Brain-wide representational drift: memory consolidation and entropic force

Felipe Yaroslav Kalle Kossio[1,2] and Raoul-Martin Memmesheimer[1]

[1]Neural Network Dynamics and Computation, Institute of Genetics, University of Bonn
[2]Department of Developmental Biology, RWTH Aachen University

September 25, 2025

Abstract

Memory engrams change on the microscopic level with time and experience as the neurons that compose them switch in a process termed representational drift. On the macroscopic level the engrams are not static either: numbers of engram neurons in different brain regions change over time, which is considered to reflect the process of memory consolidation. Here we predict a link between these two levels, using a novel statistical physics approach to engram modeling. Importantly, it is general as it makes only minimal assumptions on the engram's nature, does not rely on a specific architecture, and its fundamental implications hold for any representational drift with random component. Our first, analytically well tractable model shows that an entropic force emerges at the region level from random representational drift at the neuronal level. We refine the model by incorporating the interaction of the entropic force with biological processes that shape neuronal engrams. Here, the connectivity between the brain regions strongly influences the (quasi-)equilibrium engram distribution. The obtained distributions are in qualitative agreement with the ones of a biologically detailed drifting assembly model. Our engram description allows to predict the engram evolution in large neuronal systems such as the mouse brain. We find several predictions that are consistent across the valid tested parameter ranges, such as a strong tendency of the engram to leave the hippocampus. The results suggest that the brain operates in a regime where engrams drift, both deterministically and randomly, to allow for memory consolidation.

Introduction

One of the most important functions of the brain is to convert experiences into memories and to store these memories for significant amounts of time. Ensembles of neurons and the synapses interlinking them are believed to provide the physical substrate for the memory storage: the engram [1–6]. Memory engrams are not static: On the microscopic level the neurons that compose them change with time and experience in a process termed representational drift [3, 7–9], whose functional role is still unclear. On the macroscopic level the numbers of engram neurons in different brain regions change over time [10,11]. Furthermore, over time memories are consolidated [12–14]; consequently, the macroscopic change of engram neuron numbers is often assumed to be a neuronal correlate of memory consolidation.

A classical view of consolidation posits that the parts of a new memory engram, which are distributed throughout the brain, are crucially connected through the engram parts in the hippocampus [13, 15, 16]. The hippocampus then guides the isocortex to form direct or indirect connections between the parts itself. Over time, the memory may thereby become completely independent of the hippocampus. This view of consolidation inspired a large number of modeling studies [17–26]. These highlight a variety of network architectures and plasticity rules that may allow the transfer of the functionality of the hippocampal engram parts to the isocortex. However, there are a number of weaknesses of the classical view of consolidation and alternative possibilities exist [27].

Here we develop models connecting representational drift and the distribution of an engram across the brain's regions over the course of time. In our models the memory engrams drift by deterministic and random remodeling. Importantly, the random remodeling can induce on a macroscopic scale practically deterministic transitions between brain areas. Our models do not require the hippocampus to guide the isocortex during consolidation. The hippocampus might then just be one region among many others, which initially binds distributed engram parts possibly due to its connectivity and specific learning abilities [28]. Using an approach from statistical physics as well as detailed neural network modeling we show how the random drift of a memory engram at the microscopic level leads to the emergence of a practically deterministic entropic force [29] on the macroscopic engram, which tends to equilibrate memory coding levels across the brain regions. We further show how neuronal preferences, which can be biologically implemented via plasticity rules and govern deterministic drift, may be captured by an energy function. This allows to study their interaction with the random drift within our framework. We then introduce a detailed biological model of a drifting engram and find that its evolution characteristics qualitatively agree with those of the statistical physics models. Finally we apply our model of engram drift to the whole mouse brain to model and thereby predict macroscopic engram transformation. Our findings suggest that the brain operates in the regime that supports representational drift to allow memory consolidation.

Results

A purely-random engram drift

Engram tracking experiments often report the number of engram neurons in a particular brain region [10, 11]. We call the state specified by this macroscopic level description the macrostate. It omits the exact microscopic details of the engram structure. Consider first a single brain region and an engram within it. The macrostate is then simply the size of the engram: the number n of neurons that form it, Fig. 1. A more detailed description of the engram, listing the particular neurons forming it, specifies a microstate. The same engram macrostate can thus result from many different microstates. We imagine each engram neuron to possess strong synaptic connections with the rest of the engram [4, 5, 30], without yet explicitly modeling them.

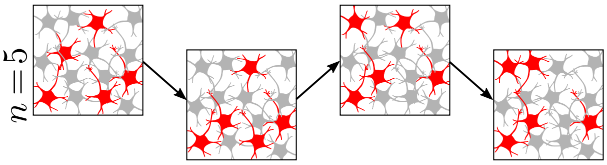

Figure 1: Engram drifting in a single brain region. Engram neurons are shown in red; non-engram neurons in gray. In our first, simple statistical model, the microstate performs a random walk with conserved overall engram size, here $n = 5$. If there is only a single brain region, the engram size defines the macrostate, which is therefore also conserved.

Experiments show that over time, some neurons may leave the engram and others join it, such that the engram drifts [3, 8, 9]. Such representational drift may be caused, for example, by noise in the plasticity rules, spontaneous synaptic turnover, or change of intrinsic properties [31–33]. The exact mechanism behind the drift is not important for our models and analysis. As a first, simple model, we assume that at each time evolution step, a random neuron leaves the engram and a random neuron joins it. Then, the engram size n and, in the case of a single region, the macrostate is conserved, Fig. 1. The drift is a random walk through the microstates. Mathematically it is a random walk on a Johnson graph [34, 35]. Since each microstate has the same number of microstates that it can transition to and arise from, each node of the corresponding graph has the same number of connections. This implies that in the long run each microstate has the same frequency of occurrence [34]. Since all microstates are equally likely, in statistical physics terms they form a microcanonical ensemble [36].

Engram drifting in two brain regions

For two brain regions, an engram has n_1 neurons in the first region and n_2 neurons in the second region, Fig. 2a; n_1 and n_2 define its macrostate $\boldsymbol{n} = (n_1, n_2)$. Random representational drift may gradually change the ensemble of engram neurons and thereby the microstate $\boldsymbol{m}$ ($m_i = 1$ if neuron i is an engram neuron and 0 otherwise). However, the macrostate now changes even if the overall engram size stays constant, Fig. 2b: this is because, for example, a neuron in Region 1 may leave the engram and a neuron in Region 2 may join it.

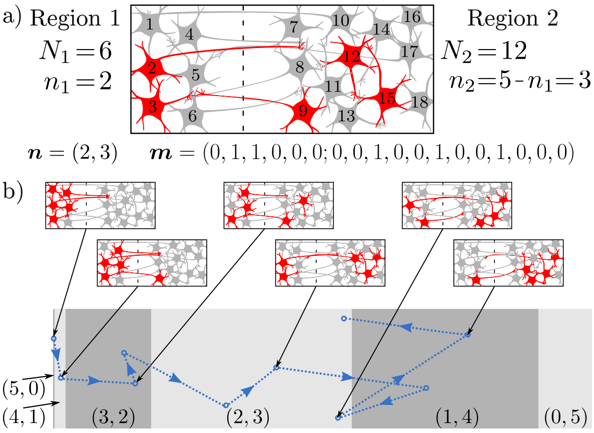

Figure 2: Engram drifting in two brain regions. (a) Example engram microstate for two regions with 6 and 12 neurons (dashed line: separation between the regions). The engram contains five neurons, shown in red; non-engram neurons are displayed in gray. To define a microstate, each neuron is assigned an index. The ith entry of the vector $\boldsymbol{m}$ characterizing the microstate is 1 if the neuron with index i is part of the engram and 0 otherwise. The two entries of the vector $\boldsymbol{n}$ characterizing the macrostate are the numbers of engram neurons in the two regions. (b) Schematic engram trajectory due to drift. (b, upper) Illustration of the realized microstates. (b, lower) Trajectory in the microstate space. Gray shaded areas indicate macrostates and their multiplicities $\Omega(\boldsymbol{n})$.

The simplest way to obtain a two-region network is to superficially define two regions in a single region network, without applying any further changes. We chose such a partition into a smaller region containing N_1 neurons and a larger one with N_2 neurons, where $N_1 + N_2 = N$ is the total network size. We consider a more biologically motivated partition in a later section. In the context of classical theories of memory consolidation (see Introduction), the smaller region may be interpreted as hippocampus and the larger one as isocortex. Assuming again that at each step a random neuron leaves the engram and a random neuron joins it, the process is a random walk in the microstate space with the same properties as in the case of a single region. In particular, in the long run each accessible microstate occurs with the same frequency. The probability of an engram being in a particular macrostate is then proportional to the "volume" that the macrostate occupies in the microstate space, more precisely to its multiplicity $\Omega(\boldsymbol{n})$, i.e. to the number of underlying microstates, Fig. 2b. This is the origin of the entropic force that distributes the engram over different brain regions in our models.

