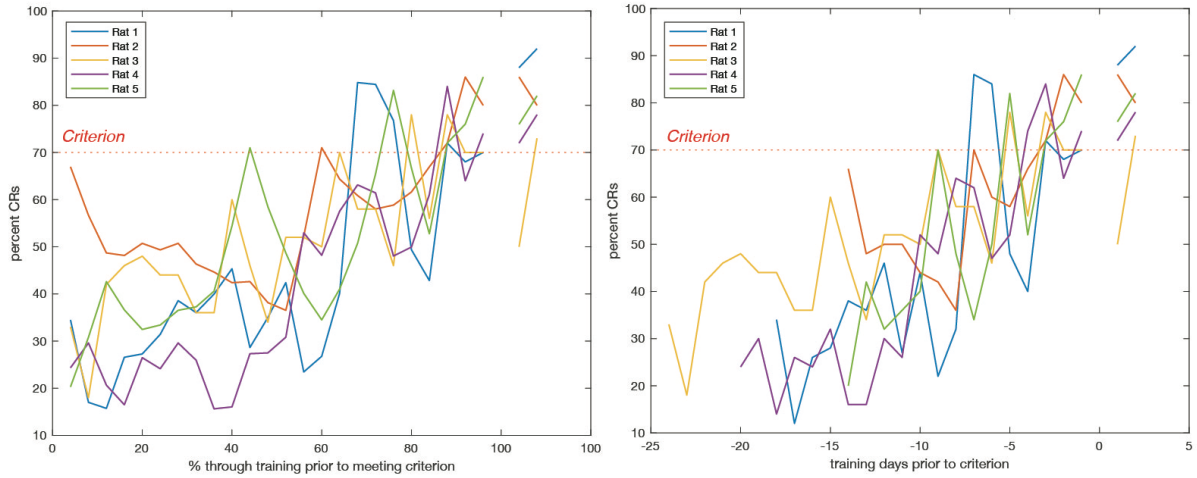
## Supplementary figures

Figure S1

**Figure S1. Photos of testing chambers.**

Photos of testing chambers. Top: Environment A, an unscented rectangular enclosure with wire floor and walls, and white lighting. Bottom: Environment B, a scented ovular enclosure with white solid floor and walls, and red lighting. Both environments were located at the same spot in the room relative to external cues. The door to the chamber was closed during testing, to accentuate the distinction between the white and red lighting.

Figure S2




1362

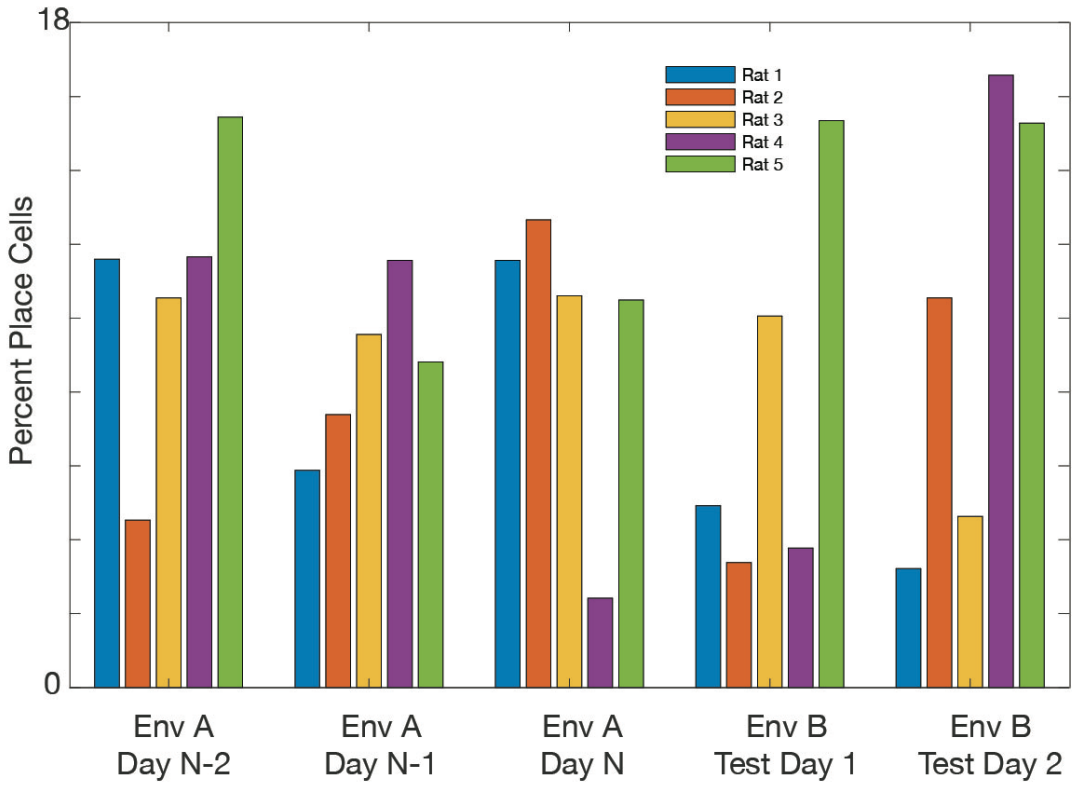
1363 **Figure S2. Learning curves for the five rats.**

1364 Substantial variability in the number of sessions required to learn the task; average number of

1365 sessions was 20 ± 4.2 (including the criterion sessions). The fastest learners (2 rats) reached

1366 criterion after 14 sessions, while the slowest rat required 24 sessions to reach criterion.

