eLife

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

Reviewed preprint posted
8 June 2023 (this version)

Sent for peer review
7 April 2023

Posted to bioRxiv
17 March 2023

Neuroscience

Drift of neural ensembles driven by slow fluctuations of intrinsic excitability

Geoffroy Delamare ✉, Yosif Zaki, Denise J Cai, Claudia Clopath

Bioengineering Department, Imperial College London, London SW7 2AZ, UK • 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 remodelling 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.

eLife assessment

This is an **important** theoretical study providing insight into how fluctuations in excitability can contribute to gradual changes in the mapping between population activity and stimulus, commonly referred to as representational drift. The authors provide **convincing** evidence that fluctuations can contribute to drift, though certain modeling choices could benefit from justification or further exploration of alternatives. Overall, this is a well-presented study that explores the question of how changes in intrinsic excitability can influence memory representations.

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). Excitability might also play a role in memory allocation [(11)–(14)], so that neurons having high excitability are preferentially allocated to memory ensembles [(14)–(16)]. Moreover, excitability is known to fluctuate over timescales from hours to days, in the amygdala (16), the hippocampus (15; 17) or the cortex (18; 19). Subsequent reactivations of a neural ensemble at different time points could therefore be biased by excitability (20), which varies at similar timescales than drift (21). Altogether, this evidence suggest that fluctuations of excitability could act as a cellular mechanism for drift (12).

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 (22) 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 [(21), (23)–(25)], 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 (12; 26; 27). 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 (26; 28)). 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).

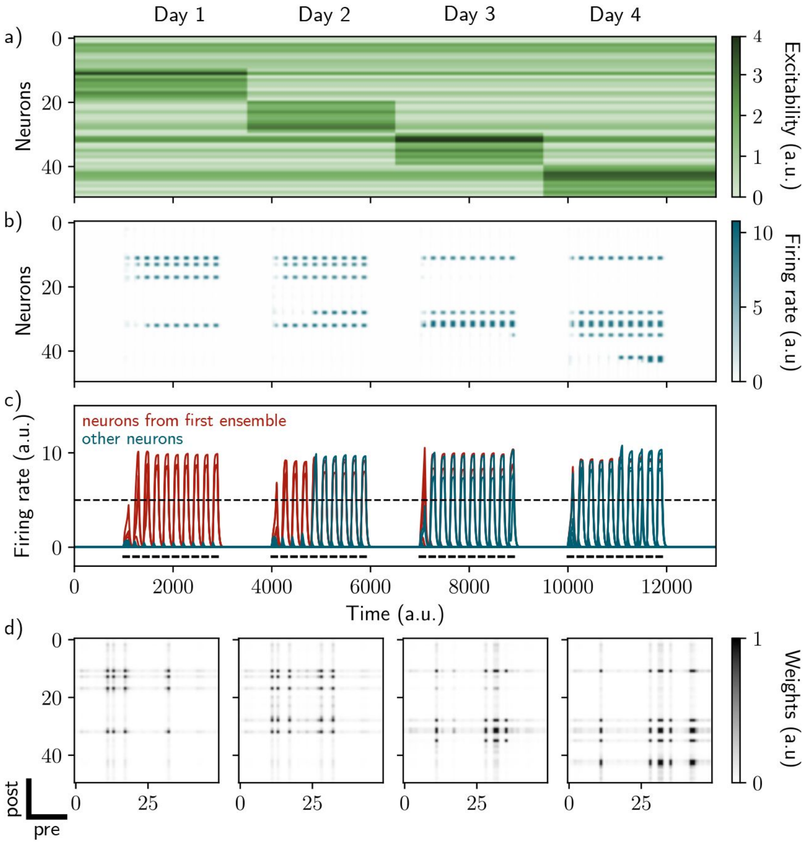


Figure 1:

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 lines correspond to the active threshold. d) Recurrent weights matrices after each of the four stimuli show the drifting assembly.

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.

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 (23). 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, 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 (23). 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).