Importantly, the entropic force leads to directed engram changes on the macroscale. To clarify this, we first study the extreme case where engram neurons are initially only in the smaller region; more realistic initial states are considered in a later section. Fig. 2b exemplarily displays engram micro- and macrostates during the initial phase of drift in a very small network. Fig. 3a shows the evolution of the engram macrostates due to drift in a larger network. We observe that the majority of the engram quickly leaves the smaller region (the hippocampus) and enters the larger one (the isocortex) until equilibrium, Fig. 3b, is reached. This general engram transformation may explain the similar experimentally observed transformation that is often identified with memory consolidation and transient initial dependence of many memories on the hippocampus [12–14, 27, 37].

Engram equilibrium

After a sufficiently long period of drift the probability that the engram is in a particular macrostate $\boldsymbol{n} = (n_1, n_2)$ is proportional to its multiplicity. We can obtain this number of corresponding microstates by computing the number of ways to select n_1 engram neurons from the N_1 available Region 1 neurons and n_2 engram neurons from the N_2 available Region 2 neurons, $\Omega(\boldsymbol{n}) = \binom{N_1}{n_1}\binom{N_2}{n_2}$. Since in our simple model the engram size is fixed

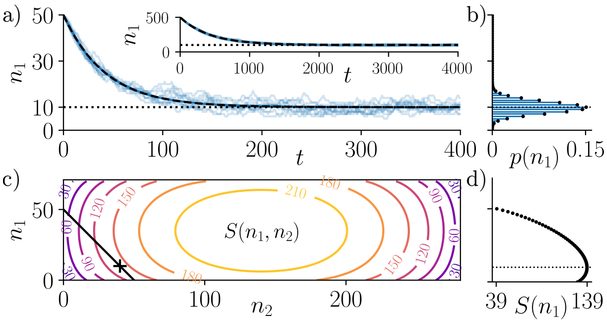

Figure 3: Drift of an engram of fixed size ($n = 50$) in a network with two regions ($N_1 = 70$, $N_2 = 280$). (a) Ten example engram drift trajectories (light blue). The engrams are initially completely in Region 1, $n_1(t = 0) = n$. They quickly enter Region 2 and eventually settle at equilibrium. The majority of engram neurons is then in Region 2. Individual trajectories stay near the ensemble average (blue) and the matching theoretical average (dashed black); the time evolution is close to deterministic despite the rather small size of the considered engram. Time t is the number of engram neuron replacements. (Inset) In a ten times larger system the relative size of fluctuations around the mean trajectory is smaller such that they are hardly visible. (b) Theoretical (black) and observed (blue, 1000 samples) equilibrium probability distributions of a macrostate with n_1 engram neurons in Region 1. (c) Entropy $S(\boldsymbol{n})$ of the macrostate $\boldsymbol{n} = (n_1, n_2)$. The black line indicates the accessible macrostates, which satisfy the constraint $n = n_1 + n_2$; the cross indicates the average of the equilibrium distribution. The microstates are discrete but dense, we use lines instead of points for visualization only. (d) Entropy of the accessible macrostates, $S(n_1) = S(n_1, n - n_1)$, i.e. essentially the entropy along the black line in (b, left). The thin dotted lines in (a,b,d) highlight the equilibrium distribution's average, which is nearly identical to the most probable macrostate.

to n, only those macrostates are accessible (can be present) that satisfy the constraint $n_1 + n_2 = n$. The multiplicity is typically a very large number. Therefore we often take its logarithm, $\ln \Omega(\boldsymbol{n})$, which can be identified with the Boltzmann entropy from statistical physics, $S(\boldsymbol{n}) = \ln \Omega(\boldsymbol{n})$, Fig. 3c,d.

The macrostate equilibrium probability, Figure 3b, is $p(\boldsymbol{n}) = \Omega(\boldsymbol{n})/\Omega(n)$ where the normalizing factor is one over the total number of accessible microstates, $\Omega(n) = \binom{N}{n}$. The expected equilibrium number of engram neurons in the first region is $\langle n_1 \rangle_{\text{eq}} = nN_1/N$. This makes sense intuitively: At a time point long after equilibration, the engram will be homogeneously distributed over the network. Thus the expected fraction of engram neurons in Region 1, $\langle n_1 \rangle_{\text{eq}}/n$ equals the fraction of neurons in Region 1, N_1/N. The drift thus tends to equalize memory coding levels, i.e. fractions of engram neurons, between the regions: As an example in our case of two regions we get $\langle n_1 \rangle_{\text{eq}}/N_1 = \langle n_2 \rangle_{\text{eq}}/N_2$.

The fluctuations at equilibrium, characterized by the coefficient of variation, are $1/\sqrt{\langle n_1 \rangle_{\text{eq}}}$ (for large engrams and networks, Methods): As the number of engram neurons in a region becomes larger, the relative size of equilibrium fluctuations decreases and the equilibrium distribution becomes more sharply peaked.

Approach to the equilibrium

We now examine the evolution of far-from-equilibrium engram macrostates in more detail. Starting in some fixed initial macrostate $n_1(0)$, the engram reaches macrostate $n_1(t)$ after t time steps. The average macrostate trajectory due to drift is

$$\langle n_1(t) \rangle = \langle n_1 \rangle_{\text{eq}} + (n_1(0) - \langle n_1 \rangle_{\text{eq}})e^{-t/\tau}, \qquad (1)$$

where $\tau = -1/\ln\left(1 - \frac{N}{n(N-n)}\right)$ (Methods) and the fluctuations around the average trajectory are small, see Fig. 3a. Therefore, the microscopically random engram drift leads on the macroscopic scale to a directed evolution towards equilibrium. For $N \gg n \gg 1$, which we expect for typical systems, the exponential relaxation to the equilibrium is determined just by the engram size, $\tau \approx n$. The emergence of practically deterministic macroscopic dynamics in a sufficiently large stochastic system is well known in statistical physics: it marks, for example, the transition to thermodynamics [36, 38]. In our system, the directed evolution suggests that there is a force acting on the engram, which drives it towards the equilibrium state. This force is the emergent entropic force [29]. It originates from the fact that the engram remodels randomly and has a large number of possible microstates.

Importantly we demonstrated the existence of this emergent force by purely statistical arguments. This reflects the fact that the exact drift mechanism is irrelevant: Random drift or drift with a random component yields an entropic force. This lets newly formed engrams that are not in equilibrium undergo a transformation on the macroscopic scale. For the simple model examined in this section, the drift yields the only force behind this transformation. In the next sections we consider also the effects of neuronal connectivity preferences and structural connectivity, which lead to more complicated long-term dynamics than an exponential relaxation to equal coding levels.

An engram energy

A hallmark of the interconnectivity of neurons in the brain is that it is generally sparser between neurons in different brain regions. This is a consequence of factors such as space and energy constraints, which limit the number of potential connections [39]. We will henceforth refer to two neurons as structurally connected if a synapse can potentially exist between them [40]. Whether a synapse exists between two structurally connected neurons is influenced by synaptic plasticity [41,42], often in an activity dependent manner. To capture effectively the interplay between neuronal activity, synaptic plasticity, and structural connectivity, we again choose a bird's eye, statistical perspective: We assign each microstate $\boldsymbol{m}$ an energy $H(\boldsymbol{m})$. Microstates with lower energy are preferred by the engram and have a higher chance of occurrence. This allows us to describe the states of our neural system as a canonical ensemble [36], i.e. the probability of a microstate is given by the Boltzmann distribution, $p(\boldsymbol{m}) \propto e^{-\beta H(\boldsymbol{m})}$. This is in contrast to our first model, where all accessible microstates were equiprobable (have the same energy). In statistical physics, the constant β is the inverse of the temperature; small β (high temperature) indicates large random fluctuations and thus a strong entropic force. Analogously, in our system, β determines the strength of the randomness of the representational drift and thus the associated entropic force. In the absence of random representational drift ($\beta \to \infty$), the engram would always drift towards microstates with lower energy.

To construct an appropriate energy function, we introduce for a system consisting of N neurons an $N \times N$ structural connectivity matrix A, with element $A_{ij} = 1$ if neuron j can potentially form a synapse to neuron i and 0 otherwise. As before, the microstate $\boldsymbol{m}$ is a vector of length N, with component $m_i = 1$ if neuron i is an engram neuron and 0 otherwise. To write an expression for the energy we need to make assumptions about the nature of an engram; consistent with previous experimental and theoretical work, we assume that it is a neuronal assembly [43–45]: a group of strongly interconnected neurons. Each assembly neuron should have sufficient but not overwhelming recurrent input from the rest of the assembly. Therefore we introduce a constant k that represents an optimal, desired number of connections from other assembly neurons. Biologically, homeostatic plasticity and limited synaptic weights may increase the probability of such a configuration [31,46] and thus implement the connectivity preference.

Furthermore, to foster an assembly, neurons should prefer to have reciprocal synapses. These are indeed more common than expected by chance in biological neural networks [47]; the preference could be implemented by symmetric forms of activity dependent plasticity [48]. We assume that all structurally permitted synapses between assembly neurons are formed. This is because we expect that biological assembly neurons tend to be coactive and that this leads to the strengthening of possible synapses. The above points lead us to assign the assembly-engram microstate the energy

$$\begin{aligned} H(\boldsymbol{m}) = \sum_{i=1}^{N} \left(\sum_{j=1}^{N} A_{ij} m_j - k \right)^2 m_i \\ + g \sum_{i,j=1}^{N} (A_{ij} - A_{ji})^2 m_i m_j . \end{aligned} \tag{2}$$

Its first term measures how much the number of synapses that each engram neuron receives from other engram neurons deviates from the desired number k. Such deviations are punished quadratically. However, this is not enough to favor a neuronal assembly: for example a circular feedforward chain [46,49] would be likewise favored by this energy term. The distinguishing feature of an assembly are its reciprocal connections between the neurons. Missing reciprocal connections are thus punished by the second term. The constant g determines the relative strength of the two terms.