Figure S3




1367

1368 **Figure S3. Percent of place cells by session for each animal.**

1369 The percentage of cells classified as place cells for each session and each animal

Figure S4

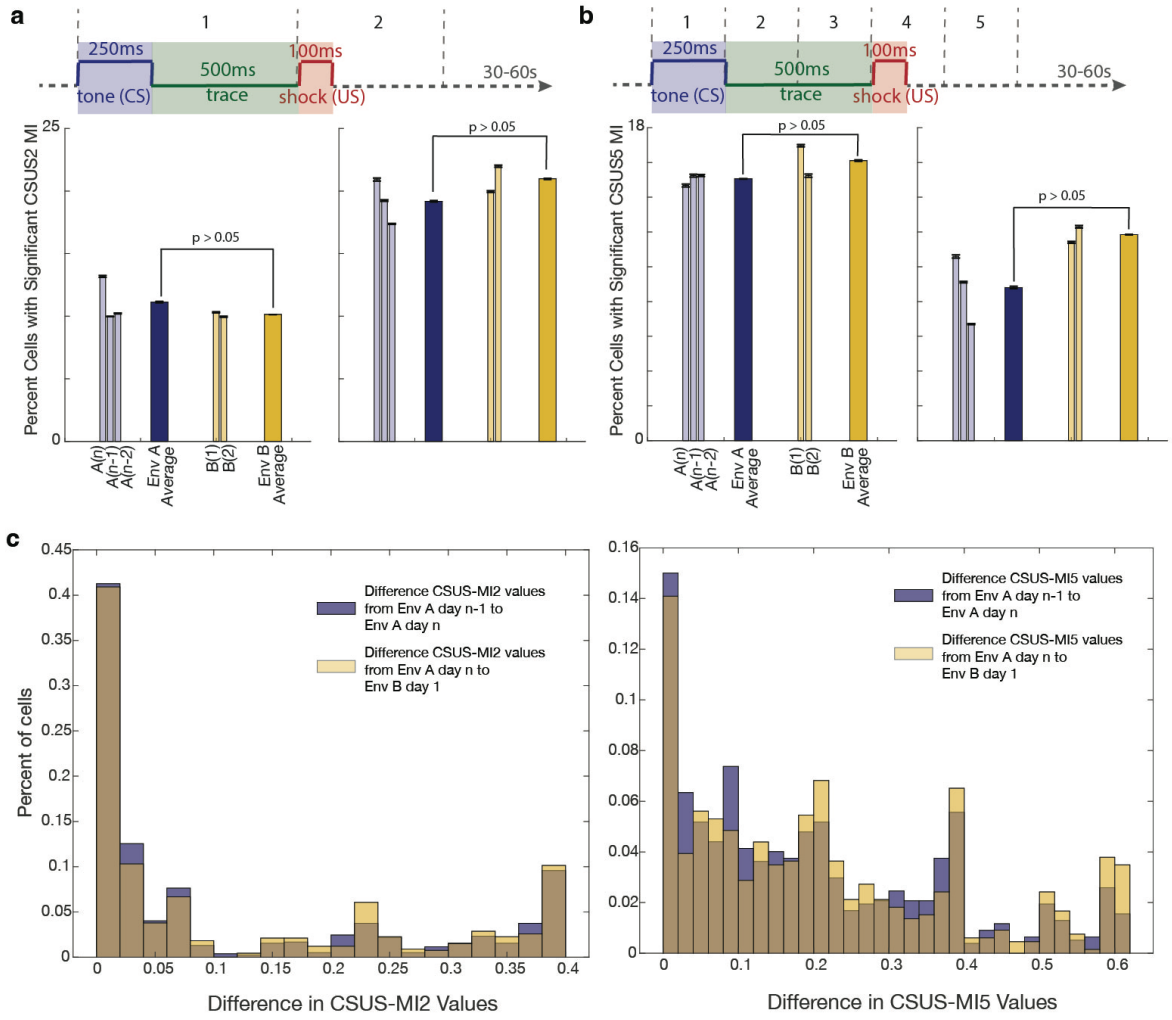



**Figure S4. CSUS-MI2 and CSUS-MI5 differences between sessions.**

**a.** The trial period was divided into two segments: the CS and trace period (750 ms) and the US and post-US period (500 ms). Mutual information (MI) was calculated for cells based on these two periods and compared to shuffled data, where period IDs were shuffled 500 times across all trials. Left: Using calcium event data, we found that 10.7% ± 4.9% of cells contained significant CSUS information related to whether the animal was in a CS or US period. Right: Using calcium traces, 19.9% ± 8.2% of cells contained significant information distinguishing the CS from the US period. No significant differences in

CSUS-MI were observed between environments A and B (double-sided t-tests, calcium events: $t(23) = 0.48$, $p > 0.05$; calcium traces: $t(23) = -0.52$, $p > 0.05$).

