) in the interval 0-100 ms relative to pattern onset in response to the three different input patterns (shown in different colors) in the initial state (i.e., before any synaptic plasticity) versus the resulting assembly sizes after the first two phases of plasticity, over 20 simulations with different random input pattern variations. A significant linear correlation ($r = 0.54$, $p < 1 \cdot 10^{-5}$; dashed line) could be found. (D) Scatter plot showing the assembly sizes versus the resulting numbers of PCUs per assembly, again over 20 simulations with different random input pattern variations. A significant linear correlation ($r = 0.78$, $p < 1 \cdot 10^{-8}$; dashed line) could be found.

computed the integral of the summed population activity from Figure 3B in the interval 0-100 ms relative to pattern onset in response to the three different input patterns from all 20 simulations, and tested its correlation with the resulting assembly sizes (Figure 3C) and the resulting numbers of PCUs in an assembly (Figure 3D). We found significant correlations, and concluded that the observed variability of assembly sizes and number of PCUs was partially caused by the randomly chosen input patterns.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3 Results

The experimental data from (Ison et al. 2015) show how the coding properties of neurons in the human medial temporal lobe (MTL) change when a new memory, consisting of the combination of two already existing memory items, is formed. We examine under what conditions their experimental data can be duplicated in a neural network model. We do not consider artificial neural network models, since their predictions are more difficult to relate to actual recordings from neurons in the brain. Rather, we investigate models for generic recurrent networks of excitatory and inhibitory spiking neurons and synapses that are subject to data-based forms of short- and long-term plasticity: diverse combinations of paired pulse depression and facilitation as short-term plasticity, and spike-timing-dependent plasticity (STDP) as long-term plasticity. Neural network models of this type provide a basis for comparing experimental data on neural responses in the human brain directly with neural responses and in the model, since both have the form of spike trains. In addition, networks of spiking neurons have – in contrast to artificial neural network models – a biologically interpretable notion of time that can be directly related to temporal aspects of the neural recordings. Finally, an understanding of how new memories are formed in generic recurrent networks of excitatory and inhibitory spiking neurons paves the way for subsequently reproducing functional memory systems in larger data-based models for brain areas.

3.1 The network model

The results of (Ison et al. 2015) are based on recordings from 613 units in the human brain. These were pooled from four areas of the MTL in the human brain: hippocampus, entorhinal cortex, amygdala, and parahippocampus. In view of the diversity of the anatomy and physiology of these brain areas we took as basis for our simulations a generic model for a recurrent network of spiking neurons, consisting of 432 excitatory and 108 inhibitory neurons with random connectivity (see *Methods* for details). Details of the short-term plasticity of synaptic connections between excitatory neurons were modelled after data from the human brain (Testa-Silva et al. 2014). Since experimental data on the short-term plasticity of synaptic connections from and to inhibitory neurons in the human brain are currently lacking, we took there standard data from the literature on non-human vertebrates. We considered long-term plasticity only for synaptic connections from input to excitatory and between excitatory neurons. Synapses from and to inhibitory neurons are also subject to long-term plasticity processes. But in view of the diversity of induction protocols and resulting plasticity results and the dependence on specific types of inhibitory neurons (Kullmann et al. 2012) it is at present difficult to point to a generic rule for long-term plasticity of these synaptic connections that is strongly supported by experimental data. Hence we chose to consider in this study long-term plasticity only for connections from input to excitatory and between

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

excitatory neurons, and asked to what extent this simple and straightforward model for synaptic plasticity would enable us to reproduce experimental data on the emergence of assembly codes and associations between them.

Generation of input patterns: We chose as external inputs three firing patterns consisting of spike trains which were generated by 200 Poisson neurons, called input neurons in the following. The firing patterns consisted of frozen Poisson spike trains over 100 ms with a firing rate of 4 Hz (Figure 3A and Figure 4). We called these three input patterns the blue, green, and red pattern. These spike patterns were superimposed on each presentation to the network with noise: with freshly generated Poisson spike trains at 3 Hz in each of the 200 input channels. This caused a biologically realistic trial-to-trial variability of network inputs (see Section 2.1.2). In total, we generated 20 such triples of input patterns. One specific set of three patterns was used for most simulations throughout this article, and we refer to this triple as the *default patterns*.

Dependence of results on specific parameter settings: Several parameters that describe the response properties of a neuron in the brain depend on the specific neuron, its genetic class and its location in the brain, and in addition on the species. Other parameters of the network model depend on its size. We chose standard values for 8 of these parameters for the results that are reported in Figure 4 to 9. Subsequently we analyze in Figure 10 how these results depend on the values of these 8 parameters.

Finer details of the results depend in addition on the specific realization of the synaptic connectivity matrix that is drawn according to given connection probabilities, the choice of initial values of synaptic weights, of delays, and UDF values from given distributions, and on the specific realization of spike input patterns that serve as inputs to the network. The results that are reported in Figures 4 to 6, 8, and 9 are all based on one specific network (with fixed connectivity and fixed initial parameters), and we refer to this network as the *standard model*. In these simulations, we also used one fixed triple of input patterns (the default patterns, see above). We analyze in Figures 3, 7, and 9 how the results for the standard model depend on the specific realization of input patterns. We consider in Figure 10 randomly generated networks (with newly drawn connectivity and initial parameters) and input patterns for a large range of parameter settings.

3.2 Generation of memory traces

The experiments of (Ison et al. 2015) studied fast formation of associations between two unrelated memory items, images of familiar faces and landscapes. These had previously become encoded through the firing of a sparse distributed set of concept cells, often referred to as a memory trace or an assembly of neurons. Hence we first need to generate in our model corresponding assemblies of neurons as neural codes for repeatedly occurring salient external inputs.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