Engram dynamics and (quasi-)equilibrium

To model the engram dynamics, we choose again in each time step randomly a neuron. If this neuron belongs to the engram, it may now stay or leave. The probability of either depends on which outcome is more favorable, i.e. on the resulting energy change. This can be interpreted as perturbing the state by preliminarily making the change and then probabilistically accepting or rejecting it. Outcomes more favorable than the current state, which result in a negative energy change, are more probable, Fig. 4a. Smaller β renders the energy change less relevant: it yields, for example, higher probabilities that the neuron leaves despite resulting less favorable states. This leads overall to stronger random drift. An alike probabilistic outcome selection happens if the neuron is originally not part of the assembly. Specifically we use the Glauber algorithm [50] for the simulation (Methods).

To illustrate the engram dynamics generated by our model, we first apply it to a network of two regions. The structural connectivity is specified by $p_{11} = p_{22} = 1$ and $p_{12} = p_{21} = p$, where p_{sr} is the probability that a structural connection from a neuron in region r to neuron in region s exists. Thus, within a single region each synaptic connection is possible and the probability of possible inter-region connections is symmetric. We again begin with the engram initially located completely within one of the regions. Fig. 4b-d displays resulting very long-lived metastable states; transient dynamics for different p are shown in Fig. S1. The very long-lived metastable states can be quasi-equilibria, which would change for time tending to infinity, or true equilibria; if we do not need to emphasize this distinction, we denote them all as quasi-equilibria for simplicity. Remarkably, we observe that for the chosen energy function we can find parameters such that the statistical model has qualitatively the same behavior as the detailed biological model introduced in the next section.

If the engram is initially in the smaller region, Fig. 4b, there is an intricate dependence of its quasi-equilibrium on the regions' interconnectivity p. It may be explained as follows: For $p \approx 0$, parameter domain I, the regions are basically separated and the engram remains in the original region for the considered long simulation times. As p increases, domain II, neurons from the larger region can join the engram. The regions are, however, still only weakly interconnected and the synaptic contribution from one region to the other remains small. Therefore, each region contains on its own a slightly smaller number of engram neurons as a single region would host to maintain the desired level of interconnectivity. The slight reduction in the number of neurons is due to the input from the other region. As p increases further, domain III, the engram can freely drift between the two regions, and, since the engram energy is smaller the more reciprocal connections it has, it moves nearly completely to one region, choosing the entropically more favorable, larger one with more available microstates. As the probability p increases further, domain IV, the engram is again found in both regions, since the energy of such microstates is not prohibitively large anymore. As p tends to 1 there is less and less distinction between the regions and the system tends to behave like one large region; the quasi-equilibrium averages are thus determined by the entropy: the numbers of engram neurons are such that the coding levels in both regions become equal as in our first model. If the engram is initially in the larger region, the number of engram neurons in this region simply decays with increasing p. In particular, there is no analog to domain II, because there are too few neurons in region 1 to expect that some have by chance sufficiently many connections with assembly neurons of Region 2. Therefore practically no neurons in Region 1 join the assembly despite the considered long simulation times and it remains in Region 2.

An engram free energy

Microstates that belong to the same macrostate can have different energies Eq. 2 due to the particular realization of the structural connectivity. In order to better understand the evolution of the assembly, we construct a simplified model by averaging the energy over the realizations of structural connectivity. For this we consider R brain regions and again denote the probability of a structural connection from a neuron in region r to a neuron in region s by p_{sr}. A macrostate is then specified by a vector

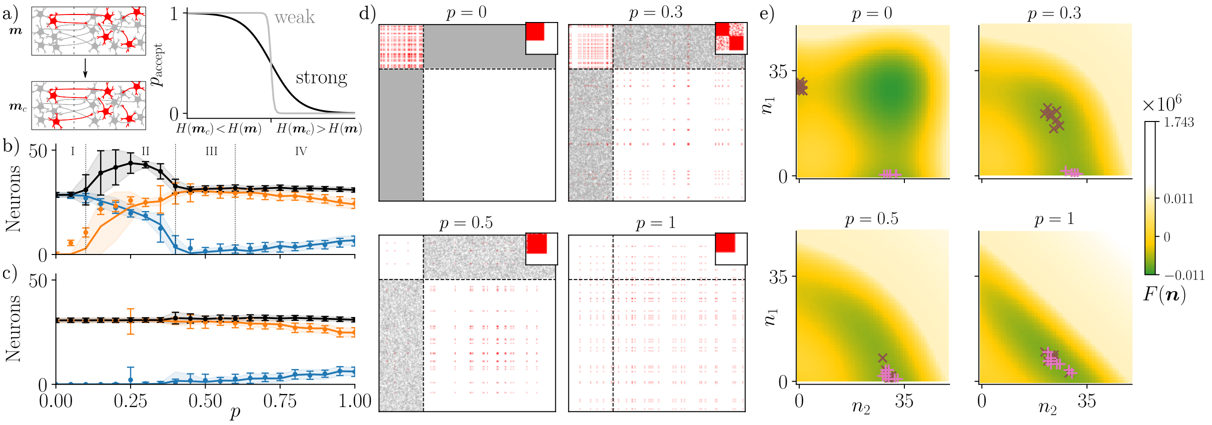

Figure 4: Engram dynamics and quasi-equilibria in the random-and-deterministic drift model ($\beta = 0.012,\ k = 28,\ g = 5.5,\ N_1 = 70,\ N_2 = 280$). (a) Energy-based simulation of engram drift. A candidate microstate is proposed by perturbing the current one, it is accepted or rejected depending on its energy. The probability of acceptance reflects the strength of representational drift. (b) Number of assembly neurons at quasi-equilibrium (mean ± std) in the first (blue) and the second (orange) region and total size (black) as a function of inter-region structural connectivity; initially the assembly is completely in Region 1. Points show mean ± std for the exact model, solid lines with shaded regions show the same for the simplified one. Dotted lines indicate parameter domains (I-IV) with different engram behavior. (c) Same as (b), but with the engram initially completely in Region 2. (d) An example of connections between assembly neurons after long time evolution, for four different levels of structural connectivity between the regions, as determined by the inter region-structural connection probability p. Initially the engrams are completely in Region 1. Connections between assembly neurons are shown in red, existing structural connections in white and absent structural connections in gray. Dashed lines indicate region boundaries. Reordering shows preserved engram structure (inset, focusing on non-empty matrix part) (e) Free energy $F(\boldsymbol{n})$ as a function of the macrostate $\boldsymbol{n}$ for the same four levels of structural connectivity between the regions as in (d). Crosses indicate macrostates of an ensemble of engrams after long evolution (brown "×": engrams initially completely in Region 1, pink "+": engrams initially completely in Region 2).

$\boldsymbol{n}$ of length R that assigns each region the number engram neurons in it. Starting with Eq. 2, the averaging yields for a microstate $\boldsymbol{m}$ that belongs to the macrostate $\boldsymbol{n}$ the energy (see Methods for details)

$$\overline{H}(\boldsymbol{n}) = \sum_{s=1}^{R}\left(\sum_{r=1}^{R} p_{sr} n_r - k\right)^2 n_s + \sum_{s,r=1}^{R} p_{sr}(1-p_{sr}) n_s n_r + 2g \sum_{s,r=1}^{R} p_{sr}(1-p_{rs}) n_s n_r - 2g \sum_{s=1}^{R} p_{ss}(1-p_{ss}) n_s, \tag{3}$$

which only depends on the macrostate. Together with our assumption that the probability of a microstate is given by the Boltzmann distribution, this implies that the probability of macrostate $\boldsymbol{n}$ is proportional to its multiplicity $\Omega(\boldsymbol{n})$ times $e^{-\beta H(\boldsymbol{n})}$, which is just the sum of the probabilities of the microstates associated to $\boldsymbol{n}$. We can rewrite this using the so-called free energy $F(\boldsymbol{n}) = \overline{H}(\boldsymbol{n}) - \ln \Omega(\boldsymbol{n})/\beta$: the probability of macrostate $\boldsymbol{n}$ is then proportional to $e^{-\beta F(\boldsymbol{n})}$. The free energy accounts for the energy and the multiplicity of the macrostate. Thereby, in biological terms, it captures the interaction between neuronal activity and plasticity rules, represented by the energy function, on the one hand and the entropic force on the other hand. For the simplified, network realization-averaged model the (quasi-)equilibrium Fig. 4a,b as well as approach to it, Fig. S1, matches well that of the exact model. The simplified model explicitly highlights the contribution of the macrostate entropy term, $S(\boldsymbol{n}) = \ln \Omega(\boldsymbol{n})$, which exclusively governed the dynamics of our very first engram model, Fig. 3, and implicitly strongly impacts the engram evolution in the full energy model. In the presence of random representational drift, the engram tends to evolve towards minima of the free energy rather than the energy, Fig. 4c, like statistical physics systems that are described by a canonical ensemble [36]. For the chosen system with symmetric structural connectivity probability, the energy, Eq. 3, is symmetric under switching the number of neurons in regions. Thus, differences in reached quasi-equilibria between the regions, Fig. 4, are due to differences in the initial conditions and in the entropic force. The initial conditions in Fig. 4b vs. c and in Fig. 4e are mirrored between Regions 1 and 2 [$(n_1, n_2) = (15, 0)$ vs. $(n_1, n_2) = (0, 15)$]. The symmetric energy function alone would thus induce mirrored temporal dynamics and mirrored quasi-equilibria: Fig. 4c would look like Fig. 4b with orange and blue colors interchanged and the sets of quasi-equilibria in Fig. 4e would be mirror symmetric along the diagonal. The entropic force breaks this mirror symmetry (due the the difference in region sizes), which manifests itself also in the strongly asymmetric free energy landscape, Fig. 4e. We note that depending on the inter-region connectivity p and the form of energy, the engram can be located in both regions or in only one. Since different engram types may have different forms of engram energy, this may explain why some engrams remain hippocampus-dependent while other do not.

