Drift of neural ensembles driven by slow fluctuations of intrinsic excitability

Geoffroy Delamare[1], Yosif Zaki[2], Denise J Cai[2], Claudia Clopath[1]

[1] Bioengineering Department, Imperial College London, London SW7 2AZ, UK
[2] Icahn School of Medicine at Mount Sinai, Department of Neuroscience, New York, New York, 10029, United States

Abstract

Representational drift refers to the dynamic nature of neural representations in the brain despite the behavior being seemingly stable. Although drift has been observed in many different brain regions, the mechanisms underlying it are not known. Since intrinsic neural excitability is suggested to play a key role in regulating memory allocation, fluctuations of excitability could bias the reactivation of previously stored memory ensembles and therefore act as a motor for drift. Here, we propose a rate-based plastic recurrent neural network with slow fluctuations of intrinsic excitability. We first show that subsequent reactivations of a neural ensemble can lead to drift of this ensemble. The model predicts that drift is induced by co-activation of previously active neurons along with neurons with high excitability which leads to remodeling of the recurrent weights. Consistent with previous experimental works, the drifting ensemble is informative about its temporal history. Crucially, we show that the gradual nature of the drift is necessary for decoding temporal information from the activity of the ensemble. Finally, we show that the memory is preserved and can be decoded by an output neuron having plastic synapses with the main region.

Introduction

In various brain regions, the neural code tends to be dynamic although behavioral outputs remain stable. Representational drift refers to the dynamic nature of internal representations as they have been observed in sensory cortical areas [1–3] or the hippocampus [4, 5] despite stable behavior. It has even been suggested that pyramidal neurons from the CA1 and CA3 regions form dynamic rather than static memory engrams [5, 6], namely that the set of neurons encoding specific memories varies across days. In the amygdala, retraining of a fear memory task induces a turnover of the memory engram [7]. Additionally, plasticity mechanisms have been proposed to compensate for drift and to provide a stable read-out of the neural code [8], suggesting that information is maintained. Altogether, this line of evidence suggest that drift might be a general mechanism with dynamical representations observed in various brain regions.

However, the mechanisms underlying the emergence of drift and its relevance for the neural computation are not known. Drift is often thought to arise from variability of internal states [2], neurogenesis [1, 9] or synaptic turnover [10] combined with noise [11, 12]. On the other hand, excitability might also play a role in memory allocation [13–16], so that neurons having high excitability are preferentially allocated to memory ensembles [16–18]. Moreover, excitability is known to fluctuate over timescales from hours to days, in the amygdala [18], the hippocampus [17, 19] or the cortex [20, 21]. Subsequent reactivations of a neural ensemble at different time points could therefore be biased by excitability [22], which varies at similar timescales than drift [23]. Altogether, this evidence suggest that fluctuations of excitability could act as a cellular mechanism for drift [14].

In this short communication, we aimed at proposing how excitability could indeed induce a drift of neural ensembles at the mechanistic level. We simulated a recurrent neural network [24] equipped with intrinsic neural excitability and Hebbian learning. As a proof of principle, we first show that slow

fluctuations of excitability can induce neural ensembles to drift in the network. We then explore the functional implications of such drift. Consistent with previous works [23, 25–27], we show that neural activity of the drifting ensemble is informative about the temporal structure of the memory. This suggest that fluctuations of excitability can be useful for time-stamping memories (*i.e.* for making the neural ensemble informative about the time at which it was form). Finally, we confirmed that the content of the memory itself can be steadily maintained using a read-out neuron and local plasticity rule, consistently with previous computational works [8]. The goal of this study is to show one possible mechanistic implementation of how excitability can drive drift.

Results

Many studies have shown that memories are encoded in sparse neural ensembles that are activated during learning and many of the same cells are reactivated during recall, underlying a stable neural representation [14, 28, 29]. After learning, subsequent reactivations of the ensemble can happen spontaneously during replay, retraining or during a memory recall task (*e.g.* following presentation of a cue [28, 30]). Here, we directly tested the hypothesis that slow fluctuations of excitability can change the structure of a newly-formed neural ensemble, through subsequent reactivations of this ensemble. To that end, we designed a rate-based, recurrent neural network, equipped with intrinsic neural excitability (Methods). We considered that the recurrent weights are all-to-all and plastic, following a Hebbian rule (Methods). The network was then stimulated following a 4-day protocol: the first day corresponds to the initial encoding of a memory and the other days correspond to spontaneous or cue-induced reactivations of the neural ensemble (Methods). Finally, we considered that excitability of each neuron can vary on a day long timescale: each day, a different subset of neurons has increased excitability (Fig. 1a, Methods).

Fluctuations of intrinsic excitability induce drifting of neural ensembles

While stimulated the naive network on the first day, we observed the formation of a neural ensemble: some neurons gradually increase their firing rate (Fig. 1b and c, neurons 10 to 20, time steps 1000 to 3000) during the stimulation. We observed that these neurons are highly recurrently connected (Fig. 1d, leftmost matrix) suggesting that they form an assembly. This assembly is composed of neurons that have a high excitability (Fig. 1a, neurons 10 to 20 have increase excitability) at the time of the stimulation. We then show that further stimulations of the network induce a remodeling of the weights. During the second stimulation for instance (Fig. 1b and c, time steps 4000 to 6000), neurons from the previous assembly (10 to 20) are reactivated along with neurons having high excitability at the time of the second stimulation (Fig. 1a, neurons 20 to 30). Moreover, across several days, recurrent weights from previous assemblies tend to decrease while others increase (Fig. 1d). Indeed, neurons from the original assembly (Fig. 1c, red traces) tend to be replaced by other neurons, either from the latest assembly or from the pool of neurons having high excitability. This is translated at the synaptic level, where weights from previous assemblies tend to decay and be replaced by new ones. Overall, each new stimulation updates the ensemble according to the current distribution of excitability, inducing a drift towards neurons with high excitability. Finally, in our model, the drift rate does not depend on the size of the original ensemble (Fig. S2, Methods).

Activity of the drifting ensemble is informative about the temporal structure of the past experience

After showing that fluctuations of excitability can induce a drift among neural ensembles, we tested whether the drifting ensemble could contain temporal information about its past experiences, as suggested in previous works [25]. Inspired by these works, we asked whether it was possible to decode relevant temporal information from the patterns of activity of the neural ensemble. We first observed that the correlation between patterns of activity after just after encoding across days decreases (Fig. 2a,

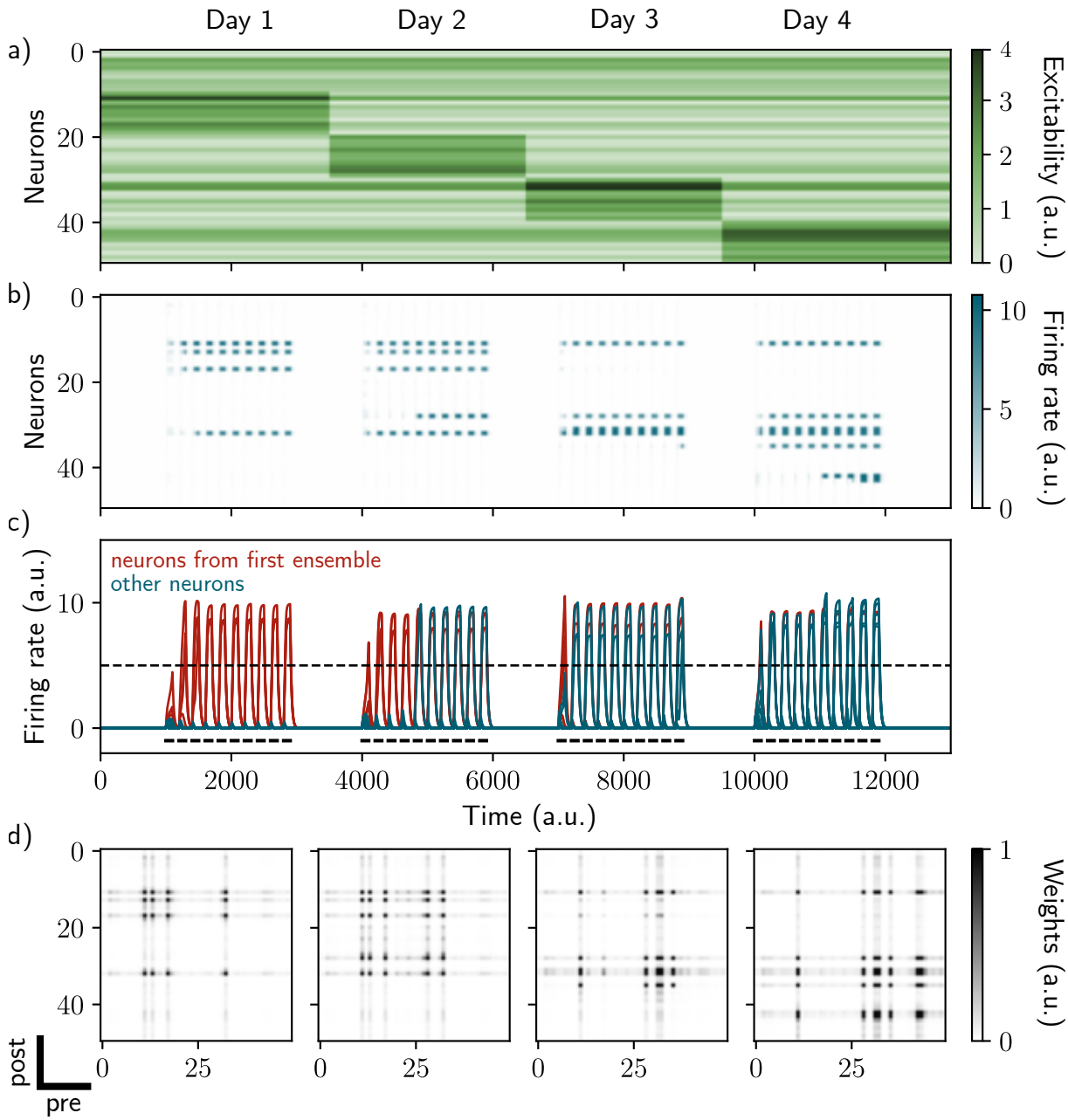

Figure 1: Excitability-induced drift of memory ensembles. a) Distribution of excitability ε_i for each neuron i, fluctuating over time. During each stimulation, a different pool of neurons has a high excitability (Methods). b) and c) Firing rates of the neurons across time. The red traces in panel c) correspond to neurons belonging to the first assembly, namely that have a firing rate higher than the active threshold after the first stimulation. The black bars show the stimulation and the dashed line shows the active threshold. d) Recurrent weights matrices after each of the four stimuli show the drifting assembly.