**b.** The trial period was divided into five equal segments, each 250 ms. MI was calculated for each cell based on these five periods. We compared the observed values to those obtained after shuffling period IDs 500 times. Left: Using calcium event data, 15.5% ± 7.8% of cells contained significant information distinguishing the five periods, compared to 10.0% ± 7.8% when using calcium trace data. No significant differences in these MI metrics were found between environments A and B (double-sided t-tests, calcium events: $t(23) = -0.32$, $p > 0.05$; calcium traces: $t(23) = -1.1$, $p > 0.05$).

**c.** Left: There was no significant difference in CSUS-MI2 values when comparing session A(n) to session A(n-1) versus session A(n) to session B(1) (Wilcoxon rank sum test: $p > 0.05$; double-sided t-test: $t(1431) = 0.86$, $p > 0.05$). Right: A small but significant difference was observed in CSUS-MI5 when comparing session A(n) to session A(n-1) versus session A(n) to session B(1) (Wilcoxon rank sum test: $p = 0.049$; double-sided t-test: $t(1431) = -2.2$, $p = 0.03$).

Figure S5

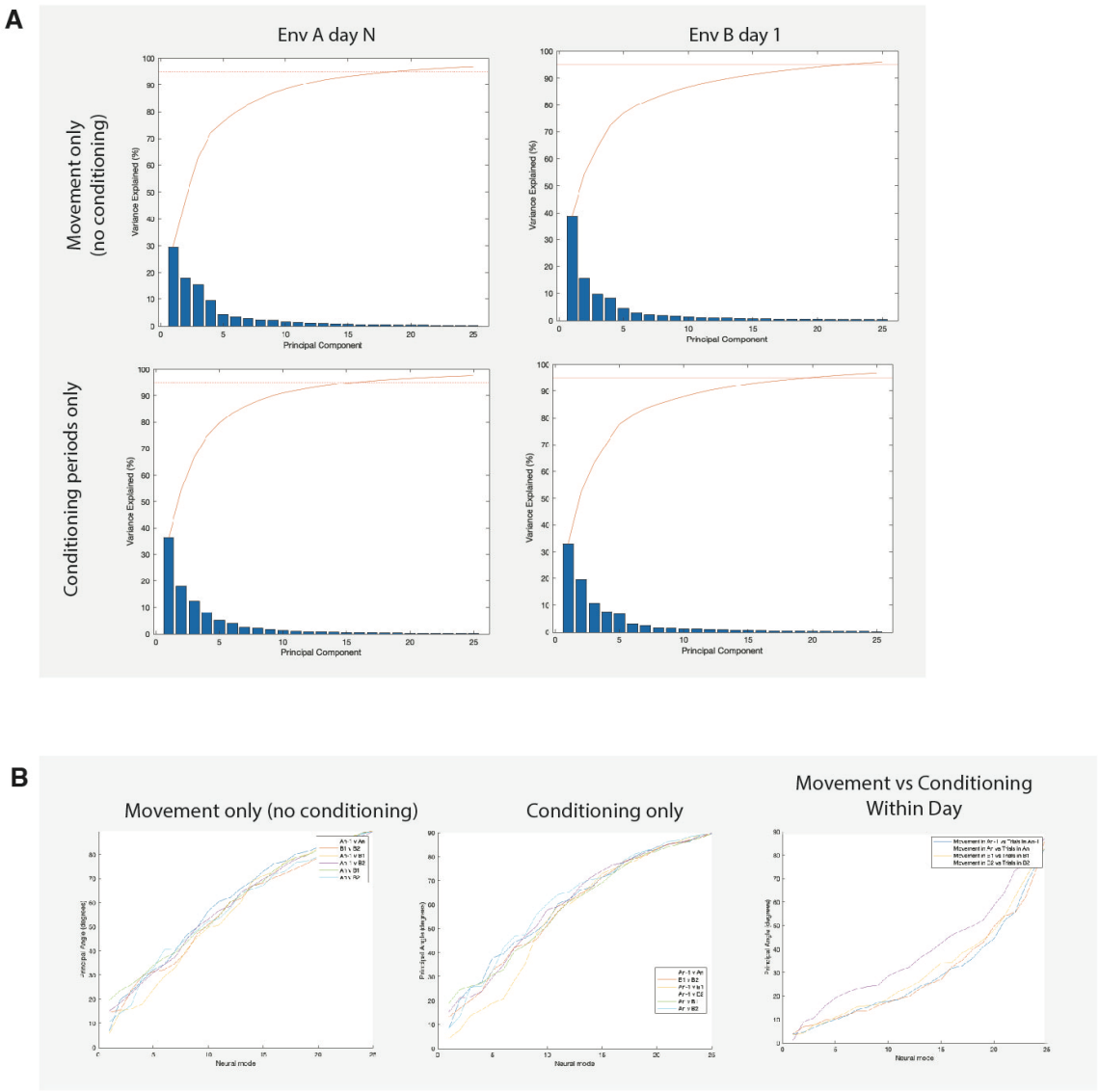
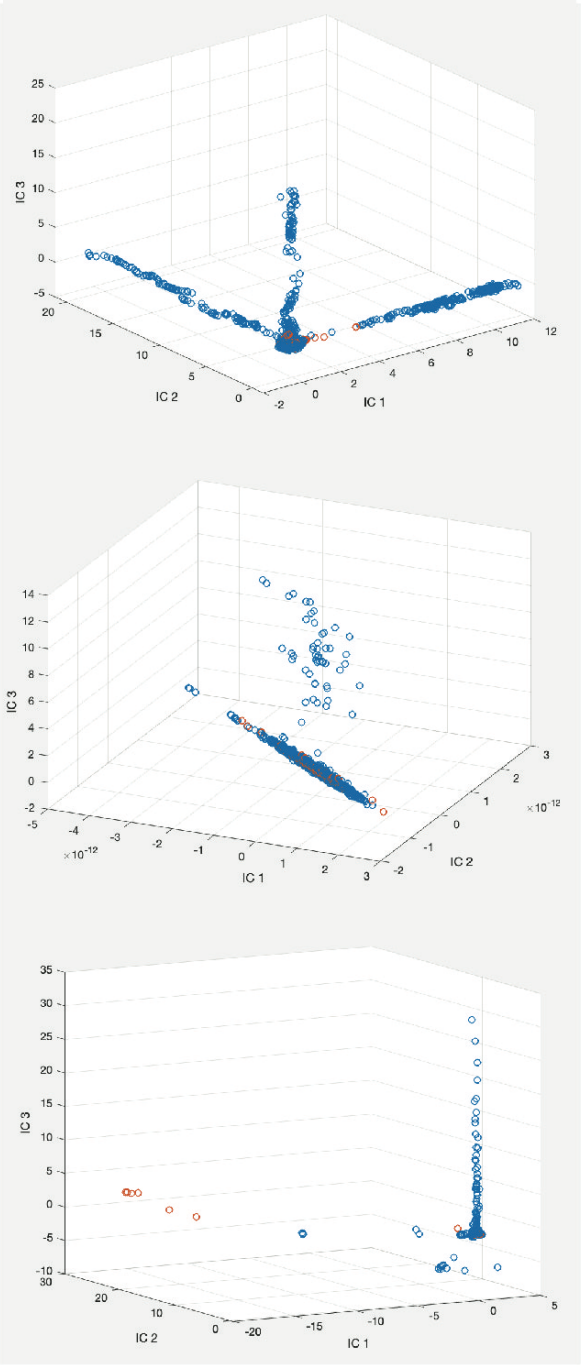