A biologically detailed model

In this section we develop a biologically detailed model of a single engram to examine its drift in a two-region network. As in the previous paragraph we model the engram as a neuronal assembly. This allows us to base our model on previous models of multiple, non-overlapping assemblies that drift in a single region [31, 32]: The network (see Methods) consists of linear Poisson ("Hawkes") neurons, describing excitatory neurons in the balanced state. We use standard spike timing-dependent plasticity (STDP) with a slight modification: The strength of the STDP depends on the firing rate of pre- and postsynaptic neurons [51]. Further there is a rate dependent weight decay of input synapses [52].

We divide the network into two regions with full intra-region structural connectivity and symmetric inter-region connectivity probability, as in the previous section. Fig. 5a shows a resulting single drifting assembly in a network with two regions. Fig. 5b displays the average number of assembly neurons in the two regions after long evolution as a function of inter-region connectivity. Trajectories towards quasi-equilibrium are shown in Fig. S2. The qualitative features match the ones of the statistical model of the previous section.

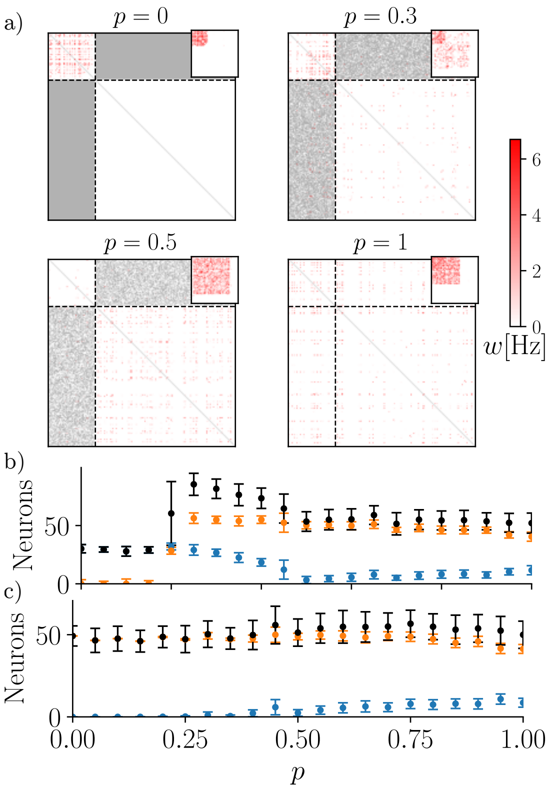

Figure 5: Quasi-equilibria in the biologically detailed engram drift model with two regions. (a) Example assembly weight matrices after long evolution for different inter-region connectivities, displayed as in Fig. 4b. (b,c) Number of assembly neurons (mean ± std) in Region 1 (blue) and 2 (orange) and total size (black), after long evolution. The assemblies are initially completely in (b) Region 1 or (c) Region 2. The dependence on inter-region connectivity is similar as in the random-and-deterministic drift model, cf. Fig. 4b,c.

Drifting engram in the mouse brain model

One advantage of our statistical model, especially the simplified, network realization-averaged version, is its relatively low computational cost. This allows to predict engram evolution in very large neural networks and even create a model of brain-wide engram drift. For this we combine brain-wide information about mouse neuron density and type [53], synaptic density [54], and connectivity [55,56], to construct a model of the mouse brain (Methods). Our model consists of 564 regions across two hemispheres. These regions are the mesoscale anatomical structures from the Allen Brain Atlas with sufficient available data. For each region we have estimated the number of excitatory neurons, N_s, and the probability of these neurons to be structurally connected to excitatory neurons in other regions, p_{sr} (Methods). The engram evolution in this network is simulated like before: we use Glauber dynamics with the network realization-averaged energy function given by Eq. 3. As initial engram macrostate we take a fear memory engram distribution across the mouse brain that was experimentally measured three days after learning [10]. In the simplified model, Eq. 3, all microstates making up this macrostate are equiprobable; we randomly pick one of these microstates accordingly. Reached quasi-equilibria for different parameter values of the energy function are shown in Fig. S3. Some parameter combinations lead to “forgetting”: the engram completely vanishes. Others lead to very large coding levels for some regions. We consider such results as pathological and the parameters as invalid (Methods). Forgetting occurs especially for high numbers of desired inputs from other engram neurons, $k = 10^4$ and $k = 10^5$, as not enough such inputs are available.

Figure 6 shows the engram dynamics and their quasi-equilibria for some prominent brain regions for two valid parameter sets. Fig. S4 analogously displays results for the remaining valid parameter sets; Figure 6a,d,e shows coding levels in further brain regions. Figure 6 illustrates that the predictions for individual regions can differ: in one case we observe a substantial increase in engram neurons in the basolateral amygdala and in another a rapid increase in the anterior cingulate area. However, there are also typical, consistent predictions across valid parameter sets: For example, the engram quickly leaves the hippocampal fields CA1-3 that form large parts of the hippocampus. This fits classical ideas of memory consolidation (Introduction). Further the overlap between the current and the initial engram quickly decays to zero: the engram quickly drifts away completely. This indicates that some compensation that adjusts the inputs and outputs of the assemblies to conserve behavior must take place, such as unsupervised compensation [31]. Finally, the simulations predict that the overall engram is conserved in the sense that its size does not change substantially during its drift. The prediction of strongly different coding levels in different regions at quasi-equilibrium differs markedly from those obtained with our purely-random drift model and shows the relevance of connectivity and engram energy.

Discussion

We have studied how the presence of a memory engram in different regions of the brain dynamically changes due to representational drift that is purely random or deterministic with a random component. The results suggest that the process of memory consolidation may rely on such representational drift. The developed approach is general in following sense: Its fundamental implications, especially the emergence of an entropic force, apply whenever representational drift has a random component. The drift mechanism is thereby irrelevant. Further the transfer of an engram between regions does not rely on a specific network architecture. Finally the approach requires only minimal, generally accepted assumptions on the engram’s nature, namely that it is formed by neurons or by neurons and their interconnections.

The presence of randomness in representational drift is generally very likely because random fluctuations are ubiquitous in biological systems [58] and, in particular, in the nervous system [59]. Furthermore, several recent modeling studies [31,32,60–62] have shown that random representational drift can occur as a consequence of experimentally observed highly noisy spiking activity and random remodeling of synapses, in conjunction with activity dependent plasticity or homeostatic plasticity or both.

We have shown how to statistically describe the

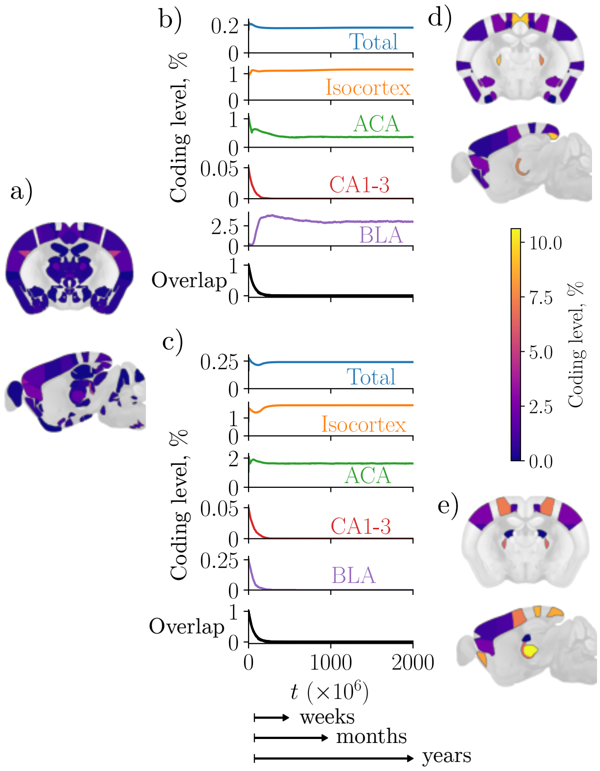

Figure 6: Fear memory engram dynamics and quasi-equilibria in the mouse brain. (a) Initial coding levels inferred from Ref. [10], coronal and sagittal sections, overlaid on Allen Mouse Brain Atlas [57] template. Coding levels for regions without engram neurons are not shown. (b) Total engram size, coding levels in selected macroscopic and mesoscopic regions and self-overlap dynamics for a valid parameter set ($\beta = 0.01,\ k = 250,\ g = 0.1$). Time t is Glauber steps; a suggested rough estimate of the timescale based on the decay of self-overlap [3, 31] is displayed below. (c) Same as in (b) but for a different valid parameter set ($\beta = 0.001,\ k = 500,\ g = 1$). (d) and (e): Same as in (a) but for the quasi-equilibria of the parameter sets used in (b) and (c), respectively. BLA - basolateral amygdala; ACA - anterior cingulate area; CA1, CA2, CA3 - hippocampal fields.

transformation of engrams over time. This accounts on the one hand for the entropic force that emerges due to random drift and the many possible engram states. On the other hand, there are forces that induce a deterministic drift. They may originate from plasticity mechanisms, which for example implement connectivity preferences; the (inhomogeneous) structural connectivity shapes them. To incorporate these forces, we introduced the concept of engram energy, which measures how beneficial an engram state is for the engram. The combination of random and deterministic drift is described by an engram free energy.

