Drift of CA1 Context Codes Across Days

Population similarity and cross-day decoding in public miniscope recordings from Zaki et al. (2025)

Tanmay Hinge
hingetanmay@gmail.com
github.com/tanmayhinge/ca1-context-drift

September 2026

Abstract

Hippocampal CA1 codes change from day to day even when the world stays the same. I asked how much of a context remains readable in CA1 population activity days later, using the public chronic one-photon calcium imaging of DANDI:000718, which holds 2 mice, up to five sessions across six days and three contexts. Cells were followed across days with the authors' CellReg tables where those exist and with a footprint matcher elsewhere, and the matcher recovers 98.5 to 100 percent of CellReg's pairs. Ensemble overlap and population similarity exceeded chance for every session pair while failing to separate same-context pairs from different-context pairs. Similarity fell steadily as the gap between two sessions grew (Spearman -0.817, $p = 0.006$ by shuffling, 10 pairs). A decoder trained on day 1 against day 3 scored 0.497 on held-out days, at chance.

1 A six-day, three-context recording in two mice

A mouse returning to a familiar room recognises it, yet the CA1 cells representing that room partly turn over between visits, a phenomenon called representational drift (Ziv et al., 2013; Rule et al., 2019). How much of a context code is still readable days later? I used the public dataset of Zaki et al. (2025), in which mice explored a neutral context on day 1, received three foot shocks in a different context on day 3, and returned on days 4, 5 and 6 to the shock context, a novel context and the neutral context. CA1 was imaged with a head-mounted miniscope, segmented with Minian (Dong et al., 2022), and freezing scored from video with ezTrack (Pennington et al., 2019).

The public subset holds 2 mice, and they are not replicates. The shock amplitudes stored in the files are 1.5 mA for Ca-EEG2-1 and 0.25 mA for Ca-EEG3-4, placing them in the two opposite groups of the original study, and only Ca-EEG3-4 has all five behavioural sessions. Results are therefore reported per animal and never averaged. Files run to 2 to 5 GB each because they carry the raw movie, so they were read over HTTP with byte-range requests, caching only the processed traces, footprints and behaviour, around 100 MB.

2 Following the same cells, and the measures applied to them

Following cells across days. Ca-EEG2-1 ships with CellReg tables (Sheintuch et al., 2017), while Ca-EEG3-4, the more informative animal, has none. I aligned each session's reference image to day 1 by phase correlation, scored nearby cell pairs by the intersection over union of their outlines, and took the best one-to-one assignment. Because one mouse has both, the matcher was scored against CellReg before being trusted on the other.

Ensembles, similarity and PCA. The planned fixed activity threshold was abandoned, because Minian keeps only cells it detected and any fixed cut therefore labels 84 to 100 percent of them active. An ensemble is instead the most active 25 percent of cells in a session, repeated at

10, 25 and 50 percent, with chance overlap from the hypergeometric distribution and intervals from resampling cells. For similarity, each session gives one activity rate per matched cell and those vectors are correlated between sessions, by Pearson and by rank, against a baseline that shuffles which cell is which. Sessions were cut to equal length, and the conditioning session contributes only its two minutes before the first shock, since activity after a shock reflects the shock and says nothing about the room. PCA was run on ten-second population vectors. Separation between two sessions is the standardised difference between their time bins along the straight line joining their means, so larger values put the sessions further apart. It was recomputed after shuffling session labels, because with fewer time bins than cells a fitted direction separates any two groups.

Decoding and behaviour. A logistic regression on one-second bins of matched cells was trained on neutral day 1 against conditioning day 3, with equal numbers of bins per class and cross-validation over contiguous time blocks. Those two sessions fall on different days, so context is confounded with day inside the training set and a high score there could reflect either. The later sessions break that confound and carry the evidence below. Freezing and ezTrack motion were measured over the same windows, and every similarity was recomputed after discarding frozen frames.