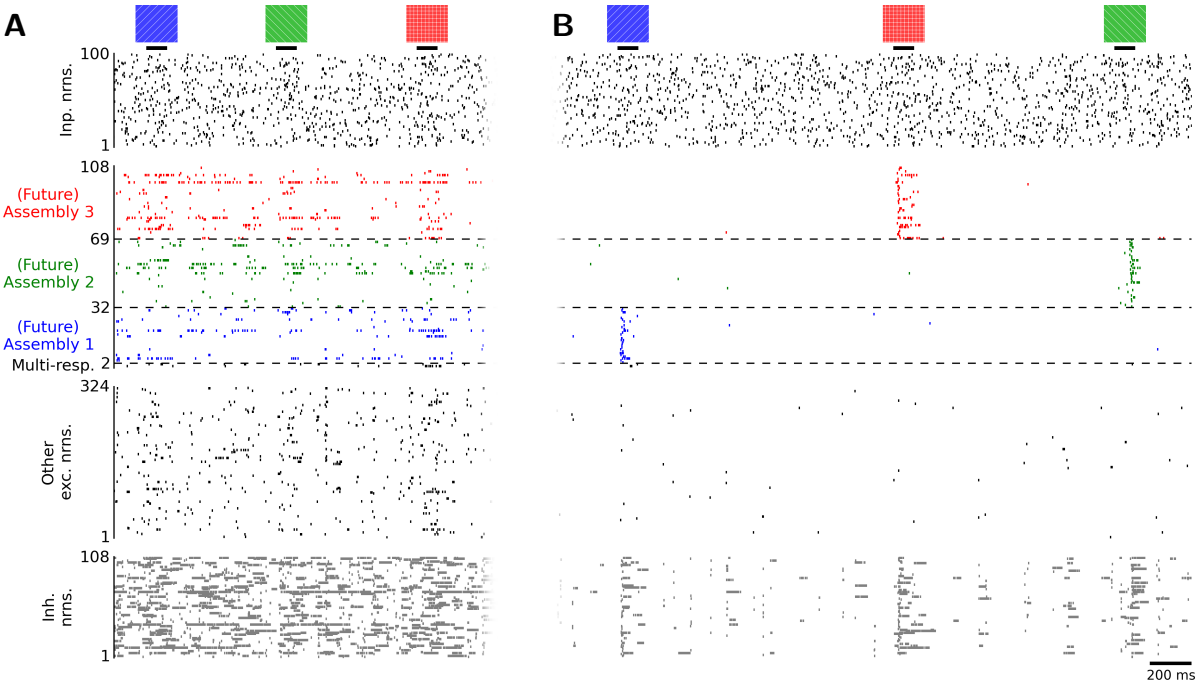

Figure 4: (A) Initial network activity, before any synaptic plasticity took place. The excitatory neurons are grouped into the assemblies to which they will belong after the first two plasticity phases. (B) Emergence of assemblies after the first two phases of plasticity. Neurons in the three assemblies responded preferentially to one of the three input patterns. Spike trains are shown (from top to bottom) for input neurons (only the first 100 of 200 are shown), assembly neurons, multi-responsive units (responding to more than one input pattern), other excitatory neurons, and inhibitory neurons. Note the different y-axis scalings to highlight neurons in assemblies.

We modelled the formation of memories, i.e., assemblies of neurons, for each of the three input firing patterns by presenting the three patterns repeatedly to the network at random time points within a continuous spike input stream over 500 s, where the occasionally occurring three input patterns with superimposed noise were interleaved with periods of pure noise inputs: freshly generated Poisson spike trains at 5 Hz, with randomly drawn durations between 0.5 s and 3 s, see the top row of Figure 4.

We created two plasticity phases of 250 s each, that mimic the complex process of the formation of memory traces in the MTL in a simple manner. In the first plasticity phase synapses from external input neurons to excitatory neurons in the recurrent network were subject to STDP. During the second plasticity phase STDP was applied to synapses between excitatory neurons in the recurrent network. The first phase of plasticity may be viewed as corresponding roughly to attention-gated plasticity in the brain for salient new sensory stimuli, while the second plasticity phase for synapses within the recurrent network mimics in a simple way internal replay and consolidation processes of the brain. We refer to the assemblies of neurons in the network that preferentially responded after these two plasticity phases to one of the three input patterns defined as in (Ison et al.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

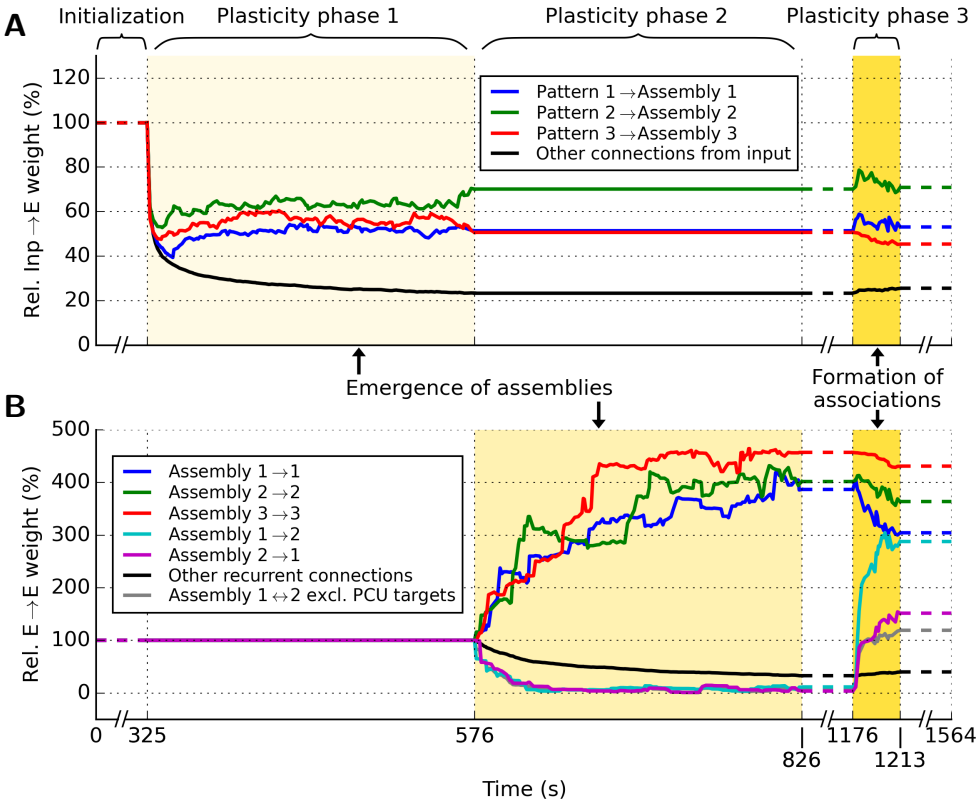

Figure 5: Mean relative weight changes of (A) connections from input to the three assemblies, taking the actual frozen input patterns into account (see text) and (B) internal connections within the three assemblies during all phases of the simulation. At the beginning, an initialization phase (0-325 s) without STDP was used. After that, two plasticity phases with STDP applied to only Inp→E (Plasticity phase 1; 325-576 s) or E→E synapses (Plasticity phase 2; 576-826 s), followed by a phase without plasticity (826-1176 s) were simulated. In the third plasticity phase (Plasticity phase 3; 1176-1213 s), both Inp→E and E→E synapses were subject to STDP, which was followed by another phase without plasticity (1213-1564 s). Colored lines indicate the assembly connections as indicated by the legends. Black lines denote all other connections, i.e., cross-connections between other pairs of assemblies and connections not related to any assembly. The gray line shows that also weights to non-PCUs increase during the third plasticity phase. Yellow background shadings indicate plasticity periods of (A) input and (B) recurrent excitatory connections.