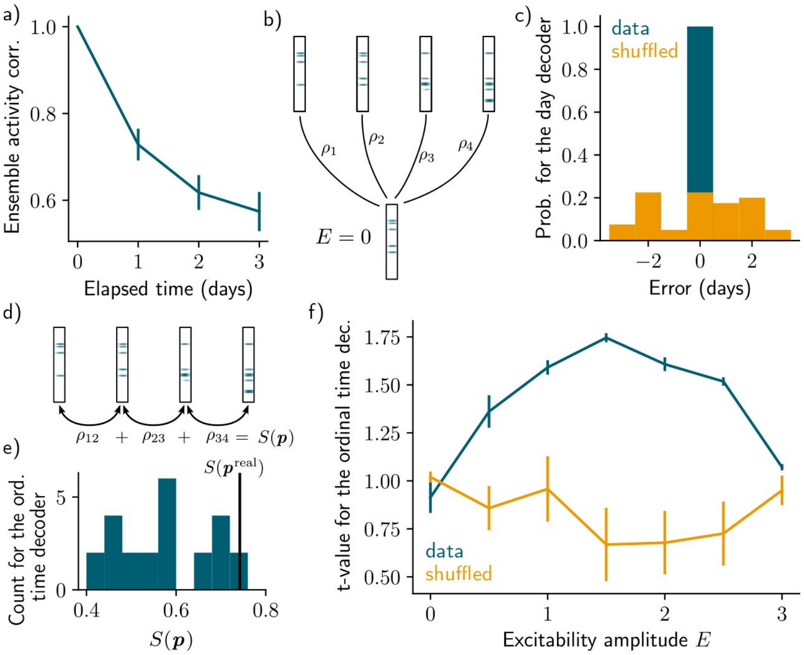


Figure 2:

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 $S(\boldsymbol{p}_{\mathrm{real}})$ 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 (23)) 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).

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).

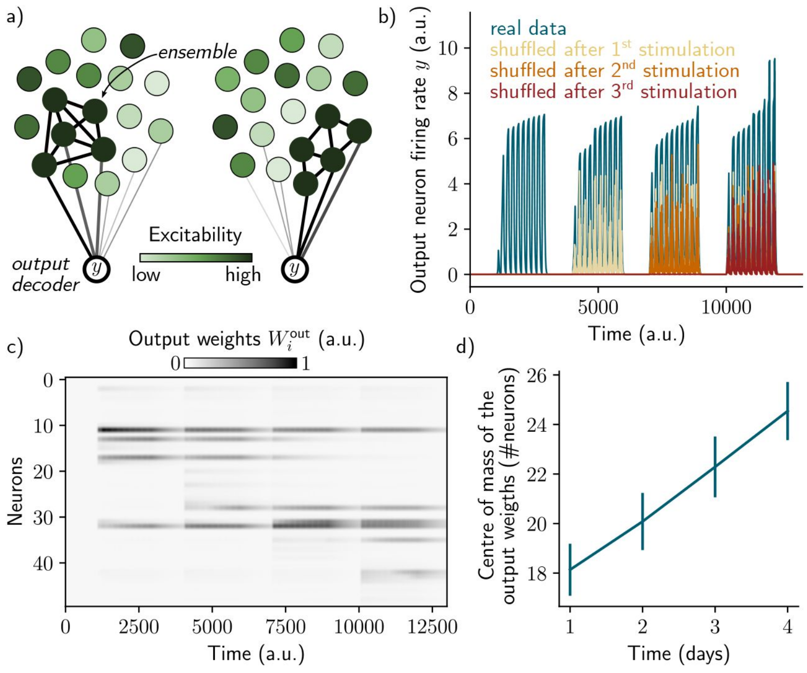


Figure 3:

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 ± s.e.m. for n = 10 simulations.

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 (23). Moreover, such gradual drifts could support complex cognitive mechanisms like mental time-travel during memory recall (23).

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 (21; 23; 24). 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; 12; 29). 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 (16), hippocampal CA1 (15) and dentate gyrus (30).

Finally, our work suggests that drift is determine both by the number of reactivations of the ensemble and the fluctuations of excitability. In particular, it is not directly related to the elapsed time between two recordings. This is consistent with growing evidence that drift is dependent more on the previous experience than on the elapsed time between different recordings (9; 23; 31). This work is also in line with the recent findings showing that fluctuations of excitability can happen for multiple reasons related to experience such as neurogenesis (9; 29; 32), sleep (33) or increase in dopamine level (34). 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:

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

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, 1] was apply to these weights. We also introduced a global inhibition term dependent on the activity of the neurons:

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

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:

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

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.