Methods), indicating that after each day, the newly formed ensemble resembles less the original one. Because the patterns of activity differ across days, they should be informative about the absolute day from which they were recorded. To test this hypothesis, we designed a day decoder (Fig. 2b, Methods), following the work of Rubin *et al.*, 2015 [25]. This decoder aims at inferring the reactivation day of a given activity pattern by comparing the activity of this pattern during training and the activity just after memory encoding without increase in excitability (Fig. 2b, Methods). We found that the day decoder perfectly outputs the reactivation day as compared to using shuffled data (Fig. 2c, blue and orange bars).

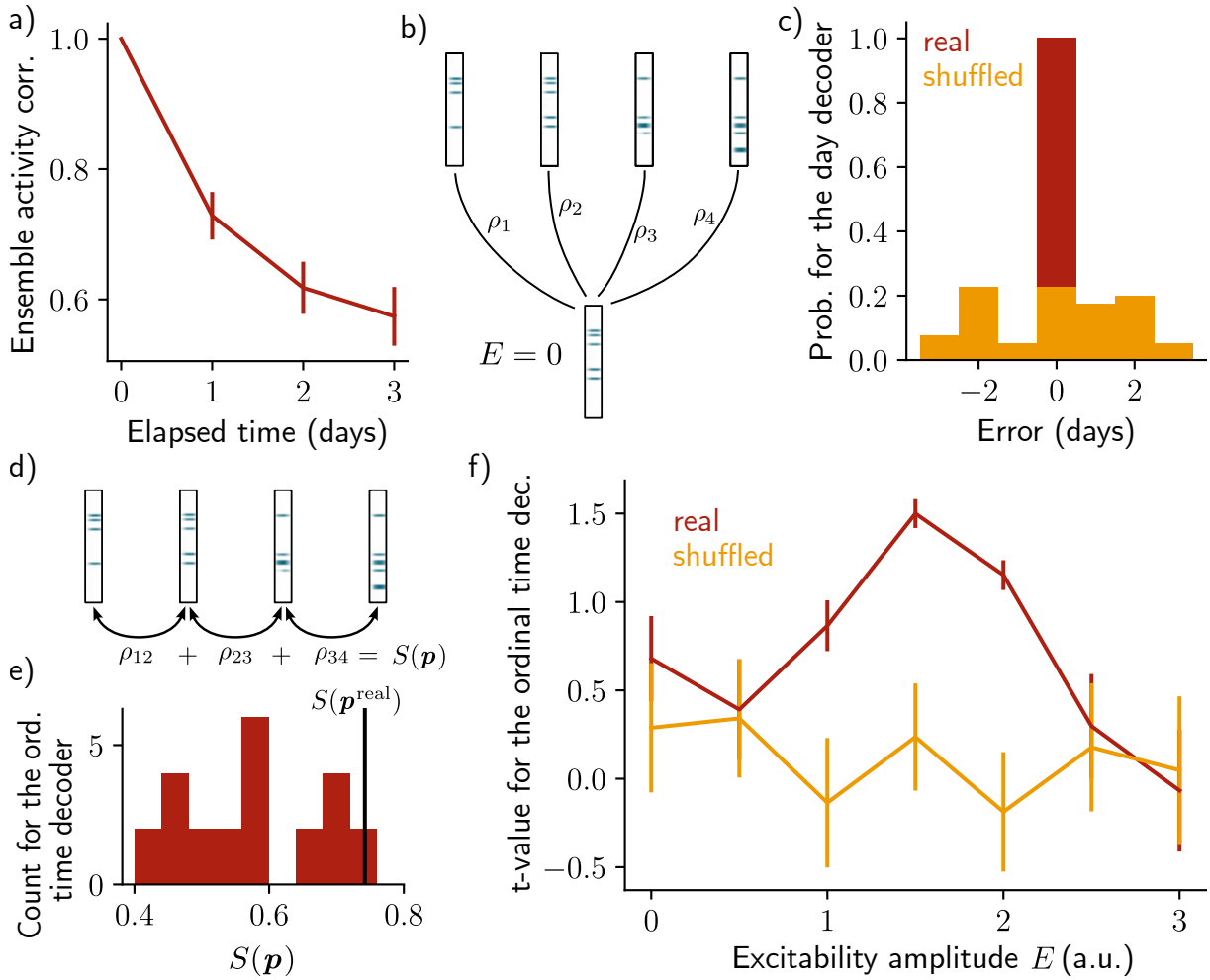

Figure 2: Neuronal activity is informative about the temporal structure of the reactivations. a) Correlation of the patterns of activity between the first day and every other days, for n = 10 simulations. Data are shown as mean ± s.e.m. b) Schema of the day decoder. The day decoder maximises correlation between the patterns of each day with the pattern from the simulation with no increase in excitability. c) Results of the day decoder for the real data (blue) and the shuffled data (orange). Shuffled data consist of the same activity pattern for which the label of each cells for every seed has been shuffled randomly. For each simulation, the error is measured for each day as the difference between the decoded and the real day. Data are shown for n = 10 simulations and for each of the 4 days. d) Schema of the ordinal time decoder. This decoder output the permutation $\boldsymbol{p}$ that maximises the sum $S(\boldsymbol{p})$ of the correlations of the patterns for each pairs of days. e) Distribution of the value $S(\boldsymbol{p})$ for each permutation of days $\boldsymbol{p}$. The value for the real permutation $S(\boldsymbol{p}_{\text{real}})$ is shown in black. f) Student's test t-value for n = 10 simulations, for the real (blue) and shuffled (orange) data and for different amplitudes of excitability E. Data are shown as mean ± s.e.m. for n = 10 simulations.

After showing that the patterns of activity are informative about the reactivation day, we took a step further by asking to whether the activity of the neurons is also informative about the order in which the memory evolved. To that end, we used an ordinal time decoder (Methods, as in Rubin et al., 2015 [25]) that uses the correlations between activity patterns for pairs of successive days, and for each possible permutation of days $\boldsymbol{p}$ (Fig. 2d, Methods). The sum of these correlations $S(\boldsymbol{p})$ differs from each permutation $\boldsymbol{p}$ and we assumed that the neurons are informative about the order at which the reactivations of the ensemble happened if the permutation maximising $S(\boldsymbol{p})$ corresponds to the real permutation $\boldsymbol{p}^{\text{real}}$ (Fig. 2e, Methods). We found that $S(\boldsymbol{p}^{\text{real}})$ was indeed statistically higher than $S(\boldsymbol{p})$ for the other permutations

E (Fig. 2f, Student's t-test, Methods). However, this was only true when the amplitude of the fluctuations of excitability E was in to a certain range. Indeed, when the amplitude of the fluctuations is null, *i.e.* when excitability is not increased ($E = 0$), the ensemble does not drift (Fig. S1a). In this case, the patterns of activity are not informative about the order of reactivations. On the other hand, if the excitability amplitude is too high ($E = 3$), each new ensemble is fully determined by the distribution of excitability, regardless of any previously formed ensemble (Fig. S1c). In this regime, the patterns of activity are not informative about the order of the reactivations either. In the intermediate regime ($E = 1.5$) , the decoder is able to correctly infer the order at which the reactivations happened, better than using the shuffled data (Fig. 2f, Fig. S1b).

Finally, we sought to test whether the results are independent on the specific architecture of the model. To that end, we defined a change of excitability as a change in the slope of the activation function, rather than of the threshold (Fig. S5, Methods). We also used sparse recurrent synaptic weights instead of the original all-to-all connectivity matrix (Fig. S4, Methods). In both cases, we observed a similar drifting behavior and were able to decode the temporal order in which the memory evolved.