2015), (see *Methods*) as the blue, green, and red assembly (see second row of Figure 4). Each of the three input patterns was presented on average 88 times to the network during the two plasticity phases of altogether 500 s. The evolution of synaptic weights under STDP of synapses from input neurons to neurons in the three assemblies and within the three assemblies are plotted in Figure 5A and in the upper three traces of Figure 5B. Input weights from a pattern k to an assembly k were computed by taking only input channels with at least one spike in pattern k into account (which was roughly one third of the input channels); $k = 1, 2, 3$.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The three assemblies that emerged were rather sparse and distributed over the whole 3D volume of the network, with their means approximately centered at the center of the volume.

Pattern responsive units (PRUs) were defined as neurons that significantly responded to at least one of the three input patterns. The previously described plasticity process gave rise to 108 PRUs, with assemblies of sizes 30, 37, and 39 for the three default input patterns. In the human MTL only a very small fraction of neurons belongs to assemblies for two or more unrelated memory items (De Falco et al. 2016). This finding was reproduced by our model; we identified after the first two phases of plasticity only two neurons as multi-responsive units (MRUs), i.e., they belonged to more than one assembly.

The emergence of assemblies as memory traces in networks of spiking neurons has already previously been modelled (Klampfl and Maass 2013, Litwin-Kumar and Doiron 2014, Zenke et al. 2015). In comparison with these preceding models we used here a simpler model that did not require a specific connectivity structure or long-term plasticity of inhibitory synapses. The neural recordings of (Ison et al. 2015) show that the firing activity of neurons that belong to an assembly tend to return to baseline activity soon after the stimulus that evoked this memory trace has been removed. This feature is duplicated in our model, see Figure 4B.

The assembly sizes and numbers of PRUs and MRUs depend on the random choice of the input patterns (and the network parameters, see Section 3.5). Twenty simulations with different random input pattern triples yielded average assembly sizes of 33.8 ± 13.4 SD (min: 9; max: 74). The total number of PRUs was found to be 105.2 ± 20.8 SD (min: 72; max: 148), the number of MRUs 3.7 ± 5.8 SD (min: 0; max: 27). Relative differences between the three assembly sizes were measures by computing their standard deviation for each simulation individually, which resulted in a mean standard deviation of 10.3 over all 20 simulations.

3.3 Emergence of associations

Associations between memory items were created in the experiments of (Ison et al. 2015) by repeatedly presenting combinations of two images from the fixed set of images, more precisely a face in front of a landscape. As a result, the assemblies for the two image components expanded: each recruited additional neurons from the other assembly. We wondered, whether analogous changes of assembly codes would emerge in our model. We repeatedly presented combinations of two of the previously used external input patterns, the blue and the green pattern of Figure 4. This combined pattern was constructed by superimposing (adding) the frozen Poisson firing patterns of the 200 external input neurons for the two input patterns, again superimposed with fresh Poisson spike trains

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

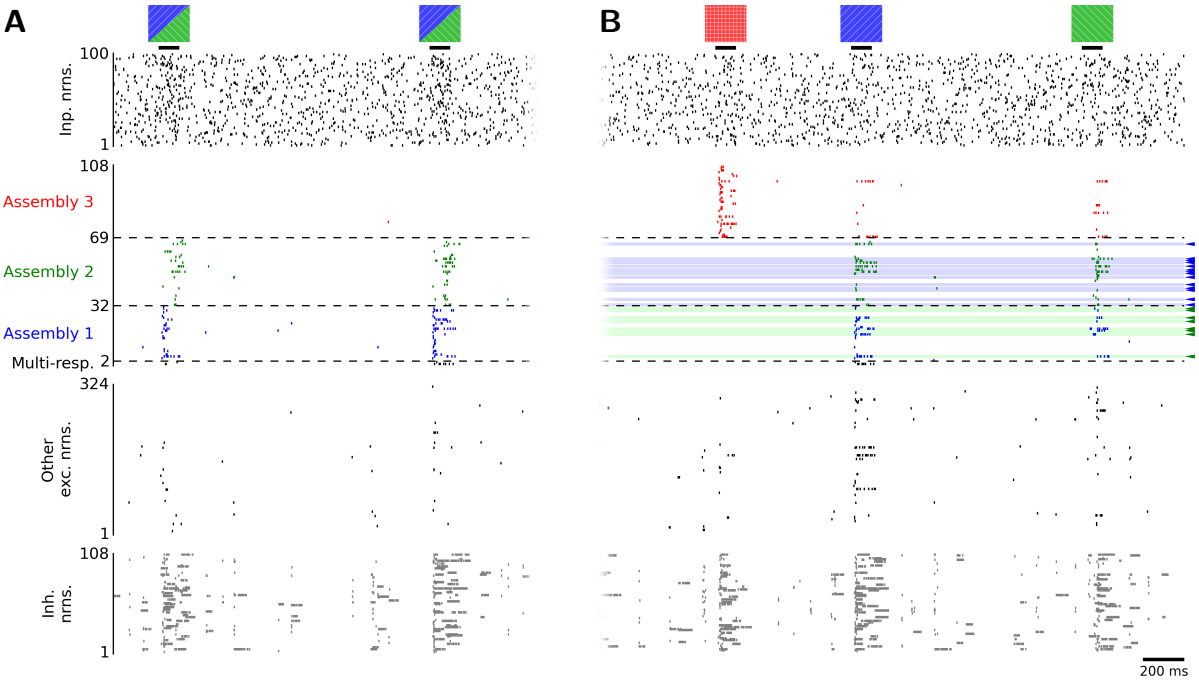

Figure 6: Emergent neural code for associations in the model. (A) In a third plasticity phase, the network was exposed to a combination of two previously shown spike patterns (blue and green). (B) After 20 presentations of this combined input pattern, the green input pattern also activated a fraction of neurons from the assembly that encoded the blue input pattern, and vice versa. Such pair-coding units (PCUs; Ison et al. 2015) are indicated by a shaded background and small arrows on the right side.

at 3 Hz. The resulting combined input pattern with superimposed noise was presented 20 times at random time points within 36 s (Figure 6A). STDP was continuously active for synapses between input and excitatory neurons as well as for recurrent excitatory synapses during this third phase of plasticity. Figure 5B shows that weights between the two associated assemblies were rapidly increasing in this phase.

We found that 8 neurons in the blue assembly (indicated by the green horizontal lines in Figure 6B) became now also members of the green assembly, and 14 members of the green assembly (indicated by the blue horizontal lines in Figure 6B) became also members of the blue assembly; according to the previously specified criterion for membership in an assembly. In analogy to the terminology of (Ison et al. 2015) we refer to these neurons in the overlaps of the two assemblies as pair-coding units (PCUs).