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 (26; 28) (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, modelling 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. 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 seed is used for each repetitions of the simulations.

Decoders

For each day d, we recorded the activity pattern V_d, which is a vector composed of the firing rate of the neurons 50 time steps after the beginning 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 (23): (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.

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:

$$d_{\text{inf}} = \arg\max_{d'}\{\text{corr}(V_d^0, V_{d'})\} \tag{4}$$

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

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 $\boldsymbol{p}^{\text{real}}$ = (1, 2, 3, 4) and corresponds to the real order of reactivations: day 1 → day 2 → day 3 → day 4. The sum of these correlations over the 3 pairs of consecutive days is expressed as:

$$S(\boldsymbol{p}) = \sum_{i=1}^{3} \text{corr}(V_{p_i}, V_{p_{i+1}}) \tag{5}$$

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

$$t = \frac{S(\boldsymbol{p}^{\text{real}}) - \mu}{\sigma/\sqrt{N}} \tag{6}$$

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

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

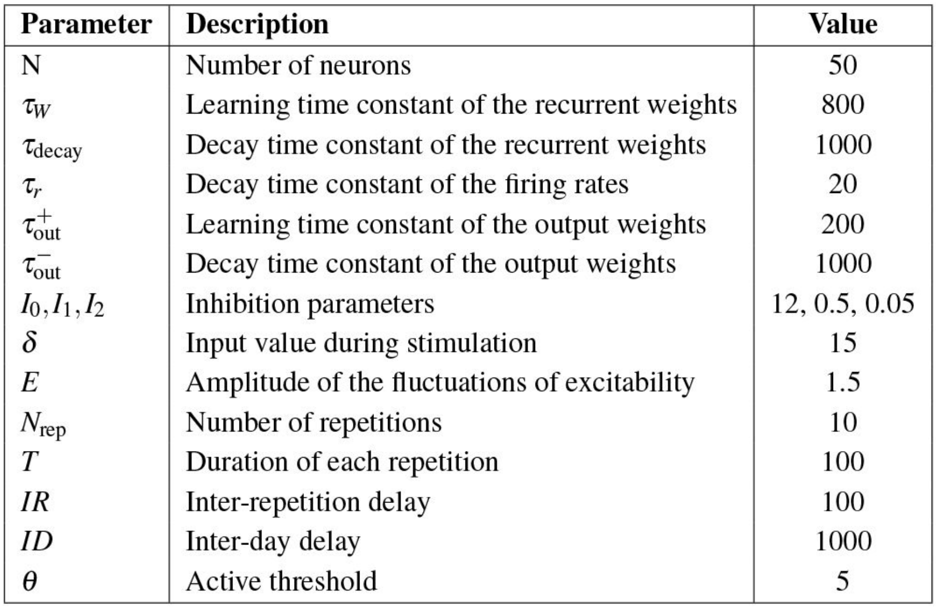
$$\frac{\mathrm{d}W_i^{\mathrm{out}}}{\mathrm{d}t} = h(\boldsymbol{W}^{\mathrm{out}}) * r_i * y/\tau_{\mathrm{out}}^{+} - W_i^{\mathrm{out}}/\tau_{\mathrm{out}}^{-} \qquad (7)$$

where τ_{out}^{+} and τ_{out}^{-} corresponds to the learning time and decay time constant, respectively. $h(\boldsymbol{W}^{\mathrm{out}})$ is a homeostatic term defined as $h(\boldsymbol{W}^{\mathrm{out}}) = 1 - \sum_{j=1}^{N} W_j^{\mathrm{out}}$ 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:

$$y = \sum_{i=1}^{N} W_i^{\mathrm{out}} r_i \qquad (8)$$

Parameters

The following parameters have been used for all simulations, when not specify otherwise. All except N are in arbitrary unit.



