



HAL
open science

A Coupled Neural Field Model for the Standard Consolidation Theory

Lisa Blum Moyse, Hugues Berry

► **To cite this version:**

Lisa Blum Moyse, Hugues Berry. A Coupled Neural Field Model for the Standard Consolidation Theory. 2024. hal-04527872

HAL Id: hal-04527872

<https://inria.hal.science/hal-04527872>

Preprint submitted on 2 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A coupled neural field model for the standard consolidation theory

Lisa Blum Moyse^{a,b,1,*}, Hugues Berry^b

^aLIRIS, CNRS UMR 5205, Villeurbanne, F-69621, France

^bAIstroSight, Inria, Hospices Civils de Lyon, Université Claude Bernard Lyon 1, Villeurbanne, F-69603, France

Abstract

The standard consolidation theory states that short-term memories located in the hippocampus enable the consolidation of long-term memories in the neocortex. In other words, the neocortex slowly learns long-term memories with a transient support of the hippocampus that quickly learns unstable memories. However, it is not clear yet what could be the neurobiological mechanisms underlying these differences in learning rates and memory time-scales. Here, we propose a novel modelling approach of the standard consolidation theory, that focuses on its potential neurobiological mechanisms. In addition to synaptic plasticity and spike frequency adaptation, our model incorporates adult neurogenesis in the dentate gyrus as well as the difference in size between the neocortex and the hippocampus, that we associate with distance-dependent synaptic plasticity. We also take into account the interconnected spatial structure of the involved brain areas, by incorporating the above neurobiological mechanisms in a coupled neural field framework, where each area is represented by a separate neural field with intra- and inter-area connections. To our knowledge, this is the first attempt to apply neural fields to this process. Using numerical simulations and mathematical analysis, we explore the short-term and long-term dynamics of the model upon alternance of phases of hippocampal replay and retrieval cue of an external input. This external input is encodable as a memory pattern in the form of a multiple bump attractor pattern in the individual neural fields. In the model, hippocampal memory patterns become encoded first, before neocortical ones, because of the smaller distances between the bumps of the hippocampal memory patterns. As a result, retrieval of the input pattern in the neocortex at short time-scales necessitates the additional input delivered by the memory pattern of the hippocampus. Neocortical memory patterns progressively consolidate at longer times, up to a point where their retrieval does not need the support of the hippocampus anymore. At longer times, perturbation of the hippocampal neural fields by neurogenesis erases the hippocampus pattern, leading to a final state where the memory pattern is exclusively evoked in the neocortex. Therefore, the dynamics of our model successfully reproduces the main features of the standard consolidation theory. This suggests that neurogenesis in the hippocampus and distance-dependent synaptic plasticity coupled to synaptic depression and spike frequency adaptation, are indeed critical neurobiological processes in memory consolidation.

Keywords: multi-bump attractors, long-term plasticity, systems consolidation memory, spike frequency adaptation, synaptic depression

1. Introduction

Memories are believed to be biologically encoded as physical traces in the brain, or engrams. It is assumed that these engrams form through the strengthening of synaptic connections in neuronal ensembles, i.e. populations of neurons involved in a memory representation or a computational task (Josselyn et al., 2015). The process through which recently encoded memories are converted into more stable long-term stored information is referred to as consolidation (Klinzing et al., 2019). This term includes two mechanisms that occur at distinct spatial and time scales: synaptic and systems consolidation. Synaptic consolidation is achieved through fast mechanisms (a few hours), like

*Corresponding author

Email addresses: lisa.blum-moyse@uni-konstanz.de (Lisa Blum Moyse), hugues.berry@inria.fr (Hugues Berry)

¹Present address: University of Konstanz, Cluster for the Advanced Study of Collective Behaviour, Konstanz, 78457, Germany

long-term synaptic plasticity (Martin et al., 2000). The most common form of long-term synaptic plasticity is qualified as Hebbian, it strengthens synapses between neurons that fire simultaneously, which stabilizes a memory trace. These processes are themselves embodied into wider and more systemic change processes invoked by systems consolidation. This second type of consolidation describes the gradual reorganization of memory patterns across different brain areas, which can endure weeks, months, or even years (Klinzing et al., 2019).

The theories of systems consolidation were introduced following neuropsychological observations of memory impairments in patients with medial temporal lobe (MTL) damages. The MTL includes the hippocampus and adjacent neocortical areas (perirhinal, entorhinal, parahippocampal) (Squire et al., 2004). MTL lesions induce anterograde amnesia that is, the inability to form new memories. This disability concerns in most cases declarative memories, memories that can be consciously recalled (Squire and Alvarez, 1995; Squire et al., 2004). These studies introduced the idea that there exist different memory systems associated with different brain areas. MTL injuries are also responsible for temporally-graded retrograde amnesia for declarative memories. Temporally-graded retrograde amnesia refers to a forgetting of memories encoded in the past, where the loss is more important for recent events. In general, the wider the damages, the more ancient the erased memories: amnesia usually affects memories ranging from 1 to 2 years in the past when lesions are restricted to the hippocampus, to a situation where all past memories are forgotten when the damages include the whole MTL and the surrounding cortical regions. These studies indicate that recent declarative memories depend initially on the MTL, and are subsequently transferred to a durable storage place, possibly cortical areas. Since then, many animal studies have examined the impact of hippocampal and cortical lesions on memory (Frankland and Bontempi, 2005).

In 1971, the initial studies on MTL damages led Marr to propose the first computational model describing systems consolidation (Marr et al., 1991). In this model, it is suggested that a new event is quickly encoded in the hippocampus and that with time this memory will be progressively “transferred” to the neocortex, through repetitive replays of the patterns during sleep. Following this pioneering work, a series of connectionist models have conceptualized the features of the standard consolidation theory and produced results in coherence with neuropsychological observations (Squire and Alvarez, 1995; McClelland et al., 1995; Káli and Dayan, 2004; Meeter and Murre, 2005; Amaral et al., 2008; Helfer and Shultz, 2020; Howard et al., 2022). These ideas have led to a contemporary model of systems memory consolidation, often referred to the standard consolidation theory (SCT) (Weingartner and Parker, 2014; Squire and Alvarez, 1995) or the standard model of systems consolidation², see figure 1 (a). According to this theory, a new experience and the set of features it contains are first encoded in different associative cortical modules. This information is then quickly transmitted and integrated by the hippocampus, which forms a compressed memory trace or pattern. Repeated offline reactivation of this hippocampal memory pattern can then activate the corresponding neocortical representations. Indeed the reactivation of memory patterns can occur during “offline” states, typically during sleep, or during “online” states, upon the encountering of an experience, or retrieval cue, related to this existing memory. These replays result in a gradual strengthening of neocortical connections, which ultimately leads to a neocortical memory pattern incorporated with pre-existing local representations and solid enough to be independent of the hippocampus. Thus, the memory patterns are expected to have a short lifetime in the hippocampus, while the ones in the neocortex can remain encoded for years. Therefore, on the one hand, the hippocampus is predicted to learn quickly online, and be essential for offline slow learning in the neocortex. According to previous connectionist models, this progressive neocortical learning would allow to prevent catastrophic interference and replacement of existing neocortical patterns with new ones (McClelland et al., 1995; Roxin and Fusi, 2013). On the other hand, while the encoded memories are kept on the long term in the neocortex, they are rapidly erased in the hippocampus. This clearance of hippocampal memories is believed to be important because of the limited storage capacity of the hippocampus (Willshaw et al., 2015).

In spite of the abundant literature on systems consolidation and its general principles, it is still unclear what are the main neurobiological mechanisms involved and how these processes intervene in the differences in learning speed and memory stability between the hippocampus and the neocortex. In particular, to our knowledge, this issue has not been addressed in the previous mathematical models of SCT. Some of the published modelling studies directly assumed larger learning rates for hippocampal or cortico-hippocampal connections compared to neocortical ones (Squire and Alvarez, 1995; McClelland et al., 1995; Káli and Dayan, 2004; Meeter and Murre, 2005; Amaral et al., 2008; Helfer

²We focus here on the widespread standard consolidation theory, although interesting alternative views exist, such as the multiple trace theory or the trace transformation theory, which will be addressed in the discussion.

and Shultz, 2020; Howard et al., 2022). However, it seems unlikely that variations in the rates of long-term synaptic plasticity would induce differences in learning rates that are of the order of several days or weeks (Lisman and Morris, 2001). We believe that structural plasticity i.e., the plasticity of the synaptic wiring diagram could be a better candidate to explain learning rate differences (Chklovskii et al., 2004). Indeed, the neocortex is a large structure comprising a very large number of neurons with sparse interconnectivity: a neuron of the neocortex is physically connected to only a small fraction of its local neighbors (Rolls and Treves, 1997). In contrast, the size of the hippocampus is much lower, which implies a smaller degree of sparsity and a larger fraction of pre-existing connections between hippocampal neurons. External stimulations can strengthen the pre-existing hippocampal synapses, leading to fast encoding. Such a fast encoding would be unlikely to happen in the much larger and sparser neocortical networks. Here, a new synaptic connection formed between two neocortical neurons would be stabilized with time through hippocampal replay via structural plasticity. This incremental process would result in memories that slowly consolidate in associative cortical modules (Frankland et al., 2001; Lisman and Morris, 2001).

Beyond their reinforcement, little is understood about the clearance of memories (Davis and Zhong, 2017). While regular hippocampal memories erasure seems a mandatory aspect in SCT models (McClelland et al., 1995; Frankland and Bontempi, 2005), usually implemented as larger forgetting rates for hippocampal or cortico-hippocampal connections than for neocortical ones (Squire and Alvarez, 1995; McClelland et al., 1995; Káli and Dayan, 2004; Meeter and Murre, 2005; Amaral et al., 2008; Helfer and Shultz, 2020; Howard et al., 2022), its neurobiological origin has scarcely been investigated. An interesting neurobiological hypothesis is that the dentate gyrus (DG) of the hippocampus, where neurogenesis occurs throughout life, would be involved in the gradual erasure of hippocampal memories (Frankland et al., 2013; Ko and Frankland, 2021). This effect can be explained by the perturbation of the hippocampal network stability as newborn neurons progressively integrate (over several weeks). Erasure is not likely to happen through the direct replacement of the pre-existing neurons of the engram by the newborn ones, since it seems that the neurons involved in an encoded pattern benefit from a survival advantage, so that they are likely to persist for longer times than the hippocampal engram (Leuner et al., 2004). Instead, erasure is thought to be due to local perturbation of the excitability of the pre-existing neurons by the newborn ones. Indeed, newborn cells are known to be highly excitable (Ge et al., 2007; Mongiat et al., 2009), so that their integration triggers offsetting mechanisms to restore local network homeostasis (Karmarkar and Buonomano, 2006). Under these processes, the excitability of pre-existing neurons is lowered or their synapses weakened by synaptic scaling, which could progressively hinder the reactivation of hippocampal engrams. This process has been implemented in mathematical models of hippocampal networks, confirming that the addition of new neurons to the DG layer induces memory degradation (Deisseroth et al., 2004; Meltzer et al., 2005; Weisz and Argibay, 2012).

We present below a novel mathematical model of SCT that incorporates a set of potential neurobiological mechanisms. In addition to spike-frequency adaptation and synaptic plasticity, we assumed that the difference of connectivity structure of the neocortex and the hippocampus is playing a role in SCT, in particular the larger size and more sparse structure of the neocortex. We also incorporate neurogenesis in the hippocampus, assuming that this newborn neurons could play a role in memory erasure. Using numerical simulations and mathematical analyses, we explore the short and long term dynamics of the model along the iterated phases of a typical consolidation process.

2. Overview of the model

Contrary to previous SCT connectionist models, we chose to represent the involved process with neural fields. To our knowledge, this is the first time that this framework is used to model SCT. This modelling approach assumes a continuum representation of large-scale biological neural networks. These models are composed of nonlinear integro-differential equations, where the integration kernels allow for spatial distributions of neural connections (Bressloff, 2012; Coombes et al., 2014). We believe that this framework is particularly suitable to model SCT for two reasons. First, the whole process involves a complex combination of mechanisms occurring at various time scales, that are challenging to capture with biophysically-detailed models, such as the Hodgkin-Huxley equations, that already describe several processes for a single neuron. Coarse-graining methods aim at producing simplified descriptions, derived from microscopic models. Thus, reduced models such as neural fields propose a good trade-off since they can express part of this complexity, while preserving some degree of mathematical analysis, even if these simplified models can become very abstract and depart from the initial microscopic description. One has to be careful with the conclusions raised using these models, but they can provide interesting indications. Second, our approach consists

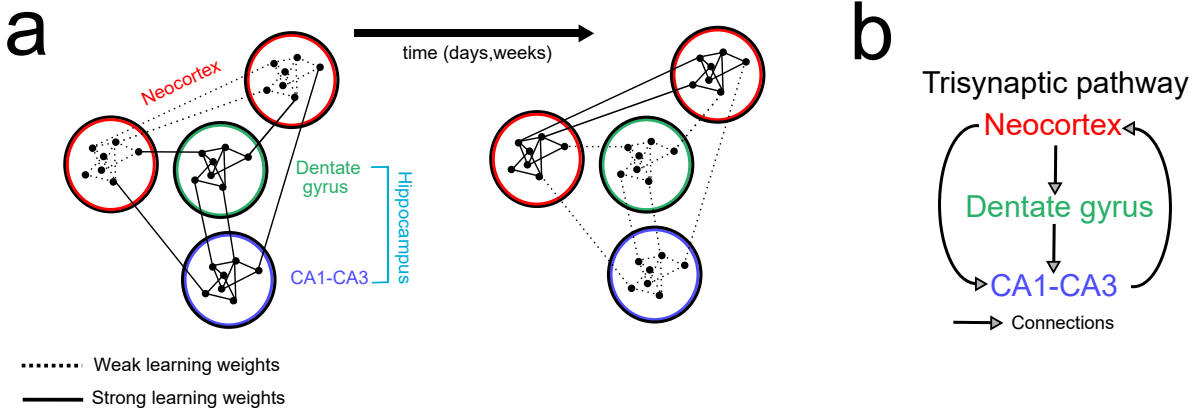


Figure 1: Interactions between the different brain areas involved in the standard consolidation theory. **(a)** Standard consolidation theory. Initially, the engram is initially present in neocortical areas (red), in a weak form (i.e., not stabilized), and in the hippocampus (dentate gyrus in green and CA1-CA3 subfields in blue), in a stable form. After some days the distribution of the memory is reorganized. Connections are consolidated in the neocortex, while neocortico-hippocampal and hippocampal connections are degraded. **(b)** A simplified scheme of the trisynaptic pathway, the circuit considered in the present model. The arrows represent the main synaptic connections between the three brain areas: neocortex, dentate gyrus, and hippocampus sub-fields CA1-CA3.

in articulating SCT, which involves interactions at a tissue level, with the underlying neurobiological mechanisms. Neural fields, while modelling large areas, aim at staying close to biological observations, and have already proven their usefulness for the exploration of neurobiological processes such as epilepsy, encoding of visual stimuli, the representation of head direction or working memory (Bressloff, 2012). Working memory corresponds to the temporary storage, on the time scale of seconds, of information in the brain. According to experimental reports, during a recall task, sensory inputs can be followed by a persistent activity within spatially defined clusters of neurons in the neocortex (Goldman-Rakic, 1995). This stationary pulse known as a bump of activity corresponds to a class of solutions in neural field theory and has also been studied for multiple bump patterns (Laing et al., 2002; Ferreira et al., 2016). Here, we propose that these bumps represent the different elements/sub-parts of an encoded memory pattern in SCT. However, in opposition to working memory, we consider here both short-term and long-term memories. This implies the introduction of synaptic plasticity in the neural field framework.