Our model allows to predict how engrams drift through the brain after learning. In general the entropic force drives the engram towards an equal distribution with identical coding levels across brain regions. This suggest that random representational drift may be employed by the brain to aid the creation of distributed memory representations. Furthermore, it allows engrams to go over potential barriers to more beneficial states. These functions add to previously suggested functions of the representational drift, which range from drift being a “bug” with no beneficial functions [63], to clearing space for memory storage [64, 65], regularization [66], sampling of solutions [67], and time stamping [68].

We derived our model from phenomenological considerations, for which the precise drift mechanisms are irrelevant. In particular, the origin of the random representational drift component is not important. Furthermore, detailed knowledge of plasticity mechanisms is not required. Rather, the allowed engram states and some preference characteristics are enough to construct the engram energy. A comparison with a biologically detailed model shows qualitative agreement between the predicted engram dynamics. Additional biological detail can be easily incorporated into our theory, like regional or temporal differences in neuronal excitability and in the magnitudes of random fluctuations.

Previous theoretical work on drifting assemblies studied drift within a single region [31–33]. Refs. [31, 32] model multiple drifting assemblies without overlaps, which completely tile the region. The drift is random and occurs due to noisy spiking activity, synaptic turnover and changes in spontaneous rate. Ref. [33] models a single assembly, which drifts due to a sequence of transient changes in excitability.

In our theory memory consolidation is a transformation of engrams from the initial non-equilibrium state towards stable (quasi-)equilibrium. This transformation is driven by both random and deterministic representational drift. Previous theoretical models of memory consolidation generally consider a few brain regions [17–26]. These models often use elaborate plasticity schemes to transfer memories from one region (usually a hippocampus model) to another (usually an isocortex model). In our model already a purely random drift can achieve such a transfer of memories from one region to another, if the size to the region to which engram is to be transferred is much larger. The deterministic drift (determined by the engram energy) modifies the transfer dynamics and resulting engram states, but preserves the general description of memory consolidation as drift.

Our study proposes an entirely novel, statistical physics-based class of models for representational dynamics. These models are both analytically and computationally well tractable. The latter allowed us to simulate the brain-wide engram transformation of fear memory in the mouse brain. This resulted in detailed experimental predictions. There is some uncertainty in the predictions resulting from uncertainty about the parameters of our energy function, which will be eliminated when the predictions at a single or a few time points are experimentally tested. Our models further enable future *in silico* experiments that predict the effects of perturbations to the engram or the neural network. For example, it is possible to systematically lesion different brain regions and examine resulting engram dynamics.

We do not explicitly take into account the contributions of inhibitory neurons to the engram. However, our approach could be extended by incorporating terms into the energy function that include inhibitory engram neurons.

We expect our approach to be applicable to various specific types of memories (e.g. episodic, semantic, motor). In particular, we expect that it will be possible to construct energy functions for specific types of engrams, for example sequentially structured ones, to predict their long term drift. Furthermore, it will be possible to add terms to the energy that account for the interaction between different engrams.

Methods

Purely-random drift model

To simulate engram evolution in our first model, Figs. 1-3, we replace in each simulation step a random engram neuron with a random non-engram neuron, in the sense that the picked engram neuron becomes a non-engram neuron and the picked non-engram neuron becomes an engram neuron. Each engram neuron has a n/N chance of being picked for replacement, and each non-engram has a $(N-n)/N$ chance of replacing it.

We can write the dynamics of the macrostate of the two-region model in the form of a Markov chain,

$$n_1(t+1) = n_1(t) + \Delta(t), \tag{4}$$

where $n_1(t)$ is the number of neurons in Region 1 at simulation step t. The increment $\Delta(t)$ can assume the values 0, 1, and -1: If an engram neuron is replaced by a non-engram neuron from the same region, we have $\Delta(t) = 0$, since the number of engram neurons in Region 1 does not change. If an engram neuron from Region 2 is replaced with a non-engram neuron from Region 1, the number of engram neurons in Region 1 increases by 1 and we have $\Delta(t) = 1$. If an engram neuron from Region 1 is replaced with a non-engram neuron from Region 2, $\Delta(t) = -1$. The increments have the following probabilities of occurrence:

$$P\big(\Delta(t)\!=\!1\big) = \frac{\big(n - n_1(t)\big)\big(N_1 - n_1(t)\big)}{n(N-n)}, \tag{5}$$

$$P\big(\Delta(t)\!=\!-1\big) = \frac{n_1(t)\big(N - N_1 - \big(n - n_1(t)\big)\big)}{n(N-n)}, \tag{6}$$

$$P\big(\Delta(t)\!=\!0\big) = 1 - P\big(\Delta(t)\!=\!1\big) - P\big(\Delta(t)\!=\!-1\big). \tag{7}$$

This is because $n(N-n)$ is the total number of ways that we can combine (and thus replace) one of the n engram neurons with one of the $N-n$ non-engram neurons; further, for example $(n-n_1(t))(N_1-n_1(t))$ is the number of ways in which we can combine (replace) one of the $n - n_1(t)$ engram neurons of Region 2 with one of the $N_1 - n_1(t)$ non-engram neurons of Region 1. We note that these considerations straightforwardly generalize to multi-region models. In fact, if we focus on the macrostate of region 1 as above in such a model, the same formulas hold as above, because we can gather all other regions into one region.

The Markov chain specified by Eqs. 4 to 7 is irreducible and aperiodic and thus has a unique equilibrium distribution [36]. We can rewrite the macrostate equilibrium probability $p(\boldsymbol{n}) = \Omega(\boldsymbol{n})/\Omega(n)$ as function of the single variable n_1, $p(n_1) = \binom{N_1}{n_1}\binom{N-N_1}{n-n_1}/\binom{N}{n}$, using the fact that $n_2 = n - n_1$. We obtain $\langle n_1\rangle = \frac{nN_1}{N}$ and $\langle n_1^2\rangle = \langle n_1\rangle\left(1 + \frac{(n-1)(N_1-1)}{N-1}\right)$ by applying the definition of expectation directly and using the Chu–Vandermonde identity. For $n_1, N_1 \gg 1$, $\langle n_1^2\rangle = \langle n_1\rangle(1 + \langle n_1\rangle)$, such that the coefficient of variation is $\langle n_1\rangle^{-1/2}$.

To derive the analytical expression for the average trajectory Eq. 1, we calculate the expected number of engram neurons in Region 1 after one simulation step; in other words, we compute the conditional expectation of $n_1(t)$ given $n_1(t-1)$,

$$\begin{aligned}\langle n_1(t)|n_1(t-1)\rangle &= \langle n_1(t-1)|n_1(t-1)\rangle \\ &\quad + \langle\Delta(t-1)|n_1(t-1)\rangle \\ &= (1-B)n_1(t-1) + A,\end{aligned} \tag{8}$$

where $B = \frac{N}{n(N-n)}$ and $A = \frac{N_1}{N-n}$, and we have used Equations (5) to (7) to directly average the second term. We obtain for $u > t$

$$\begin{aligned}\langle n_1(u)|n_1(t)\rangle &= \sum_{n_1(u-1)=0}^{n} \langle n_1(u)|n_1(u-1)\rangle \\ &\quad \times P(n_1(u-1)|n_1(t)) \\ &= (1-B)\langle n_1(u-1)|n_1(t)\rangle + A,\end{aligned} \tag{9}$$

where we used the Markov property, which implies that $n_1(u)$ depends on $n_1(t)$ only via $n_1(u-1)$, and Eq. 8; the detailed derivation is given in Supplemental section S1. Starting with some initial state $n_1(0)$ and applying t times the recurrence relation given by Eq. 9 yields Eq. 1.