Parameter	Description	Value
N	Number of neurons	50
τ_W	Learning time constant of the recurrent weights	800
τ_{decay}	Decay time constant of the recurrent weights	1000
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, I_1, I_2	Inhibition parameters	12, 0.5, 0.05
δ	Input value during stimulation	15
E	Amplitude of the fluctuations of excitability	1.5
N_{rep}	Number of repetitions	10
T	Duration of each repetition	100
IR	Inter-repetition delay	100
ID	Inter-day delay	1000
θ	Active threshold	5

Supplementary

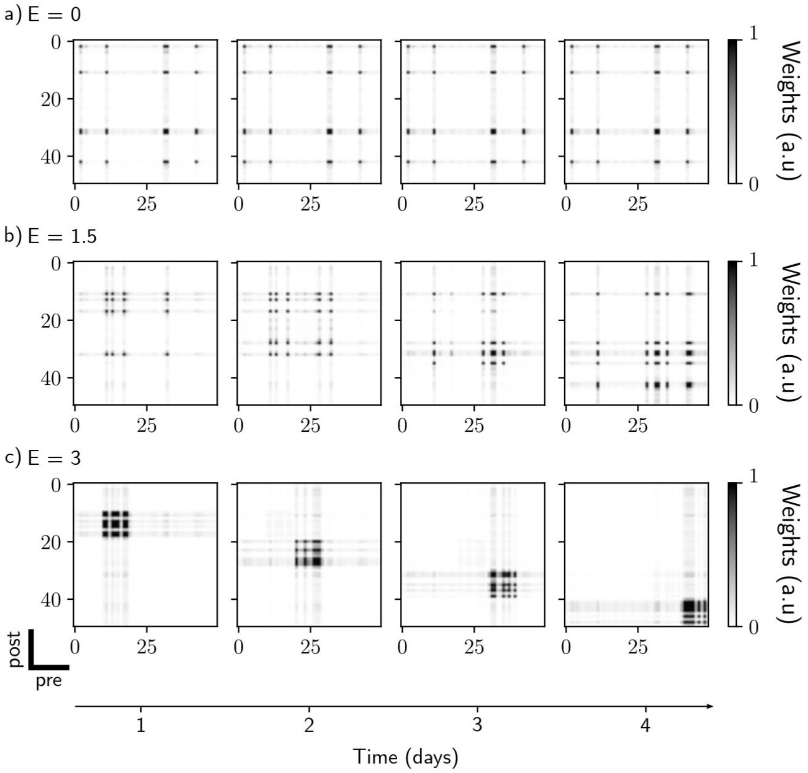


Figure S1:

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.

References

1. Driscoll L. N. , Pettit N. L. , Minderer M. , Chettih S. N. , Harvey C. D. (2017) **Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex** *Cell* **170**:986–999
https://doi.org/10.1016/j.cell.2017.07.021

2. Sadeh S. , Clopath C. , Palmer S. E. , Frank M. J. , Ziv Y. (2022) **Contribution of behavioural variability to representational drift** *eLife* **11**
https://doi.org/10.7554/eLife.77907

3. Driscoll L. N. , Duncker L. , Harvey C. D. (2022) **Representational drift: Emerging theories for continual learning and experimental future directions** *Current Opinion in Neurobiology* **76**
https://doi.org/10.1016/j.conb.2022.102609

4. Ziv Y. , et al. (2013) **Long-term dynamics of CA1 hippocampal place codes** *Nature Neuroscience* **16**:264–266
https://doi.org/10.1038/nn.3329

5. Hainmueller T. , Bartos M. (2018) **Parallel emergence of stable and dynamic memory engrams in the hippocampus** *Nature* **558**:292–296
https://doi.org/10.1038/s41586-018-0191-2

6. Spalla D. , Cornacchia I. M. , Treves A. (2021) **Continuous attractors for dynamic memories** *eLife* **10**
https://doi.org/10.7554/eLife.69499

7. Cho H.-Y. , et al. (2021) **Turnover of fear engram cells by repeated experience** *Current Biology* **31**:5450–5461
https://doi.org/10.1016/j.cub.2021.10.004

8. Rule M. E. , O'Leary T. (2022) **Self-healing codes: How stable neural populations can track continually reconfiguring neural representations** *Proceedings of the National Academy of Sciences* **119**
https://doi.org/10.1073/pnas.2106692119

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

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

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

12. Mau W. , Hasselmo M. E. , Cai D. J. (2020) **The brain in motion: How ensemble fluidity drives memoryupdating and flexibility** *eLife* **9**
https://doi.org/10.7554/eLife.63550