**Figure S5. PCA computations for session A(n) and session B(1).**

Only cells present in both sessions were used. Principal component analysis (PCA) revealed that approximately 15-25 principal components (PCs) are needed to account for 95% of the variance in the data. When using the complete cell population (not shown), more than 25 PCs are required to achieve the same variance. Across and within all sessions and representations (spatial and task), the principal angles between manifolds remain highly similar.

Figure S6




1402

1403 **Figure S6. ICA computations across different segments of a session.**

1404 Top: Independent component analysis (ICA) was for data over an entire session, using three

1405 independent components (ICs). Blue dots represent non-trial times, while red dots represent trial

1406 times. Middle: ICA computed over the last two-thirds of the same session shows variability in

ICs across segments within a session. Bottom: ICA computed over the second half of the session shows additional variability in the results of the analysis depending on how the session is divided. These results indicate that independent components are highly variable over the course of one session and are sensitive to how the session is partitioned.

Figure S7

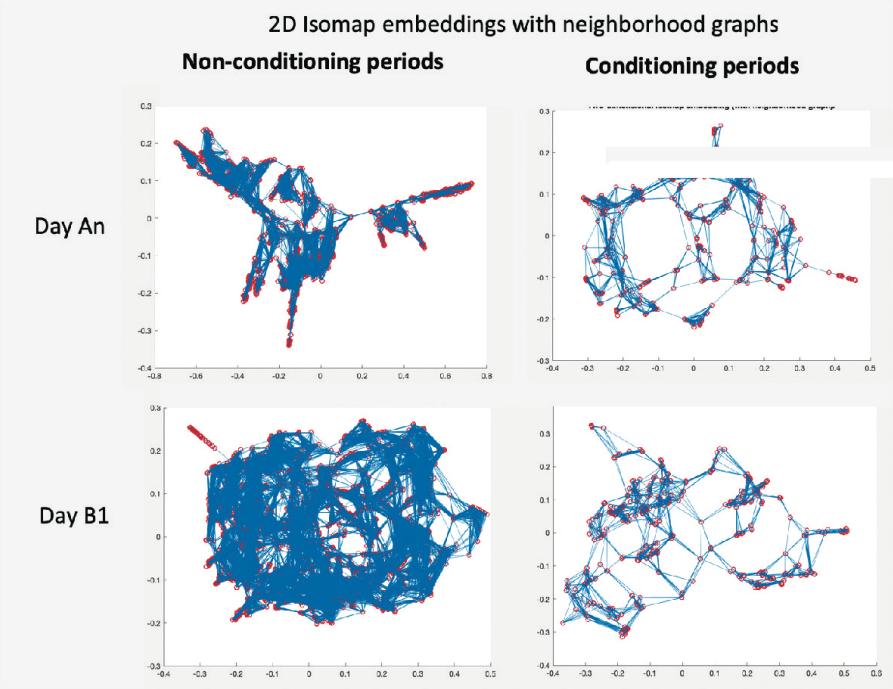



**Figure S7. Isomap computations for session A(n) and session B(1).**

Isomap computations suggest that approximately five neural modes are sufficient to achieve a residual variance of 5-10%. However, the shape of the Isomap embedding does not correlate with any discernable properties of neural activity or behavior, suggesting limited interpretability of the embedding structure.