Random-and-deterministic drift model

At each simulation step we construct a candidate microstate, $\boldsymbol{m_c}$, by modifying the current microstate $\boldsymbol{m}$ as follows: We pick a random neuron. If this neuron is an engram neuron, we remove it from the engram; if the picked neuron is a non-engram neuron, we add it to the engram. The candidate microstate is accepted with probability $1/(1 + e^{\beta(H(\boldsymbol{m_c})-H(\boldsymbol{m})})$, otherwise, with probability $1/(1 + e^{\beta(H(\boldsymbol{m})-H(\boldsymbol{m_c})})$, the system stays in the current microstate. An analogous procedure is applied for the simplified model; in this model the energies of different microstates are equal if they belong to the same macrostate.

To obtain Eq. 3 we average Eq. 2 over the realizations of the A_{ij}

$$\begin{aligned}\overline{H}(\boldsymbol{m}) = \sum_s \sum_{i\in s} &\overline{\left(\sum_r \sum_{j\in r} A_{ij}m_j - k\right)^2} m_i \\ &+ g \sum_{s,r} \sum_{\substack{i\in s\\ j\in r}} \overline{(A_{ij} - A_{ji})^2} m_i m_j,\end{aligned} \tag{10}$$

where indices i and j run over neurons and s and r over regions, $k \in q$ signifies that neuron k is in region q. To convert from microstates to macrostates we use $n_s = \sum_{i\in s} m_i$. We allow autapses, i.e. it is possible that $A_{ii} = 1$. Expanding terms of Eq. 10, averaging using the fact that A_{ij} are independent Bernoulli random variables, and rearranging leads to Eq. 3; the detailed derivation is given in Supplemental section S2.

Biologically detailed assembly model

The network consists of N linear Poisson neurons that spike with instantaneous rates $f_i(t)$, for $i = 1, 2, ..., N$. The spike rate of a linear Poisson neuron is incremented instantaneously at the arrival of an input spike by an amount proportional to the synaptic weight. In the absence of input spikes the rate decays exponentially with time constant τ to the background value f_{sp}. The dynamics are thus given by

$$\tau\frac{d}{dt}f_i(t) = f_{\mathrm{sp}} - f_i(t) + \tau\sum_{j=1}^{N} w_{ij}(t^-)\sum_{t_j}\delta(t - t_j),$$

where $w_{ij}(t^-)$ is the strength of the synapse from neuron j to neuron i just before time t (where it may change in jump-like manner, see below); t_j are the spike times of neuron j. The synaptic weights are positive, with maximum possible weight w_{max}: $0 < w_{ij} < w_{\mathrm{max}}$.

The synaptic strengths are modified at each pre- and each postsynaptic spike by STDP that depend also on the rates of pre- and post synaptic neurons

$$\Delta w_{ij}^{\mathrm{STDP}} = (A_p e^{-|t_i - t_j|/\tau_p} + A_d e^{-|t_i - t_j|/\tau_d}) \times (f_i(t_j)\Theta(t_j - t_i) + f_j(t_i)\Theta(t_i - t_j)) \,,$$

where A_p, A_d, τ_p, τ_d and Θ is the Heaviside step function. In addition, each synaptic weight decays at a constant rate. Finally if a neuron j spikes, its output synaptic weights are instantaneously decremented by an amount proportional to its frequency,

$$\frac{d}{dt} w_{ij}(t) = -d_f \sum_{t_j} f_j(t^-)\delta(t - t_j) - d_{\mathrm{sp}} f_{\mathrm{sp}},$$

where d_f and d_{sp} are constants, and the sum is over all spikes times t_j of neuron j.

Mouse brain model

The regions in our model are "summary structures" - non-overlapping mesoscale brain regions defined by the Allen Mouse Brain Atlas [57]. We use subscripts r and s, ranging from 1 to R, to index these mesoscopic regions. The Allen Mouse Brain Atlas also defines 12 macroscopic brain regions [57]: isocortex, olfactory areas, hippocampal formation, cortical subplate, striatum, pallidum, thalamus, hypothalamus, midbrain, pons, medulla, and cerebellum. We use this macroscopic structuring to obtain reference regions for the estimation of unknown data (synaptic densities) and to assign long range connectivity properties (see below).

We obtain each mesoscopic region's volume V_r and number of excitatory N_r and inhibitory N_r^I neurons from Ref. [53]. The structural connectivity in the model is based on the mesoscale connectome obtained by Ref. [56] and refined by Ref. [55]: These studies measured connection strengths by anterograde tracing such that both excitatory and inhibitory projections are accounted for. We obtain the volume-normalized connection strength between two regions, D_{rs}^{norm}, from ref. [55] and convert it to the connection strength $D_{rs} = D_{rs}^{\mathrm{norm}} V_r V_s$. We are interested in the excitatory connection strengths, D_{rs}^E, only. We assume that all long-range (inter-region) connections are excitatory, $D_{rs}^E = D_{rs}$ if $r \neq s$, with the exception of those that are part of the subcortical brain regions and cerebellum, which have prominent inhibitory projections [69]. We therefore assume that the number of excitatory inter-region projections is proportional to the number of excitatory neurons $D_{rs}^E = \frac{N_s}{N_s + N_s^I} D_{rs}$, if the region s belongs to the subcortical regions, i.e. to the macroscopic regions cortical subplate, striatum, pallidum, thalamus, hypothalamus, midbrain, pons, medulla, or if the region belongs to the cerebellum excluding the cerebellar cortex. The cerebellar cortex has only inhibitory outputs [70], and we therefore set $D_{rs}^E = 0$ if s is a region in the cerebellar cortex. For the intra-region connections, we estimate the connection strength due to excitatory projections D_{rr}^E by assuming that the fraction of excitatory intra-region connections equals the fraction of excitatory neurons, which yields $D_{rr}^E = \frac{N_r}{N_r + N_r^I} D_{rr}$. We assume that the number of excitatory synapses from region s to region r, S_{rs}, is proportional to the strength of connections as measured by the anterograde tracing $S_{rs} = \sigma_r D_{rs}^E$. We estimate the region-specific proportionality constant σ_r, in a method similar to the one used in Ref. [71]: Summing the number of excitatory synapses from all input regions s to a region r, must yield the total number of excitatory synapses in that region,

$$\sum_{s=1}^{R} S_{rs} = \sigma_r \sum_{s=1}^{R} D_{rs}^E = \rho_r^{\mathrm{syn}} V_r, \tag{11}$$

where ρ_r^{syn}, is the density of excitatory synapses in region r, which we obtain for some regions from Ref. [54]. For regions for which it is not available, we estimate ρ_r^{syn} by taking the average over regions in the same macroscopic region with known densities. Solving Eq. 11 for σ_r and inserting the result into the defining equation for S_{rs} yields

$$S_{rs} = \frac{\rho_r^{\mathrm{syn}} V_r}{\sum_{s=1}^{R} D_{rs}^E} D_{rs}^E. \tag{12}$$

Assuming that all $N_r + N_r^I$ neurons in a region r are statistically equivalent, the probability that a particular synaptic input from a region s is present is given by the number of synaptic inputs from region s, S_{rs}, divided by the number $(N_r + N_r^I)N_s$ of in principle possible connections. For our simulations we consider only the excitatory neurons. The probability p_{rs}^{syn} that an excitatory neuron in region

r receives a synapse from an excitatory neuron in region s, is then also given by

$$P_{rs}^{\mathrm{syn}} = \frac{S_{rs}}{(N_r + N_r^I)N_s}, \tag{13}$$

for region r with excitatory neurons ($N_r > 0$), otherwise $p_{rs}^{\mathrm{syn}} = 0$. Here we accounted for the fact that we only considered output connections of excitatory neurons, but did not distinguish whether they end at excitatory or inhibitory neurons. The total number of possible connections that we divide by is therefore the product of the number of excitatory neurons in the sending and the total number of neurons in the receiving area.

The probability p_{rs}^{syn} calculated above relates to the already existing synapses. We assume that the probability of a structural (potential) connection is proportional to it, $p_{rs} = \lambda p_{rs}^{\mathrm{syn}}$. The proportionality constant λ depends on the filling fraction and the number of synapses formed between two neurons. The filling fraction estimates the ratio of existing synapses to all potential ones without major axonal or dendritic remodeling. It is estimated to be 0.26 in some areas of mouse isocortex, and ranges between 0.12 and 0.34 for different brain areas and model organisms [40]. There is typically more than one synapse between connected neurons. For example, for the rat isocortex on average 5.5 [72] and 4.7 [73] synaptic contacts between neurons were reported. Multiple synapses were also reported for the pairs of neurons form the mouse isocortex [74]. Taking into account multiple synapses between pairs of neurons on average compensates the effect of the filling fraction, and we therefore set $\lambda = 1$.

We interpret p_{rs} as the probability that a synapse may in principle exist between two neurons, i.e. that these neurons are structurally connected. p_{rs} and N_r specify our brain model; it covers 282 brain regions (per hemisphere).