In its simplest form, the neural field equation is the following integro-differential equation (Amari, 1977):

$$\frac{\partial u}{\partial t}(x, t) = -u(x, t) + \int_{-\infty}^{\infty} dy w(x-y) f(u(y, t) - \kappa) + I_{\text{ext}}(x, t) \quad (1)$$

This equation describes the evolution of the average synaptic current $u(x, t)$ at spatial location x of the continuous neural tissue and time t . The membrane time constant of around 10 ms, is often not explicitly written. $I_{\text{ext}}(x, t)$ represents an external input, $w(X)$ is the synaptic weight distribution with distance X , and f is the firing rate function. This term represents the effective mean firing rate as a function of an increasing input received by a neuron. Theoretical studies on stochastic conductance-based models have shown that f could typically be described as a sigmoid function (Izhikevich, 2006):

$$f(u - \kappa) = \frac{1}{1 + e^{-\beta_f(u - \kappa)}} \quad (2)$$

where β_f is the gain and κ is the threshold.

For a synaptic weight kernel chosen as

$$w(X) = \left(1 - \frac{|X|}{\sigma}\right) e^{-\frac{|X|}{\sigma}}, \quad (3)$$

This kernel describes a local excitation, and lateral inhibition connectivity. This organization has been found in orientation tuning in the visual cortex (Ben-Yishai et al., 1995), and has been used early on in neural fields stud-

ies (Amari, 1977). In particular, its use to model encoding of visual stimuli and working memory makes it suitable for our modelling purposes, see the introduction section. It is however a phenomenological choice, and its application to the SCT processes we aim to describe here would need to be verified.

eq.(1) admits a stationary bump solution (Amari, 1977) $u_0(x)$, that can be derived at the limit of infinite gains, where the firing rate function converges to a heaviside function (i.e. $f(u - \kappa) = \Theta(u - \kappa)$ when $\beta_f \rightarrow \infty$):

$$u_0(x) = \int_{-a}^a dy w(x - y) = W(x + a) - W(x - a) \quad (4)$$

with $W(X) = X e^{-\frac{|X|}{\sigma}}$. The bump width, $2a$, is determined considering the boundary conditions: $u_0(\pm a) = \kappa$.

Equation (1) constitutes the basis of our model. We extend it here in multiple ways:

- We introduce spike frequency adaptation and synaptic depression to modulate the average synaptic current. Both mechanisms have already been implemented and studied in detail in a neural field context in the works of Kilpatrick and Bressloff (Kilpatrick and Bressloff, 2010b,a). Spike frequency adaptation describes the attenuation of firing rate after a prolonged period of firing, often due to a calcium-activated potassium current. This process can be implemented as an activity-dependent increase of the threshold κ of eq.(1) (Coombes and Owen, 2005). Synaptic depression is related to presynaptic resource depletion. This process can be accounted for in the neural fields equation by introducing an activity-dependent factor q in the nonlocal, integral term of eq.(1) (Kilpatrick and Bressloff, 2010b,a).
- We add long-term synaptic plasticity by introducing a Hebbian plasticity expression in the kernel w of eq. (3). A core hypothesis of our model is that the differences of learning rates between the neocortex and the hippocampus are related to the distance separating distinct elements of an engram, here the bumps of the multiple bump solution. Thus, our Hebbian plasticity rule is modulated by the distance between neurons. In addition to this distance-dependent learning rule, our plasticity equation also includes a slow decay term, that accounts for the degradation of memories in the absence of reactivation.
- We add a feedback mechanism for the threshold κ that accounts for neurogenesis. As evoked above, the insertion of highly excitable newborn neurons can lead to a decrease of the intrinsic excitability of mature pre-existing neurons. We implement this process by a specific activity-dependent regulation of the threshold κ .
- Another particularity of our model is the inter-connection of three individual neural fields, that represent the major regions of the brain involved in the interactions between the hippocampus and the neocortex. These three regions are classically described as part of a particular structure, the trisynaptic pathway (Basu and Siegelbaum, 2015), see figure 1 (b), that connects the neocortex (C), to the hippocampus split here into the dentate gyrus (D) and the CA1-CA3 subfields (H), merged here for the sake of simplicity. Hence, our model features three neural fields. In each neural field, the kernel w of eq. (3) accounts both for synaptic connections within the neural field and between the two others, in agreement with the trisynaptic pathway. Beyond biological realism related to the trisynaptic pathway, we found that it was necessary to distinguish two separate hippocampal neural fields in the model. Indeed, with a single hippocampal neural field, the highly excitable newborn neurons affect not only the hippocampus but also the neocortical neural field, which erases memory in the neocortex too and prevents retrieval.

3. Methods

Our model consists of three connected $1d$ neural fields, each one representing a brain area. The different variables are indexed with α to refer to the area/neural field they describe, the neocortex ($\alpha = C$), dentate gyrus ($\alpha = D$) and CA regions ($\alpha = H$). In each area, the memory engram takes the form of a double bump attractor pattern, centered around positions $Z^\alpha = \{A^\alpha, B^\alpha\}$, with widths $2a$ (see the ‘‘External stimulation phases’’ subsection for further details). Our model is given by the following differential equations.

Activity. The evolution of activity $u^\alpha(x, t)$ in region / neural field α , at position x and time t , is described by

$$\frac{\partial u^\alpha}{\partial t}(x, t) = -u^\alpha(x, t) + I^\alpha(x, t) + I_{\text{ext}}^\alpha(x, t), \quad (5)$$

where $I_{\text{ext}}^\alpha(x, t)$ corresponds to an external input and $I^\alpha(x, t)$ is the average synaptic current.

Average synaptic current. $I^\alpha(x, t)$ integrates the neural population activity, averaged over the domain of integration Γ^α , that comprises the activity of the considered neural field α and that of the neural fields β to which it is connected:

$$I^\alpha(x, t) = \sum_{\beta \in E} G^{\alpha\beta} \int_{\Gamma^\beta} dy q^\beta(y, t) w_{\text{tot}}^{\alpha\beta}(x, y, t) f(u^\beta(y, t) - \kappa_{\text{tot}}^\beta(y, t)) \quad (6)$$

where $E = \{C, D, H\}$ is the set of field indexes. The constants $G^{\alpha\beta}$ are the elements of the adjacency matrix that defines the directed connectivity of the trisynaptic circuit (fig. 1 (b)). $G^{\alpha\beta} > 0$ means that the β field sends connections the α field while $G^{\alpha\beta} = 0$ means there is no connection from β to α . Moreover, to equilibrate the total currents received by each field, we fix $G^{\alpha\beta} = 1$ for fields α which receive connection from a unique field β , while for other connections $G^{\alpha\beta} = 0.5$, see table 2.

In eq. (6), $\kappa_{\text{tot}}^\alpha(x, t)$ is the adaptive threshold, $q^\alpha(x, t)$ is the synaptic scaling factor and $w_{\text{tot}}^{\alpha\beta}(x, y, t)$ describes the synaptic weights. The activity-dependent expressions of those three terms are given below.

Adaptive thresholds. The threshold $\kappa_{\text{tot}}^\alpha(x, t)$ is composed of two parts:

$$\kappa_{\text{tot}}^\alpha(x, t) = \kappa^\alpha(x, t) + \kappa_n^\alpha(x, t). \quad (7)$$

$\kappa^\alpha(x, t)$ implements spike-frequency adaptation while $\kappa_n^\alpha(x, t)$ accounts for the effects of neurogenesis.

For spike-frequency adaptation, we mostly follow ref. (Kilpatrick and Bressloff, 2010b,a; Coombes and Owen, 2005):

$$\tau_\kappa \frac{\partial \kappa^\alpha}{\partial t}(x, t) = -(\kappa^\alpha(x, t) - \kappa_{\text{in}}) + \eta_\kappa f_a(u^\alpha(x, t) - \kappa_{\text{in}}) \quad (8)$$

In the original studies of this equation, the baseline threshold in the first term of the right-hand side and the threshold of the firing function f_a of the second term have different values (Coombes and Owen, 2005; Kilpatrick and Bressloff, 2010b,a). However, detailed biophysical models suggest that it is more biologically realistic to use the same parameter in the two terms, here κ_{in} (Benda and Herz, 2003). f_a is a firing function smoother than $f(u - \kappa) = 1 / (1 + e^{-\beta_f(u - \kappa)})$ in eq. (6), i.e. $f_a(u - \kappa) = 1 / (1 + e^{-\beta_{f_a}(u - \kappa)})$ with $\beta_{f_a} < \beta_f$ (see below for a justification). We refer to the smoothness of such sigmoid functions as the small derivative of their slope. The characteristic time scale of the process τ_κ has been experimentally found to lie between 40 and 120 ms (Madison and Nicoll, 1984).

The neurogenesis term $\kappa_n^\alpha(x, t)$ is expected to emulate the progressive adaptation mechanism responsible for the local decrease of excitation in reaction to the integration of highly excitable newborn DG neurons. For simplicity, we use the same evolution equation as for spike frequency adaptation, however with a much slower timescale ($\tau_{\kappa_n} \gg \tau_\kappa$) and a lower adaptation threshold ($\theta_n < \kappa_{\text{in}}$):

$$\tau_{\kappa_n} \frac{\partial \kappa_n^\alpha}{\partial t}(x, t) = -\kappa_n^\alpha(x, t) + \eta_{\kappa_n} \Theta(u^\alpha(x, t) - \theta_n) \quad (9)$$

Here we use a heaviside function Θ for simplicity instead of the smoother firing rate function of eq. (8). We leave for future works the evaluation of the impact of this simplification.

To emulate the high excitability of newborn D neurons, the threshold of the neurons located around the areas involved in the engram is lowered to a value $g_n \kappa_{\text{in}}$ with $g_n < 1$ in eq. (8). We only consider neighbour neurons because more distant neurons would have a very small effect. The spatial domain of these newborn neurons is therefore

$$\bigcup_{z \in Z^D} ([z - a - \delta_n, z - a] \cup [z + a, z + a + \delta_n])$$

with δ_n the width of newborn neurons subarea.

Synaptic scaling. The evolution of the synaptic scaling factor $q^\alpha(x, t)$ emulates synaptic depression. Here again we mostly follow the implementation proposed by Kilpatrick and Bressloff (2010) (Kilpatrick and Bressloff, 2010b,a):

$$\frac{\partial q^\alpha}{\partial t}(x, t) = \frac{1 - q^\alpha(x, t)}{\alpha_q} - m_q^\alpha(x, t)q^\alpha(x, t)f_a(u^\alpha(x, t) - \kappa_{\text{tot}}^\alpha(x, t)) \quad (10)$$

In contrast to the original equation however, we introduce an activity dependent term for synaptic depletion that allows to keep bump durations roughly constant when learning weights evolve and/or neurogenesis is applied:

$$m_q^\alpha(x, t) = \beta_q u^\alpha(x, t) \quad (11)$$

Synaptic weights and plasticity. The kernel $w_{\text{tot}}^{\alpha\beta}(x, y, t)$ of eq. 6 is composed of two terms:

$$w_{\text{tot}}^{\alpha\beta}(x, y, t) = w(x - \Delta^{\alpha\beta}(x) - y) + \gamma^{\alpha\beta} \eta_s^{\alpha\beta}(x, t) s^{\alpha\beta}(x, y, t) \quad (12)$$

- $w(X)$ is given by eq. (3) above. It represents a permanent (learning-independent) connectivity between a post-synaptic neuron at position x and a presynaptic neuron at position y . However, compared to the expression used in classical neural field formulations, we introduce a position-dependent shift $\Delta^{\alpha\beta}(x) = (B^\alpha - B^\beta) \text{sgn}(x) = -(A^\alpha - A^\beta) \text{sgn}(x)$. This shift emulates the difference in size of the neocortical and hippocampal fields. Using this trick, despite the intrinsic difference of distances between pattern locations ($|A^C - B^C| > |A^D - B^D|$ and $|A^D - B^D| = |A^H - B^H|$), $\Delta^{\alpha\beta}(x)$ allows to force the permanent connection between A^C and A^H (or A^D), B^C and B^H (or B^D). For instance, for $x = A^C$, $w(A^C - \Delta^{CH}(A^C) - A^H) = w(A^C - \Delta^{CC}(A^C) - A^C) = w(0) = 1$, i.e. the weight between A^C and A^C is the same as between A^C and A^H .

This phenomenological description of inter-field interactions for bumps separated by different distances is a highly simplified modelling choice. It is here made possible, as there are only two bumps at defined positions in each field. In a case where bumps could be anywhere on the fields, this space shift would not be a satisfying description, and it would be more suitable to implement directly a larger neocortical field.

- Synaptic plasticity rules in neural field models can take the form of an immediate modulation of the kernel depending on the neuronal activities (Abbassian et al., 2012). However in the case of long-lasting synaptic modifications it is interesting to introduce an additional differential equation with a larger characteristic timescale. A rate-based plasticity rule preserving temporal correlations has been introduced for connections between pre- and post-synaptic neural fields (Robinson, 2011), and has been used in particular for topographic maps (Gale et al., 2022; Detorakis and Rougier, 2012). However such temporal correlations are not important for our model, and we preferred a simpler Hebbian expression that preserves mathematical and computational tractability. Therefore, we implement here synaptic plasticity by the second term of eq. (12): $\gamma^{\alpha\beta} \eta_s^{\alpha\beta}(x, t) s^{\alpha\beta}(x, y, t)$, where the $\gamma^{\alpha\beta}$ s are positive constants, assumed to be larger for intra than for inter-field connections, $\gamma^{\alpha\alpha} > \gamma^{\alpha\beta}$ for $\alpha \neq \beta$, to represent stronger interactions within one brain area.

$s^{\alpha\beta}(x, y, t)$ is a variable that represents the learning weight. Its dynamics is given by two additive terms:

$$\frac{\partial s^{\alpha\beta}}{\partial t}(x, y, t) = L^{\alpha\beta}(x, y, t) - F^{\alpha\beta}(x, y, t) \quad (13)$$

The first term $L^{\alpha\beta}(x, y, t)$ is a bounded, distance-dependent, Hebbian (Gerstner and Kistler, 2002) learning rule:

$$L^{\alpha\beta}(x, y, t) = (1 - s^{\alpha\beta}(x, y, t))d(x, y) \times f(u^\alpha(x, t) - \kappa_{\text{tot}}^\alpha(x, t))f(u^\beta(y, t) - \kappa_{\text{tot}}^\beta(y, t)), \quad (14)$$

where distance modulates the learning rate via the factor:

$$d(x, y) = \frac{A_d}{\sigma_d} e^{-|x-y|/\sigma_d}. \quad (15)$$

With eq. (15), closer x and y positions yield larger $d(x, y)$, thus faster learning rates. The second term $F^{\alpha\beta}$ is a forgetting, decay term, which occurs much more slowly than the first term ($c_0 \ll 1$).