A read-out neuron can track the drifting ensemble

So far, we showed that the drifting ensemble contains information about its history, namely about the days and the order at which the subsequent reactivations of the memory happened. However, we have not shown that we could use the neural ensemble to actually decode the memory itself, in addition to its temporal structure. To that end, we introduced a decoding output neuron connected to the recurrent neural network, with plastic weights following a Hebbian rule (Methods). As shown by Rule *et al.*, 2022 [8], the goal was to make sure that the output neuron can track the ensemble even if it is drifting. This can be down by constantly decreasing weights from neurons that are no longer in the ensemble and increasing those associated with neurons joining the ensemble (Fig. 3a). We found that the output neuron could steadily decode the memory (*i.e.* it has a higher firing than in the case where the output weights are randomly shuffled; Fig. 3b, blue trace for the real output and white, orange and red traces for the shuffled weights). This is due to the fact that weights are plastic under Hebbian learning, as shown by Rule *et al.* 2022 [8]. We confirmed that this was induced by a change in the output weights across time (Fig. 3c). In particular, the weights from neurons that no longer belong to the ensemble are decreased while weights from newly recruited neurons are increased, so that the center of mass of the weights distribution drifts across time (Fig. 3d). Finally, we found that the quality of the read-out decreases with the rate of the drift (Fig. S3, Methods).

Two memories drift independently

Finally, we tested whether the network is able to encode two different memories and whether excitability could make two ensembles drift. On each day, we stimulated a random half of the neurons (context A) and the other half (context B) sequentially (Methods). We found that, day after day, the two ensembles show a similar drift than when only one ensemble was formed (Fig. S6). In particular, the correlation between the patterns activity on the first day and the other days decay in a similar way (Fig. S7.a). For both contexts, the temporal order of the reactivations can be decoded for a certain range of excitability amplitude (Fig. S7.b). Finally, we found that using two output decoders allowed us to decode both memories independently . The output weights associated to both ensembles are remodeled to follow the drifting ensembles, but are not affected by the reactivation of the other ensemble (Fig. S7.c). Indeed, both neurons are able to "track" the reactivation of their associated ensemble while not responding to the other ensemble (Fig. S7.d).

Discussion

Overall, our model suggests a potential cellular mechanisms for the emergence of drift that can serve a computational purpose by "time-stamping" memories while still being able to decode the memory across time. Although the high performance of the day decoder was expected, the performance of the ordinal time decoder is not trivial. Indeed, the patterns of activity of each day are informative about the distribution of excitability and therefore about the day at which the reactivation happened. However, the ability for the neural ensemble to encode the order of past reactivations requires drift to be gradual (*i.e.* requires consecutive patterns of activity to remain correlated across days). Indeed, if the amplitude of excitability is too low ($E = 0$) or too high ($E = 3$), it is not possible to decode the order at which the successive reactivations happened. This result is consistent with the previous works showing gradual change in neural representations, that allows for decoding temporal information of the ensemble [25].

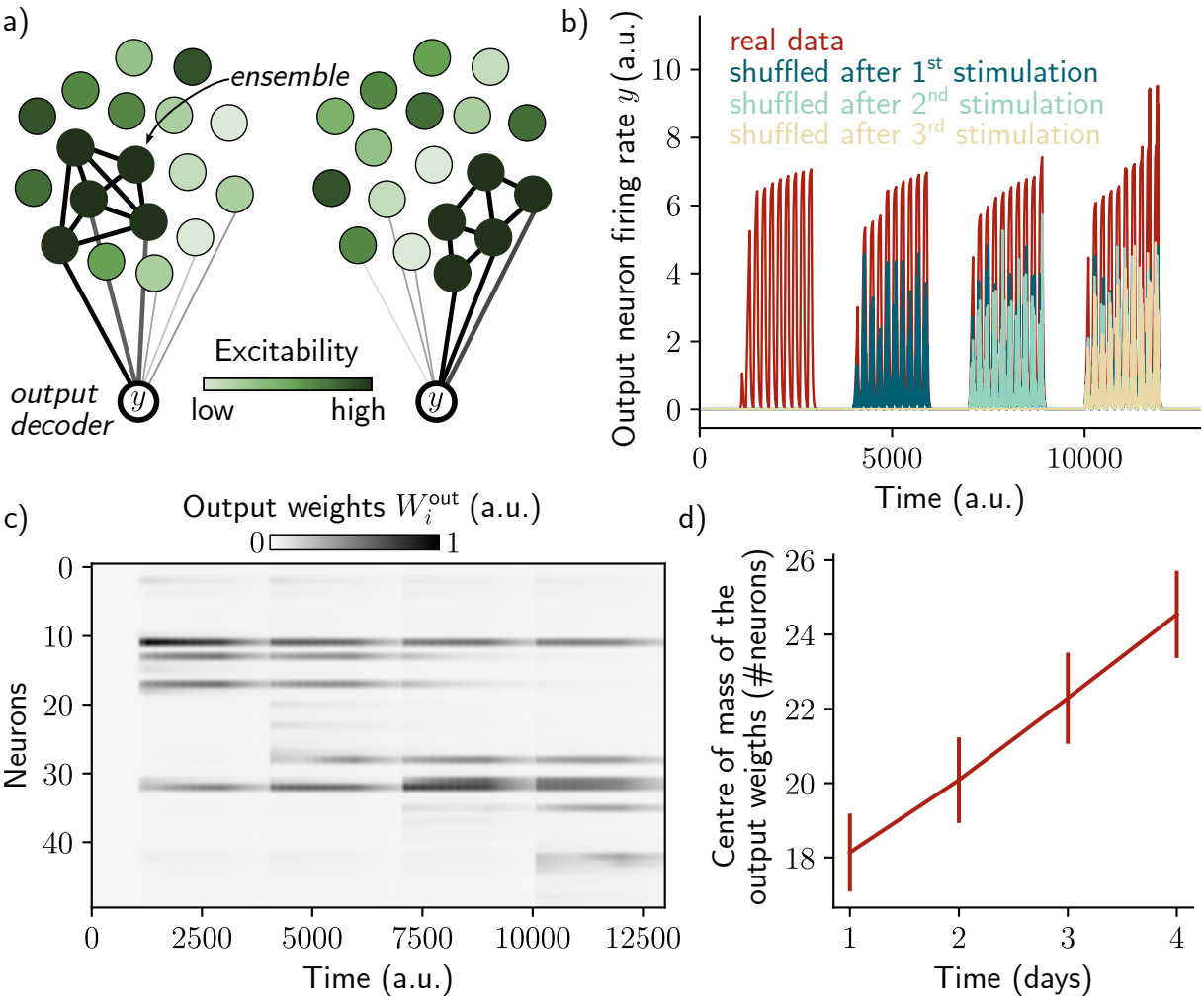

Figure 3: A single output neuron can track the memory ensemble through Hebbian plasticity. a) Conceptual architecture of the network: the read-out neuron y in red "tracks" the ensemble by decreasing synapses linked to the previous ensemble and increasing new ones to linked to the new assembly. b) Output neuron's firing rate across time. The blue trace correspond to the real output. The white, orange and red traces correspond to the cases where the output weights were randomly shuffled for every time points after presentation of the first, second and third stimulus, respectively. c) Output weights for each neuron across time. d) Center of mass of the distribution of the output weights (Methods) across days. The output weights are centered around the neurons that belong to the assembly at each day. Data are shown as mean $\pm$ s.e.m. for n $= 10$ simulations.

Moreover, such gradual drifts could support complex cognitive mechanisms like mental time-travel during memory recall [25].

In our model, drift is induced by co-activation of the previously formed ensemble and neurons with high excitability at the time of the reactivation. The pool of neurons having high excitability can therefore "time-stamps" memory ensembles by biasing allocation of these ensembles [23, 25, 26]. We suggest that such time-stamping mechanism could also help link memories that are temporally close and dissociate those which are spaced by longer time [3, 14, 31]. Indeed, the pool of neurons with high excitability varies across time so that any new memory ensemble is allocated to neurons which are shared with other ensembles formed around the same time. This mechanism could be complementary to the learning-induced increase in excitability observed in amygdala [18], hippocampal CA1 [17] and dentate gyrus [32].

Finally, we intended to model drift in the firing rates, as opposed to a drift in the turning curve, of the neurons. Recent studies suggest that drifts in the mean firing rate and tuning curve arise from two different mechanisms [33, 34]. Experience drives a drift in neurons turning curve while the passage of time drives a drift in neurons firing rate. In this sense, our study is consistent with these findings by providing a possible mechanism for a drift in the mean firing rates of the neurons driven a dynamical excitability. Our work suggests that drift can depend on any experience having an impact on excitability dynamics such as exercise as previously shown experimentally [9, 35] but also neurogenesis [9, 31, 36], sleep [37] or increase in dopamine level [38]. Overall, our work is a proof of principle which highlights the importance of considering excitability when studying drift, although further work would be needed to test this link experimentally.

Methods

Recurrent neural network with excitability

Our rate-based model consists of a single region of N neurons (with firing rate r_i, $1 \leq i \leq N$). All-to-all recurrent connections W are plastic and follow a Hebbian rule given by:

$$\frac{\mathrm{d}W_{ij}}{\mathrm{d}t} = r_i * r_j / \tau_W - W_{ij} / \tau_{\text{decay}} \tag{1}$$

where i and j correspond to the pre- and post-synaptic neuron respectively. τ_W and τ_{decay} are the learning and the decay time constants of the weights, respectively. A hard bound of $[0, c]$ was applied to these weights. We also introduced a global inhibition term dependent on the activity of the neurons:

$$I = I_0 + I_1 \sum_{i=1}^{N} \boldsymbol{r}_i + I_2 \sum_{i=1}^{N} \boldsymbol{r}_i^2 \tag{2}$$

Where I_0, I_1 and I_2 are positive constants. All neurons receive the same input, $\Delta(t)$, during stimulation of the network (Fig. 1c, black bars). Finally, excitability is modeled as a time-varying threshold ε_i of the input-output function of each neuron i. The rate dynamics of a neuron i is given by:

$$\tau_r \frac{\mathrm{d}r_i}{\mathrm{d}t} + r_i = \text{ReLU}\left(\Delta(t) + \sum_{j=1}^{N} W_{ij} r_j - I + \varepsilon_i(t)\right) \tag{3}$$

where τ_r is the decay time of the rates and ReLU is the rectified linear activation function. We considered that a neurons is active when its firing rate reaches the active threshold θ.