Figure S8

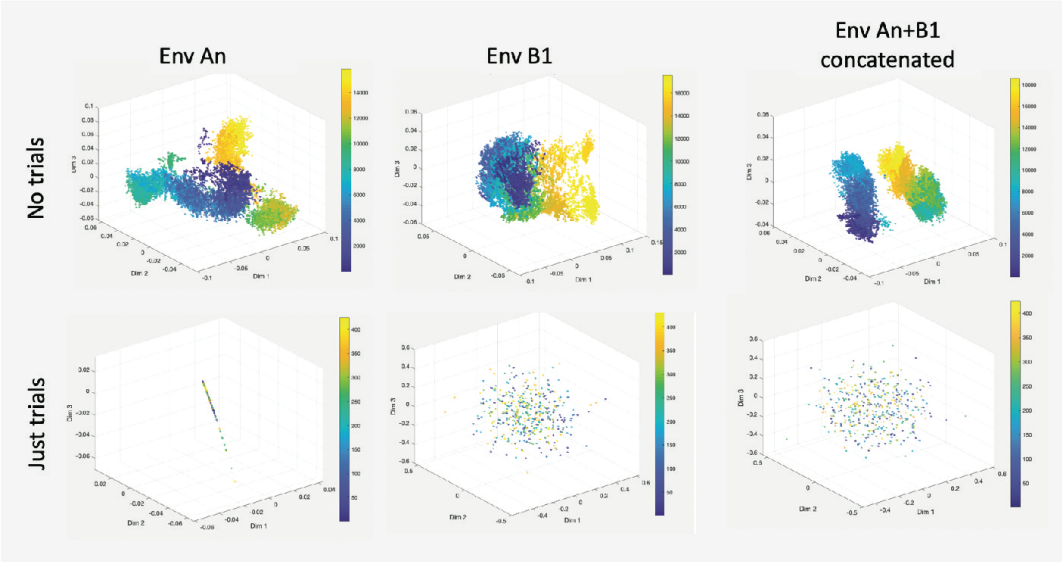



**Figure S8. MIND outputs for sessions A(n), B(1), and concatenated sessions.**

Top row: MIND embeddings during movement, excluding trial periods, with color bars representing frames. The temporal structure of the data is well captured, with clear separation between A(n) and B(1). Bottom row: MIND embeddings during conditioning periods are highly unstable. Small changes in parameters result in substantial shifts in the embedding structure, transitioning from a linear structure (left) to an undefined, unstable cloud (middle and right).