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

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

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

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

17. Grosmark A. D. , Buzsá ki G. (2016) **Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences** *Science* **351**:1440–1443
https://doi.org/10.1126/science.aad1935

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

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

20. Mau W. , et al. (2022) **Ensemble remodeling supports memory-updating**
https://doi.org/10.1101/2022.06.02.494530

21. Mau W. , et al. (2018) **The Same Hippocampal CA1 Population Simultaneously Codes Temporal Informa-tion over Multiple Timescales** *Current Biology* **28**:1499–1508
https://doi.org/10.1016/j.cub.2018.03.051

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

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

24. Clopath C. , Bonhoeffer T. , Hübener M. , Rose T. (2017) **Variance and invariance of neuronal longterm representations** *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**
https://doi.org/10.1098/rstb.2016.0161

25. Miller A. M. P. , Frankland P. W. , Josselyn S. A. (2018) **Memory: Ironing Out a Wrinkle in Time** *Current Biology* **28**
https://doi.org/10.1016/j.cub.2018.03.053

26. Josselyn S. A. , Tonegawa S. (2020) **Memory engrams: Recalling the past and imagining the future** *Science* **367**
https://doi.org/10.1126/science.aaw4325

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

28. Ká li S. , Dayan P. (2004) **Off-line replay maintains declarative memories in a model of hippocampalneocortical interactions** *Nature Neuroscience* **7**:286–294
https://doi.org/10.1038/nn1202

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

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

31. Rule M. E. , O'Leary T. , Harvey C. D. (2019) **Causes and consequences of representational drift** *Current Opinion in Neurobiology. Computational Neuroscience* **58**:141–147
https://doi.org/10.1016/j.conb.2019.08.005

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

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

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

Author information

Geoffroy Delamare

Bioengineering Department, Imperial College London, London SW7 2AZ, UK

For correspondence: g.delamare21@imperial.ac.uk

ORCID iD: 0000-0001-6217-4370

Yosif Zaki

Icahn School of Medicine at Mount Sinai, Department of Neuroscience, New York, New York, 10029, United States
ORCID iD: 0000-0002-8167-0182

Denise J Cai

Icahn School of Medicine at Mount Sinai, Department of Neuroscience, New York, New York, 10029, United States
ORCID iD: 0000-0002-7729-0523

Claudia Clopath

Bioengineering Department, Imperial College London, London SW7 2AZ, UK
ORCID iD: 0000-0003-4507-8648

Editors

Reviewing Editor
Lisa Giocomo
Stanford School of Medicine, United States of America

Senior Editor
Laura Colgin
University of Texas at Austin, United States of America

Reviewer #1 (Public Review):

Current experimental work reveals that brain areas implicated in episodic and spatial memory have a dynamic code, in which activity representing familiar events/locations changes over time. This paper shows that such reconfiguration is consistent with underlying changes in the excitability of cells in the population, which ties these observations to a physiological mechanism.

Delamare et al. use a recurrent network model to consider the hypothesis that slow fluctuations in intrinsic excitability, together with spontaneous reactivations of ensembles, may cause the structure of the ensemble to change, consistent with the phenomenon of representational drift. The paper focuses on three main findings from their model: (1) fluctuations in intrinsic excitability lead to drift, (2) this drift has a temporal structure, and (3) a readout neuron can track the drift and continue to decode the memory. This paper is relevant and timely, and the work addresses questions of both a potential mechanism (fluctuations in intrinsic excitability) and purpose (time-stamping memories) of drift.

The model used in this study consists of a pool of 50 all-to-all recurrently connected excitatory neurons with weights changing according to a Hebbian rule. All neurons receive the same input during stimulation, as well as global inhibition. The population has heterogeneous excitability, and each neuron's excitability is constant over time apart from a transient increase on a single day. The neurons are divided into ensembles of 10 neurons each, and on each day, a different ensemble receives a transient increase in the excitability of each of its neurons, with each neuron experiencing the same amplitude of increase. Each day for four days, repetitions of a binary stimulus pulse are applied to every neuron.

The modeling choices focus in on the parameter of interest-the excitability-and other details are generally kept as straightforward as possible. That said, I wonder if certain aspects may be overly simple. The extent of the work already performed, however, does serve the intended purpose, and so I think it would be sufficient for the authors to comment on these choices rather than to take more space in this paper to actually implement these choices. What might happen were more complex modeling choices made? What is the justification for the choices that are made in the present work?