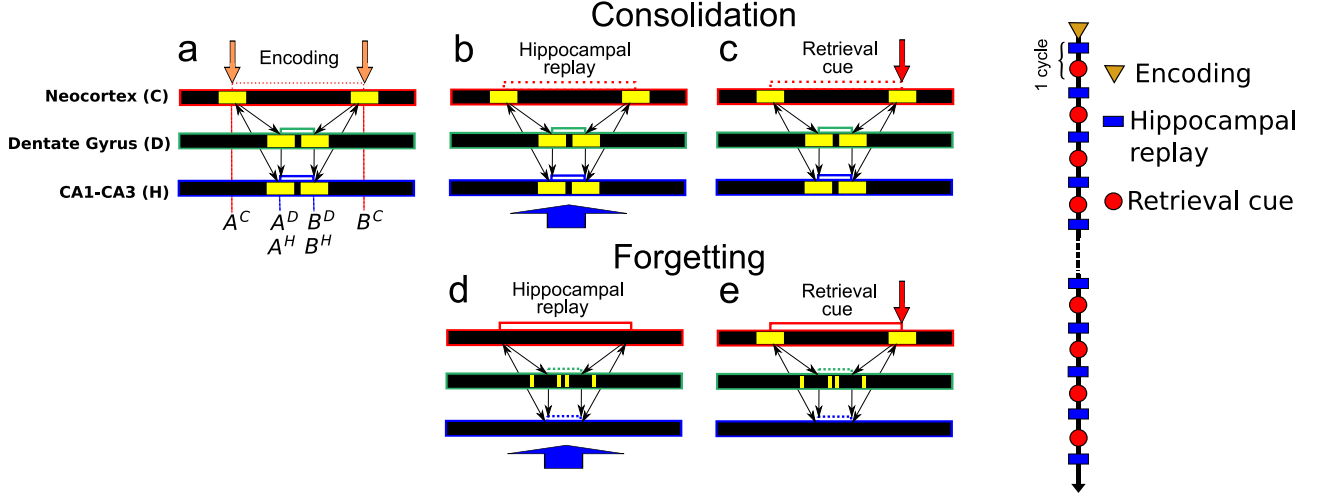


Figure 2: Temporal sequence describing the consolidation process according to standard consolidation theory. **(Left)** Different steps of the computational model. **(a)** The first diagram represents the encoding phase, where two external signals arrive on the neocortical field at two distinct positions, A^C and B^C . The activated neocortical neurons that now encode the memory pattern fire, which activates other neurons in the two other hippocampal fields (D, H), at positions A^D and B^D (and A^H and B^H). After the initial encoding, cycles of hippocampal replay and retrieval cue steps are repeated, which allows consolidation. **(b)** The hippocampal replay step occurs during sleep. We model it by a spatially uniform external stimulation on the CA region (H). **(c)** The retrieval cue is a partial external input centered on only A^C or B^C in the neocortex. **(d,e)** Following experimental suggestions (Frankland et al., 2013), we model the effect of neurogenesis in the dentate gyrus (D) by reducing the threshold of newborn neurons located in the neighborhood of the pattern (see text). **(Right)** Summary of the steps for the whole process. After the encoding phase, cycles of hippocampal replay followed by retrieval cue step are iterated. The time between two steps there is a long waiting period, 100 times longer than the duration of a pattern reactivation.

$$F^{\alpha\beta}(x, y, t) = \begin{cases} \frac{c_0}{s^{\alpha\beta}(x, y, t)} & \text{if } s^{\alpha\beta}(x, y, t) > 0 \\ 0 & \text{if } s^{\alpha\beta}(x, y, t) = 0 \end{cases} \quad (16)$$

Introducing the $s^{\alpha\beta}(x, y, t)$ term in the right-hand side of eq. (12) generates a discontinuity in the activity (Fotouhi et al., 2015). To maintain the continuity in $u^{\alpha\beta}(x, t)$ expression, we therefore add a continuity function $\eta_s^{\alpha\beta}(x, t)$ in eq.(12) :

$$\eta_s^{\alpha\beta}(x, t) = \left[\int_{\Gamma^\beta} dy w(x - \Delta^{\alpha\beta}(x) - y) \Theta(s^{\alpha\beta}(x, y, t)) - \kappa_{in} \right]_+ \quad (17)$$

with the rectification function $[v]_+ = v$ if $v > 0$, 0 otherwise.

When $s^{\alpha\beta}(x, y, t) > 0$, i.e. for neurons that are part of the memory pattern, the integral in eq.(17) corresponds to the received bump current of one field, which is κ_{in} at boundaries. This guarantees continuity at boundaries, see the supplementary materials section Appendix A for more details.

External stimulation phases. To emulate the repetition of reactivations of hippocampal memory patterns and presentations of the pattern to the neocortex that are central to the SCT, we submitted the model to a sequence composed of hundreds of alternation of two phases, see figure 2. The initial encoding is simulated by the application of an external input on the neocortex that represents the experience to be memorised. This external input can be encoded as a memory pattern in the three neural fields in the form of a double bump attractor pattern. At this initial encoding stage, the external current is non-zero only in the neocortex (C), where it assumes a rectangular function, located around positions A^C and B^C , with widths equal to $2a$. These characteristics define the information contained in the memory pattern. The external input to the neocortex during the initial encoding stage thus is:

$$I_{\text{ext}}^C(x, t) = G^{\text{CExt}} \sum_{z \in Z^C} \left[\Theta(x - (z - a)) - \Theta(x - (z + a)) \right] i^C(t) \quad (18)$$

Symbol	Definition
Time varying functions	
u^α	Activity
f	Firing function
f_a	Firing function for κ^α and q^α equations
I^α	Average synaptic current
I_{ext}^α	External current
$s^{\alpha\beta}$	Learning weight
$\eta_s^{\alpha\beta}$	Continuity function
κ_{tot}^α	Threshold
κ^α	Threshold - adaptation
κ_n^α	Threshold - neurogenesis
q^α	Synaptic scaling factor
Time independent functions	
w	Permanent weight
W	Integral of the permanent weight
d	Distance function

Table 1: Main functions used for the neural field model

where $i^C(t)$ is an indicator function equals to 1 during the encoding phase, 0 otherwise. After this initial encoding, we iterate cycles composed of a hippocampal replay (HR)– assumed to occur offline, during sleep– followed by a retrieval cue (RC) – assumed to occur online, during a waking period. In this sequence, one cycle corresponds approximately to one day.

HR steps are simulated in the model by the application of an homogeneous external input on the CA region (H), that is assumed to emulate slow-wave sleep (Klinzing et al., 2019). Hence, during HR steps, the external current targets the H region only, with $I_{ext}^H(x, t)$ uniform over the neural field space, $I_{ext}^H(x, t) = G^{HExt}$, $\forall x$. An RC step corresponds to the application of a partial external input on the neocortex only, that stimulates only one of the two bumps of the encoded pattern. Stimulation of a single bump represents an experience partially similar to the encoded pattern (that exhibits 2 bumps). Memory retrieval is achieved if the stimulation of this single bump activates the totality of the 2-bump encoded memory. Therefore, the external current during RC steps assumes a form similar to eq. (18), except that Z^C is restricted to A^C or B^C (in alternation). In every steps, the external input is maintained for a short period of time (see table 2).

Numerical simulations. We solved the model eq. (5) to (17) on a one-dimensional spatial domain with periodic boundaries, that is, on ring of length total length $x_{m1} - x_{m2} = 60$ centered at $x = 0$. The equations were solved using a forward explicit Euler method with spatial increment $dx = 0.08$ and time increment $dt = 0.1$. Numerical integration of the integral was carried out as a sum of rectangles of width dx , times dx .

The code used to generate the results of this article is freely available online on GitLab: <https://gitlab.inria.fr/lisa.blum-moyse/a-coupled-neural-field-model-for-the-standard-consolidation-theory>.

4. Results

As illustrated in fig. 2, we submitted our coupled neural field model, defined by eq. (5) to (17) to cycles of repeated external inputs. We first study below the consolidation mechanism, then the forgetting effect of neurogenesis is tested on consolidated memory patterns. Figure 2 is used as a visual support throughout this results section and as a guide to the different phases of the process.

For convenience, throughout the paper we refer to neurons located within the excited region around A^α ($[A^\alpha - a, A^\alpha + a]$) or B^α ($[B^\alpha - a, B^\alpha + a]$), as respectively A^α or B^α neurons. Furthermore, to monitor the temporal evolution of synaptic learning along the repetition cycles, we show the evolution of the learning weights $s^{\alpha\beta}(x, y, t)$ instead of the total weights $w_{tot}^{\alpha\beta}(x, y, t)$, eq. (12), because they constitute a relevant and concise measure of synaptic plasticity, since they are bounded between 0 and 1.

Parameter	Value	Definition
Structure and positions		
E	-	Set of field indexes
C	-	Neocortex field index
D	-	Dentate gyrus field index
H	-	CA regions field index
A^C	-16	Position of the C left pattern location
B^C	16	Position of the C right pattern location
A^D	-10	Position of the D left pattern location
B^D	10	Position of the D right pattern location
A^H	-10	Position of the H left pattern location
B^H	10	Position of the H right pattern location
a	0.9	Bump widths in all fields
G^{CC}	1	Amplitude of the current from C to C
G^{DD}	1	Amplitude of the current from D to D
G^{HH}	1	Amplitude of the current from H to H
G^{CH}	1	Amplitude of the current from H to C
G^{CD}	0	Amplitude of the current from D to C
G^{HD}	0.5	Amplitude of the current from D to H
G^{HC}	0.5	Amplitude of the current from C to H
G^{DC}	1	Amplitude of the current from C to D
G^{DH}	0	Amplitude of the current from H to D
σ	1.5	Width of the permanent weights
Firing rate		
β_f	250	Gain of the firing function
β_{f_a}	50	Gain of the firing function for κ^α and q^α equations
Learning kernels		
γ	1.5	Constant for intra-field learning weights ($\gamma^{\alpha\alpha} = \gamma$)
cr	0.2	Factor for the constant for inter-field learning weights ($\gamma^{\alpha\beta} = cr \times \gamma, \alpha \neq \beta$)
A_d	3	Amplitude of the distance function
σ_d	9	Width of the distance function
c_0	8.10^{-7}	Decay rate
Thresholds		
κ_{in}	0.54	Baseline threshold
τ_κ	0.8	Time scale for spike frequency adaptation
η_κ	0.54	Strength for spike frequency adaptation
τ_{κ_n}	1000	Time scale of the slow adaptation related to neurogenesis
η_{κ_n}	0.5	Strength of the slow adaptation related to neurogenesis
θ_n	0.001	Baseline threshold for slow adaptation related to neurogenesis
Synaptic scaling		
α_q	800	Time scale of synaptic resources recovery
β_q	0.01	Inverse of the time scale of synaptic resources depletion
External currents		
G^{CExt}	1.5	Amplitude of the external current in C
G^{HExt}	0.87	External current value in H
-	1.8	Duration of the external stimulation in C
-	1.2	Duration of the external stimulation in H
Neurogenesis		
δ_n	0.32	Width of the newborn neurons area
g_n	0.1	Factor of newborn neurons reduced threshold
Numerical parameters		
x_{m1}	-30	Left extremity of a field
x_{m2}	30	Right extremity of a field
dx	0.08	Spatial increment
dt	0.1	Time increment

Table 2: Parameters values used for the neural field model. Because of the abstract nature of the model, units were not taken into account here. See the section 5 for a discussion on the timescale parameters.

4.1. Consolidation

Encoding. The process is initiated by the encoding step, where an external signal stimulates two distinct areas around the positions A^C and B^C , in the neocortex (C), see the schematized process in figure 2 (a) and the corresponding numerical simulation in figure 3 (a). The stimulated cortical (C) neurons then induce firing in the dentate gyrus (D), and, finally, in the CA regions (H) where the H neurons receive inputs from C and D. The interconnection of these neural fields which are quickly activated one after the other modifies the total currents in the activity equation eq.(5). This variation takes place quickly, shortly after the external input. Since spike frequency adaptation is active at this timescale, the firing threshold increases from its initial value κ_{in} to a value $\kappa_{in} + \eta_{\kappa}$. This term is responsible for the disappearance of the bumps after a time (fig. 3 (a)). Spike frequency adaptation also maintains the bump width, which characterizes the information of the memory pattern (see section Appendix A for the relation between these threshold values and their width). However, numerical exploration of the model showed that adaptation was not able to properly and quickly enough adjust to preserve the bump width, unless its firing function (f_a in eq. (8)) is smoother than the firing function f of the activity equation eq.(5), i.e. $\beta_{f_a} < \beta_f$.

As long as the bump pattern is present, the firing of the corresponding neurons leads to an increase of the learning weights through eq. (14). Therefore, during the encoding phase, all the weights increase between neurons located in an engram location ($[A^\alpha - a, A^\alpha + a]$ and $[B^\alpha - a, B^\alpha + a]$) because of Hebbian synaptic plasticity (see eq.(14)). However as illustrated in figure 4a, the synaptic weights between the neocortical neurons A^C and B^C increase much more slowly than the intra-hippocampal weights $A^H - B^H$. This difference in learning rate is due to the distance-modulated learning rate of eq. (15). Especially since $|A^C - B^C| > |A^D - B^D|$ ($|A^H - B^H| = |A^D - B^D|$), we get $d(A^D - B^D) > d(A^C - B^C)$ in eq. (15). In other words, consolidation is slower in the neocortex (C) because the distance between the two bumps of memory pattern is larger than in the hippocampus. Therefore, at the end of the encoding phase, the synaptic weights between the two bumps of the memory pattern in the neocortex are much smaller than those between the two parts of the hippocampal patterns, or between the neocortex and the hippocampus (see figure B.1 in the supplementary material).

Hippocampal replay. Repeated reactivations of the memory pattern during sleep, which originates in the hippocampus, consolidate the cortical encoding is through see the schematized hippocampal replays in figure 2 (b) and the numerical simulations in figure 3 (b). During hippocampal replay, an external signal stimulates all neurons of the CA regions (H) neural field. However, only those neurons involved in the pattern locations fire for a long time. Indeed, the current they receive has a higher value, since their learning weights are stronger. This firing in H leads to pattern reactivations in D and C, see fig. 3 (b). The two parts of the memory pattern being active, the weights between A and B neurons in each neural field grow. When the HR is repeated for a sufficient number of times, this also increases the intra-cortical weights between A^C and B^C , as shown in fig. 4a, thus stabilising the cortical engram.