Figure S9

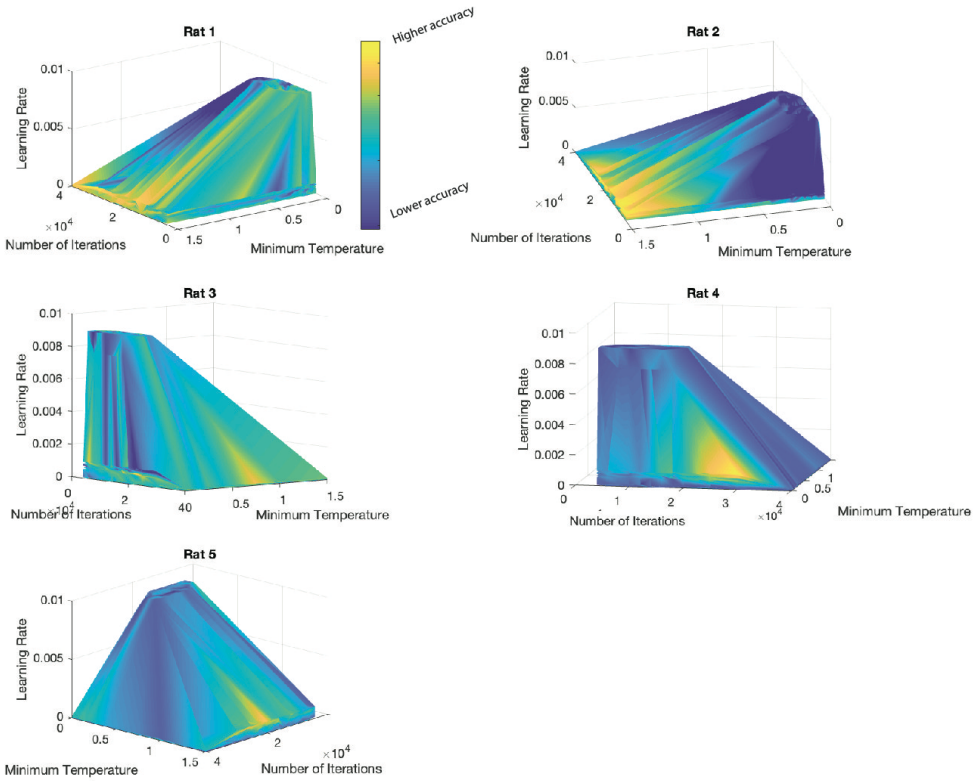
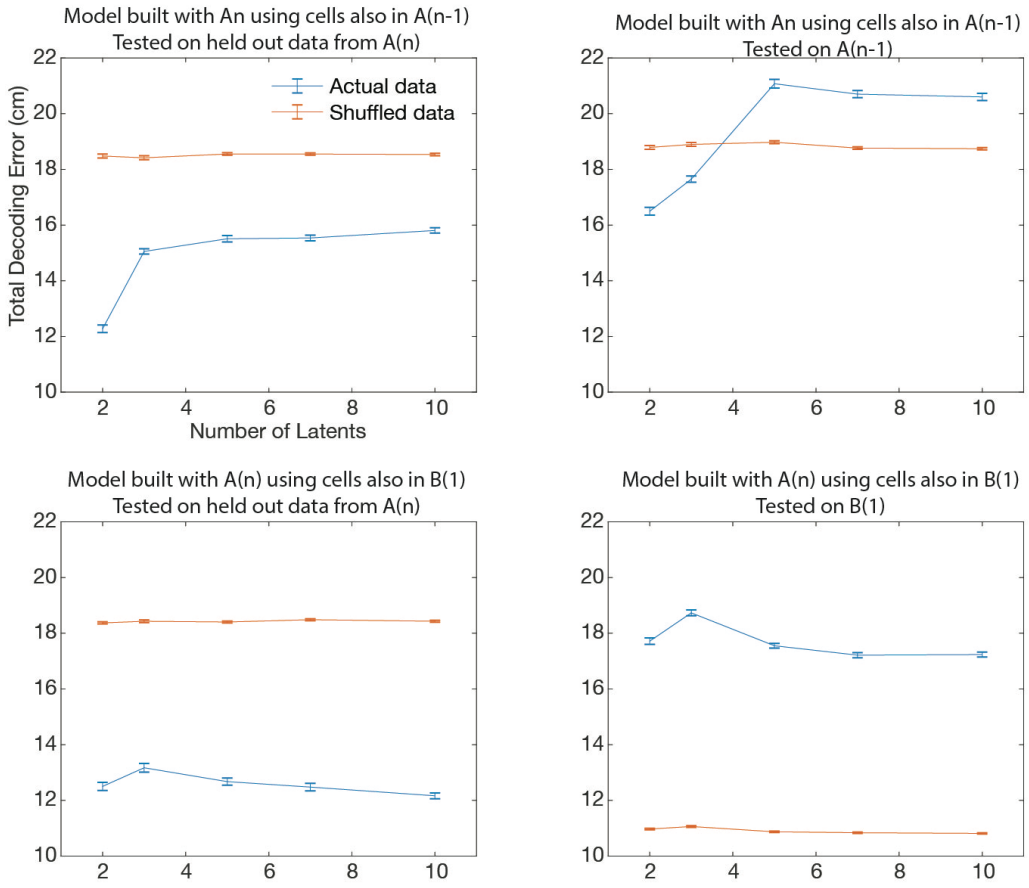



**Figure S9. Grid search over decoding parameters for position.**

A grid search was performed over three parameters: minimum temperature, learning rate, and number of iterations, for decoding position. Models were trained using cells from session A(n) that also appeared in session A(n-1). The figure shows decoding accuracy for session A(n-1) using the models trained on data from session A(n). Yellow areas indicate higher decoding accuracy.

Figure S10




1433

1434 **Figure S10. Position decoding error as a function of the number of latents (Rat 5).**

1435 This figure shows the decoding error for position as the number of latents increases. Left panels:

1436 model trained on data from A(n), tested on held out data from A(n). Upper left panel: As the