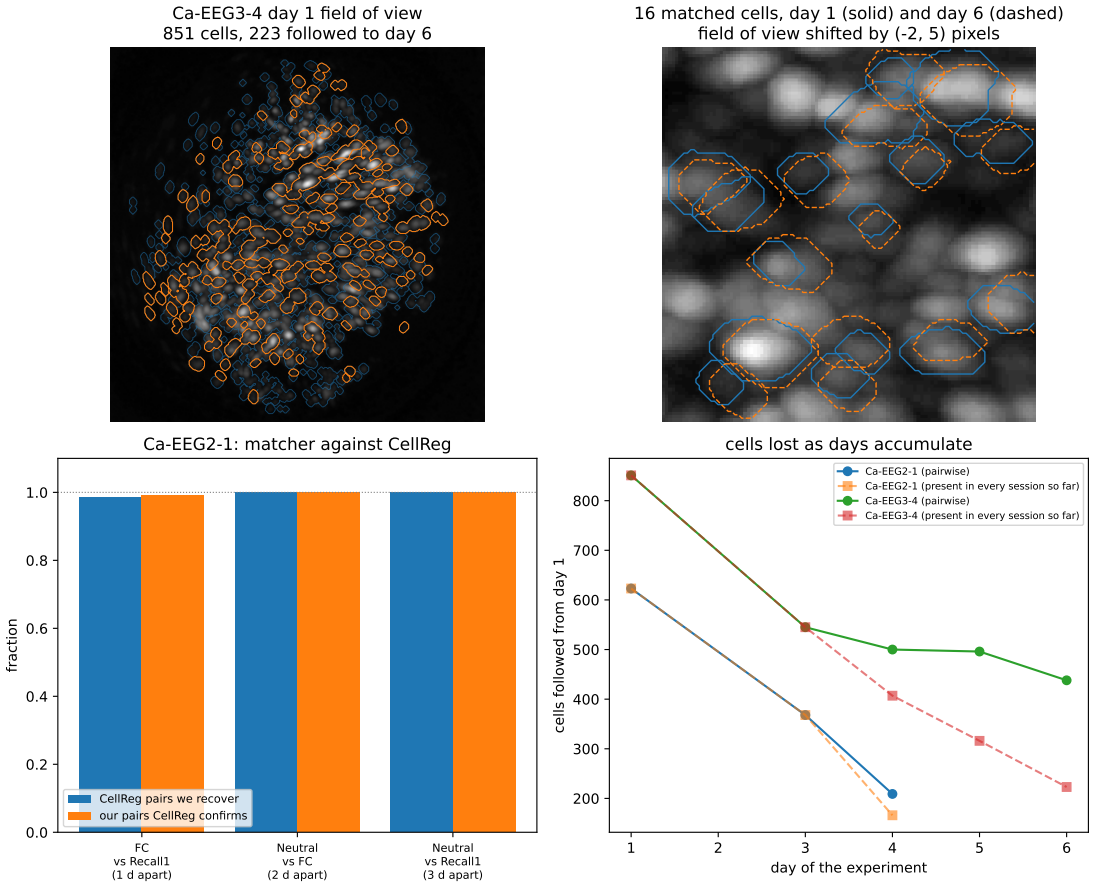

Figure 1: Cells can be followed across days, and the matcher can be checked. The first panel shows the day 1 field of view of Ca-EEG3-4, with the 223 cells followed to day 6 outlined. The second shows matched cells on day 1 (solid) and day 6 (dashed) after aligning the two fields of view. The third reports that on Ca-EEG2-1, where CellReg exists, the matcher recovers 98.5 to 100 percent of its pairs while CellReg confirms 99.2 to 100 percent of ours. The fourth shows cells being lost steadily as days accumulate.

3 The code drifts with elapsed time and does not identify the context

Cross-day matching is reliable over the gaps where it could be checked. On Ca-EEG2-1 the matcher recovers 98.5 to 100 percent of the pairs that at least 10 of the archive's 19 independent CellReg runs agree on, and CellReg confirms 99.2 to 100 percent of ours among cells it tracked (Figure 1). It returns roughly twice as many pairs as that consensus. Only the 85 consensus cells are independently confirmed, and the extra pairs, which CellReg never rejected, are used here on that weaker footing. On this basis the matcher follows 166 cells across the three sessions of Ca-EEG2-1 and 223 across all five sessions and six days of Ca-EEG3-4.

Overlap and similarity exceed chance and match across contexts. Ensemble overlap runs from 1.351 to 2.276 times chance, and population correlations from 0.23 to 0.572 against a shuffled band of ±0.148 (Figure 2), so cross-day matching carries real signal. The diagnostic comparison fails. For Ca-EEG3-4, returning to the *same* neutral room after five days gives $r = 0.304$ (95% interval 0.179 to 0.428), which sits at the level of a *novel* room after four days (0.295, 0.176 to 0.405) and below the shock room after three (0.392). Ensemble overlap agrees, at 1.636 times chance for the same context (1.219 to 2.107) against 1.707 for the novel one (1.295 to 2.185), two intervals that overlap almost entirely.