Retrieval cue. Since the learning weights between A^α and B^α grow faster in the hippocampus, neocortical memory in the early steps of the process can be retrieved only with the hippocampal neurons, as illustrated in the numerical simulation of figure 3 (c). During this step, only the one bump location is stimulated, A^C or B^C , in alternation. Retrieval is achieved when this stimulation of the stimulated bump, e.g., B^C leads to the activation of non-stimulated one (here, A^C). However at the initial stages of the process, the activation of A^C cannot happen directly from B^C to A^C , it goes via the hippocampus. Indeed, at these initial stages, the weights between A^C and B^C are still small, the activation of the B^C neurons during retrieval first activates the B^D neurons; the B^H neurons are then activated by B^C and B^D , which activates the A^H neurons because the $B^H - A^H$ learning weights are quickly large at the initial stages of the process (fig. 4a). Finally, it is the A^H neurons that activate the A^C neurons via the strong $A^C - A^H$ weights, leading to the whole pattern recovery via the complete reactivation of A^C , see figure 3 (c). Here again, recovering the activity in the two bumps of the neocortical memory pattern ($[A^C - a, A^C + a]$ and $[B^C - a, B^C + a]$), contributes to progressively increase the intra-cortical weights $A^C - B^C$ and thus consolidate the neocortical engram, see figure 4 (a) and supplementary figure B.3 (b)(c).

To illustrate further that the neocortex is dependent on the hippocampal neural fields for its engram retrieval in the initial stages of consolidation, we performed the simulation shown in fig. 4a-c of a simulated hippocampal lesion. Deletion of the whole connectivity $w_{tot}^{\alpha\beta}$ stops transmission between the fields, and forbids any retrieval in the hippocampal fields in the retrieval cue step for example. We thus choose here to study the case of a ‘‘partial’’ lesion, keeping the permanent connectivity intact. The hippocampal lesion was simulated by cancelling the learning

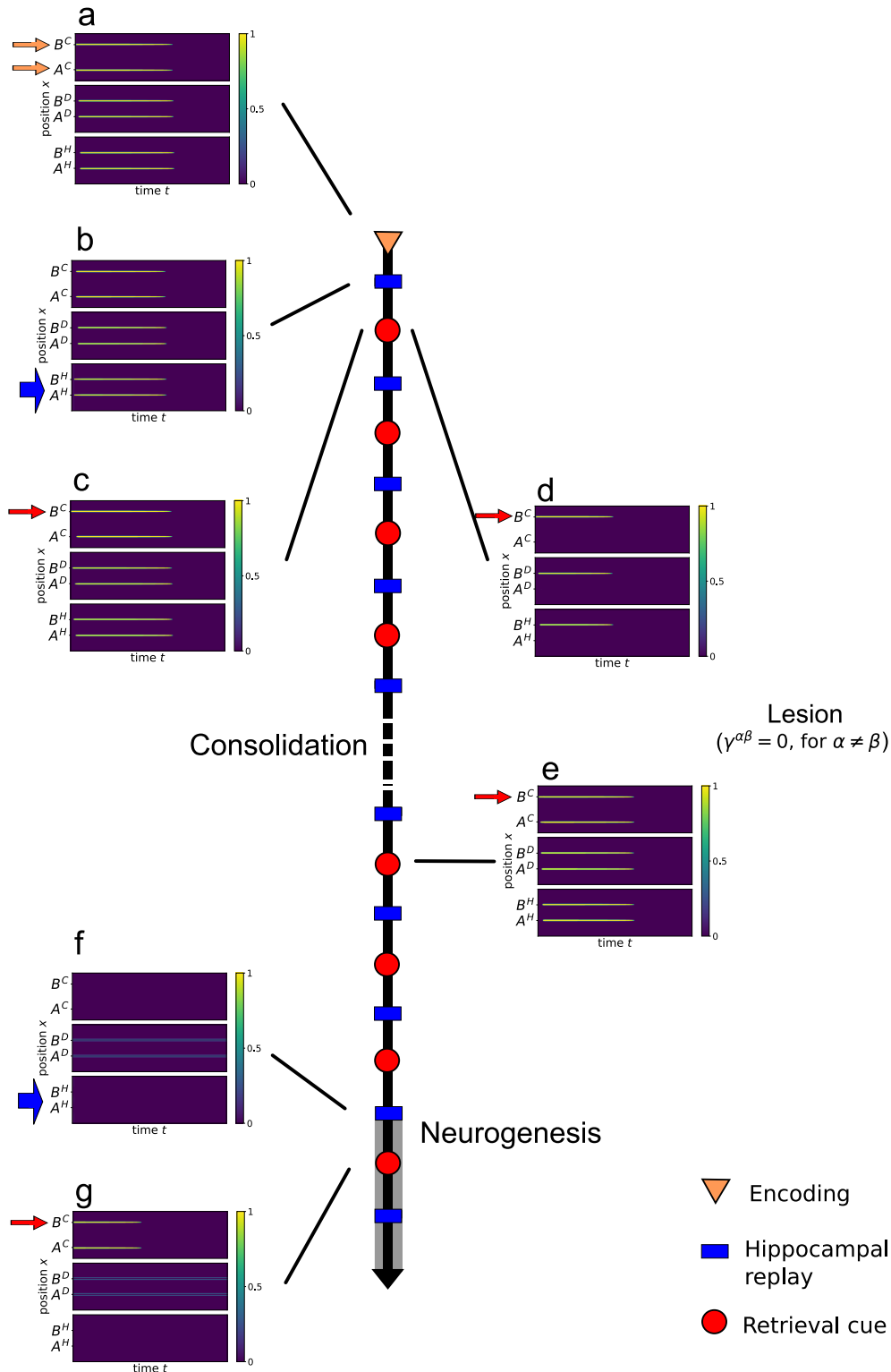


Figure 3: Snapshots of firing rates in the three fields, neocortex (C), dentate gyrus (D) and CA regions (H) along the repetitions of the temporal sequence. (a) The encoding step shows a first pattern activation in the three fields. The initial stimuli around positions A^C and B^C in the neocortical field activate C neurons which in turn activate D and H neurons, yielding bump attractors in all neural fields. (b) The hippocampal replay step allows engram reactivation. We model hippocampal reactivation by sending a uniform stimulus over space on neural field H. (c) The retrieval cue step allows engrams reactivation. This step consists in stimulating only the A^C or B^C location of the neocortex (here B^C). Retrieval is exhibited by the subsequent activation of A^C . (d,e) A “lesion” between the hippocampal and neocortical neural fields is simulated by setting $\gamma^{\alpha\beta} = 0$ for all $\alpha \neq \beta$ in eq.(12). (d) At early stages, the activation of the B^C neurons during retrieval fails to activate the A^C neurons if hippocampal learning is not functional. (e) However, at the end of the consolidation process, pattern retrieval in the neocortex is possible even in the absence of a functional hippocampus. (f,g) HR and RC when neurogenesis is effective. The thin lines in the D field evidence the continuous firing of those new neurons. (f) Neurogenesis prevents the reactivation of hippocampal engrams during the HR phase because of the increased thresholds caused by the excitable newborn D neurons. (g) Neurogenesis also prevents the reactivation of hippocampal engrams during neocortical RC.

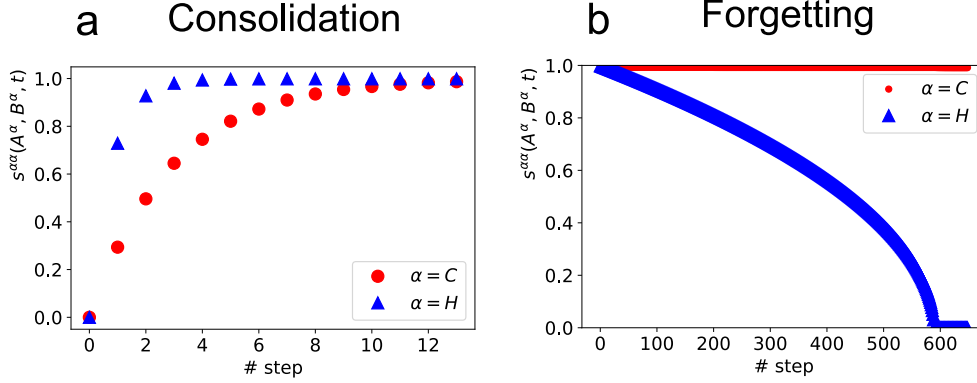


Figure 4: Evolution of the learning weights $s^{CC}(A^C, B^C, t)$ (red circles) and $s^{HH}(A^H, B^H, t)$ (blue triangle) at the beginning of every consolidation step. $s^{DD}(A^D, B^D, t)$ displays the same dynamics as $s^{HH}(A^H, B^H, t)$. (a) During the consolidation phase, the $A^H - B^H$ weights rise faster than the $A^C - B^C$ ones: the intra-hippocampal weights reach the maximal value 1 after approximately 4 steps, while this is achieved in 12 steps for neocortical ones. (b) When neurogenesis is effective, the hippocampal pattern cannot be retrieved anymore (fig. 3 (f) (g)), so that $s^{HH}(A^H, B^H, t)$ slowly decreases. On the opposite, since the whole neocortical pattern is retrieved independently of the hippocampal fields (fig. 3 (g)), $s^{CC}(A^C, B^C, t)$ remain at the maximum value 1.

weights in the hippocampal fields, between them and with the neocortical field, that is, we set $\gamma^{\alpha\beta} = 0$ for $\alpha \neq \beta$ in eq.(12). In this configuration, the complete neocortical pattern cannot be retrieved: activating B^C neocortical neurons now fails to activate A^C neurons (figure 3 (d)). Similarly, the weights between A^α and B^α neurons fail to consolidate (supplementary figure B.4 (b)(c)). Hence, during the initial stages of the consolidation process, memory cannot be retrieved without a functional hippocampus in our model.

End of consolidation. After approximately 6 cycles (equivalent to six days) of RC and HR steps, the learning weights are fully consolidated everywhere, see figure 4 (a). In particular, strong weights now connect A^C and B^C neurons in the neocortex, which was the longest process to establish. From this stage on, simulation of a retrieval cue step under an hippocampal “lesion”, as presented above, exhibits a complete neocortical pattern retrieval: the activation of B^C neurons is enough to activate the A^C bump, even in the absence of a functional hippocampal learning, see figure 3 (e). Hence, in opposition to the initial stages of the consolidation process where neocortical pattern retrieval was not possible without the hippocampus learning weights, retrieval becomes independent of the hippocampal fields at the end of the consolidation process, since the $A^C - B^C$ weights have become strong enough.

4.2. Forgetting

While neocortical memories, once consolidated, can persist for years, the SCT stipulates that hippocampal patterns disappear quickly. To explore this erasure part of the dynamics, we followed the neurogenesis hypothesis (see Methods). We focus on neurons that are located in the neighbourhood of the bump pattern and assume that this region of the neural field is where newborn neurons integrate. In the model, we actually reproduce this process by imposing to this region a tenfold decrease of the firing threshold κ_{in} in eq. (8). Since these neurons fire continuously for a long time even in the absence of input (see fig. 3 (f)(g)), the slow adaptation term $\kappa_n^\alpha(x, t)$ of eq. (9) in the neighboring regions of the two hippocampal neural fields (D and H) reaches its stationary value η_{κ_n} . This saturation of the thresholds of the hippocampal pattern neurons prevents the formation of bump attractors when external currents are applied. As a result, with neurogenesis, the HR and the RC steps both fail to evoke the two bumps pattern in the hippocampal neural fields, see figure 3 (f)(g). Neurogenesis thus induces both a reduced excitability of the pattern neurons and a failure to reactive hippocampal memory patterns. The neurons of the pattern in the hippocampal fields consequently exhibit a long-lasting period of silence: the $L^{\alpha\beta}(x, y, t)$ terms in the equation of their synaptic plasticity (eq. (13)) progressively decreases, while the forgetting term $F^{\alpha\beta}(x, y, t)$ grows. After some steps, $s^{\alpha\beta}\partial t(x, y, t)$ eventually vanishes, a point at which the hippocampal pattern is effectively erased.

Therefore, with neurogenesis, hippocampal memory patterns failed to be retrieved, contrary to neocortical ones during retrieval cues since the C memory pattern at this stage of the consolidation process is strong enough to be retrieved entirely. After approximately 300 cycles (equivalent to roughly 300 days) our model exhibits a complete

erasure of hippocampal memory pattern while the neocortical one is still strong and retrieved independently of the hippocampus (fig. 4 (b) and supplementary figure B.5 (c)(d)).

5. Discussion

Summary. Previous computational models have reproduced SCT processes and addressed a range of related fundamental questions (Squire and Alvarez, 1995; McClelland et al., 1995; Káli and Dayan, 2004; Meeter and Murre, 2005; Amaral et al., 2008; Helfer and Shultz, 2020; Howard et al., 2022). However, to our knowledge, little attention has been paid to the underlying neurobiological processes and how they are responsible for the differences in learning rates and the erasure of memories between the neocortex and the hippocampus. We proposed in this paper a computational model of SCT that embarks two main hypotheses from the neurobiological literature:

- The slower consolidation in the neocortex could be explained by its larger structure, implying longer durations to connect remote areas of the same memory pattern.
- The forgetting of hippocampal memories could be due to adult neurogenesis in the dentate gyrus and perturbation of memory retrieval.

To model SCT, we adopted an original neural field approach, with original components compared to classic neural field models; such as interactions between three coupled neural fields or long-term synaptic plasticity. Our numerical simulations reproduced the main features of the standard consolidation theory, through a set of emerging processes that we summarize on figure 5. After the initial encoding step, the connections between the elements of the neocortical memory pattern (bumps) are weaker than those in the hippocampus, because of the larger distance between them. Therefore, at this stage, pattern retrieval in the neocortex needs neocortex-hippocampus connections (see figure 5 (a)): during neocortical pattern retrieval, neural activity is first transmitted from the stimulated neocortical element, here B^C , to its equivalent area in the hippocampus (B^D and B^H , steps 1-2 in fig. 5 (a)). This subsequently activates the other parts of the hippocampal memory pattern as an intra-hippocampal process (step 3, A^D and A^H). These activations of A^D and A^H eventually feedback to the neocortical neural field to retrieve the activation of A^C (step 4). However as the consolidation process progresses, hippocampal replays as well as retrieval cue steps induce bumps of firing and thus learning weight consolidation in the three neural fields, even for connections located more remotely and in particular in the neocortex. At this stage, then, the connections between A^C and B^C have become large enough that the activation of B^C during retrieval directly activates A^C , as a pure intra-cortical process (steps 1-2 of fig. 5 (b)).

In parallel with the learning dynamics, we explored the effect of neurogenesis in the dentate gyrus on the stability of hippocampal memories. We modelled the insertion of highly excitable newborn neurons by a reduction of the threshold of a small fraction of dentate gyrus neurons in the neighborhood of the pattern neurons. The sustained firing of newborn neurons was found to prevent the reactivation of both dentate gyrus and CA memory patterns, due to local threshold adaptations in nearby pattern neurons. This inhibits hippocampal memory retrieval as soon as neurogenesis is significant, i.e. at long timescales, where the neocortical engram is the only one to be retrieved during the retrieval cue steps (see fig. 5 (b)).

An important prediction of our model is that the structure of the trisynaptic pathway could play a crucial role in SCT. Indeed, we found out that a crucial element in our model is that the neural field where neurogenesis takes place (here the dentate gyrus) does not project connections to the neocortex, but to a third neural field (here the CA1-CA3 subfield region). In the presence of direct connections from the neurogenesis neural field to the neocortical one, the insertion of the newborn neurons erases the memory patterns not only in the hippocampus, but also in the neocortex. Hence, in our model, the connectivity structure of the trisynaptic pathway, with its three individual regions and a partially feed-forward connectivity (no connection from the dentate gyrus to the neocortex, in particular), allows neurogenesis to erase hippocampal memory patterns while sparing neocortical ones.