Instead of 21 out of 51 neurons in their data that preferred one of the two image components and responded after learning with an increased firing rate also to the non-preferred (NP) stimulus, we found in our network model that 22 out of 67 neurons (8 in assembly 1 and 14 in assembly 2, for the default input patterns) in the blue and green assembly responded after the third plasticity phase with an increased firing rate to the NP

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

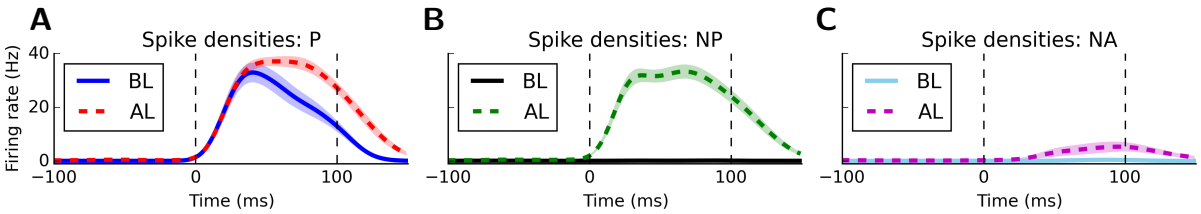

Figure 7: Mean spike densities of all PCUs (estimated with a Gaussian kernel with $\sigma = 10\,\text{ms}$) before (BL) and after (AL) learning, i.e., the third phase of plasticity, over 13 simulations with random variations of the input patterns which had sufficiently large assembly sizes (see selection criterion in text) in response to their (A) preferred (P), (B) non-preferred (NP), and (C) non-associated (NA) pattern. The stimulus onset was at $t = 0\,\text{ms}$, dashed vertical lines indicate the pattern presentation period. Shaded areas represent the SEM. After learning, a significantly increased firing rate in response to the NP stimulus, but not to the NA stimulus could be observed. Compare with Figure 5A-C of (Ison et al. 2015).

stimulus. This fraction of around 33 % is not a constant of our model, but depends on several parameters and accidental features, such as input patterns, noise rates, network connectivity, and plasticity parameters.

We again assessed the stability of the results for different random input patterns applied to our standard model. We selected simulations which had sufficiently large assemblies for reporting statistics on PCUs. A minimum assembly size of at least 5 % of the population (22 neurons) was required, which excluded 7 of the 20 earlier simulations. In the remaining 13 simulations, we found an average number of 18.0 ± 7.8 SD (min: 4; max: 32) PCUs. We found that after the formation of associations, 19.1 ± 5.8 (SD) previously non-responsive units became responsive to a single input pattern and 11.2 ± 3.9 (SD) units to the combined pattern respectively. These units were identified by comparing the numbers of pattern responsive units before versus after the formation of associations, and including not only responses to the three separate but to the combined pattern as well. Over all 20 simulations we found a significant linear correlation between the assembly sizes and the resulting numbers of PCUs per assembly ($r = 0.78$, $p < 1 \cdot 10^{-8}$), as shown in Figure 3D.

Since the neurons in our model produce spike trains, we can directly compare changes in firing responses of neurons before and after the induction of the association between the data (Figure 4D of Ison et al. 2015) and our model (see Figure 7). We employ here the terminology of (Ison et al. 2015), where one says that the blue input pattern is the preferred ("P") stimulus for neurons in the blue assembly, whereas the green pattern is the non-preferred ("NP") stimulus for neurons in this assembly (analogous for the green assembly).

We find that the neural coding properties of PCUs in the blue and green assemblies change through repeated representations of combined input patterns in a way that is very similar to the data of (Ison et al. 2015): The firing response remains significant

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

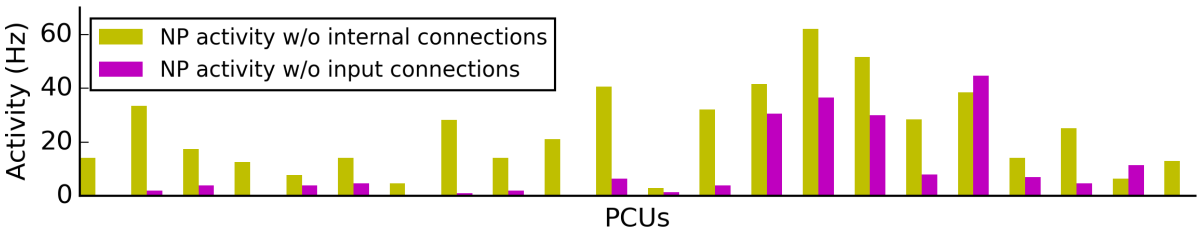

Figure 8: Sources of synaptic inputs to PCUs for the NP stimulus. Yellow bars: Firing rates of PCUs after the internal connections from the assembly of the NP stimulus were disabled. Magenta bars: Firing rates after the input connections were disabled.

for the preferred stimulus, changes from insignificant to significant for the non-preferred stimulus, and remains insignificant for the red ("non-associated" or "NA") input pattern that was not part of the combined stimulus (Figure 7). This increased firing rate was specific to the input patterns that participated in the combined stimulus, and no increased firing rate was observed for the non-associated input pattern (Figure 6B and Figure 7C).

A neuron can become in two different ways a member of the assembly for the previously non-preferred stimulus during the presentation of the combined patterns: By increasing its weights from input neurons that are highly active during the NP stimulus, or by increasing its weights from neurons in the assembly for the NP stimulus. A latency analysis in (Ison et al. 2015) arrived at the conclusion that a combination of both effects occurred. In our model we can ask directly how much synaptic input a neuron that starts to respond to the NP stimulus after the association induction gets from the external input neurons, and how much from the original assembly for the NP stimulus. So we ran simulations where (i) all internal connections from the assembly of the NP or (ii) all connections from the input were disabled. The resulting PCU activities in response to the NP simulus can be found in Figure 8. The figure shows that also in the model a combination of both effect occurs, with the contribution from the external input neurons being somewhat stronger.

3.4 Neuronal and functional learning curves

A key point of the experimental data of (Ison et al. 2015) was that the overlap of assemblies emerged at about the same trial when the association between the two memory traces became functional, i.e., when the subject was able to select in an interjected multiple choice test the correct background that had previously been shown in conjunction with a face.

We asked whether the same effect would occur in our model, i.e., whether the overlap between the associated blue and green assembly would emerge at about the same trial

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

when a downstream network would be able to detect a functional association between the two assemblies. In addition, we were able to investigate in the model a question that could not be probed through recordings from the human MTL: will weights of synaptic connections between neurons in the two assemblies increase significantly through the association process, and will a significant weight increase appear at about the same time as the overlap between the two assemblies? If this is the case, it suggests that besides the emergent overlap also this weight increase is related to the emergent functionality of the association. We chose as models for this readout the two arguably simplest options:

Readout A: Integration of evidence by counting spikes in the associated and the non-associated assembly, with a subsequent symbolic multiple choice test where the assembly with the higher spike count is selected as associated assembly.

Readout B: Linear readout neurons are trained for each of the assemblies to fire whenever this assembly is activated through the corresponding external input. Then we test whether the readout neuron for the green assembly (with the same weights) signals that the green assembly is activated when – after association pairing – the external input for the blue assembly is injected.