In Fig. S4, we applied a random binary mask to the recurrent weights in order to set 50% of the synapses at 0. A new mask was randomly sampled for each simulation.

In Fig. S5, we modeled excitability as a change of the slope of the activation function (ReLU) instead of a change of the threshold as previously used (Fig. S5.a):

$$\tau_r \frac{\mathrm{d}r_i}{\mathrm{d}t} + r_i = \varepsilon_i(t) * \mathrm{ReLU}\left(\Delta(t) + \sum_{j=1}^{N} W_{ij} r_j - I\right) \tag{4}$$

Protocol

We designed a 4-day protocol, corresponding to the initial encoding of a memory (1st day) and subsequent random or cue-induced reactivations of the ensemble [28, 30] (2nd, 3rd and 4th day). Each stimulation consists of N_{rep} repetitions of interval T spaced by a inter-repetition delay IR. $\Delta(t)$ takes the value δ during these repetitions and is set to 0 otherwise. The stimulation is repeated four times, modeling four days of reactivation, spaced by an inter-day delay ID. Excitability ε_i of each neuron i is sampled from a chi-squared distribution of parameter 1. In Fig. S5, excitability ε_i is sampled from the square of a normal distribution of mean 0.4 and standard deviation 0.2. Neurons 10 to 20, 20 to 30, 30 to 40 and 40 to 50 then receive an increase of excitability of amplitude E, respectively on days 1, 2, 3 and 4 (Fig. 1a). A different random seed is used for each repetitions of the simulations. When two memories were modeled (Fig. S6,S7), we stimulated a random half of the neurons (context A) and the other half (context B) successively (Fig. S6.a), every day.

Decoders

For each day d, we recorded the activity pattern V_d, which is a vector composed of the firing rate of the neurons at the end of the last repetition of stimulation. To test the decoder, we also stimulated the network while setting the excitability at baseline ($E = 0$), and recorded the resulted pattern of activity V_d^0 for each day d. We then designed two types of decoders, inspired by previous works [25]: (1) a day decoder which infers the day at which each stimulation happened and (2) an ordinal time decoder which infers the order at which the reactivations occurred. For both decoders, the shuffled data was obtained by randomly shuffling the day label of each neuron. When two memories were modeled (Fig. S6,S7), the patterns of activity were taken at the end of the stimulations by context A and B, and the decoders were used independently on each memory.

1. The day decoder aims at inferring the day at which a specific patterned of activity occurred. To that end, we computed the Pearson correlation between the pattern with no excitability V_d^0 of the day d and the patterns of all days d' from the first simulation $V_{d'}$. Then, the decoder outputs the day d_{inf} that maximises the correlation:

$$d_{\mathrm{inf}} = \arg\max_{d'} \{\mathrm{corr}(V_d^0, V_{d'})\} \tag{5}$$

The error was defined as the difference between the inferred and the real day $d_{\mathrm{inf}} - d$.

2. The ordinal time decoder aims at inferring the order at which the reactivations happened from the patterns of activity V_d of every days d. To that end, we computed the pairwise correlations of each consecutive days, for the 4! possible permutations of days $\boldsymbol{p}$. The real permutation is called $\boldsymbol{p}^{\mathrm{real}} = (1,2,3,4)$ and corresponds to the real order of reactivations: day 1 $\rightarrow$ day 2 $\rightarrow$ day 3 $\rightarrow$ day 4. The sum of these correlations over the 3 pairs of consecutive days is expressed as:

$$S(\boldsymbol{p}) = \sum_{i=1}^{3} \mathrm{corr}(V_{p_i}, V_{p_{i+1}}) \tag{6}$$

We then compared the distribution of these quantities for each permutation $\boldsymbol{p}$ to that of the real permutation $S(\boldsymbol{p}^{\mathrm{real}})$ (Fig. 2). The patterns of activity are informative about the order of reactivations if $S(\boldsymbol{p}^{\mathrm{real}})$

corresponds to the maximal value of $S(\boldsymbol{p})$. To compare $S(\boldsymbol{p}^{\text{real}})$ with the distribution $S(\boldsymbol{p})$, we performed a Student's t-test, where the t-value is defined as :

$$t = \frac{S(\boldsymbol{p}^{\text{real}}) - \mu}{\sigma/\sqrt{N}} \tag{7}$$

where μ and σ corresponds to the mean and standard deviation of the distribution $S(\boldsymbol{p})$, respectively.

The drift rate Δ (Fig. S2 and Fig. S3) was computed as:

$$\Delta = \sum_{i=2}^{4} (1 - \text{corr}(V_1, V_i)) \tag{8}$$

Memory read-out

To test if the network is able to decode the memory at any time point, we introduced a read-out neuron with plastic synapses to neurons from the recurrent network, inspired by previous computational works [8]. The weights of these synapses are named $\boldsymbol{W}^{\text{out}} = (W_i^{\text{out}})_{1 \leq i \leq N}$ and follow the Hebbian rule defined as:

$$\frac{\mathrm{d}W_i^{\text{out}}}{\mathrm{d}t} = h(\boldsymbol{W}^{\text{out}}) * r_i * y/\tau_{\text{out}}^{+} - W_i^{\text{out}}/\tau_{\text{out}}^{-} \tag{9}$$

where τ_{out}^{+} and τ_{out}^{-} corresponds to the learning time and decay time constant, respectively. $h(\boldsymbol{W}^{\text{out}})$ is a homeostatic term defined as $h(\boldsymbol{W}^{\text{out}}) = 1 - \sum_{j=1}^{N} W_j^{\text{out}}$ which decreases to 0 throughout learning. h takes the value 1 before learning and 0 when the sum of the weights reaches the value 1. y is the firing rate of the output neuron defined y as:

$$y = \sum_{i=1}^{N} W_i^{\text{out}} r_i \tag{10}$$

The read-out quality index Q (Fig. S3) was defined as:

$$Q = \langle \sum_{d=2}^{4} y_d / y_d^{\text{shuffle}} \rangle_{N_{\text{shuffle}}} \tag{11}$$

where y_d corresponds to the value of y taken at the end of the last repetition of day d, and y_d^{shuffle} the equivalent with shuffled outputs weights. $\langle ... \rangle_{N_{\text{shuffle}}}$ indicates the average over $N_{\text{shuffle}} = 10$ simulations.

In Fig. S6,S7, two output decoders y_k, $k \in \{1,2\}$, with corresponding weights $\boldsymbol{W}_{\boldsymbol{k}}^{\text{out}} = (W_{i,k}^{\text{out}})_{1 \leq i \leq N}$ are defined as:

$$y_k = \sum_{i=1}^{N} W_{i,k}^{\text{out}} r_i + \beta_k \tag{12}$$

and follow the Hebbian rule defined as:

$$\frac{\mathrm{d}W_{i,k}^{\text{out}}}{\mathrm{d}t} = h(\boldsymbol{W}_{i,k}^{\text{out}}) * r_i * y_k/\tau_{\text{out}}^{+} - W_{i,k}^{\text{out}}/\tau_{\text{out}}^{-} \tag{13}$$

Then, we aimed at allocating y_1 and y_2 to the first and the second ensemble (context A and B), respectively. To that end, we used supervised learning on the first day by adding a current β_k to the output neurons which is positive when the corresponding context is on:

$$\begin{aligned} \beta_0 &= 0.1 \text{ if } 1000 < t < 3000, \text{ 0 otherwise} \\ \beta_1 &= 0.1 \text{ if } 4000 < t < 6000, \text{ 0 otherwise} \end{aligned} \tag{14}$$

The shuffled traces were obtained by randomly shuffling the output weights $\boldsymbol{W}^{\text{out}}$ or $\boldsymbol{W}_{\boldsymbol{k}}^{\text{out}}$ for each ensemble k.

Table of parameters

The following parameters have been used for the simulations. When unspecified, the defaults values were used. All except N are in arbitrary unit. Fig. S5 corresponds to the change from a threshold-based to a slope-based excitability. Fig. S6,S7 corresponds to the stimulation of two ensembles. Fig. S4 corresponds to the sparsity simulation.