We sketch in the supplementary material section Appendix A.1 an analytical framework to study the multi-bump pattern solutions in our model and their stability with regards to spike frequency adaptation, synaptic depression and long-term synaptic plasticity. This analysis is challenging because of the spatial and temporal dimensions of the bump pattern, that demands considering complicated spatio-temporal perturbations (contraction, expansion, shifts) and because of the presence of three interconnected neural fields. Because of this, we could not complete the analysis of all the possible perturbations. However, our preliminary analysis shows that the presence of learning weights

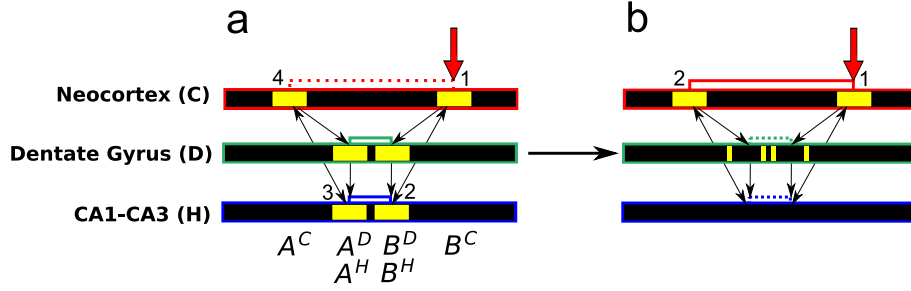


Figure 5: Evolution from a dependent to an independent neocortical pattern retrieval in the connected neural fields. The retrieval cue is a partial signal on only A^C or B^C in the neocortex, here B^C (vertical red arrow). (a) At the beginning of the whole process, when the neocortex is dependent on the hippocampus, the activity is transmitted following $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$. A^C is activated via the hippocampal fields. (b) However at the end of the process, the cortical weights between A^C and B^C are strong enough to activate each other directly ($1 \rightarrow 2$), independently of the hippocampus, into which pattern retrieval is prevented by neurogenesis in the dentate gyrus.

(long-term synaptic plasticity) during the consolidation process stabilizes the multi-bump solution towards contraction perturbations whereas it has no effect on expansion perturbations. Future work will be needed to interpret these results in neurobiological terms, in particular regarding shift perturbations, but we believe that the possibility to produce such analytical evaluation even in the case of interacting complex neurobiological processes, is one of the main interest of neural fields.

Future directions. Despite its complexity, our computational model is still highly simplified in regards to the neurobiology and could be improved in several ways. The modelling of the learning weights (eq. (12) to (17)) could be refined to achieve expressions with a stronger biophysical support, such as the adaptation of STDP for neural fields (Robinson, 2011), while maintaining the continuum in time and space. In particular the phenomenological continuity function $\eta_s^{\alpha\beta}(x, t)$ (eq. (17)) could be re-evaluated, and replaced by another mechanism providing continuity of the bump solutions, an important feature in neural field models. Furthermore, to approach biological reality, it is important to question the pertinence of the use of neural fields for hippocampus networks since neural fields have been developed to model neocortical networks (Coombes, 2005). Here, we used an homogeneous external input to induce hippocampal replays, thus emulating slow-wave sleep. This could be replaced by a more realistic spontaneous reactivation mechanism (Klinzing et al., 2019).

Moreover, the timescales used in our model are questionable. We take for reference the time constant for the neural activity $u^\alpha(x, t)$ which has been fixed to 1 in our model, but which is estimated to be around 10 ms experimentally (Abbott et al., 1997). In our model the synaptic resources recovery rate α_q in synaptic depression eq.(10), should lie between 20 and 80 (since experimentally estimated to be between 200 and 800 ms (Abbott et al., 1997)). However, the numerical value used in our simulations is $\alpha_q = 800$. Likewise, the spike frequency timescale τ_κ should be between 4 and 12 (experimentally estimated to be between 40 and 120 ms (Madison and Nicoll, 1984)). But in our numerical simulations we used $\tau_\kappa = 0.8$. These discrepancies between the experimental values and those used for our parameters can be explained by the complexity of the process to model. Indeed multiple timescales are present in the model, including those mentioned above, the activity-dependent synaptic resources depletion rate $m_q^\alpha(x, t)$ (equations(10)(11)), the timescale for the threshold related to neurogenesis τ_{κ_n} (eq.(9)), the distance-dependent learning rate $d(x, y)$ (equations(14)(15)), the forgetting rate c_0 (eq.(16)), the duration of the stimulations $I_{\text{ext}}^\alpha(x, t)$ (eq.(18) for $I_{\text{ext}}^C(x, t)$). The values of these time-scale parameters must be reevaluated in future works, to better approach biological reality.

Furthermore, additional phenomenological mechanisms could be accounted for. In the equation for the threshold related to neurogenesis $\kappa_n^\alpha(x, t)$, eq.(9), we used a Heaviside function for simplicity. But the effects of a smoother firing rate function would need to be studied in the future. In addition, the phenomenological introduction of an activity dependent synaptic resources depletion rate $m_q^\alpha(x, t) = \beta_q u^\alpha(x, t)$ in equations(10)(11) was found convenient in our model because it allows to get a roughly constant duration for the bump patterns even with changing synaptic weights or facing neurogenesis. However further work must be carried out to study whether this mechanism has a physiological meaning or should be removed from the model.

This model succinctly studied the neurogenesis mechanism leading to the erasure of memories, and would need to be further developed to be fully understood as it has been in previous modelling works (Deisseroth et al., 2004; Meltzer et al., 2005; Weisz and Argibay, 2012; Becker, 2005). In particular, its integration in a neural field model in coordination with the different homeostasis processes could be investigated in more detail. Another interesting hypothesis for forgetting is evoked by experiments of silencing of engrams (Tonegawa et al., 2018; Josselyn and Tonegawa, 2020).

Finally our model aimed at providing a proof of principle and a selection for the main ingredients to include in the model. It would be interesting to test whether similar results can be obtained with a more realistic neural network, such as a stochastic spiking neural network. Such a model would allow more complex pattern configurations and more accurate dynamics rules for the learning weights. More detailed attention to anatomical properties of the circuits would also be of great interest (Pyka and Cheng, 2014) and more realistic synaptic plasticity equations could be implemented (Tomé et al., 2022). Thus, it would be compelling to derive a neural field model from such a microscopic model, and to compare its components with those developed in this paper.

Here, we focused on the widespread standard consolidation theory. However, there exist other models and questions around systems memory consolidation concepts. For instance, the multiple trace theory suggests that some of the hippocampus patterns are conserved in the long term. This theory follows observations of hippocampal damages that produced temporally-graded retrograde amnesia only for semantic memories, but not for episodic ones (Nadel and Moscovitch, 1997). The trace transformation theory posits a selection process that depends on the circumstances at retrieval, and permits the persistence of both the neocortical or the hippocampal memory (Winocur et al., 2010). On the other hand, the more recent concept of active systems consolidation studies in greater detail the influence of sleep on consolidation (Klinzing et al., 2019). Another interesting related phenomenon which has already been modelled in some connectionist models (Amaral et al., 2008; Helfer and Shultz, 2020), is systems memory reconsolidation. In this approach, it is the replay of an already consolidated memory, which can involve again the hippocampus (Nader and Einarsson, 2010).

In addition to these theories, some models suggest that the prefrontal cortex might play a key role in memory organization (Laroche et al., 2000). On the one hand, the prefrontal cortex is involved in the processing and integration of ancient neocortical memories and seems to inhibit hippocampal activity when new information is too similar to an already stored neocortical pattern (Frankland and Bontempi, 2005; Preston and Eichenbaum, 2013). This has been included in a recent connectionist model (Hwu and Krichmar, 2020). On the other hand, the prefrontal cortex has also been identified to be a location for working memory (Joaquin, 2010). A link between working and long-term memory has been introduced in a modelling study, with the hippocampus modelled as intermediate-term memory (Fiebig and Lansner, 2014). Tonegawa et al. also studied the role of the prefrontal cortex and the basolateral amygdala in memory reorganization (Tonegawa et al., 2018). The implementation of all these proposals into a coupled neural field model similar to the one used here could open perspectives in the domain.

Finally, the preliminary stability analysis results shown in supplementary materials Appendix A have not been addressed numerically through the different perturbations. It remains an important issue to address in the future, in particular because our mathematical analysis only provides conditions for instability, not for stability. Furthermore, for the analysis of the synaptic depression, we used $m_q^\alpha(x, t) = \beta_q$, but it will be important to study the case $m_q^\alpha(x, t) = \beta_q u^\alpha(x, t)$, as used in our numerical simulations.

References

- Abbassian, A.H., Fotouhi, M., Heidari, M., 2012. Neural Fields with Fast Learning Dynamic Kernel. *Biological Cybernetics* 106, 15–26. doi:10.1007/s00422-012-0475-9.
- Abbott, L.F., Varela, J.A., Sen, K., Nelson, S.B., 1997. Synaptic Depression and Cortical Gain Control. *Science* 275, 221–224. doi:10.1126/science.275.5297.221.
- Amaral, O.B., Osan, R., Roesler, R., Tort, A.B., 2008. A Synaptic Reinforcement-Based Model for Transient Amnesia Following Disruptions of Memory Consolidation and Reconsolidation. *Hippocampus* 18, 584–601. doi:10.1002/hipo.20420.
- Amari, S.i., 1977. Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields. *Biological Cybernetics* 27, 77–87. doi:10.1007/BF00337259.
- Basu, J., Siegelbaum, S.A., 2015. The Corticohippocampal Circuit, Synaptic Plasticity, and Memory. *Cold Spring Harbor Perspectives in Biology* 7, a021733. doi:10.1101/cshperspect.a021733.

- Becker, S., 2005. A Computational Principle for Hippocampal Learning and Neurogenesis. *Hippocampus* 15, 722–738. doi:10.1002/hipo.20095.
- Ben-Yishai, R., Bar-Or, R.L., Sompolinsky, H., 1995. Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences* 92, 3844–3848. doi:10.1073/pnas.92.9.3844.
- Benda, J., Herz, A.V.M., 2003. A Universal Model for Spike-Frequency Adaptation. *Neural Computation* 15, 2523–2564. doi:10.1162/089976603322385063.
- Bressloff, P.C., 2012. Spatiotemporal Dynamics of Continuum Neural Fields. *Journal of Physics A: Mathematical and Theoretical* 45, 033001. doi:10.1088/1751-8113/45/3/033001.
- Chklovskii, D.B., Mel, B.W., Svoboda, K., 2004. Cortical Rewiring and Information Storage. *Nature* 431, 782–788. doi:10.1038/nature03012.
- Coombes, S., 2005. Waves, Bumps, and Patterns in Neural Field Theories. *Biological Cybernetics* 93, 91–108. doi:10.1007/s00422-005-0574-y.
- Coombes, S., beim Graben, P., Potthast, R., 2014. Tutorial on Neural Field Theory, in: Coombes, S., beim Graben, P., Potthast, R., Wright, J. (Eds.), *Neural Fields: Theory and Applications*. Springer, Berlin, Heidelberg, pp. 1–43. doi:10.1007/978-3-642-54593-1_1.
- Coombes, S., Owen, M.R., 2005. Bumps, Breathers, and Waves in a Neural Network with Spike Frequency Adaptation. *Physical Review Letters* 94, 148102. doi:10.1103/PhysRevLett.94.148102.
- Davis, R.L., Zhong, Y., 2017. The Biology of Forgetting—A Perspective. *Neuron* 95, 490–503. doi:10.1016/j.neuron.2017.05.039.
- Deisseroth, K., Singla, S., Toda, H., Monje, M., Palmer, T.D., Malenka, R.C., 2004. Excitation-Neurogenesis Coupling in Adult Neural Stem/Progenitor Cells. *Neuron* 42, 535–552. doi:10.1016/S0896-6273(04)00266-1.
- Detorakis, G.I., Rougier, N.P., 2012. A Neural Field Model of the Somatosensory Cortex: Formation, Maintenance and Reorganization of Ordered Topographic Maps. *PLoS ONE* 7, e40257. doi:10.1371/journal.pone.0040257.
- Ferreira, F., Erhlagen, W., Bicho, E., 2016. Multi-Bump Solutions in a Neural Field Model with External Inputs. *Physica D: Nonlinear Phenomena* 326, 32–51. doi:10.1016/j.physd.2016.01.009.
- Fiebig, F., Lansner, A., 2014. Memory Consolidation from Seconds to Weeks: A Three-Stage Neural Network Model with Autonomous Reinstatement Dynamics. *Frontiers in Computational Neuroscience* 8. doi:10.3389/fncom.2014.00064.
- Fotouhi, M., Heidari, M., Sharifitabar, M., 2015. Continuous Neural Network with Windowed Hebbian Learning. *Biological Cybernetics* 109, 321–332. doi:10.1007/s00422-015-0645-7.
- Frankland, P.W., Bontempi, B., 2005. The Organization of Recent and Remote Memories. *Nature Reviews Neuroscience* 6, 119–130.
- Frankland, P.W., Köhler, S., Josselyn, S.A., 2013. Hippocampal Neurogenesis and Forgetting. *Trends in Neurosciences* 36, 497–503. doi:10.1016/j.tins.2013.05.002.
- Frankland, P.W., O'Brien, C., Ohno, M., Kirkwood, A., Silva, A.J., 2001. α -CaMKII-dependent Plasticity in the Cortex Is Required for Permanent Memory. *Nature* 411, 309–313. doi:10.1038/35077089.
- Gale, N., Rodger, J., Small, M., Eglon, S., 2022. Analysis of Activity Dependent Development of Topographic Maps in Neural Field Theory with Short Time Scale Dependent Plasticity. *Mathematical Neuroscience and Applications Volume 2*, 8390. doi:10.46298/mna.8390.
- Ge, S., Yang, C.h., Hsu, K.s., Ming, G.l., Song, H., 2007. A Critical Period for Enhanced Synaptic Plasticity in Newly Generated Neurons of the Adult Brain. *Neuron* 54, 559–566. doi:10.1016/j.neuron.2007.05.002.
- Gerstner, W., Kistler, W.M., 2002. Mathematical Formulations of Hebbian Learning. *Biological Cybernetics* 87, 404–415. doi:10.1007/s00422-002-0353-y.
- Goldman-Rakic, P., 1995. Cellular Basis of Working Memory. *Neuron* 14, 477–485. doi:10.1016/0896-6273(95)90304-6.
- Helfer, P., Shultz, T.R., 2020. A Computational Model of Systems Memory Consolidation and Reconsolidation. *Hippocampus* 30, 659–677. doi:10.1002/hipo.23187.
- Howard, M.D., Skorheim, S.W., Pilly, P.K., 2022. A Model of Bi-Directional Interactions between Complementary Learning Systems for Memory Consolidation of Sequential Experiences. *Frontiers in Systems Neuroscience* 16, 972235. doi:10.3389/fnsys.2022.972235.
- Hwu, T., Krichmar, J.L., 2020. A Neural Model of Schemas and Memory Encoding. *Biological Cybernetics* 114, 169–186. doi:10.1007/s00422-019-00808-7.
- Izhikevich, E.M., 2006. *Dynamical Systems in Neuroscience. The Geometry of Excitability and Bursting*. doi:10.7551/mitpress/2526.001.0001.
- Joaquin, F., 2010. Cognitive Functions of the Prefrontal Cortex. *Frontiers in Human Neuroscience* 4. doi:10.3389/conf.fnins.2010.14.00001.
- Josselyn, S.A., Köhler, S., Frankland, P.W., 2015. Finding the Engram. *Nature Reviews Neuroscience* 16, 521–534. doi:10.1038/nrn4000.
- Josselyn, S.A., Tonegawa, S., 2020. Memory Engrams: Recalling the Past and Imagining the Future. *Science* 367, eaaw4325. doi:10.1126/science.aaw4325.
- Káli, S., Dayan, P., 2004. Off-Line Replay Maintains Declarative Memories in a Model of Hippocampal-Neocortical Interactions. *Nature Neuroscience* 7, 286–294. doi:10.1038/nn1202.
- Karmarkar, U.R., Buonomano, D.V., 2006. Different Forms of Homeostatic Plasticity Are Engaged with Distinct Temporal Profiles: Differential Expression of Homeostatic Plasticity. *European Journal of Neuroscience* 23, 1575–1584. doi:10.1111/j.1460-9568.2006.04692.x.
- Kilpatrick, Z.P., Bressloff, P.C., 2010a. Effects of Synaptic Depression and Adaptation on Spatiotemporal Dynamics of an Excitatory Neuronal Network. *Physica D: Nonlinear Phenomena* 239, 547–560. doi:10.1016/j.physd.2009.06.003.
- Kilpatrick, Z.P., Bressloff, P.C., 2010b. Stability of Bumps in Piecewise Smooth Neural Fields with Nonlinear Adaptation. *Physica D: Nonlinear Phenomena* 239, 1048–1060. doi:10.1016/j.physd.2010.02.016.
- Klinzing, J.G., Niethard, N., Born, J., 2019. Mechanisms of Systems Memory Consolidation during Sleep. *Nature Neuroscience* 22, 1598–1610. doi:10.1038/s41593-019-0467-3.
- Ko, S.Y., Frankland, P.W., 2021. Neurogenesis-Dependent Transformation of Hippocampal Engrams. *Neuroscience Letters* 762, 136176. doi:10.1016/j.neulet.2021.136176.
- Laing, C.R., Troy, W.C., Gutkin, B., Ermentrout, G.B., 2002. Multiple Bumps in a Neuronal Model of Working Memory. *SIAM Journal on Applied Mathematics* 63, 62–97. doi:10.1137/S0036139901389495.