The two specific modeling choices I question are (1) the excitability dynamics and (2) the input stimulus. The ensemble-wide synchronous and constant-amplitude excitability increase, followed by a return to baseline, seems to be a very simplified picture of the dynamics of intrinsic excitability. At the very least, justification for this simplified picture would benefit the reader, and I would be interested in the authors' speculation about how a more complex and biologically realistic dynamics model might impact the drift in their network model. Similarly, the input stimulus being binary means that, on the single-neuron level, the only type of drift that can occur is a sort of drop-in/drop-out drift; this choice excludes the possibility of a neuron maintaining significant tuning to a stimulus but changing its preferred value. How would the use of a continuous input variable influence the results.

Result (1): Fluctuations in intrinsic excitability induce drift
The two choices highlighted above appear to lead to representations that never recruit the neurons in the population with the lowest baseline excitability (Figure 1b: it appears that only 10 neurons ever show high firing rates) and produce networks with very strong

bidirectional coupling between this subset of neurons and weak coupling elsewhere (Figure 1d). This low recruitment rate need may not necessarily be problematic, but it stands out as a point that should at least be commented on. The fact that only 10 neurons (20% of the population) are ever recruited in a representation also raises the question of what would happen if the model were scaled up to include more neurons.

Result (2): The observed drift has a temporal structure
The authors then demonstrate that the drift has a temporal structure (i.e., that activity is informative about the day on which it occurs), with methods inspired by Rubin et al. (2015). Rubin et al. (2015) compare single-trial activity patterns on a given session with full-session activity patterns from each session. In contrast, Delamare et al. here compare full-session patterns with baseline excitability (E = 0) patterns. This point of difference should be motivated. What does a comparison to this baseline excitability activity pattern tell us? The ordinal decoder, which decodes the session order, gives very interesting results: that an intermediate amplitude E of excitability increase maximizes this decoder's performance. This point is also discussed well by the authors. As a potential point of further exploration, the use of baseline excitability patterns in the day decoder had me wondering how the ordinal decoder would perform with these baseline patterns.

Result (3): A readout neuron can track drift
The authors conclude their work by connecting a readout neuron to the population with plastic weights evolving via a Hebbian rule. They show that this neuron can track the drifting ensemble by adjusting its weights. These results are shown very neatly and effectively and corroborate existing work that they cite very clearly.

Overall, this paper is well-organized, offers a straightforward model of dynamic intrinsic excitability, and provides relevant results with appropriate interpretations. The methods could benefit from more justification of certain modeling choices, and/or an exploration (either speculative or via implementation) of what would happen with more complex choices. This modeling work paves the way for further explorations of how intrinsic excitability fluctuations influence drifting representations.

Reviewer #2 (Public Review):

In this computational study, Delamare et al identify slow neuronal excitability as one mechanism underlying representational drift in recurrent neuronal networks and that the drift is informative about the temporal structure of the memory and when it has been formed. The manuscript is very well written and addresses a timely as well as important topic in current neuroscience namely the mechanisms that may underlie representational drift.

The study is based on an all-to-all recurrent neuronal network with synapses following Hebbian plasticity rules. On the first day, a cue-related representation is formed in that network and on the next 3 days it is recalled spontaneously or due to a memory-related cue. One major observation is that representational drift emerges day-by-day based on intrinsic excitability with the most excitable cells showing highest probability to replace previously active members of the assembly. By using a day-decoder, the authors state that they can infer the order at which the reactivation of cell assemblies happened but only if the excitability state was not too high. By applying a read-out neuron, the authors observed that this cell can track the drifting ensemble which is based on changes of the synaptic weights across time. The only few questions which emerged and could be addressed either theoretically or in the discussion are as follows:

1. Would the similar results be obtained if not all-to-all recurrent connections would have been molded but more realistic connectivity profiles such as estimated for CA1 and CA3?
2. How does the number of excited cells that could potentially contribute to an engram influence the representational drift and the decoding quality?
3. How does the rate of the drift influence the quality of readout from the readout-out neuron?