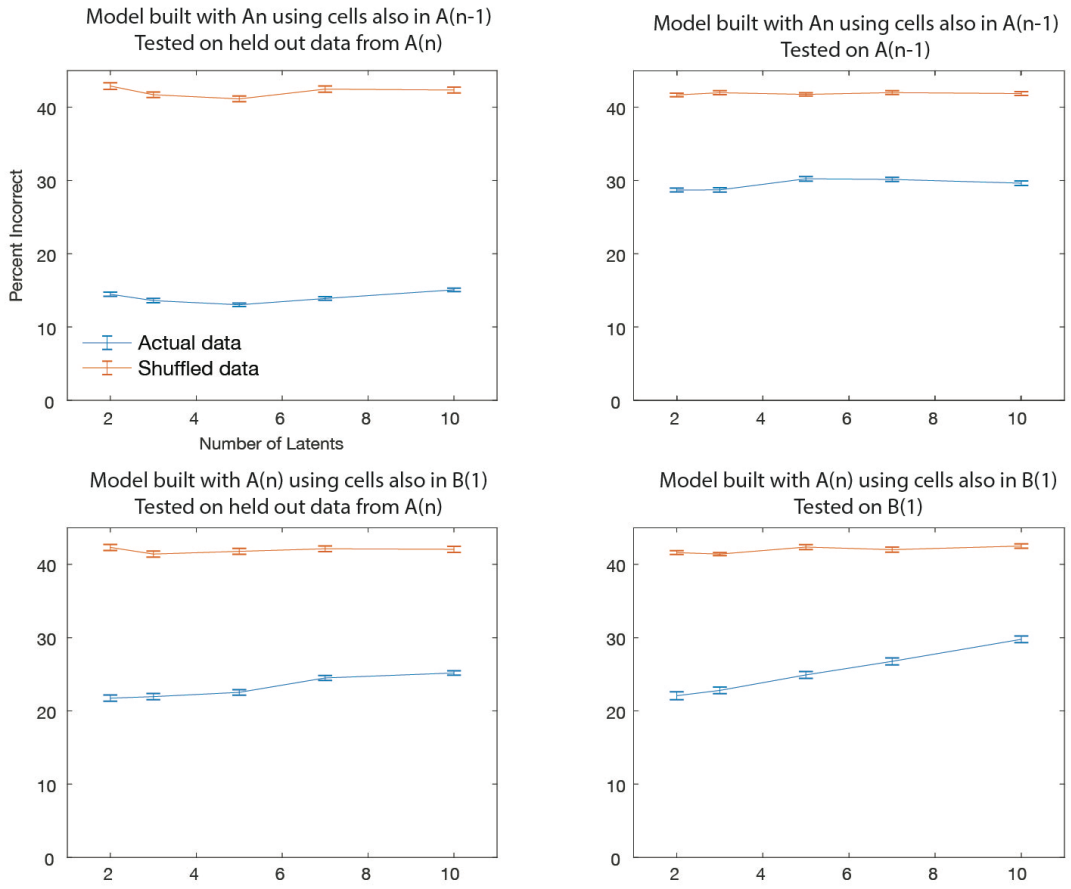
1437 number of latents increases, the model's ability to decode a different session within the same

1438 environment decreases. This effect is not consistent across all rats; see lower left panel. Right

1439 panels: model trained on data from A(n), tested on A(n-1) (top) or B(1) (bottom). Performance is

1440 particularly bad when the model is tested on a different environment (lower right panel).

Figure S11




1441

1442 **Figure S11. CSUS2 decoding accuracy with increasing number of latents (Rat 3).**

1443 Percent of incorrect decoding for the CSUS2 model as the number of latents increases. Left

1444 panels: model trained on data from A(n), tested on held out data from A(n). As the number of

1445 latents increases, the model's ability to decode a different session within the same environment

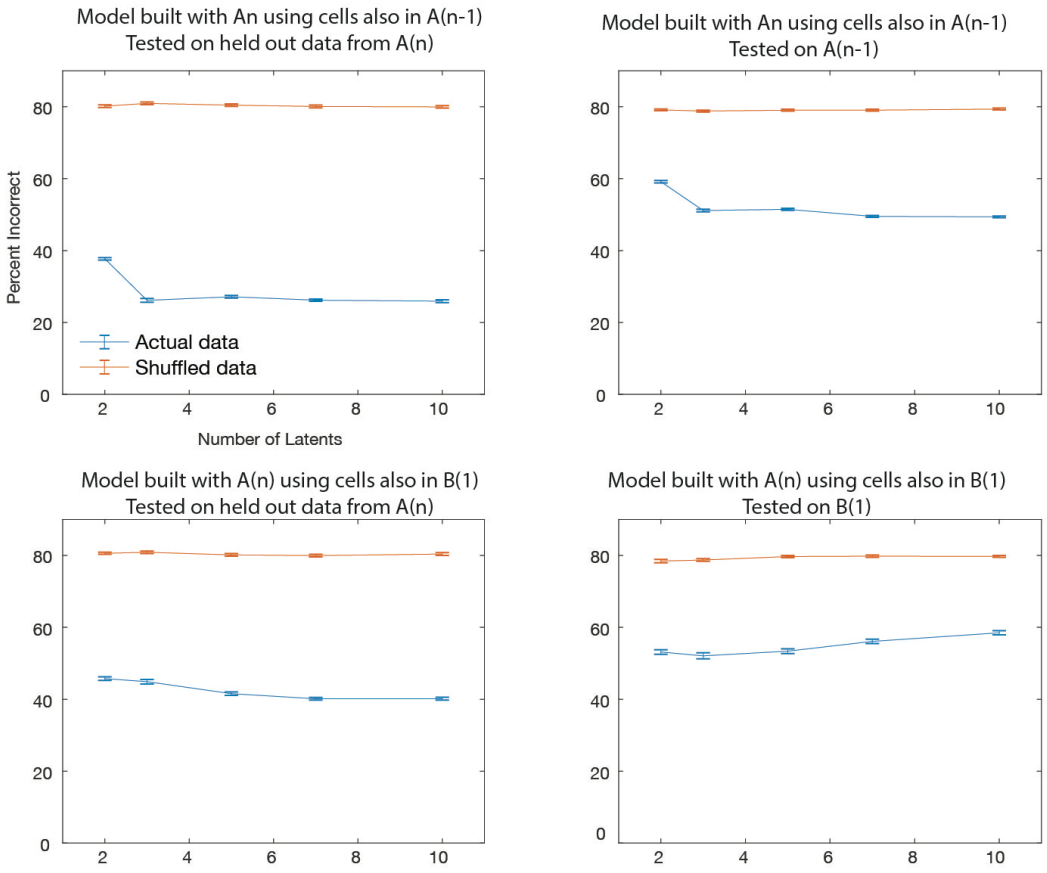
1446 remains stable or slightly decreases. Right panels: model trained on data from A(n), tested on

1447 A(n-1) (top) or B(1) (bottom). Performance remains stable for a different session in the same

1448 environment (upper right panel) but deteriorates with increasing number of latents when the

1449 model is tested in a different environment (lower right panel). Each model was run 100 times.