The results are shown in Figure 9. Solid curves show averages over 13 experiments with random input pattern variations applied to the standard model. Small circles show results for the default input patterns. One finds a strong correlation between the size of the overlap between the two associated assemblies (panel A), the sum of weights of synaptic connections between the two assemblies (panel B), and the performance of readouts A and B (panels C and D). Readout A, which models a multiple choice test, shows good performance after just two presentations of the combined external stimulus, and both the size of assembly overlaps and the sum of weights of synaptic connections also exhibit a significant increase after two trials. Whereas readout A jumps after two trials immediately to its final performance level, similarly as in Figure 5 of (Ison et al. 2015), one sees a further increase in overlap sizes, sum of weights of synaptic connections, and the performance of readout B during the subsequent 18 trials. Hence also in our simple neural network model the size of the overlap between assemblies reflects on the level of neural coding the current functionality of the association quite well. In addition, the average of weights of synaptic connections between the two assemblies is also highly correlated with the current functionality of the association.

One should point out that just carrying out the interleaved tests of associations by the human subjects (interleaved Tasks 2 and 3 in Ison et al. 2015) may have enhanced learning processes beyond the learning from passive presentation of combined stimuli. This may have sped up the learning of associations in (Ison et al. 2015), but is difficult to reproduce with a simple readout model.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

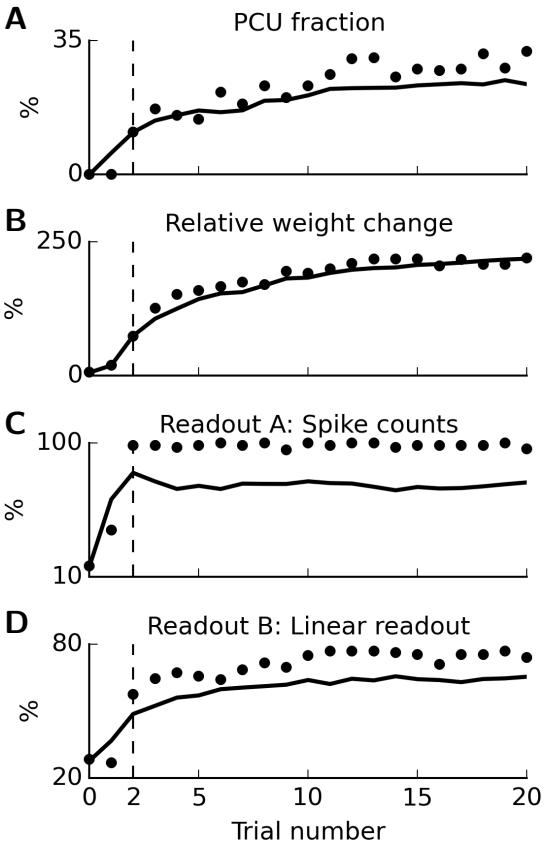

Figure 9: Neuronal and functional changes in the recurrent network during 20 presentations ("trials") of the combined input pattern (trial 0: initial state). Solid lines represent average results over 13 simulations with different random input pattern variations. Markers depict results for the default input patterns. (A) Mean fraction of PCUs relative to their corresponding assembly sizes. (B) Mean relative weight changes of connections between the two associated assemblies. (C) Mean readout performance using spike counts. (D) Mean readout performance using a linear readout.

3.5 Functional impact of network parameters and choices of input patterns

As described in Section 3.1, the previously discussed results of the model were based on a standard setting for 8 parameters that affect network activity, more specifically the balance between excitation and inhibition. These parameters affect the size of the network response to single and combined external stimuli, and thereby also the impact of STDP for synaptic connections between excitatory neurons. In this way they also affect the number of PRUs, i.e., the number of neurons that become members of an assembly, and the number of PCUs, i.e., the resulting size of the overlap of the two assemblies for which combinations of the corresponding patterns had been presented. Table 3 summarizes these 8 selected network parameters and their qualitative impact

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Symbol	Parameter description	Standard value	Impact of value increase on #PRUs	Impact of value increase on #PCUs
d_{EE}	Mean synaptic delays for E→E connections	5 ms	decreasing	decreasing
d_{I}	Mean synaptic delays for E→I, I→I, and I→I connections	2 ms	increasing	increasing
f_0	Steady-state firing rate used for weight re-scaling	5 Hz	increasing	increasing
g_{E}	Inverse scaling factor between inhibition and excitation	1/550	increasing	increasing
$\alpha_{\mathrm{E,offset}}$	Excitability offset of excitatory neurons	$-250 \cdot 10^{-3}$	mixed	mixed
$\alpha_{\mathrm{I,offset}}$	Excitability offset of inhibitory neurons	$-150 \cdot 10^{-3}$	increasing	increasing
w_{scale}	Global scaling factor for initial weights	1.0	mixed	mixed
p_{EE}	E→E connection probability	50 %	increasing	saturating

Table 3: List of selected network parameters with a short description and their standard values, together with their qualitative impact on the resulting numbers of PRUs and PCUs for increasing parameter values. In detail, "increasing" ("decreasing") means that the number of PRUs or PCUs tends to increase (decrease) if the corresponding parameter is increased, while in case of "saturating" an increase can be seen first which seems not to exceed a certain upper limit. "Mixed" means that no clear trend can be observed.

on the resulting numbers of PRUs and PCUs if their values are changed. Details of the functional impact of network parameters can be seen in Figure 10.

Figure 10A and B show the dependence of the number of PRUs and PCUs on two of the parameters: g_{E} regulates the scaling between recurrent excitation and inhibition, and $\alpha_{\mathrm{E,offset}}$ regulates the excitability of excitatory neurons. The black star marks the standard values of these two parameters. One sees that the dependence of the number of PRUs and PCUs on these parameters is highly correlated. Figure 10C and D depict their dependence on 7 of the parameters in logarithmic scale. One sees that their number peaks roughly for the standard setting of $\alpha_{\mathrm{E,offset}}$ and w_{scale} (standard setting = 100 %), whereas both numbers tend to grow with the parameters d_{I}, f_0, g_{E}, $\alpha_{\mathrm{I,offset}}$, and are negatively correlated with d_{EE}. Altogether these control experiments show that reasonable assembly sizes and sizes of overlaps for associate assemblies emerge for a wide range of parameter settings. Figure 10E shows that the number of PRUs and PCUs are significantly correlated over all settings of these 7 parameters that we have tested.