Param.	Description	Default	Fig. S5	Fig. S6,S7	Fig. S4
N	Number of neurons	50	-	-	-
τ_W	Learning time constant of the recurrent weights	800	700	-	-
τ_{decay}	Decay time constant of the recurrent weights	1000	800	4000	-
τ_r	Decay time constant of the firing rates	20	-	-	-
τ^+_{out}	Learning time constant of the output weights	200	-	-	-
τ^-_{out}	Decay time constant of the output weights	1000	-	-	-
I_0	First inhibition parameter	12	4	8	7
I_1	Second inhibition parameter	0.5	0.7	0.8	0.8
I_2	Third inhibition parameter	0.05	-	-	-
δ	Input current during stimulation	15	-	12	20
E	Amplitude of the fluctuations of excitability	1.5	0.5	-	-
N_{rep}	Number of repetitions	10	-	-	-
T	Duration of each repetition	100	-	-	-
IR	Inter-repetition delay	100	-	-	-
ID	Inter-stimulation delay	1000	-	-	-
θ	Active threshold	5	1	-	-
c	Cap on the recurrent weights	1	.5	-	-

References

1. Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N. & Harvey, C. D. Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. *Cell* **170,** 986–999.e16. doi:10.1016/j.cell.2017.07.021 (Aug. 2017).
2. Sadeh, S. & Clopath, C. Contribution of behavioural variability to representational drift. *eLife* **11** (eds Palmer, S. E., Frank, M. J. & Ziv, Y.) e77907. doi:10.7554/eLife.77907 (Aug. 2022).
3. Driscoll, L. N., Duncker, L. & Harvey, C. D. Representational drift: Emerging theories for continual learning and experimental future directions. *Current Opinion in Neurobiology* **76,** 102609. doi:10.1016/j.conb.2022.102609 (Oct. 2022).
4. Ziv, Y. *et al.* Long-term dynamics of CA1 hippocampal place codes. *Nature Neuroscience* **16,** 264–266. doi:10.1038/nn.3329 (Mar. 2013).
5. Hainmueller, T. & Bartos, M. Parallel emergence of stable and dynamic memory engrams in the hippocampus. *Nature* **558,** 292–296. doi:10.1038/s41586-018-0191-2 (June 2018).
6. Spalla, D., Cornacchia, I. M. & Treves, A. Continuous attractors for dynamic memories. *eLife* **10,** e69499. doi:10.7554/eLife.69499 (Sept. 2021).
7. Cho, H.-Y. *et al.* Turnover of fear engram cells by repeated experience. *Current Biology* **31,** 5450–5461.e4. doi:10.1016/j.cub.2021.10.004 (Dec. 2021).
8. Rule, M. E. & O'Leary, T. Self-healing codes: How stable neural populations can track continually reconfiguring neural representations. *Proceedings of the National Academy of Sciences* **119,** e2106692119. doi:10.1073/pnas.2106692119 (Feb. 2022).

9. Rechavi, Y., Rubin, A., Yizhar, O. & Ziv, Y. Exercise increases information content and affects long-term stability of hippocampal place codes. *Cell Reports* **41,** 111695. doi:10.1016/j.celrep.2022.111695 (Nov. 2022).

10. Attardo, A., Fitzgerald, J. E. & Schnitzer, M. J. Impermanence of dendritic spines in live adult CA1 hippocampus. *Nature* **523,** 592–596. doi:10.1038/nature14467 (July 2015).

11. Kalle Kossio, Y. F., Goedeke, S., Klos, C. & Memmesheimer, R.-M. Drifting assemblies for persistent memory: Neuron transitions and unsupervised compensation. *Proceedings of the National Academy of Sciences* **118,** e2023832118. doi:10.1073/pnas.2023832118 (Nov. 2021).

12. Manz, P. & Memmesheimer, R.-M. *Purely STDP-based assembly dynamics: stability, learning, overlaps, drift and aging* Mar. 2023. doi:10.1101/2022.06.20.496825.

13. Zhou, Y. *et al.* CREB regulates excitability and the allocation of memory to subsets of neurons in the amygdala. *Nature Neuroscience* **12,** 1438–1443. doi:10.1038/nn.2405 (Nov. 2009).

14. Mau, W., Hasselmo, M. E. & Cai, D. J. The brain in motion: How ensemble fluidity drives memory-updating and flexibility. *eLife* **9,** e63550. doi:10.7554/eLife.63550 (Dec. 2020).

15. Rogerson, T. *et al.* Synaptic tagging during memory allocation. *Nature Reviews Neuroscience* **15,** 157–169. doi:10.1038/nrn3667 (Mar. 2014).

16. Silva, A. J., Zhou, Y., Rogerson, T., Shobe, J. & Balaji, J. Molecular and Cellular Approaches to Memory Allocation in Neural Circuits. *Science* **326,** 391–395. doi:10.1126/science.1174519 (Oct. 2009).

17. Cai, D. J. *et al.* A shared neural ensemble links distinct contextual memories encoded close in time. *Nature* **534,** 115–118. doi:10.1038/nature17955 (June 2016).

18. Rashid, A. J. *et al.* Competition between engrams influences fear memory formation and recall. *Science* **353,** 383–387. doi:10.1126/science.aaf0594 (July 2016).

19. Grosmark, A. D. & Buzsáki, G. Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. *Science* **351,** 1440–1443. doi:10.1126/science.aad1935 (Mar. 2016).

20. Huber, R. *et al.* Human Cortical Excitability Increases with Time Awake. *Cerebral Cortex* **23,** 1–7. doi:10.1093/cercor/bhs014 (Feb. 2013).

21. Levenstein, D., Buzsáki, G. & Rinzel, J. NREM sleep in the rodent neocortex and hippocampus reflects excitable dynamics. *Nature Communications* **10,** 2478. doi:10.1038/s41467-019-10327-5 (June 2019).

22. Mau, W. *et al. Ensemble remodeling supports memory-updating* June 2022. doi:10.1101/2022.06.02.494530.

23. Mau, W. *et al.* The Same Hippocampal CA1 Population Simultaneously Codes Temporal Information over Multiple Timescales. *Current Biology* **28,** 1499–1508.e4. doi:10.1016/j.cub.2018.03.051 (May 2018).

24. Delamare, G., Feitosa Tomé, D. & Clopath, C. Intrinsic neural excitability induces time-dependent overlap of memory engrams. *bioRxiv,* 2022.08.27.505441. doi:10.1101/2022.08.27.505441 (Jan. 2022).

25. Rubin, A., Geva, N., Sheintuch, L. & Ziv, Y. Hippocampal ensemble dynamics timestamp events in long-term memory. *eLife* **4** (ed Eichenbaum, H.) e12247. doi:10.7554/eLife.12247 (Dec. 2015).

26. Clopath, C., Bonhoeffer, T., Hübener, M. & Rose, T. Variance and invariance of neuronal long-term representations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372,** 20160161. doi:10.1098/rstb.2016.0161 (Mar. 2017).

27. Miller, A. M. P., Frankland, P. W. & Josselyn, S. A. Memory: Ironing Out a Wrinkle in Time. *Current Biology* **28,** R599–R601. doi:10.1016/j.cub.2018.03.053 (May 2018).

28. Josselyn, S. A. & Tonegawa, S. Memory engrams: Recalling the past and imagining the future. *Science* **367,** eaaw4325. doi:10.1126/science.aaw4325 (Jan. 2020).

29. Poo, M.-m. *et al.* What is memory? The present state of the engram. *BMC Biology* **14,** 40. doi:10.1186/s12915-016-0261-6 (May 2016).

30. Káli, S. & Dayan, P. Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions. *Nature Neuroscience* **7,** 286–294. doi:10.1038/nn1202 (Mar. 2004).

31. Aimone, J. B., Wiles, J. & Gage, F. H. Potential role for adult neurogenesis in the encoding of time in new memories. *Nature Neuroscience* **9,** 723–727. doi:10.1038/nn1707 (June 2006).

32. Pignatelli, M. *et al.* Engram Cell Excitability State Determines the Efficacy of Memory Retrieval. *Neuron* **101,** 274–284.e5. doi:10.1016/j.neuron.2018.11.029 (Jan. 2019).

33. Geva, N., Deitch, D., Rubin, A. & Ziv, Y. Time and experience differentially affect distinct aspects of hippocampal representational drift. *Neuron.* doi:10.1016/j.neuron.2023.05.005 (June 2023).

34. Khatib, D. *et al.* Active experience, not time, determines within-day representational drift in dorsal CA1. *Neuron* **111,** 2348–2356.e5. doi:10.1016/j.neuron.2023.05.014 (Aug. 2023).

35. De Snoo, M. L., Miller, A. M. P., Ramsaran, A. I., Josselyn, S. A. & Frankland, P. W. Exercise accelerates place cell representational drift. *Current Biology* **33,** R96–R97. doi:10.1016/j.cub.2022.12.033 (Feb. 2023).

36. Tran, L. M. *et al.* Adult neurogenesis acts as a neural regularizer. *Proceedings of the National Academy of Sciences* **119,** e2206704119. doi:10.1073/pnas.2206704119 (Nov. 2022).

37. Levenstein, D., Watson, B. O., Rinzel, J. & Buzsáki, G. Sleep regulation of the distribution of cortical firing rates. *Current Opinion in Neurobiology. Neurobiology of Sleep* **44,** 34–42. doi:10.1016/j.conb.2017.02.013 (June 2017).

38. Chowdhury, A. *et al.* A locus coeruleus-dorsal CA1 dopaminergic circuit modulates memory linking. *Neuron.* doi:10.1016/j.neuron.2022.08.001 (Aug. 2022).