- Laroche, S., Davis, S., Jay, T.M., 2000. Plasticity at Hippocampal to Prefrontal Cortex Synapses: Dual Roles in Working Memory and Consolidation. *Hippocampus* 10, 438–446. doi:10.1002/1098-1063(2000)10:4<438::AID-HIP10>3.0.CO;2-3.
- Leuner, B., Mendolia-Loffredo, S., Kozorovitskiy, Y., Samburg, D., Gould, E., Shors, T.J., 2004. Learning Enhances the Survival of New Neurons beyond the Time When the Hippocampus Is Required for Memory. *The Journal of Neuroscience* 24, 7477–7481. doi:10.1523/JNEUROSCI.0204-04.2004.
- Lisman, J., Morris, R., 2001. Why Is the Cortex a Slow Learner? *Nature* 411, 248–249. doi:10.1038/35077185.
- Madison, D.V., Nicoll, R.A., 1984. Control of the Repetitive Discharge of Rat CA 1 Pyramidal Neurones in Vitro. *The Journal of Physiology* 354, 319–331. doi:10.1113/jphysiol.1984.sp015378.
- Marr, D., Willshaw, D., McNaughton, B., 1991. Simple Memory: A Theory for Archicortex, in: Vaina, L. (Ed.), *From the Retina to the Neocortex*. Birkhäuser Boston, Boston, MA, pp. 59–128. doi:10.1007/978-1-4684-6775-8_5.
- Martin, S.J., Grimwood, P.D., Morris, R.G.M., 2000. Synaptic Plasticity and Memory: An Evaluation of the Hypothesis. *Annual Review of Neuroscience* 23, 649–711. doi:10.1146/annurev.neuro.23.1.649.
- McClelland, J.L., McNaughton, B.L., O'Reilly, R.C., 1995. Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory. *Psychological Review* 102, 419–457. doi:10.1037/0033-295X.102.3.419.
- Meeter, M., Murre, J.M.J., 2005. Tracelink: A Model of Consolidation and Amnesia. *Cognitive Neuropsychology* 22, 559–587. doi:10.1080/02643290442000194.
- Meltzer, L.A., Yabaluri, R., Deisseroth, K., 2005. A Role for Circuit Homeostasis in Adult Neurogenesis. *Trends in Neurosciences* 28, 653–660. doi:10.1016/j.tins.2005.09.007.
- Mongiat, L.A., Espósito, M.S., Lombardi, G., Schinder, A.F., 2009. Reliable Activation of Immature Neurons in the Adult Hippocampus. *PLoS ONE* 4, e5320. doi:10.1371/journal.pone.0005320.
- Nadel, L., Moscovitch, M., 1997. Memory Consolidation, Retrograde Amnesia and the Hippocampal Complex. *Current Opinion in Neurobiology* 7, 217–227. doi:10.1016/S0959-4388(97)80010-4.
- Nader, K., Einarsson, E.Ö., 2010. Memory Reconsolidation: An Update. *Annals of the New York Academy of Sciences* 1191, 27–41. doi:10.1111/j.1749-6632.2010.05443.x.
- Preston, A.R., Eichenbaum, H., 2013. Interplay of Hippocampus and Prefrontal Cortex in Memory. *Current Biology* 23, R764–R773. doi:10.1016/j.cub.2013.05.041.
- Pyka, M., Cheng, S., 2014. Pattern Association and Consolidation Emerges from Connectivity Properties between Cortex and Hippocampus. *PLoS ONE* 9, e85016. doi:10.1371/journal.pone.0085016.
- Robinson, P.A., 2011. Neural Field Theory of Synaptic Plasticity. *Journal of Theoretical Biology* 285, 156–163. doi:10.1016/j.jtbi.2011.06.023.
- Rolls, E., Treves, A., 1997. *Neural Networks and Brain Function*. Oxford University Press. doi:10.1093/acprof:oso/9780198524328.001.0001.
- Roxin, A., Fusi, S., 2013. Efficient Partitioning of Memory Systems and Its Importance for Memory Consolidation. *PLoS Computational Biology* 9, e1003146. doi:10.1371/journal.pcbi.1003146.
- Squire, L.R., Alvarez, P., 1995. Retrograde Amnesia and Memory Consolidation: A Neurobiological Perspective. *Current Opinion in Neurobiology* 5, 169–177. doi:10.1016/0959-4388(95)80023-9.
- Squire, L.R., Stark, C.E., Clark, R.E., 2004. The Medial Temporal Lobe. *Annual Review of Neuroscience* 27, 279–306. doi:10.1146/annurev.neuro.27.070203.144130.
- Tomé, D.F., Sadeh, S., Clopath, C., 2022. Coordinated Hippocampal-Thalamic-Cortical Communication Crucial for Engram Dynamics underneath Systems Consolidation. *Nature Communications* 13, 840. doi:10.1038/s41467-022-28339-z.
- Tonegawa, S., Morrissey, M.D., Kitamura, T., 2018. The Role of Engram Cells in the Systems Consolidation of Memory. *Nature Reviews Neuroscience* 19, 485–498. doi:10.1038/s41583-018-0031-2.
- Weingartner, H., Parker, E.S. (Eds.), 2014. *Memory Consolidation*. 0 ed., Psychology Press. doi:10.4324/9781315802626.
- Weisz, V.I., Argibay, P.F., 2012. Neurogenesis Interferes with the Retrieval of Remote Memories: Forgetting in Neurocomputational Terms. *Cognition* 125, 13–25. doi:10.1016/j.cognition.2012.07.002.
- Willshaw, D.J., Dayan, P., Morris, R.G.M., 2015. Memory, Modelling and Marr: A Commentary on Marr (1971) 'Simple Memory: A Theory of Archicortex'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370, 20140383. doi:10.1098/rstb.2014.0383.
- Winocur, G., Moscovitch, M., Bontempi, B., 2010. Memory Formation and Long-Term Retention in Humans and Animals: Convergence towards a Transformation Account of Hippocampal–Neocortical Interactions. *Neuropsychologia* 48, 2339–2356. doi:10.1016/j.neuropsychologia.2010.04.016.

Appendix A. Supplementary text

Appendix A.1. Effect of learning weights on bump solutions and stability analyses

This subsection focuses on the effect of the learning weights on the existence and the stability of stationary bumps solutions. We assume that the width of the bump is conserved across the neural fields, i.e. $a^C = a^D = a^H = a$, which is verified numerically. To carry out an analytical treatment of existence and stability of the bump patterns, we consider the case where all learning weights are at their equilibrium value, after the consolidation process, with the decay term $F^{\alpha\beta}(x, y, t)$ neglected, since $c_0 \ll \min(d(x, y))$. In fact this leads to a dimension reduction of the system, since bump profiles will all be identical. Which is not the case for transient learning weights due to the differences of timescales within and between neural fields. Furthermore, the threshold $\kappa_n^\alpha(x, t)$ is not considered here, due to its large time scale. It could be used as a slow-varying parameter in continuation bifurcation analysis. The effects of spike frequency adaptation and synaptic depression are studied separately below.

Throughout this analysis we will use the results of Kilpatrick and Bressloff (Kilpatrick and Bressloff, 2010b,a), that are restricted to a single field, but we adapt them to our model composed of three connected neural fields, with learning weights.

Appendix A.1.1. Spike frequency adaptation

Existence of stationary bumps solution. As is usual in the field of neural fields, we replace all the continuous firing functions f and f_a by discrete Heaviside functions ($\beta_f \rightarrow \infty, \beta_{f_a} \rightarrow \infty$).

In this case a stationary bump solution of our model ($u_0^\alpha(x), \kappa_0^\alpha(x), s_0^{\alpha\beta}(x, y)$) satisfies:

$$u_0^\alpha(x) = \begin{cases} \sum_{\beta \in E} G^{\alpha\beta} J^{\alpha\beta}(x, a) & \text{if } x \notin R[u_0^\alpha] \\ \sum_{\beta \in E} G^{\alpha\beta} [J^{\alpha\beta}(x, a)\xi_s - (\xi_s - 1)\kappa_{in}] & \text{if } x \in R[u_0^\alpha] \end{cases}, \quad (\text{A.1})$$

$$\kappa_0^\alpha(x) = \begin{cases} \kappa_{in} & \text{if } x \notin R[\kappa_0^\alpha] \\ \kappa_{in} + \eta_\kappa & \text{if } x \in R[\kappa_0^\alpha] \end{cases}, \quad (\text{A.2})$$

and

$$s_0^{\alpha\beta}(x, y) = \begin{cases} 0 & \text{if } x \notin R[u_0^\alpha] \text{ or } y \notin R[u_0^\beta] \\ 1 & \text{if } x \in R[u_0^\alpha], y \in R[u_0^\beta] \end{cases} \quad (\text{A.3})$$

with

$$J^{\alpha\beta}(x, a) = \sum_{z \in Z^\beta} \left[W(x - \Delta^{\alpha\beta}(x) - (z - a)) - W(x - \Delta^{\alpha\beta}(x) - (z + a)) \right]. \quad (\text{A.4})$$

$R[u_0^\alpha]$ is the excited region for u_0^α , which corresponds also to the location where learning weights have a non-zero value, defined as

$$R[u_0^\alpha] = \bigcup_{z \in Z^\alpha} [z - a, z + a] \quad (\text{A.5})$$

The bump boundaries are defined by the threshold conditions:

$$u_0^\alpha(z \pm a) = \kappa_{in} + \eta_\kappa = 2 \times 2ae^{-2a/s} \quad (\text{A.6})$$

$R[\kappa_0^\alpha]$, the excited region for κ_0^α is different

$$R[\kappa_0^\alpha] = \bigcup_{z \in Z^\alpha} [z - b, z + b] \quad (\text{A.7})$$

with b implied in the relation

$$u_0^\alpha(z \pm b) = \kappa_{in} = 2ae^{-2a/s} \quad (\text{A.8})$$

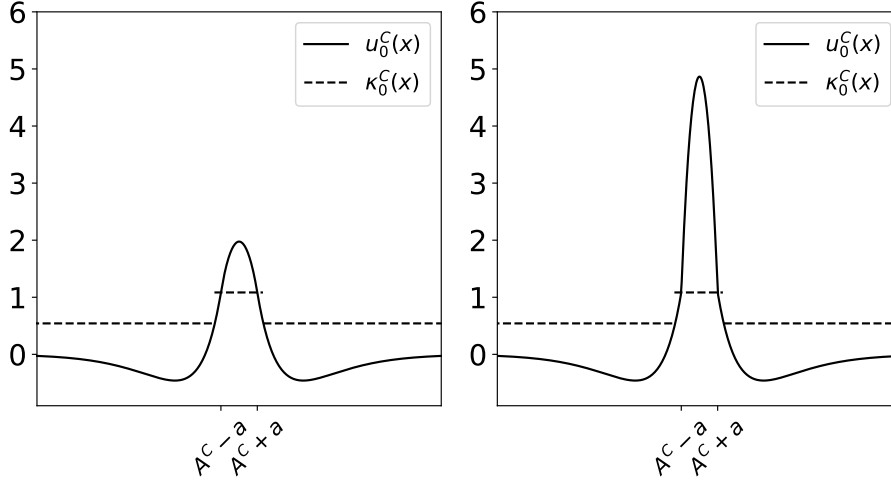


Figure A.1: Shape of a bump in the C neural field in two extreme cases: (*left*) in the absence of learning, i.e. for $s_0^{\alpha\beta}(x, y) = 0$ along the whole field and (*right*) with synaptic modifications at their maximum values, i.e. for $s_0^{\alpha\beta}(x, y)$ set to 1 in the pattern areas.

Please note that $b > a$. The factor ξ_s is defined as:

$$\xi_s = 1 + \gamma(1 + cr)4a \quad (\text{A.9})$$

The second term in the right-hand side of this equation represents the sum over the neural fields and the bumps of the integral of the learning weights. The term ξ_s accounts for the learning weights in the neural fields model. In the absence of learning, $\xi_s = 1$, otherwise $\xi_s > 1$. We will study the effect of varying ξ_s on the existence of bumps and on their stability. We introduce the total weights defined as the sum of the synaptic plasticity term with learning weight at equilibrium and the permanent weights.

$$w_{\text{tot}}^{\alpha\beta}(x, y) = w(x - \Delta^{\alpha\beta}(x) - y) + \gamma^{\alpha\beta} \eta_s^{\alpha\beta}(x) s_0^{\alpha\beta}(x, y) \quad (\text{A.10})$$

where

$$\eta_s^{\alpha\beta}(x) = \left[J^{\alpha\beta}(x, a) - \kappa_{\text{in}} \right]_+ \quad (\text{A.11})$$

Two bump profiles in the C neural field with all the $s_0^{\alpha\beta}(x, y)$ set to 0 or to 1, are presented in figures A.1 (Left) and (Right), respectively. The figure confirms that the continuity of the solution is preserved even with learning weights. The bump boundaries a are not modified by the introduction of synaptic plasticity, but inside the $R[u_0^C]$ zone the bump amplitude is larger with synaptic plasticity.