We obtain the initial state of the engram in region s as

$$n_s(0) = n_s^{\mathrm{RE}}(1 - n_s^{\mathrm{HC}}/N_s), \tag{14}$$

where n_s^{HC} and n_s^{RE} are the average numbers of excitatory neurons that were labeled by cFos in the home cage and during fear memory recall three days after fear conditioning, respectively [10]. The rationale behind this formula is as follows: The probability that a neuron is active in the home cage is n_s^{HC}/N_s and the probability that a neuron is active during recall is n_s^{RE}/N_s. We expect that there is a background of spuriously active neurons that we need to subtract from n_s^{RE} to get the true engram size. Due to the lack of information about these neurons and the expectation that spurious activity may be similar in both the recall and the homecage measurement, we use as a rough estimate that the number of spuriously active neurons equals the number of neurons that lie in the overlap of the sets of neurons that are active during recall and in the homecage. Assuming (somewhat contradictorily) that the probability that a neuron is active during recall and in the home cage are independent because the two events have little in common, the probability of a neuron to be spuriously active during a measurement is then $n_s^{\mathrm{HC}}/N_s \cdot n_s^{\mathrm{RE}}/N_s$. This yields as the probability of a neuron to be a true recall engram neuron $n_s(0)/N_s = n_s^{\mathrm{RE}}/N_s - n_s^{\mathrm{RE}}/N_s \cdot n_s^{\mathrm{HC}}/N_s$ and thus for the expected number of recall engram neurons Eq. 14. We assume that both excitatory and inhibitory neurons are labeled by cFos with the same probability, and scale the experimental data by the fraction of excitatory neurons in the region to obtain only the excitatory engram.

We repeated simulations five times for each parameter set (β, k, g). We consider a set of parameters as pathological, if it led to a coding level of more than 20% in a region with more than 1000 excitatory neurons, at any simulation step, for the majority of realizations.

Acknowledgements

We thank Simon Altrogge for enlightening discussions and insightful ideas, the German Federal Ministry of Education and Research (BMBF) for support via the Bernstein Network (Bernstein Award 2014, 01GQ1710), and the German Research Society (DFG) for support via a Research Grant (430156448).

References

[1] Leon G Reijmers, Brian L Perkins, Naoki Matsuo, and Mark Mayford. Localization of a sta-

ble neural correlate of associative memory. *Science*, 317(5842):1230–1233, 2007.

[2] Xu Liu, Steve Ramirez, Petti T Pang, Corey B Puryear, Arvind Govindarajan, Karl Deisseroth, and Susumu Tonegawa. Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, 484(7394):381–385, 2012.

[3] Laura A DeNardo, Cindy D Liu, William E Allen, Eliza L Adams, Drew Friedmann, Lisa Fu, Casey J Guenthner, Marc Tessier-Lavigne, and Liqun Luo. Temporal evolution of cortical ensembles promoting remote memory retrieval. *Nature neuroscience*, 22(3):460–469, 2019.

[4] Chaery Lee, Byung Hun Lee, Hyunsu Jung, Chiwoo Lee, Yongmin Sung, Hyopil Kim, Jooyoung Kim, Jae Youn Shim, Ji-il Kim, Dong Il Choi, et al. Hippocampal engram networks for fear memory recruit new synapses and modify pre-existing synapses in vivo. *Current Biology*, 33(3):507–516, 2023.

[5] Marios Abatis, Rodrigo Perin, Ruifang Niu, Erwin van den Burg, Chloe Hegoburu, Ryang Kim, Michiko Okamura, Haruhiko Bito, Henry Markram, and Ron Stoop. Fear learning induces synaptic potentiation between engram neurons in the rat lateral amygdala. *Nature Neuroscience*, pages 1–9, 2024.

[6] Yosif Zaki and Denise J Cai. Memory engram stability and flexibility. *Neuropsychopharmacology*, pages 1–9, 2024.

[7] Nitzan Geva, Daniel Deitch, Alon Rubin, and Yaniv Ziv. Time and experience differentially affect distinct aspects of hippocampal representational drift. *Neuron*, 111(15):2357–2366, 2023.

[8] Carl E Schoonover, Sarah N Ohashi, Richard Axel, and Andrew JP Fink. Representational drift in primary olfactory cortex. *Nature*, 594(7864):541–546, 2021.

[9] Laura N Driscoll, Noah L Pettit, Matthias Minderer, Selmaan N Chettih, and Christopher D Harvey. Dynamic reorganization of neuronal activity patterns in parietal cortex. *Cell*, 170(5):986–999, 2017.

[10] Dheeraj S. Roy, Young-Gyun Park, Minyoung E. Kim, Ying Zhang, Sachie K. Ogawa, Nicholas DiNapoli, Xinyi Gu, Jae H. Cho, Heejin Choi, Lee Kamentsky, Jared Martin, Olivia Mosto, Tomomi Aida, Kwanghun Chung, and Susumu Tonegawa. Brain-wide mapping reveals that engrams for a single memory are distributed across multiple brain regions. *Nature Communications*, 13(1):1799, Apr 2022.

[11] Paul W Frankland, Bruno Bontempi, Lynn E Talton, Leszek Kaczmarek, and Alcino J Silva. The involvement of the anterior cingulate cortex in remote contextual fear memory. *Science*, 304(5672):881–883, 2004.

[12] Yadin Dudai, Avi Karni, and Jan Born. The consolidation and transformation of memory. *Neuron*, 88(1):20–32, 2015.

[13] Larry R. Squire, Lisa Genzel, John T. Wixted, and Richard G. Morris. Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, 7(8), 2015.

[14] Susumu Tonegawa, Mark D. Morrissey, and Takashi Kitamura. The role of engram cells in the systems consolidation of memory. *Nature Reviews Neuroscience*, 19(8):485–498, Aug 2018.

[15] G. Buzsáki. Two-stage model of memory trace formation: A role for “noisy” brain states. *Neuroscience*, 31:551–570, 1989.

[16] Jens G Klinzing, Niels Niethard, and Jan Born. Mechanisms of systems memory consolidation during sleep. *Nat. Neurosci.*, 22(10):1598–1610, October 2019.

[17] P Alvarez and L R Squire. Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of the National Academy of Sciences*, 91(15):7041–7045, 1994.

[18] Szabolcs Káli and Peter Dayan. Hippocampally-dependent consolidation in a hierarchical model of neocortex. In *Proceedings of the 13th International Conference on Neural Information Processing Systems*,

NIPS'00, page 22–23, Cambridge, MA, USA, 2000. MIT Press.

[19] Gayle M. Wittenberg, Megan R. Sullivan, and Joe Z. Tsien. Synaptic reentry reinforcement based network model for long-term memory consolidation. *Hippocampus*, 12(5):637–647, 2002.

[20] Martin Pyka and Sen Cheng. Pattern association and consolidation emerges from connectivity properties between cortex and hippocampus. *PLOS ONE*, 9(1):1–14, 01 2014.

[21] Florian Fiebig and Anders Lansner. Memory consolidation from seconds to weeks: a three-stage neural network model with autonomous reinstatement dynamics. *Frontiers in Computational Neuroscience*, 8, 2014.

[22] Peter Helfer and Thomas R. Shultz. A computational model of systems memory consolidation and reconsolidation. *Hippocampus*, 30(7):659–677, 2020.

[23] Michiel W. H. Remme, Urs Bergmann, Denis Alevi, Susanne Schreiber, Henning Sprekeler, and Richard Kempter. Hebbian plasticity in parallel synaptic pathways: A circuit mechanism for systems memory consolidation. *PLOS Computational Biology*, 17(12):1–37, 12 2021.

[24] Dhairyya Singh, Kenneth A. Norman, and Anna C. Schapiro. A model of autonomous interactions between hippocampus and neocortex driving sleep-dependent memory consolidation. *Proceedings of the National Academy of Sciences*, 119(44):e2123432119, 2022.

[25] Douglas Feitosa Tomé, Sadra Sadeh, and Claudia Clopath. Coordinated hippocampal-thalamic-cortical communication crucial for engram dynamics underneath systems consolidation. *Nat. Commun.*, 13(1):840, February 2022.

[26] Brandon J Bhasin, Jennifer L Raymond, and Mark S Goldman. Synaptic weight dynamics underlying memory consolidation: Implications for learning rules, circuit organization, and circuit function. *Proc. Natl. Acad. Sci. U. S. A.*, 121(41):e2406010121, October 2024.

[27] Morris Moscovitch and Asaf Gilboa. Has the concept of systems consolidation outlived its usefulness? identification and evaluation of premises underlying systems consolidation. *Faculty Reviews*, 11, 2022.

[28] Zhenrui Liao and Attila Losonczy. Learning, fast and slow: Single-and many-shot learning in the hippocampus. *Annual Review of Neuroscience*, 47, 2024.

[29] Ingo Müller. *A history of thermodynamics: the doctrine of energy and entropy*. Springer Science & Business Media, 2007.

[30] Mu-ming Poo, Michele Pignatelli, Tomás J Ryan, Susumu Tonegawa, Tobias Bonhoeffer, Kelsey C Martin, Andrii Rudenko, Li-Huei Tsai, Richard W Tsien, Gord Fishell, et al. What is memory? the present state of the engram. *BMC biology*, 14:1–18, 2016.

[31] Yaroslav Felipe Kalle Kossio, Sven Goedeke, Christian Klos, and Raoul-Martin Memmesheimer. Drifting assemblies for persistent memory: Neuron transitions and unsupervised compensation. *Proceedings of the National Academy of Sciences*, 118(46):e2023832118, 2021.