Supplementary

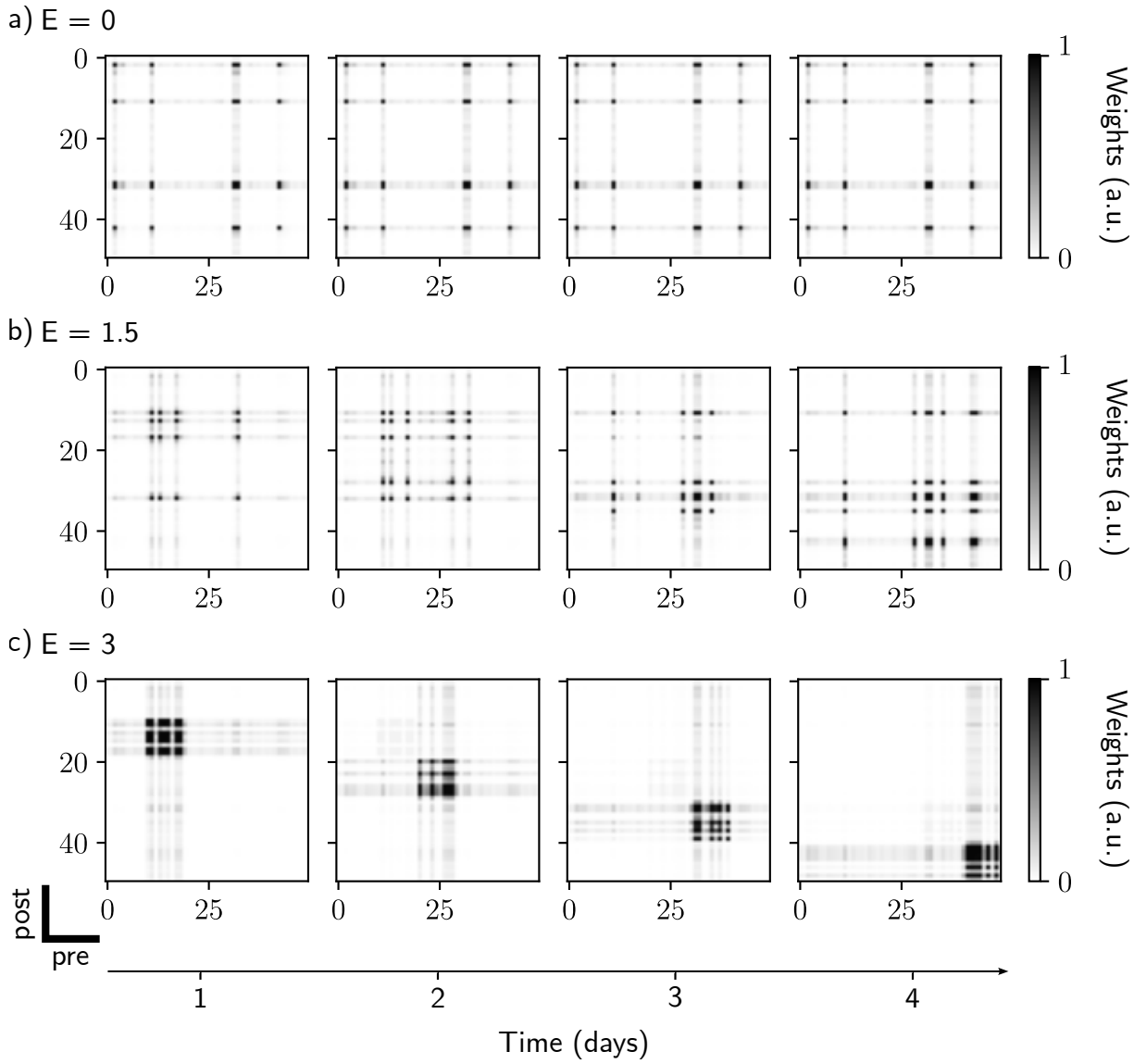

Figure S1: Comparison of drifting behavior for different values of excitability amplitude. a) $E = 0$, no drift. A neural assembly is initially formed during the first stimulation and later reactivated every subsequent day. b) $E = 1.5$, partial drift. The ensemble is gradually modified during each new stimulation. c) $E = 3$, full drift. Each new stimulation leads to formation of a new ensemble, containing neurons that have high excitability during this time.

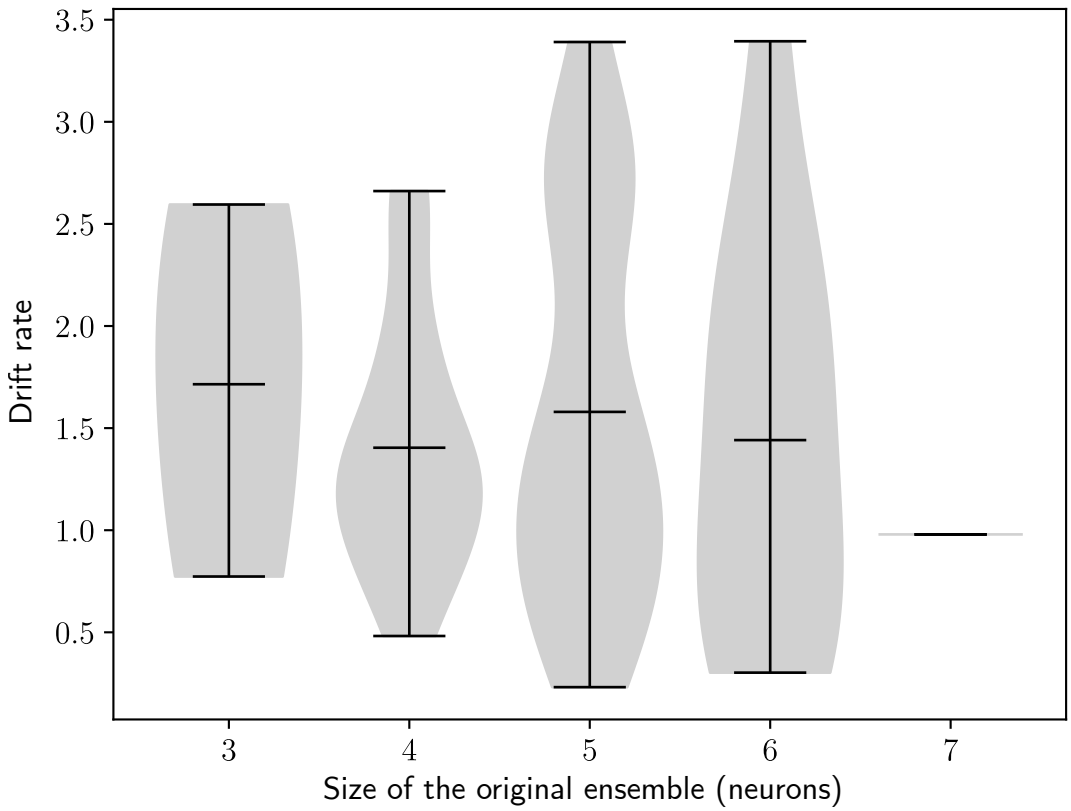

Figure S2: The rate of the drift does not depend on the size of the initial engram. Drift rate against the size of the original engram. Each dot shows one simulation (Methods). n = 100 simulations.

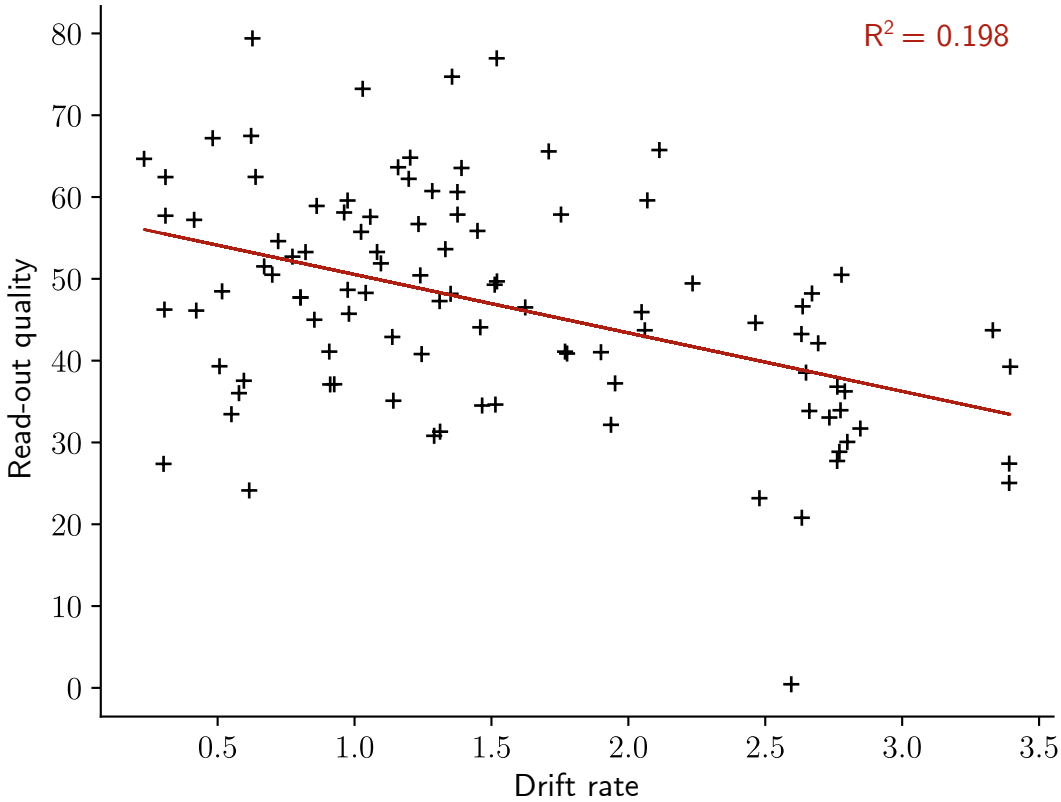

Figure S3: The quality of the read-out decreases with the rate of the drift. Read-out quality computed on the firing rate of the output neuron against the rate of the drift (Methods). Each dot shows one simulation. n = 100 simulations.