Stability of the bumps. Following the computations of Kilpatrick and Bressloff (2010) (Kilpatrick and Bressloff, 2010b), we develop eq. (5) with a spatio-temporal perturbation, $u^\alpha(x, t) = u_0^\alpha(x) + \epsilon \phi_u^\alpha(x, t)$, where $\phi_u^\alpha(x, t)$ are smooth perturbations and $\epsilon \ll 1$. Since we truncate the results at first order, no learning weights nor threshold perturbations will appear. The linear stability analysis we derive here thus does not reflect the underlying translation invariance of the system.

We then assume time-space separability of the perturbation: $\phi_u^\alpha(x, t) = e^{\mu t} \psi_u^\alpha(x)$. The calculations are detailed in Ref. (Kilpatrick and Bressloff, 2010b), the modifications of our analysis compared to this study is mainly that the value above which the threshold starts increasing and the baseline threshold are both equal to κ_{in} , which simplifies the analysis. Moreover, in our model, the interconnections between the three neural fields and the learning weights add a 2 factor multiplying the permanent weights, and induce a modification of the spatial derivative of $u_0^\alpha(x)$. In fact, the

factor π_u takes into account different left and right derivatives (see below).
The general perturbation equation can be written as:

$$(\lambda + 1)\psi_u^\alpha(x) = \sum_{\beta \in E} G^{\alpha\beta} \sum_{z \in Z^\beta} \left[w_{\text{tot}}^{\alpha\beta}(x, z - a) \frac{\psi_u^\beta(z - a)}{|u_0'^\beta(z - a)|} + w_{\text{tot}}^{\alpha\beta}(x, z + a) \frac{\psi_u^\beta(z + a)}{|u_0'^\beta(z + a)|} \right] \quad (\text{A.12})$$

with

$$\frac{1}{|u_0'^\alpha(z + \sigma a)|} = \begin{cases} \pi_u & \text{if } \psi_u^\alpha(z + \sigma a) > 0 \\ \frac{\pi_u}{\xi_s} & \text{if } \psi_u^\alpha(z + \sigma a) < 0 \end{cases} \quad (\text{A.13})$$

and

$$\pi_u = \frac{1}{2(w(0) - w(2a))} \quad (\text{A.14})$$

The essential spectrum is located at $\lambda = -1$. The discrete spectrum is obtained by setting $x = z \pm a$, with $z = A^\alpha$ or B^α . At these boundaries $\eta_s^\alpha(x) = 0$, erasing the learning weights terms. Furthermore the distance between two bumps, even in the nearest case in D, H , is large enough so that the values of the non-learned, permanent weights between them can be neglected. Moreover, all learning weights are equal. Therefore, we can assume that $u_0^\alpha(x)$ and $\kappa_0^\alpha(x)$ are equal at their bump boundaries for all the neural field. We also consider that $\psi_u^\alpha(x)$ and $\psi_\kappa^\alpha(x)$ exhibit identical values at each boundary. This simplification allows the calculations below. Moreover, within the linear regime, infinitesimal changes in u^α will only perturb the threshold in a neighborhood of $x = z \pm b$, so that $\phi_\kappa^\alpha(z \pm a, t) = 0$. There are four classes of solutions which determine the discrete spectrum: expansion, contraction, leftward shift and rightward shift of the stationary bump solution.

Expansion case:. $\forall \beta \in E, \forall z \in Z^\beta, \forall \sigma \in \{-1, 1\}, \psi_u^\beta(z + \sigma a) > 0$

$$\left[2\pi_u[w(0) - w(2a)] - (\lambda + 1) \right] \left[2\pi_u[w(0) + w(2a)] - (\lambda + 1) \right] = 0 \quad (\text{A.15})$$

$$\lambda_\pm = \frac{w(0) \pm w(2a)}{w(0) - w(2a)} - 1 \quad (\text{A.16})$$

These eigenvalues are independent of ξ_s . With our parameter values, we find that $\lambda_- = 0$ and $\lambda_+ > 0$, uncovering a degenerate case for which we cannot conclude on the stability.

Contraction case:. $\forall \beta \in E, \forall z \in Z^\beta, \forall \sigma \in \{-1, 1\}, \psi_u^\beta(z + \sigma a) < 0$

$$\lambda_\pm = \frac{1}{\xi_s} \frac{w(0) \pm w(2a)}{w(0) - w(2a)} - 1 \quad (\text{A.17})$$

For $\xi_s > 1$, $\lambda_\pm < 0$. Figure A.2 presents the evolution of the contraction eigenvalues with γ ($\xi_s = 1 + \gamma(1 + cr)4a$). With our parameter values the bump is stable to contraction perturbation, with increasing stability when γ increases.

Leftward shift case:. This case corresponds to a class of solutions where $\forall \beta \in E, \forall z \in Z^\beta, \psi_u^\beta(z + a) < 0$ and $\psi_u^\beta(z - a) > 0$.

$$\left[2\pi_u w(0) - (\lambda + 1) \right] \left[2\pi_u / \xi_s w(0) - (\lambda + 1) \right] - 1 / \xi_s \left[2\pi_u w(2a) \right]^2 = 0 \quad (\text{A.18})$$

$$\lambda_\pm = 2\pi_u \frac{(1 + 1/\xi_s)w(0) \pm \sqrt{(1 + 1/\xi_s)^2 w(0)^2 - 4[w(0)^2 - 1/\xi_s w(2a)^2]}}{2} - 1 \quad (\text{A.19})$$

These eigenvalues are real for small ξ_s values (complex eigenvalues violate the initial assumption that λ are real), where $\lambda_\pm < 0$ for $\xi_s > 1$. When ξ_s increases, λ_+ decreases and λ_- increases, see figure A.3. Therefore, we cannot conclude on the effect of learning weights on shift perturbations.

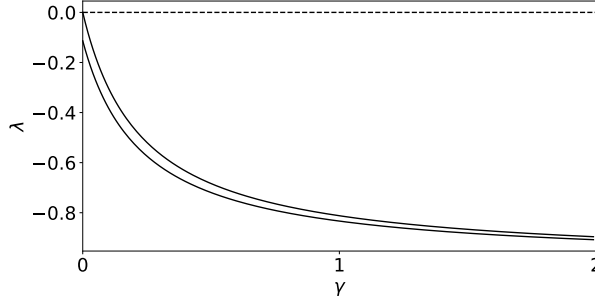


Figure A.2: Learning weights stabilize the bump in case of a contraction perturbation. Eigenvalues λ decrease as γ increases (which increases ξ_s , as $\xi_s = 1 + \gamma(1 + cr)4a$).

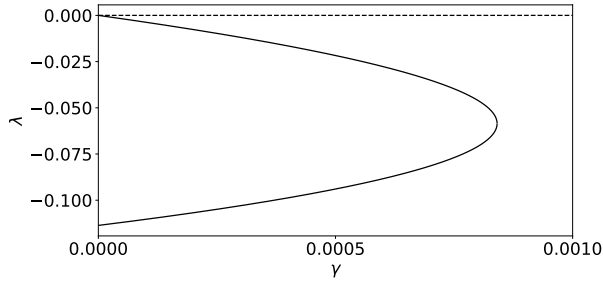


Figure A.3: It is not possible to conclude on the influence of learning weights in case of a shift perturbation. These eigenvalues are real for small γ values, where $\lambda_{\pm} \leq 0$. λ_+ decreases and λ_- increases as γ increases (which increases ξ_s , as $\xi_s = 1 + \gamma(1 + cr)4a$).

Rightward shift case:. This case corresponds to a class of solutions where $\forall \beta \in E, \forall z \in Z^\beta, \psi_u^\beta(z + a) > 0$ and $\psi_u^\beta(z - a) < 0$.

The spectrum associated with rightward shifts is identical to that of leftward shifts due to the reflection symmetry of the system.

To conclude, the introduction of synaptic plasticity / learning weights stabilizes the system through contraction, but has no effects on stability in case of an expansion perturbation. We cannot conclude on their influence on the shift behaviour since it increases one eigenvalue but diminishes the other one. This method can only provide sufficient conditions for instability but not for stability of a bump. So there could be unstable modes not detected by this analysis, although numerical simulations suggest that this is not the case for the parameters of the model.

Appendix A.1.2. Synaptic depression

We now study the effect of the learning weights on the existence of bumps and stability in presence of synaptic depression. With the parameters values used in our model, there is no stable / stationary bump solution when synaptic depression is added. Synaptic depression therefore shuts down the neural bump activities. We still develop the analysis, which could be used in the future to study the effect of the learning weights on an alternative system, but we do not compute values of eigenvalues and thus, we do not conclude here on stability behaviours. We consider the original equation with $m_q^\alpha(x, t) = \beta_q$.

Existence of stationary bumps solution. After replacement of the firing functions by heaviside functions, a stationary bump solution $(u_0^\alpha(x), q_0^\alpha(x), s_0^{\alpha\beta}(x, y))$ is found to satisfy:

$$u_0^\alpha(x) = \begin{cases} \frac{1}{1+\alpha_q\beta_q} \sum_{\beta \in E} G^{\alpha\beta} J^{\alpha\beta}(x, a) & \text{if } x \notin R[u_0^\alpha] \\ \frac{1}{1+\alpha_q\beta_q} \sum_{\beta \in E} G^{\alpha\beta} [J^{\alpha\beta}(x, a)\xi_s - (\xi_s - 1)\kappa_{in}] & \text{if } x \in R[u_0^\alpha] \end{cases} \quad (\text{A.20})$$

The excited region and bumps boundaries are defined by the same conditions described in equations(A.5) and(A.6).

$$q_0^\alpha(x) = \begin{cases} 1 & \text{if } x \notin R[u_0^\alpha] \\ \frac{1}{1+\alpha_q\beta_q} & \text{if } x \in R[u_0^\alpha] \end{cases} \quad (\text{A.21})$$

The equilibrium solution of the learning weight factor $s_0^{\alpha\beta}(x, y)$ was defined in(A.3). We remind that for our parameter values, this bump solution does not exist.

Stability of the bumps. Following Kilpatrick and Bressloff (Kilpatrick and Bressloff, 2010b), we develop eq.(5) and(8) with $u^\alpha(x, t) = u_0^\alpha(x) + \epsilon\phi_u^\alpha(x, t)$ and $q^\alpha(x, t) = q_0^\alpha(x) + \epsilon\phi_q^\alpha(x, t)$, where $\psi_u^\alpha(x, t)$, $\phi_q^\alpha(x, t)$ are smooth perturbations and $\epsilon \ll 1$. With a method similar to the one developed in the previous subsection Appendix A.1.1, we find the general equation:

$$(\lambda + 1)\psi_u^\alpha(x) = \sum_{\beta \in E} G^{\alpha\beta} \sum_{z \in Z^\beta} \sum_{\sigma \in \{-1, 1\}} w_{\text{tot}}^{\alpha\beta}(x, z + \sigma a) \left[1 - \frac{\beta_q \Theta(\psi_u^\beta(z + \sigma a))}{(\lambda + 1/\alpha_q + \beta_q)} \right] G(\psi_u^\beta(z + \sigma a)) \frac{\psi_u^\beta(z + \sigma a)}{|u_0'^\beta(z + \sigma a)|} \quad (\text{A.22})$$

$\frac{1}{|u_0'^\alpha(z + \sigma a)|}$ is defined as in eq.(A.13), but with π_u defined here as:

$$\pi_u = \frac{1 + \alpha_q\beta_q}{2(w(0) - w(2a))} \quad (\text{A.23})$$

The function $G(X)$ is written as:

$$G(X) = \begin{cases} 1 & \text{if } X > 0 \\ (1 + \alpha_q\beta_q)^{-1} & \text{if } X < 0 \end{cases} \quad (\text{A.24})$$

Here again, there are four classes of solutions which determine the discrete spectrum: expansion, contraction, leftward shift and rightward shift of the stationary bump solution.

Expansion case: This case corresponds to a class of solutions where $\forall \beta \in E, \forall z \in Z^\beta, \forall \sigma \in \{-1, 1\}, \psi_u^\beta(z + \sigma a) > 0$. With the simplified equation, since all $\psi_u^\beta(z + \sigma a)$ have the same sign, all $\psi_u^\beta(z + \sigma a)$ are equal and all $\psi_u^\beta(z + \sigma a) > 0$. Finally the eigenvalue λ satisfies the equation

$$(\lambda + \alpha_q^{-1} + \beta_q)(\lambda + 1) = (\lambda + \alpha_q^{-1})(1 + \alpha_q\beta_q)\Omega \quad (\text{A.25})$$

With

$$\Omega = \frac{w(0) + w(2a)}{w(0) - w(2a)} \quad (\text{A.26})$$

The solutions of this equation are:

$$\lambda_\pm = \frac{\Omega(1 + \alpha_q\beta_q) - (1 + \alpha_q^{-1} + \beta_q) \pm \sqrt{(\Omega(1 + \alpha_q\beta_q) - (1 + \alpha_q^{-1} + \beta_q))^2 + 4(\Omega - 1)(\alpha_q^{-1} + \beta_q)}}{2} \quad (\text{A.27})$$

These eigenvalues are independent of ξ_s , thus stability does not depend on γ .

Contraction case: This case corresponds to a class of solutions where $\forall \beta \in E, \forall z \in Z^\beta, \forall \sigma \in \{-1, 1\}, \psi_u^\beta(z + \sigma a) < 0$. With the simplified equation, since all $\psi_u^\beta(z + \sigma a)$ have the same sign, all $\psi_u^\beta(z + \sigma a)$ are equal and all $\psi_u^\beta(z + \sigma a) < 0$. The solution is

$$\lambda = \Omega/\xi_s - 1 \quad (\text{A.28})$$

Assuming $\Omega > 0$ and $\xi_s \geq 1$, the stability is increased by learning weights when γ increases (since $\xi_s = 1 + \gamma(1 + cr)4a$).

Leftward shift case:. This case corresponds to a class of solutions where $\forall \beta \in E, \forall z \in Z^\beta, \psi_u^\beta(z+a) < 0$ and $\psi_u^\beta(z-a) > 0$. With the simplified equation, since all $\psi_u^\beta(z+a)$ (respectively $\psi_u^\beta(z-a)$) have the same sign, all $\psi_u^\beta(z+a)$ ($\psi_u^\beta(z-a)$) are equal.