Reviewer #3 (Public Review):

The authors explore an important question concerning the underlying mechanism of representational drift, which despite intense recent interest remains obscure. The paper explores the intriguing hypothesis that drift may reflect changes in the intrinsic excitability of neurons. The authors set out to provide theoretical insight into this potential mechanism.

They construct a rate model with all-to-all recurrent connectivity, in which recurrent synapses are governed by a standard Hebbian plasticity rule. This network receives a global input, constant across all neurons, which can be varied with time. Each neuron also is driven by an "intrinsic excitability" bias term, which does vary across cells. The authors study how activity in the network evolves as this intrinsic excitability term is changed.

They find that after initial stimulation of the network, those neurons where the excitability term is set high become more strongly connected and are in turn more responsive to the input. Each day the subset of neurons with high intrinsic excitability is changed, and the network's recurrent synaptic connectivity and responsiveness gradually shift, such that the new high intrinsic excitability subset becomes both more strongly activated by the global input and also more strongly recurrently connected. These changes result in drift, reflected by a gradual decrease across time in the correlation of the neuronal population vector response to the stimulus.

The authors are able to build a classifier that decodes the "day" (i.e. which subset of neurons had high intrinsic excitability) with perfect accuracy. This is despite the fact that the excitability bias during decoding is set to 0 for all neurons, and so the decoder is really detecting those neurons with strong recurrent connectivity, and in turn strong responses to the input. The authors show that it is also possible to decode the order in which different subsets of neurons were given high intrinsic excitability on previous "days". This second result depends on the extent by which intrinsic excitability was increased: if the increase in intrinsic excitability was either too high or too low, it was not possible to read out any information about past ordering of excitability changes.

Finally, using another Hebbian learning rule, the authors show that an output neuron, whose activity is a weighted sum of the activity of all neurons in the network, is able to read out the activity of the network. What this means specifically, is that although the set of neurons most active in the network changes, the output neuron always maintains a higher firing rate than a neuron with randomly shuffled synaptic weights, because the output neuron continuously updates its weights to sample from the highly active population at any given moment. Thus, the output neuron can readout a stable memory despite drift.

Strengths:
The authors are clear in their description of the network they construct and in their results. They convincingly show that when they change their "intrinsic excitability term", upon stimulation, the Hebbian synapses in their network gradually evolve, and the combined synaptic connectivity and altered excitability result in drifting patterns of activity in response to an unchanging input (Fig. 1, Fig. 2a). Furthermore, their classification analyses

(Fig. 2) show that information is preserved in the network, and their readout neuron successfully tracks the active cells (Fig. 3). Finally, the observation that only a specific range of excitability bias values permits decoding of the temporal structure of the history of intrinsic excitabililty (Fig. 2f and Figure S1) is interesting, and as the authors point out, not trivial.

Weaknesses:

1. The way the network is constructed, there is no formal difference between what the authors call "input", Δ(t), and what they call "intrinsic excitability" Ɛ_i(t) (see Equation 3). These are two separate terms that are summed (Eq. 3) to define the rate dynamics of the network. The authors could have switched the names of these terms: Δ(t) could have been considered a global "intrinsic excitability term" that varied with time and Ɛ_i(t) could have been the external input received by each neuron i in the network. In that case, the paper would have considered the consequence of "slow fluctuations of external input" rather than "slow fluctuations of intrinsic excitability", but the results would have been the same. The difference is therefore semantic. The consequence is that this paper is not necessarily about "intrinsic excitability", rather it considers how a Hebbian network responds to changes in excitatory drive, regardless of whether those drives are labeled "input" or "intrinsic excitability".

2. Given how the learning rule that defines input to the readout neuron is constructed, it is trivial that this unit responds to the most active neurons in the network, more so than a neuron assigned random weights. What would happen if the network included more than one "memory"? Would it be possible to construct a readout neuron that could classify two distinct patterns? Along these lines, what if there were multiple, distinct stimuli used to drive this network, rather than the global input the authors employ here? Does the system, as constructed, have the capacity to provide two distinct patterns of activity in response to two distinct inputs?

Impact:
Defining the potential role of changes in intrinsic excitability in drift is fundamental. Thus, this paper represents a potentially important contribution. Unfortunately, given the way the network employed here is constructed, it is difficult to tease apart the specific contribution of changing excitability from changing input. This limits the interpretability and applicability of the results.