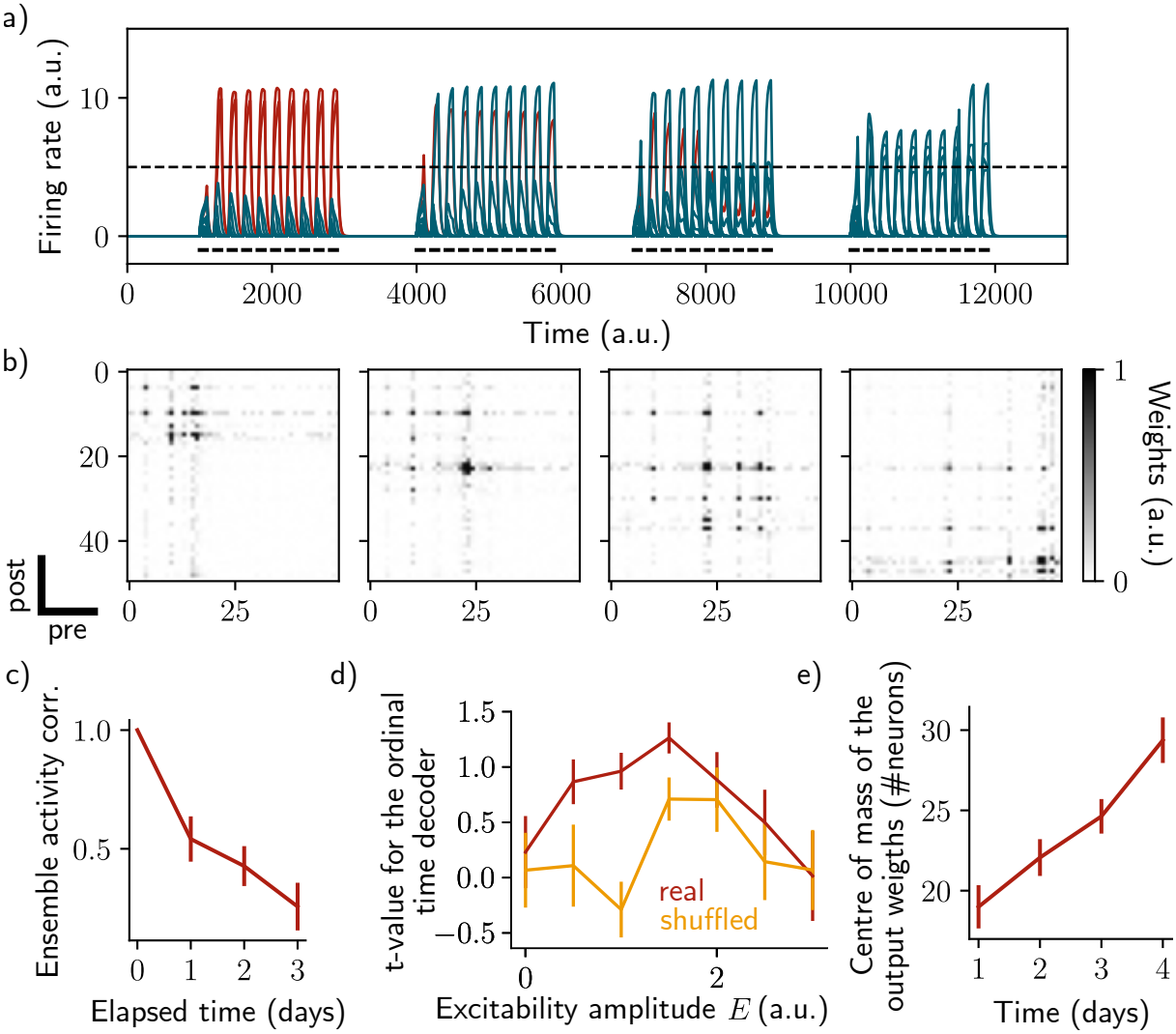

Figure S4: Sparse recurrent connectivity shows similar drifting behavior as all-to-all connectivity. The same simulation protocol as Fig. 1 was used while the recurrent weights matrix was made 50% sparse (Methods). a) Firing rates of the neurons across time. The red traces correspond to neurons belonging to the first assembly, namely that have a firing rate higher than the active threshold after the first stimulation. The black bars show the stimulation and the dashed line shows the active threshold. b) Recurrent weights matrices after each of the four stimuli show the drifting assembly. c) Correlation of the patterns of activity between the first day and every other days. d) Student's test t-value of the ordinal time decoder, for the real (blue) and shuffled (orange) data and for different amplitudes of excitability E. e) Center of mass of the distribution of the output weights (Methods) across days. c-e) Data are shown as mean $\pm$ s.e.m. for n $=$ 10 simulations.

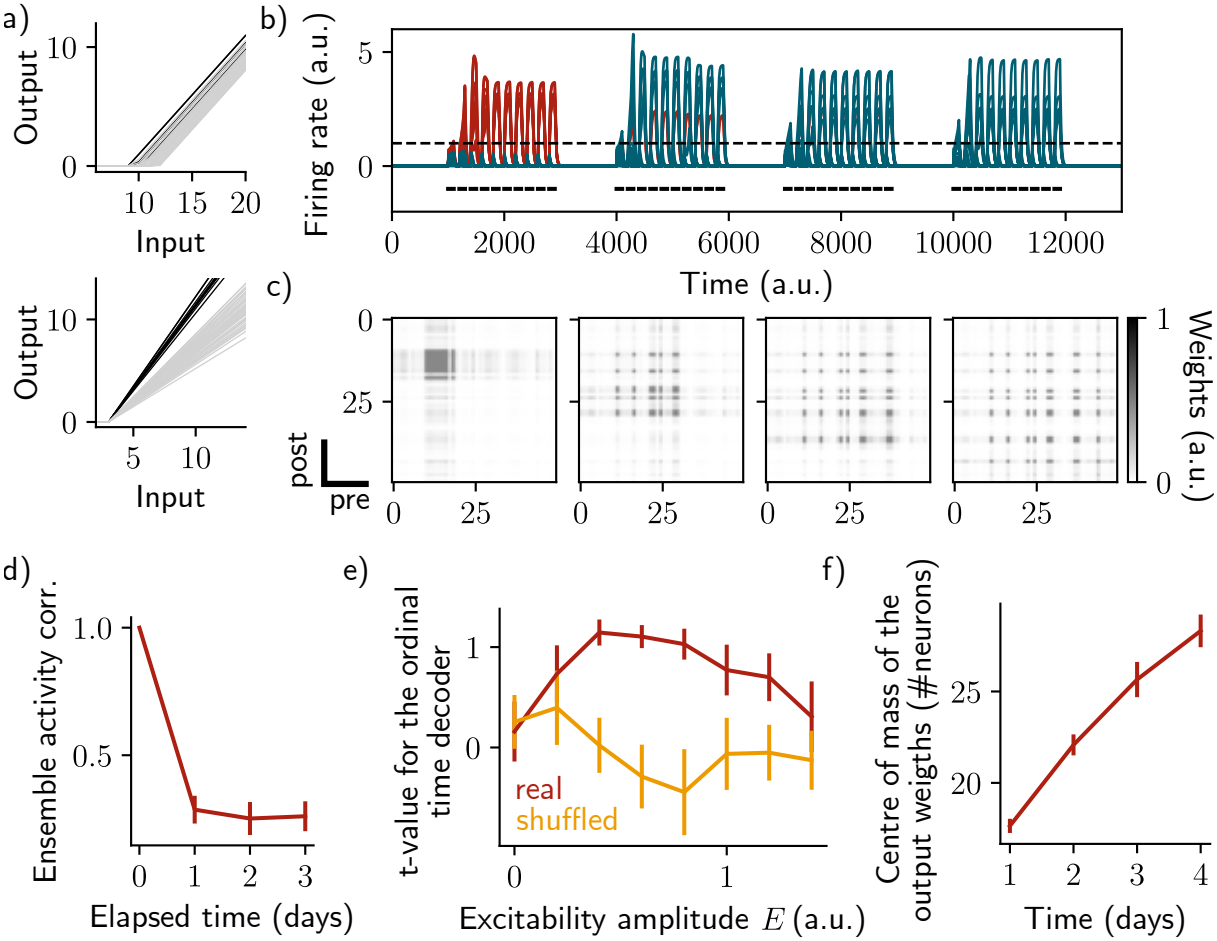

Figure S5: Change of excitability as a variable slope of the input-output function shows similar drifting behavior as considering a change in the threshold. The same simulation protocol as Fig. 1 was used while the excitability changes were modeled as a change in the activation function slope (Methods). a) Schema showing two different ways of defining excitability, as a threshold (top) or slope (bottom) of the activation function. Each line shows one neuron and darker lines correspond to neurons with increased excitability. b) Firing rates of the neurons across time. The red traces correspond to neurons belonging to the first assembly, namely that have a firing rate higher than the active threshold after the first stimulation. The black bars show the stimulation and the dashed line shows the active threshold. c) Recurrent weights matrices after each of the four stimuli show the drifting assembly. d) Correlation of the patterns of activity between the first day and every other days. e) Student's test t-value of the ordinal time decoder, for the real (blue) and shuffled (orange) data and for different amplitudes of excitability E. f) Center of mass of the distribution of the output weights (Methods) across days. d-f) Data are shown as mean $\pm$ s.e.m. for n $=$ 10 simulations.