[32] Paul Manz and Raoul-Martin Memmesheimer. Purely stdp-based assembly dynamics: Stability, learning, overlaps, drift and aging. *PLOS Computational Biology*, 19(4):1–24, 04 2023.

[33] Geoffroy Delamare, Yosif Zaki, Denise J Cai, and Claudia Clopath. Drift of neural ensembles driven by slow fluctuations of intrinsic excitability. *Elife*, 12, May 2024.

[34] László Lovász. Random walks on graphs. *Combinatorics, Paul erdos is eighty*, 2(1-46):4, 1993.

[35] Derek Allan Holton and John Sheehan. *The Petersen Graph*, volume 7. Cambridge University Press, 1993.

[36] James P Sethna. *Statistical mechanics: entropy, order parameters, and complexity*, volume 14. Oxford University Press, USA, 2021.

[37] James L. McGaugh. Memory–a century of consolidation. *Science*, 287(5451):248–251, 2000.

[38] Marco Baldovin, Lorenzo Caprini, and Angelo Vulpiani. Irreversibility and typicality: A simple analytical result for the ehrenfest model. *Physica A: Statistical Mechanics and its Applications*, 524:422–429, 2019.

[39] Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nature reviews neuroscience*, 13(5):336–349, 2012.

[40] Armen Stepanyants, Patrick R Hof, and Dmitri B Chklovskii. Geometry and structural plasticity of synaptic connectivity. *Neuron*, 34(2):275–288, 2002.

[41] Jeffrey C Magee and Christine Grienberger. Synaptic plasticity forms and functions. *Annual review of neuroscience*, 43(1):95–117, 2020.

[42] Noam E Ziv and Naama Brenner. Synaptic tenacity or lack thereof: spontaneous remodeling of synapses. *Trends in neurosciences*, 41(2):89–99, 2018.

[43] Anthony Holtmaat and Pico Caroni. Functional and structural underpinnings of neuronal assembly formation in learning. *Nature neuroscience*, 19(12):1553–1562, 2016.

[44] György Buzsáki. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, 68(3):362–385, 2010.

[45] Alwyn Scott. *Neuroscience: A mathematical primer*. Springer, 2002.

[46] Neta Ravid Tannenbaum and Yoram Burak. Shaping neural circuits by high order synaptic interactions. *PLOS Computational Biology*, 12(8):e1005056, 2016.

[47] Sen Song, Per Jesper Sjöström, Markus Reigl, Sacha Nelson, and Dmitri B Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.*, 3(3):e68, March 2005.

[48] Rajiv K Mishra, Sooyun Kim, Segundo J Guzman, and Peter Jonas. Symmetric spike timing-dependent plasticity at CA3-CA3 synapses optimizes storage and recall in autoassociative networks. *Nat. Commun.*, 7(1):11552, May 2016.

[49] Ila R Fiete, Walter Senn, Claude ZH Wang, and Richard HR Hahnloser. Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron*, 65(4):563–576, 2010.

[50] Roy J. Glauber. Time-dependent statistics of the ising model. *Journal of Mathematical Physics*, 4(2):294–307, 02 1963.

[51] Robert Froemke, Dominique Debanne, and Guo-Qiang Bi. Temporal modulation of spike-timing-dependent plasticity. *Frontiers in Synaptic Neuroscience*, 2, 2010.

[52] Gina G. Turrigiano. The self-tuning neuron: Synaptic scaling of excitatory synapses. *Cell*, 135(3):422–435, 2008.

[53] Dimitri Rodarie, Csaba Verasztó, Yann Roussel, Michael Reimann, Daniel Keller, Srikanth Ramaswamy, Henry Markram, and Marc-Oliver Gewaltig. A method to estimate the cellular composition of the mouse brain from heterogeneous datasets. *PLOS Computational Biology*, 18(12):e1010739, 2022.

[54] Andrea Santuy, Laura Tomás-Roca, José-Rodrigo Rodríguez, Juncal González-Soriano, Fei Zhu, Zhen Qiu, Seth GN Grant, Javier DeFelipe, and Angel Merchan-Perez. Estimation of the number of synapses in the hippocampus and brain-wide by volume electron microscopy and genetic labeling. *Scientific Reports*, 10(1):14014, 2020.

[55] Joseph E Knox, Kameron Decker Harris, Nile Graddis, Jennifer D Whitesell, Hongkui Zeng, Julie A Harris, Eric Shea-Brown, and Stefan Mihalas. High-resolution data-driven model of the mouse connectome. *Network Neuroscience*, 3(1):217–236, 2018.

[56] Seung Wook Oh, Julie A. Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M. Henry, Marty T. Mortrud, Benjamin Ouellette, Thuc Nghi Nguyen, Staci A. Sorensen, Clifford R. Slaughterbeck, Wayne Wakeman, Yang Li, David Feng, Anh Ho, Eric Nicholas, Karla E. Hirokawa, Phillip Bohn,

Kevin M. Joines, Hanchuan Peng, Michael J. Hawrylycz, John W. Phillips, John G. Hohmann, Paul Wohnoutka, Charles R. Gerfen, Christof Koch, Amy Bernard, Chinh Dang, Allan R. Jones, and Hongkui Zeng. A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214, Apr 2014.

[57] Quanxin Wang, Song-Lin Ding, Yang Li, Josh Royall, David Feng, Phil Lesnar, Nile Graddis, Maitham Naeemi, Benjamin Facer, Anh Ho, et al. The allen mouse brain common coordinate framework: a 3d reference atlas. *Cell*, 181(4):936–953, 2020.

[58] Lev S Tsimring. Noise in biology. *Reports on Progress in Physics*, 77(2):026601, jan 2014.

[59] A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nature reviews neuroscience*, 9(4):292–303, 2008.

[60] Michael E Rule and Timothy O'Leary. Self-healing codes: How stable neural populations can track continually reconfiguring neural representations. *Proc. Natl. Acad. Sci. U. S. A.*, 119(7):e2106692119, February 2022.

[61] Joel Bauer, Uwe Lewin, Elizabeth Herbert, Julijana Gjorgjieva, Carl E Schoonover, Andrew J P Fink, Tobias Rose, Tobias Bonhoeffer, and Mark Hübener. Sensory experience steers representational drift in mouse visual cortex. *Nat. Commun.*, 15(1):9153, October 2024.

[62] Jens-Bastian Eppler, Thomas Lai, Dominik Aschauer, Simon Rumpel, and Matthias Kaschube. Representational drift reflects ongoing balancing of stochastic changes by hebbian learning. January 2025.

[63] Paul Masset, Shanshan Qin, and Jacob A Zavatone-Veth. Drifting neuronal representations: Bug or feature? *Biological cybernetics*, 116(3):253–266, 2022.

[64] William Mau, Michael E Hasselmo, and Denise J Cai. The brain in motion: How ensemble fluidity drives memory-updating and flexibility. *Elife*, 9:e63550, 2020.

[65] Laura N Driscoll, Lea Duncker, and Christopher D Harvey. Representational drift: Emerging theories for continual learning and experimental future directions. *Current Opinion in Neurobiology*, 76:102609, 2022.

[66] Michael E Rule, Timothy O'Leary, and Christopher D Harvey. Causes and consequences of representational drift. *Current Opinion in Neurobiology*, 58:141–147, 2019. Computational Neuroscience.

[67] David Kappel, Stefan Habenschuss, Robert Legenstein, and Wolfgang Maass. Network plasticity as bayesian inference. *PLoS computational biology*, 11(11):e1004485, 2015.

[68] Alon Rubin, Nitzan Geva, Liron Sheintuch, and Yaniv Ziv. Hippocampal ensemble dynamics timestamp events in long-term memory. *elife*, 4:e12247, 2015.

[69] Jocelyn Urrutia-Piñones, Camila Morales-Moraga, Nicole Sanguinetti-González, Angelica P Escobar, and Chiayu Q Chiu. Long-range gabaergic projections of cortical origin in brain function. *Frontiers in Systems Neuroscience*, 16:841869, 2022.

[70] Hannsjörg Schröder, Natasha Moser, and Stefan Huggenberger. *Neuroanatomy of the mouse: An introduction.* Springer Nature, 2020.

[71] Michael W Reimann, Michael Gevaert, Ying Shi, Huanxiang Lu, Henry Markram, and Eilif Muller. A null model of the mouse whole-neocortex micro-connectome. *Nature communications*, 10(1):3903, 2019.

[72] Henry Markram, Joachim Lübke, Michael Frotscher, Arnd Roth, and Bert Sakmann. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *The Journal of physiology*, 500(2):409–440, 1997.

[73] Eyal Gal, Michael London, Amir Globerson, Srikanth Ramaswamy, Michael W Reimann, Eilif Muller, Henry Markram, and Idan Segev. Rich cell-type-specific network topology in neocortical microcircuitry. *Nature neuroscience*, 20(7):1004–1013, 2017.

[74] Narayanan Kasthuri, Kenneth Jeffrey Hayworth, Daniel Raimund Berger, Richard Lee Schalek, José Angel Conchello, Seymour Knowles-Barley, Dongil Lee, Amelio Vázquez-Reina, Verena Kaynig, Thouis Raymond Jones, et al. Saturated reconstruction of a volume of neocortex. *Cell*, 162(3):648–661, 2015.