Figure 10F shows that the number of PRUs is also strongly correlated with the connection probability between pairs of excitatory neurons in the recurrent network, whereas the fraction of PCUs remains in the range between 10 % and 15 % for a large range of

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

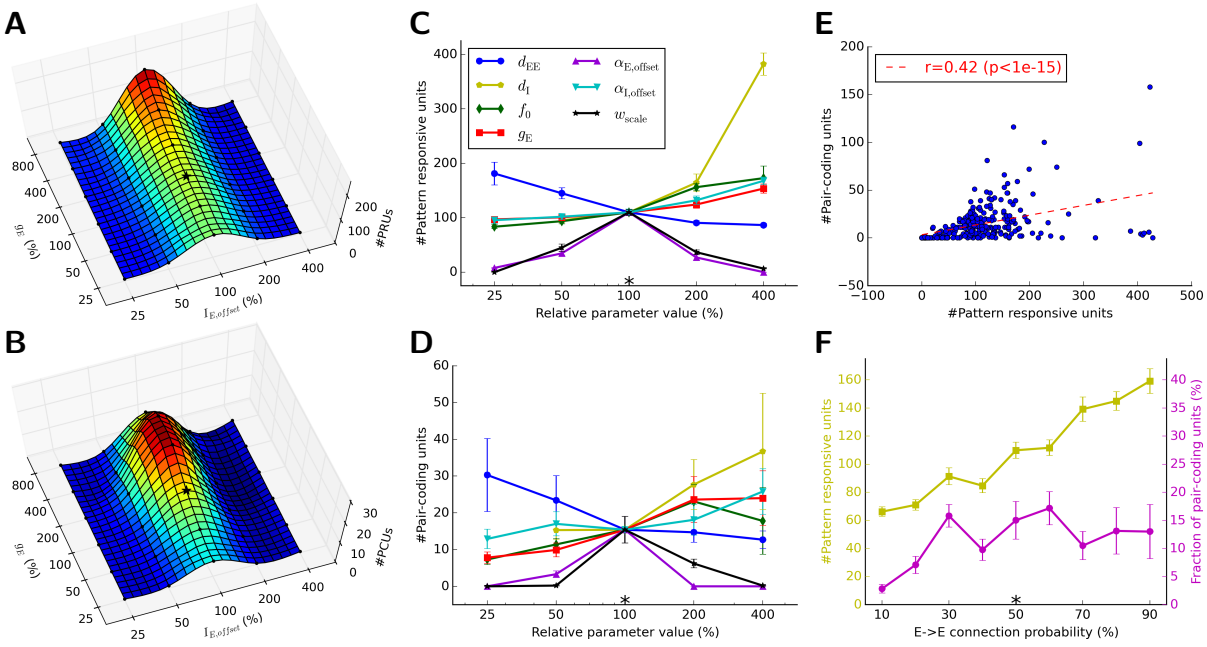

Figure 10: Impact of various network parameters on the numbers of PRUs and PCUs. Mean values were estimated over 10 simulations with different random seeds. Error bars represent the SEM. The * symbols mark the standard parameter values. (A, B) Mean numbers of PRUs and PCUs when parameters g_{E} and $\alpha_{\mathrm{E,offset}}$ are co-varied in logarithmic steps from 25 % up to 800 % relative to their respective standard values, estimated over 10 simulations with different random seeds. Intermediate points were interpolated. (C, D) Mean numbers of PRUs and PCUs (estimated over 10 simulations) for variations of the network parameters d_{EE}, d_{I}, f_0, g_{E}, $\alpha_{\mathrm{E,offset}}$, $\alpha_{\mathrm{I,offset}}$ and w_{scale}. The parameter values were varied in logarithmic steps from 25 % to 400 % relative to the respective standard values. (E) Over all parameter tests shown in (C, D), the number of PCUs was found to be positively correlated with the number of PRUs ($r = 0.42$, $p < 1 \cdot 10^{-15}$). (F) Mean number of PRUs (yellow) and PCUs (magenta) for different levels of recurrent excitatory connectivity. The number of PRUs was found to be increasing with connectivity while an increased number of PCUs can be seen for connectivities of 30 % and above.

connection probabilities. A proper choice of this connection probability is difficult for a model, because its impact on the number of PCUs depends on the size of the neural network model. From the functional perspective it is essential from how many neurons in the same and the associated assembly a neuron in an assembly receives synaptic connections. If this number, which depends on the connection probability and the network size, is too low, few PCUs are likely to emerge.

A simple calculation shows the following for the subarea CA3 of the hippocampus, which is estimated to consist of 2.83 million pyramidal cells in humans (Andersen et al. 2007) and has an estimated connection probability between pyramidal cells of 0.92 % in rodents (Guzman et al. 2016): An assembly for a memory item in the human MTL was estimated to consist of between 0.2 and 1 % of the pyramidal cells in the MTL (Waydo et al. 2006).

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

If one assumes that this also holds for area CA3, one arrives at an estimate of 5660 to 28300 for the number of neurons in an assembly in area CA3. Thus with a connection probability of 0.92 %, each neuron in one assembly receives on average synaptic input from 52 to 260 neurons in any other assembly. This suggests that models of different network sizes should have a connection probability that scales this number of presynaptic neurons from another assembly into a comparable range, so that they can contribute significantly to its firing probability. In our small neural network model this number of presynaptic neurons from another assembly had an average value of 18, but in order to achieve that we had to increase the connection probability between excitatory neurons to an unrealistically large value of 50 %.

As described in Section 3.1, the results of the model also depend on the particular random seed for the generation of the input patterns, for the small neural network size that we are considering. An analysis of this effect is presented in Section 2.5.2 in *Methods*. Figure 3A-C therein shows that the firing response that a randomly selected input pattern has in the recurrent network before any synaptic plasticity is significantly correlated with the assembly size that emerges for this external input pattern through STDP ($r = 0.54$, $p < 1 \cdot 10^{-5}$). This also affects the number of PCUs (Figure 3D). Brains may use additional control mechanisms that regulate the size of network responses to arbitrary stimuli. We propose that this normalization of response sizes also normalizes the sizes of memory traces that emerges when a stimulus becomes significant or occurs repeatedly.

4 Discussion

We have shown that associations between memory traces, expressed through assemblies of neurons, emerge through STDP in a generic recurrent network of excitatory and inhibitory spiking neurons. Furthermore, these associations become expressed through the emergence of overlaps between the assemblies, corroborating recent results of neural recordings from the human MTL (Ison et al. 2015). These overlaps emerge in about the same trial when the association becomes functional from a computational perspective, i.e., when a readout detects the indirect activation of an associated assembly. This result also agrees with the experimental data from (Ison et al. 2015).

In the model we could monitor in addition the evolution of synaptic weights between the two assemblies during the emergence of their association. We found that the average of these weights increases simultaneously with the overlap between the two assemblies. We propose that these increased weights, that also concern neurons that do no come to lie in the overlap of the two assemblies, contribute to the emergence of the recall capability. Future experiments will have to test to what extent this is valid for the brain.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

On the side we have also introduced a new method for modelling the emergence of assemblies in generic recurrent networks of excitatory and inhibitory neurons, that just relies on the plasticity of synapses between excitatory neurons, but assumes that this plasticity is gated through saliency. No special circuit connectivity like in (Klampfl and Maass 2013) was required for that. This method appears also to be of interest for creating models for neocortical circuits that exhibit a rich repertoire of assemblies and assembly sequences as found in neural recordings (Luczak and MacLean 2012).