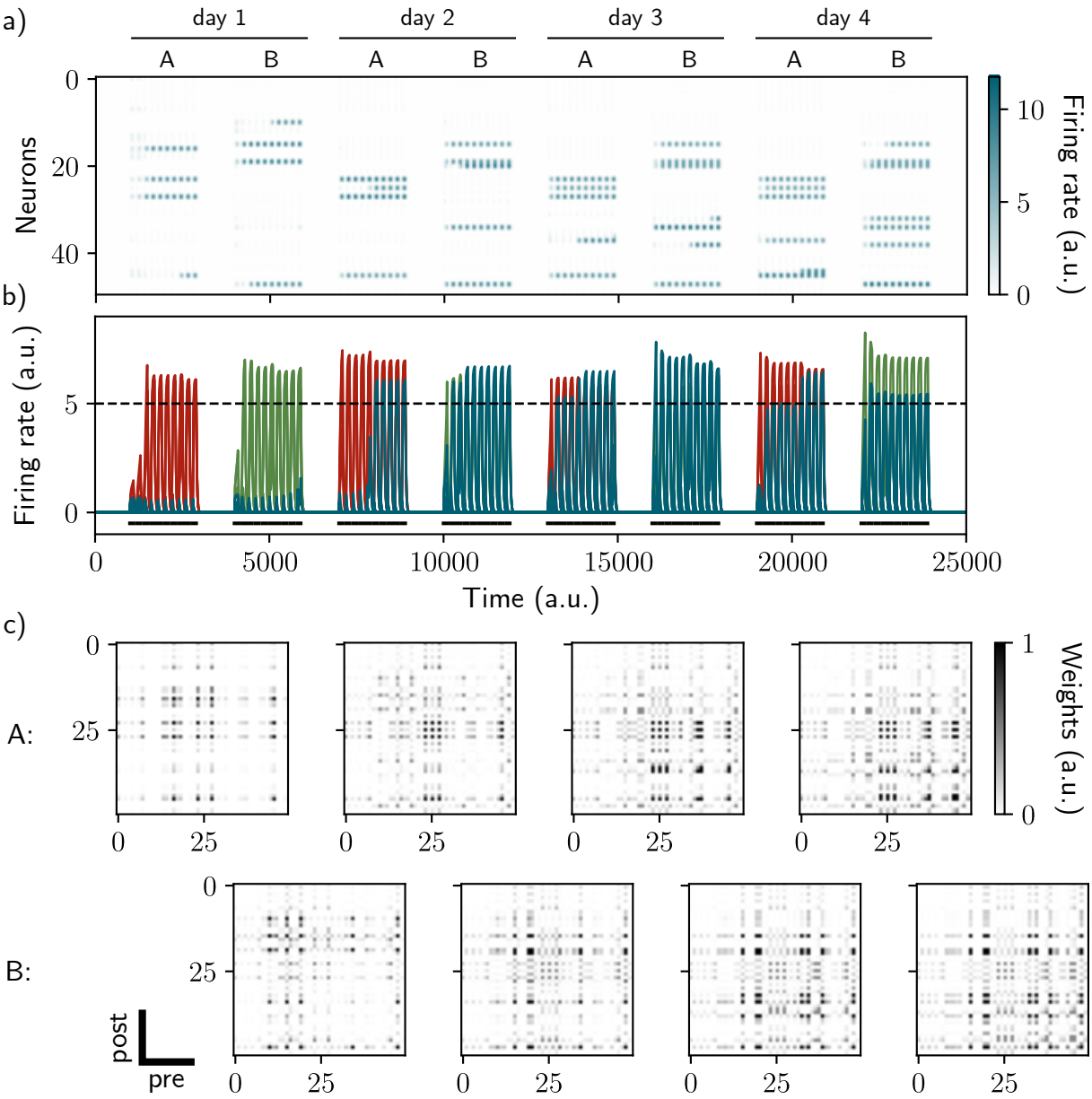

Figure S6: Two distinct ensembles can be encoded and drift independently. a) and b) Firing rates of the neurons across time. The red traces in panel b) correspond to neurons belonging to the first assembly and the green traces to the second assembly on the first day. They correspond to neurons having a firing rate higher than the active threshold after the first stimulation of each assembly. The black bars show the stimulation and the dashed line shows the active threshold. c) Recurrent weights matrices after each of the eight stimuli showing the drifting of the first (top) and second (bottom) assembly.

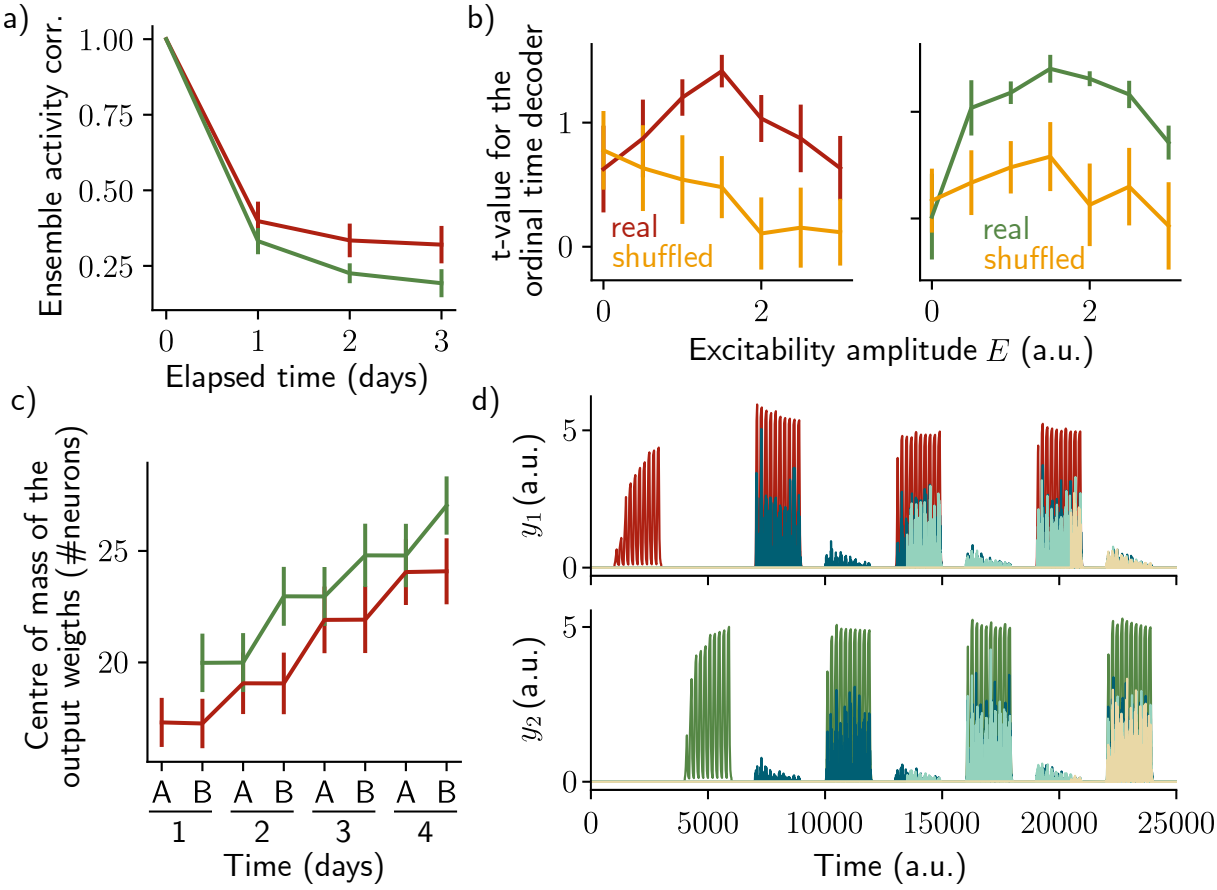

Figure S7: The two ensembles are informative about their temporal history and can be decoded using two output neurons. a) Correlation of the patterns of activity between the first day and every other days, for the first assembly (red) and the second assembly (green). b) Student's test t-value of the ordinal time decoder, for the first (red, left) and second ensemble (green, right) for different amplitudes of excitability E. Shuffled data are shown in orange. c) Center of mass of the distribution of the output weights (Methods) across days for the first ($\boldsymbol{W}_1^{\text{out}}$, red) and second ($\boldsymbol{W}_2^{\text{out}}$, green) ensemble. a-c) Data are shown as mean $\pm$ s.e.m. for n $=$ 10 simulations. d) Output neurons firing rate across time for the first ensemble (y_1, top) and the second ensemble (y_2, bottom). The red and green traces correspond to the real output. The dark blue, light blue and yellow traces correspond to the cases where the output weights were randomly shuffled for every time points after presentation of the first, second and third stimulus, respectively.