Elapsed time predicts similarity across all ten session pairs. Over the 10 session pairs of Ca-EEG3-4, similarity falls as the gap between two sessions grows, with Spearman −0.817 and $p = 0.006$ from 10000 shuffles of which similarity belongs to which gap. That correlation is the basis for reading these results as drift in time. The PCA agrees. Its first two components hold only 7.1 percent of the variance for Ca-EEG3-4 and 13.6 percent for Ca-EEG2-1, and Ca-EEG3-4 needs 26 components for half of it, so distances in such a plot are untrustworthy and the sessions overlap there. Separation along a fitted direction is large for every pair, while shuffled session labels already give 1.996 to 3.226. What survives that control orders by elapsed time, its largest excess of 2.286 belonging to the two visits to the *same* room five days apart.

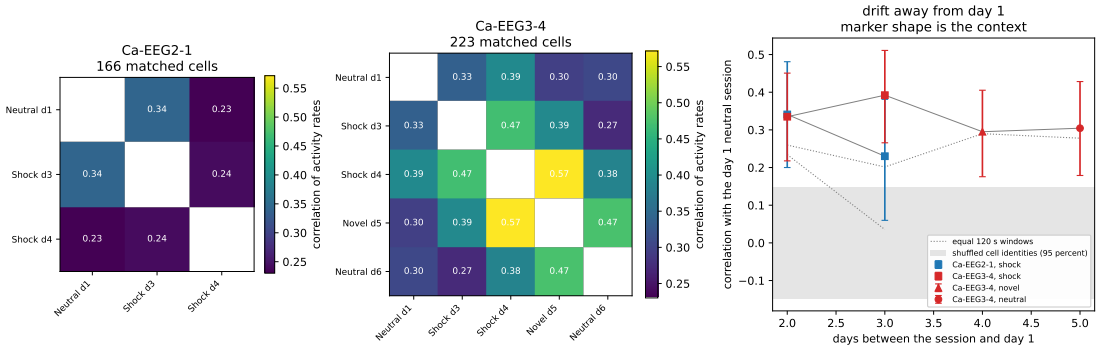

Figure 2: Population similarity follows the calendar and ignores the context. Each matrix gives the correlation between per-cell activity rates for every session pair, per animal, over matched cells. The right panel plots similarity to the day 1 neutral session against days elapsed, with marker shape giving the context, a grey band for the shuffled baseline and a dotted line repeating the curve with every session cut to 120 s. The day 6 return to the neutral context sits at the level of the day 5 novel context.

The decoder separates the training days and stays at chance on new ones. Within the two training days, balanced accuracy is 0.738 and 0.587 for the two mice, against an empirical chance of 0.498 and 0.503 from shuffled labels (Figure 3). Those two sessions fall on different days, so that score is compatible with a code for the context and with a code for the day. Slow drift inside a single recording is ruled out, since telling the first half of a session from its second half gives 0.453 to 0.573 across all eight sessions. On days the decoder never saw, its answers collapse onto one class. For Ca-EEG3-4 it calls shock in 3.7 percent of the day 4 shock-context

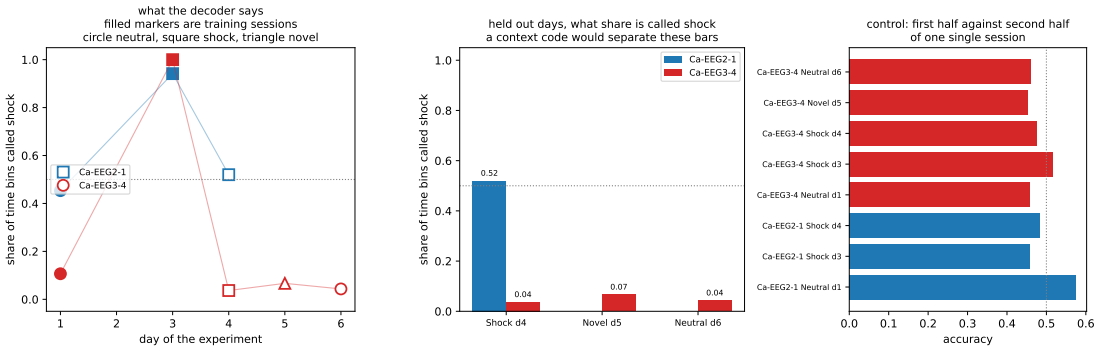

Figure 3: A context learned on one day is not recognised on later days. The left panel gives the share of each session's time bins called "shock", with training sessions filled. The middle panel repeats that share for the held-out days alone, where a context code would put the shock day high and the neutral day low. The right panel is the control, telling the first half of a single session from its second half, which sits at chance for all eight sessions.