Figure S12




1450

1451 **Figure S12. CSUS5 decoding accuracy with increasing number of latents (Rat 5).**

1452 Same as Figure S11, but for CSUS5; the conditioning period has been divided into five time bins

1453 instead of two. The percent of incorrect decoding is shown as the number of latents increases.

1454 Left panels: model trained on data from A(n), tested on held out data from A(n). As the number

1455 of latents increases, the model's ability to decode a different session within the same environment

1456 remains stable or slightly increases. Right panels: model trained on data from A(n), tested on

1457 A(n-1) (top) or B(1) (bottom). Performance remains stable for a different session in the same

1458 environment (upper right panel) but deteriorates with increasing number of latents when the

1459 model is tested in a different environment (lower right panel). Each model was run 100 times.

1460

Figure S13

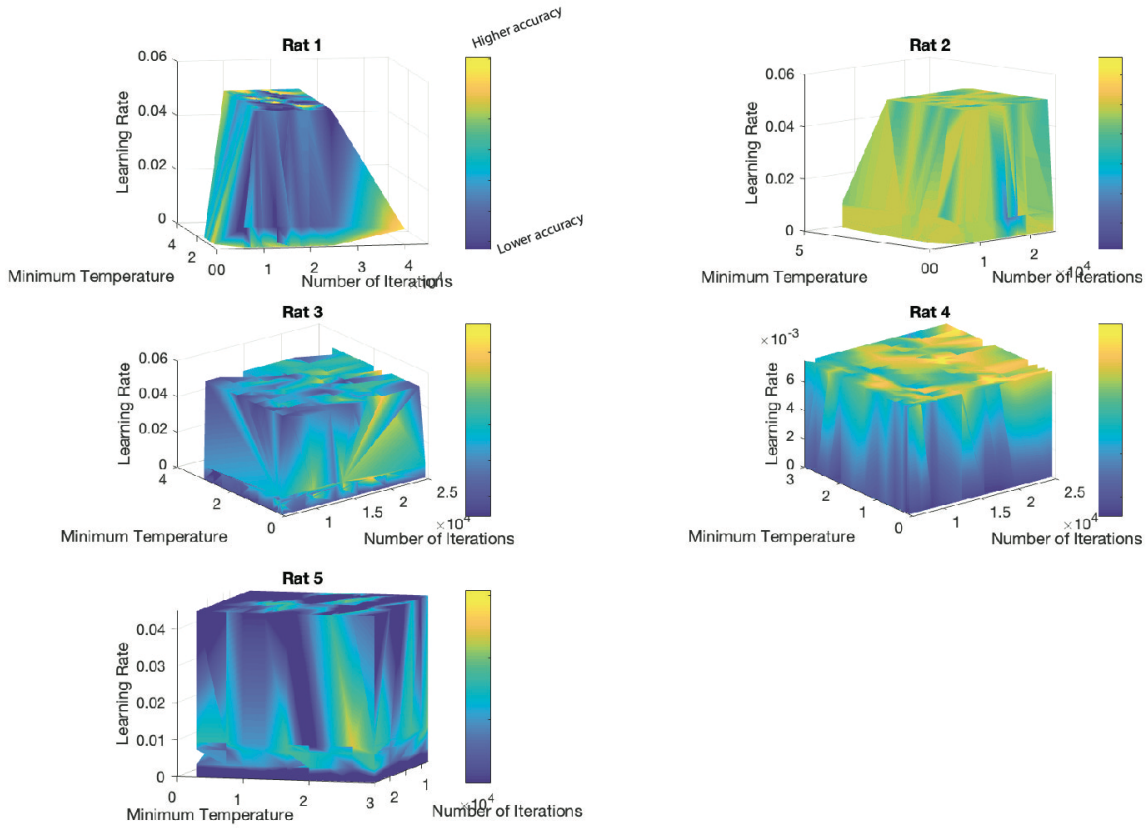



**Figure S13. Grid search over decoding parameters for conditioning.**

A grid search was performed over three parameters: minimum temperature, learning rate, and number of iterations for conditioning decoding. Models were trained using cells from session A(n) that also appeared in session B(1). The figure shows decoding accuracy for CSUS2 session B(1) using the models trained on data from session A(n). Yellow areas indicate higher decoding accuracy. For Rats 1 and 4, the 'Euclidean' distance with 'constant' temperature mode was used. For Rats 2 and 5, 'cosine' distance with 'constant' temperature mode was used. For Rat 3, 'cosine' distance with 'auto' temperature mode was used.