Although we have made an effort to base parameters of the model on data from neurophysiology and neuroanatomy, there remained some parameters whose value for the human MTL can not be extracted from experimental data and scaled to our small-size network. We have investigated how the emergence of assemblies as memory traces and associations between these assemblies depend on those parameters. We found that the sizes of the assemblies and the sizes of the overlaps between them are correlated, and vary monotonously with several of the parameters. The values of some of these parameters are likely to depend on the network activity and concentrations of neuromodulators, and may be regulated by the brain in a flexible manner, thereby providing a possible method for regulating the sizes of assemblies and strengths of associations in an adaptive manner. The experimental data from (Waydo et al. 2006, De Falco et al. 2016) suggest that the sizes of memory traces and overlaps of associated memory traces do in fact vary on the time scale of days and longer. For example, overlaps between associated memory traces were reported to consist first of about 40 % of the neurons in the two assemblies, and shrink to about 4 % in the long term (De Falco et al. 2016).

A nice overview of different types of abstract models for the formation of associations is given in (Kahana et al. 2008, Kahana 2012). One type of models proposes that during the association of memory items a new memory trace is created, together with links between the memory items and the new memory trace for their combination. Recall of a memory item from an associated memory item proceeds in these models via a detour over the new representation of the combined memory. Closely related are abstract models which are based on the idea that each memory item is represented by a vector in a high-dimensional space, and that the association between two memory items is carried out through an algebraic operation on the vectors for the memory components, such as vector addition, concatenation, or convolution (Fodor and Pylyshyn 1988, Smolensky 1990, Kanerva 1994, Rizzuto and Kahana 2001, Plate 2003). One possibility for relating these abstract models to neural networks is to assume that each vector in the high-dimensional space is represented via population coding through the activity of neurons that each have some preferred direction in this high-dimensional space. Their joint activity can then represent a vector in the high-dimensional space as a weighted sum, where the preferred direction of each neuron is multiplied by its current firing rate (Eliasmith 2013). This neural coding assumption implies that the activity of neurons for each memory item (vector) does not change through the formation of an association, but that a new neural activity pattern is created that represents the combined memory. Also mathematically tractable models for the formation of combined memories through an operation called

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

JOIN (Valiant 2000a;b; 2005) or PJOIN (Papadimitriou and Vempala 2015) have these properties.

In contrast, the neural recordings from the human MTL (Ison et al. 2015) show that the neural representations of the memory items (assemblies) themselves change during the formation of an association through the formation of overlaps. On the other hand only 3 out of 641 recorded units became in their data freshly recruited for a combined memory. Hence if new memory traces are formed for combined memories in the human brain, they are likely to be represented primarily by neurons outside of the MTL. These experimental data are more consistent with a second type of memory models that were referred to as chaining or buffer models in (Kahana et al. 2008, Kahana 2012). Characteristic for these models is that links between the memory items are strengthened during the formation of an association, and that no new neural representation is created as trace for the combined memory. The neural network model that we are proposing is a model of this type.

From a more general perspective one should keep in mind that there exists a fundamental difference between a simple model for a neural circuit, such as the one we have considered, and the complexity and diversity of a family of neural circuits in the brain that are involved in the formation of memory associations. In addition, experimental studies of associations in the brain usually involve associations between memory items that belong to specific semantic categories and are in specific spatial, temporal, or semantic relations to each other. For example, one memory item may precede the other in a sequential presentation, or one may represent a temporal or spatial context for the other (Eichenbaum 2017). In addition, different areas of the human MTL are known to play specific roles in memory organization and retrieval. In order to reflect these additional aspects of the formation of associations, one has to expand the simple neural network model that we have presented into one where different areas of the MTL and their role in the representation of time and space are reflected. In addition, the complex relationship between the MTL and the neocortex, as well as large-scale dynamic activity pattern such as ripples and oscillations (Ritter et al. 2015), need to be taken into account. But we would like to argue that such more complex future models will benefit from the fundamental insight about links between memory function and the physiology of generic neurons and synapses that emerges from our simple model. In addition, a new generation of more abstract theoretically tractable neural network models for memory associations that are consistent with newer experimental data can be based on this simple model.

The intricate web of associations between memory items in the human brain and its capability to continuously integrate new information into this web are truly amazing from a computational perspective. Hence insight into the organization of this system is not only of primary importance in brain research, but may also provide guidance for the organization of new memory systems in future neuromorphic hardware, where memory storage and processing units are co-located in order to achieve a brain-like superior energy efficiency.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

Written under partial support by the Human Brain Project of the European Union #604102 and #720270, and the Austrian Science Fund (FWF): I 3251-N33.

References

P. Andersen, R. Morris, D. Amaral, T. Bliss, and J. O'Keefe. *The Hippocampus Book*. Oxford University Press, 2007.

D. Aronov, R. Nevers, and D. W. Tank. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719–722, 2017.

B. V. Atallah and M. Scanziani. Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. *Neuron*, 62(4):566–577, 2009.

G. Calfa, W. Li, J. M. Rutherford, and L. Pozzo-Miller. Excitation/inhibition imbalance and impaired synaptic inhibition in hippocampal area CA3 of Mecp2 knockout mice. *Hippocampus*, 25(2):159–168, 2015.

E. De Falco, M. J. Ison, I. Fried, and R. Quian Quiroga. Long-term coding of personal and universal associations underlying the memory web in the human brain. *Nature Communications*, 7:13408, 2016.

H. Eichenbaum. Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron*, 44(1):109–120, 2004.

H. Eichenbaum. Memory: Organization and control. *Annual Review of Psychology*, 68: 19–45, 2017.

C. Eliasmith. *How to build a brain: A neural architecture for biological cognition*. Oxford University Press, 2013.

J. Eppler, M. Helias, E. Muller, M. Diesmann, and M.-O. Gewaltig. PyNEST: A convenient interface to the NEST simulator. *Frontiers in Neuroinformatics*, 2(12), 2009.

J. A. Fodor and Z. W. Pylyshyn. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1):3–71, 1988.

M.-O. Gewaltig and M. Diesmann. NEST (NEural Simulation Tool). *Scholarpedia*, 2 (4):1430, 2007.

A. Gupta, Y. Wang, and H. Markram. Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science*, 287(5451):273–278, 2000.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

S. J. Guzman, A. Schlögl, M. Frotscher, and P. Jonas. Synaptic mechanisms of pattern completion in the hippocampal CA3 network. *Science*, 353(6304):1117–1123, 2016.

B. Haider, M. Häusser, and M. Carandini. Inhibition dominates sensory responses in awake cortex. *Nature*, 493(7430):97, 2013.

M. J. Ison, R. Quian Quiroga, and I. Fried. Rapid encoding of new memories by individual neurons in the human brain. *Neuron*, 87(1):220–230, 2015.