bins, 6.7 percent of the day 5 novel bins and 4.3 percent of the day 6 neutral bins, so the two days with a correct answer pool to a balanced accuracy of 0.497, at chance. Day 4 of Ca-EEG2-1 gives 52 percent, also at chance.

Behaviour separates the two mice and leaves the neural results standing. Ca-EEG2-1 froze 63.5 percent of its return to the shock context against 10.5 percent in the neutral context on day 1, while Ca-EEG3-4 reached 8.5 percent there, consistent with the shock-intensity difference of Zaki et al. (2025) with one animal per condition (freezing figure in the repository). Recomputing every similarity from moving frames only changes it by at most 0.081, and behavioural and neural change are related in the intuitive direction without being resolvable ($\rho = -0.455$, $p = 0.187$, 10 pairs).

4 What these two mice cannot tell us

2 mice from two shock conditions is a case study, and only one has the three-context design, so the central negative result rests on 223 cells in a single mouse. The matcher was validated only over the session gaps Ca-EEG2-1 provides, which reach three days, while the comparison carrying the argument spans five, and matching quality may fall over that longer gap. The ten pairs behind the Spearman correlation come from five sessions and so share sessions with one another, which is why its p value comes from shuffling. Sessions run five to ten minutes with no position tracking, so a place code could be present and invisible to session-averaged measures. The ensemble size is a choice forced by a fixed threshold labelling nearly every cell active, and the direction of the result holds at three sizes while its magnitude does not. Two-dimensional projections distort distances, which is why nothing rests on the PCA plot. Failed transfer also cannot distinguish absent information from information a linear decoder on one-second bins is unable to read.

5 The offline recordings are the obvious next analysis

The analysis this dataset was built for comes next, whether the neutral and shock ensembles co-reactivate during the homecage rest recordings, already cross-registered to the encoding sessions. Aligning activity to what the animal was doing at each moment, using all of the offline recordings with their EEG-scored sleep states, and comparing the observed drift against a sequence model trained on the same context sequence would each test whether elapsed time is genuinely the organising variable.

6 Where the code, figures and data live

Code, figures and numbers sit at github.com/tanmayhinge/ca1-context-drift. Every figure regenerates from `scripts/` and every number here is read from `results/summary.json` through generated macros, so no value is typed by hand. The data are DANDI:000718 (Zaki et al., 2026), used under CC-BY-4.0 and never redistributed here.

References

Zhe Dong, William Mau, et al. Minian, an open-source miniscope analysis pipeline. *eLife*, 11:e70661, 2022. doi: 10.7554/eLife.70661.

Zachary T. Pennington, Zhe Dong, et al. eztrack: An open-source video analysis pipeline for the investigation of animal behavior. *Scientific Reports*, 9:19979, 2019. doi: 10.1038/s41598-019-56408-9.

Michael E. Rule, Timothy O'Leary, and Christopher D. Harvey. Causes and consequences of representational drift. *Current Opinion in Neurobiology*, 58:141–147, 2019. doi: 10.1016/j.conb.2019.08.005.

Liron Sheintuch, Alon Rubin, et al. Tracking the same neurons across multiple days in Ca^{2+} imaging data. *Cell Reports*, 21(4):1102–1115, 2017. doi: 10.1016/j.celrep.2017.10.013.

Yosif Zaki, Zachary T. Pennington, et al. Offline ensemble co-reactivation links memories across days. *Nature*, 637:145–155, 2025. doi: 10.1038/s41586-024-08168-4.

Yosif Zaki, Zachary T. Pennington, and Denise J. Cai. Offline ensemble co-reactivation links memories across days (dandi:000718, version 0.260825.1902). DANDI Archive, 2026. URL `https://dandiarchive.org/dandiset/000718`.

Yaniv Ziv, Laurie D. Burns, et al. Long-term dynamics of CA1 hippocampal place codes. *Nature Neuroscience*, 16(3):264–266, 2013. doi: 10.1038/nn.3329.