The resulting equation is

$$\begin{aligned} & \left[\Gamma_{\beta_q}(\lambda) - \pi_u(\lambda + \alpha_q^{-1})2w(0) \right] \left[\Gamma_{\beta_q}(\lambda) - \pi_u/\xi_s(\lambda + \alpha_q^{-1})2w(0) \right] - 1/\xi_s \left[\pi_u(\lambda + \alpha_q^{-1})2w(2a) \right]^2 \\ & + \frac{\lambda \alpha_q \beta_q / \xi_s \gamma_u}{1 + \alpha_q \beta_q} \left[\Gamma_{\beta_q}(\lambda) - \pi_u(\lambda + \alpha_q^{-1})2w(0) \right] 2w(0) - \left[\pi_u(\lambda + \alpha_q^{-1})2w(2a) \right] 2w(2a) = 0, \end{aligned} \quad (\text{A.29})$$

with

$$\Gamma_{\beta_q}(\lambda) = (\lambda + \alpha_q^{-1} + \beta_q)(\lambda + 1) \quad (\text{A.30})$$

The roots of this equation can be searched for numerically, and must satisfy the conditions $\psi_u^\beta(z+a) < 0$ and $\psi_u^\beta(z-a) > 0$. But, since we do not compute them for any set of parameters, we cannot conclude on the effect of learning weights on stability in this case.

Rightward shift case:. This case corresponds to a class of solutions where $\forall \beta \in E, \forall z \in Z^\beta, \psi_u^\beta(z+a) > 0$ and $\psi_u^\beta(z-a) < 0$. Due to the symmetry of the system, the spectrum associated with rightward shifts is identical to that of leftward shifts.

As a conclusion, even if we do not compute the eigenvalues with a given set of parameters, it was still possible to estimate the stability behaviours, provided that equilibrium bump solutions exist. As in the previous subsection Appendix A.1.1, we found that learning weights stabilize the system through contraction, but have no effects on stability in case of an expansion perturbation. However, we cannot conclude on their influence on the shift behaviour since we did not compute eigenvalues values.

All in all, this analysis provided an analytical framework to study the three neural fields model with spike frequency adaptation, synaptic depression and learning weights.

Appendix B. Supplementary figures

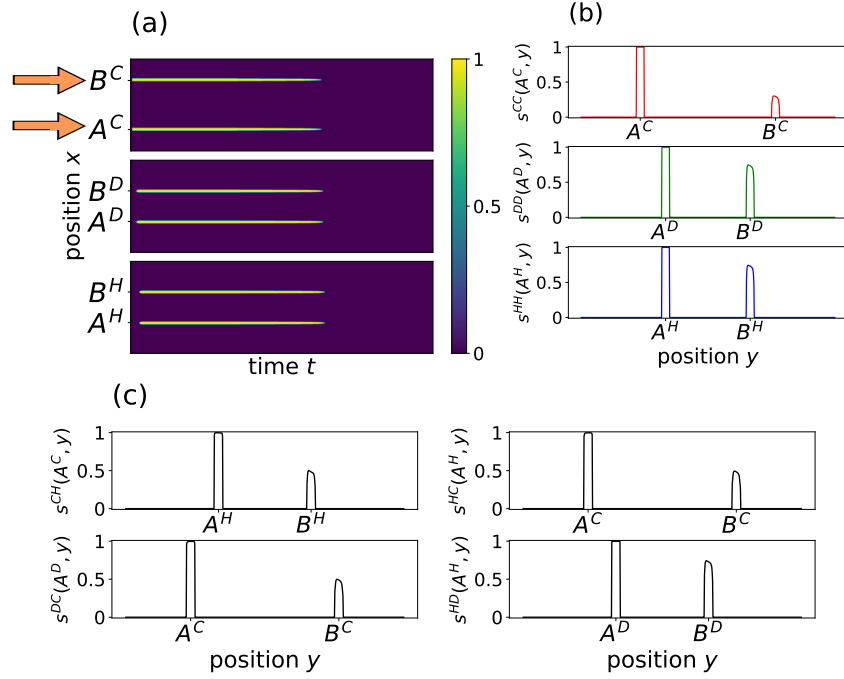


Figure B.1: The encoding phase shows faster learning in hippocampal fields than in neocortical ones. **(a)** Firing rates in the three fields, neocortex (C), dentate gyrus (D) and CA regions (H). The initial stimuli are two bumps around positions A^C and B^C in the neocortical field. They activate C neurons which in turn activate D neurons. Then since C and D neurons fire, H neurons can be activated. In the end, neurons shutdown, due to the depletion of synaptic resources. **(b)** Intra and **(c)** inter-field learning weights at the end of the encoding phase. Learning weights within the A^α part of each pattern are at their maximum. Indeed, neurons inside this location are very close to each other, so their weights grow fast. On the opposite the weights between the A^α and B^α parts are smaller, since the distance is more important. Especially these cross weights are still smaller in C than in D and H fields, since the distance between A^C and B^C is larger than those between A^D and B^D (or A^H and B^H). Thus in our model, distance is the main reason why the neocortex is a slow learner and the hippocampus a fast learner.

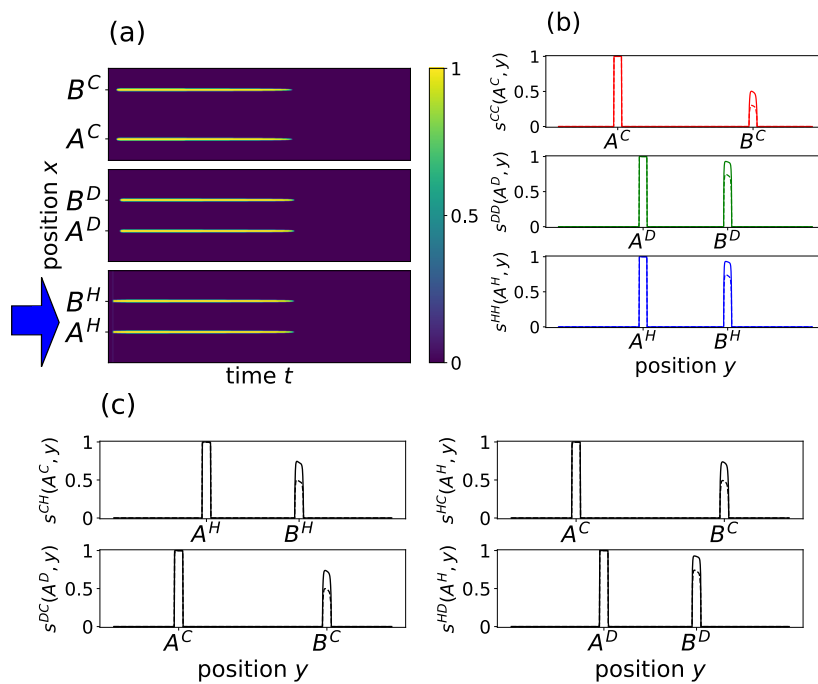


Figure B.2: The hippocampal replay step allows engrams reactivation. **(a)** Firing rates in the three fields, neocortex (C), dentate gyrus (D) and CA regions (H). We modeled hippocampal reactivation by sending a spatially uniform stimulus on H. All H neurons receive this input, but only those involved in the pattern fire for a long time, because of their non zero learning weights. Activated H neurons lead to firing in C neurons, which then activate D neurons. **(b)** Intra and **(c)** inter-field learning weights at the end of the hippocampal replay phase (solid lines) compared to learning weights at the end of the encoding phase (dotted lines). Weights consolidate thanks to pattern reactivation, in particular in the neocortex.

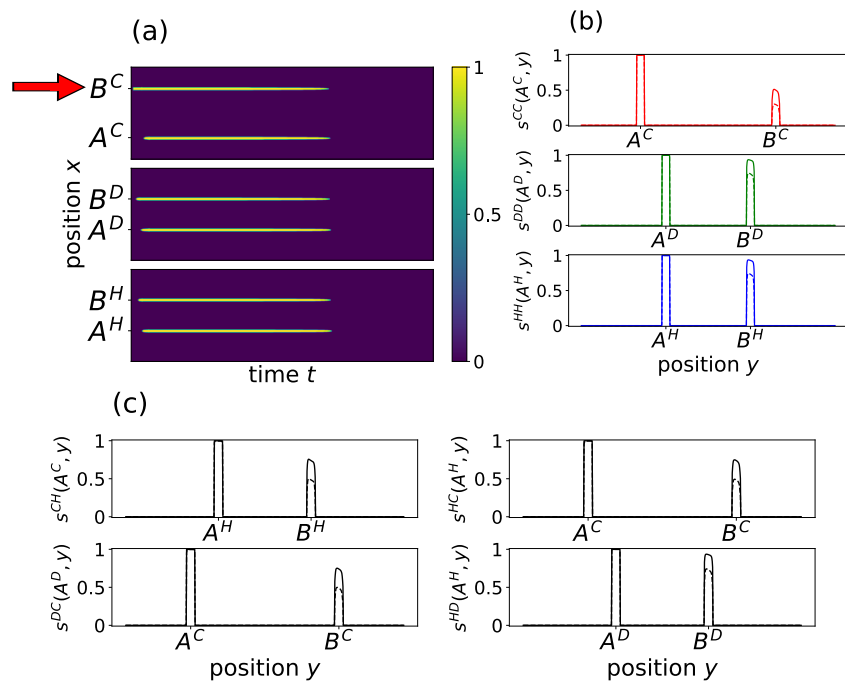


Figure B.3: The retrieval cue step allows engrams reactivation. This step consists in stimulating only the A^C or B^C location of the neocortex (here B^C). Retrieval is exhibited by subsequent activation of A^C . **(a)** Firing rates in the three fields, neocortex (C), dentate gyrus (D) and CA regions (H). A signal is sent on only B^C in the neocortical field. The weights between A^C and B^C are still small, C is still dependent on H to recover the whole pattern. Thus, B^C neurons first activate B^D then B^H neurons, which will activate A^H neurons thanks to the strong enough $B^H - A^H$ learning weights. Finally, the A^H neurons activate A^C neurons via the strong $A^C - A^H$ weights, leading to the whole pattern recovery. **(b)** Intra and **(c)** inter-field learning weights at the end of the retrieval cue process (solid lines) compared to learning weights at the end of the encoding phase (dotted lines). Weights are consolidated thanks to pattern reactivation, in particular in the neocortex.

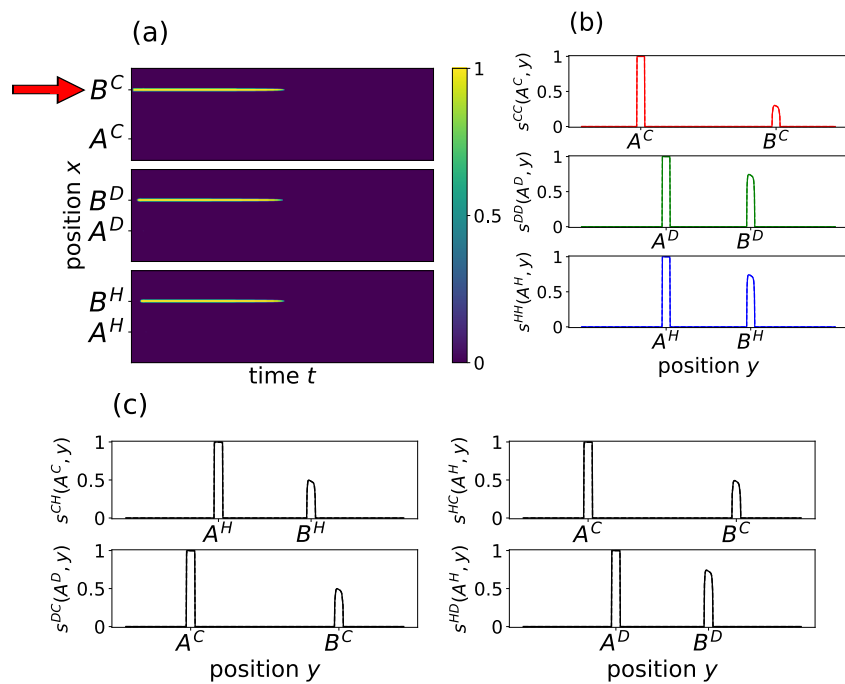


Figure B.4: Neocortical engram reactivation depends on hippocampal fields during the retrieval cue steps. The retrieval cue stimulation with hippocampal “lesion” is obtained by cancelling the learning weights amplitudes in the hippocampal fields, between them and the neocortical field ($\gamma^{\alpha\beta} = 0$ for all α, β except γ^{CC}). (a) Firing rates in the three fields, neocortex (C), dentate gyrus (D) and CA regions (H). Only the B^H neurons, directly stimulated by the partial retrieval cue, fire. Memory cannot be retrieved without a functional hippocampus. (b) The intra and (c) inter-field learning weights between the A^α and B^α cannot be consolidated in absence of firing of the whole pattern.

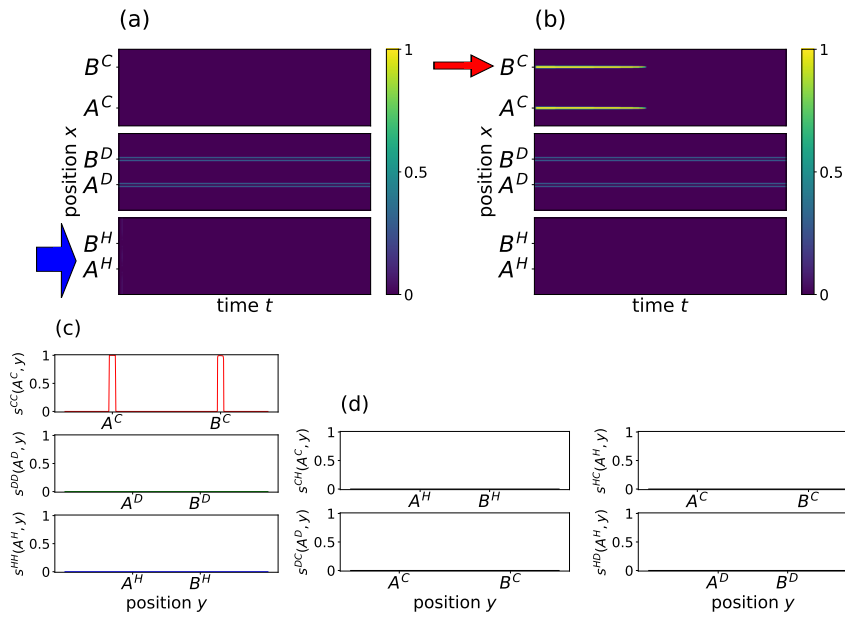


Figure B.5: Neurogenesis in the dentate gyrus prevents the reactivation of hippocampal engrams. **(a)** Firing rates in the three fields, neocortex (C), dentate gyrus (D) and CA regions (H), during the hippocampal replay phase. Neurons in the neighbourhood of the pattern in the dentate gyrus field are modeled as highly excitable newborn neurons (lower threshold, see the methods subsection 3). The thin lines in the D field show the continuous firing of those new neurons. As a result, the thresholds of the neurons in the vicinity adapt and stay high. Therefore, the neurons of the pattern stay silent in response to the replay pattern. **(b)** During neocortical retrieval cues, C neurons fire since there are no direct connections from D to C. However nothing fires in the hippocampal fields. The retrieval cue activates the whole pattern in C independently of the hippocampal fields. **(c)** Intra and **(d)** inter-field learning weights after 325 cycles with neurogenesis (solid lines). Learning weights are fully consolidated in the neocortex and have disappeared in hippocampal fields.