R. Jolivet, A. Rauch, H.-R. Lüscher, and W. Gerstner. Predicting spike timing of neocortical pyramidal neurons by simple threshold models. *Journal of Computational Neuroscience*, 21(1):35–49, 2006.

Z. Jonke. Stochastic computations and learning in networks of spiking neurons: Simulation framework, analysis and theory, PhD thesis, Graz University of Technology, 2013.

S. A. Josselyn, S. Köhler, and P. W. Frankland. Finding the engram. *Nature Reviews Neuroscience*, 16(9):521, 2015.

M. J. Kahana. *Foundations of Human Memory*. Oxford University Press, 2012.

M. J. Kahana, M. W. Howard, and S. M. Polyn. *Associative retrieval processes in episodic memory*. semanticscholar.org, 2008.

P. Kanerva. The spatter code for encoding concepts at many levels. *In* ICANN'94, Springer, London (226–229), 1994.

S. Klampfl and W. Maass. Emergence of dynamic memory traces in cortical microcircuit models through STDP. *The Journal of Neuroscience*, 33(28):11515–11529, 2013.

D. M. Kullmann, A. W. Moreau, Y. Bakiri, and E. Nicholson. Plasticity of inhibition. *Neuron*, 75(61):951–962, 2012.

A. Litwin-Kumar and B. Doiron. Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature Communications*, 5(5319), 2014.

X. Liu, S. Ramirez, P. T. Pang, C. B. Puryear, A. Govindarajan, K. Deisseroth, and S. Tonegawa. Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, 484(7394):381–385, 2012.

A. Luczak and J. N. MacLean. Default activity patterns at the neocortical microcircuit level. *Frontiers in Integrative Neuroscience*, 6, 2012.

H. Markram, Y. Wang, and M. Tsodyks. Differential signaling via the same axon of neocortical pyramidal neurons. *Proceedings of the National Academy of Sciences*, 95 (9):5323–5328, 1998.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

H. Markram, E. Muller, S. Ramaswamy, M. W. Reimann, M. Abdellah, C. A. Sanchez, A. Ailamaki, L. Alonso-Nanclares, N. Antille, S. Arsever, et al. Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2):456–492, 2015.

D. Marr. Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, pages 23–81, 1971.

J. L. McClelland, B. L. McNaughton, and R. C. O'reilly. 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, 1995.

M. McCloskey and N. J. Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of Learning and Motivation*, 24:109–165, 1989.

B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M.-B. Moser. Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, 7(8):663, 2006.

J. F. Miller, M. Neufang, A. Solway, A. Brandt, M. Trippel, I. Mader, S. Hefft, M. Merkow, S. M. Polyn, J. Jacobs, et al. Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. *Science*, 342(6162):1111–1114, 2013.

R. G. Morris, E. Moser, G. Riedel, S. Martin, J. Sandin, M. Day, and C. O'Carroll. Elements of a neurobiological theory of the hippocampus: the role of activity-dependent synaptic plasticity in memory. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 358(1432):773–786, 2003.

M. Moscovitch. Memory and working with memory: Evaluation of a component process model and comparisons with other models. *Memory Systems*, 1994(369–394):224, 1994.

E. I. Moser, E. Kropff, and M.-B. Moser. Place cells, grid cells, and the brain's spatial representation system. *Annual Review of Neuroscience*, 31, 2008.

B. Nessler, M. Pfeiffer, L. Buesing, and W. Maass. Bayesian computation emerges in generic cortical microcircuits through spike-timing-dependent plasticity. *PLOS Computational Biology*, 9(4):e1003037, 2013.

K. A. Norman and R. C. O'Reilly. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychological Review*, 110(4):611, 2003.

C. H. Papadimitriou and S. S. Vempala. Cortical learning via prediction. In *Proceedings of The 28th Conference on Learning Theory (COLT)*, volume 40, pages 1–21, 2015.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

D. Pecevski and W. Maass. Learning probabilistic inference through STDP. *eNeuro*, 2016.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

T. A. Plate. *Holographic Reduced Representation: Distributed Representation for Cognitive Structures.* Stanford, CA, 2003.

R. Quian Quiroga. Neuronal codes for visual perception and memory. *Neuropsychologia*, 83:227–241, 2016.

R. Quian Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried. Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102–1107, 2005.

P. Ritter, J. Born, M. Brecht, H. R. Dinse, U. Heinemann, B. Pleger, D. Schmitz, S. Schreiber, A. Villringer, and R. Kempter. State-dependencies of learning across brain scales. *Frontiers in Computational Neuroscience*, 9, 2015.

D. S. Rizzuto and M. J. Kahana. An autoassociative neural network model of paired-associate learning. *Neural Computation*, 13:2075–2092, 2001.

E. T. Rolls, A. Treves, and E. T. Rolls. *Neural networks and brain function*, volume 572. Oxford University Press, Oxford, 1998.

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

P. Smolensky. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, 46(1–2):159–216, 1990.

L. R. Squire, C. E. Stark, and R. E. Clark. The medial temporal lobe. *Annual Review of Neuroscience*, 27:279–306, 2004.

D. Sussillo, T. Toyoizumi, and W. Maass. Self-tuning of neural circuits through short-term synaptic plasticity. *Journal of Neurophysiology*, 97(6):4079–4095, 2007.

G. Testa-Silva, M. B. Verhoog, D. Linaro, C. P. De Kock, J. C. Baayen, R. M. Meredith, C. I. De Zeeuw, M. Giugliano, and H. D. Mansvelder. High bandwidth synaptic communication and frequency tracking in human neocortex. *PLOS Biology*, 12(11): e1002007, 2014.

S. Tonegawa, X. Liu, S. Ramirez, and R. Redondo. Memory engram cells have come of age. *Neuron*, 87(5):918–931, 2015.

bioRxiv preprint doi: https://doi.org/10.1101/188938; this version posted September 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

L. G. Valiant. A neuroidal architecture for cognitive computation. *Journal of the ACM*, 47(5):854–882, 2000a.

L. G. Valiant. *Circuits of the Mind*. Oxford University Press, 2000b.

L. G. Valiant. Memorization and association on a realistic neural model. *Neural Computation*, 17(3):527–555, 2005.

S. Waydo, A. Kraskov, R. Quian Quiroga, I. Fried, and C. Koch. Sparse representation in the human medial temporal lobe. *The Journal of Neuroscience*, 26(40):10232–10234, 2006.

S. Wirth, M. Yanike, L. M. Frank, A. C. Smith, E. N. Brown, and W. A. Suzuki. Single neurons in the monkey hippocampus and learning of new associations. *Science*, 300 (5625):1578–1581, 2003.

E. R. Wood, P. A. Dudchenko, and H. Eichenbaum. The global record of memory in hippocampal neuronal activity. *Nature*, 397(6720):613, 1999.

F. Zenke, E. J. Agnes, and W. Gerstner. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications*, 6(6922):1–13, 2015.