

How reward and experience shape neural population dynamics in mouse visual cortex

John Jairo Madrid-Carvajal¹ , Rohit Jeswanth^{2†} , Ishatpreet Singh^{3†}, Katja Kaurinkoski^{4†} , and Dimitra Maoutsa⁵ 

¹Institute of Cognitive Science, University of Osnabrück, Germany, ²Cognitive Science Lab, International Institute of Information Technology, Hyderabad, India, ³Centre for Neuroscience, Indian Institute of Science, Bengaluru, India, ⁴HiLIFE-Neuroscience Center, University of Helsinki, Finland, ⁵Neuromatch Inc.

[†]These authors contributed equally to this work.


Abstract

Learning reshapes cortical activity, but it remains unclear whether population-level changes primarily reflect exposure to sensory statistics or reward-driven assignment of behavioral relevance. We analysed calcium imaging data from mice exposed to the same visual environment with or without reward, and used tensor component analysis to separate within-trial dynamics from across-trial learning-related structure. Rewarded learning produced a distinct reorganisation of population geometry. After learning, fewer components were sufficient to reconstruct neural activity, and variance became more concentrated in the dominant modes. This compression was not merely a non-specific reduction in variability, but was task-aligned: the leading variance-explaining components also carried strong stimulus discriminability. These results suggest that reward does not simply enhance sensory selectivity, but reorganises visual cortical population geometry so that behaviorally relevant stimulus dimensions are embedded in the dominant modes of population activity.

Keywords unsupervised learning, reward-driven learning, tensor component analysis, population activity

Published Jun 10, 2026

Source Repository

 [impact-scholars/madrid-carvajal-2026-reward-shapes-geometry](https://github.com/impact-scholars/madrid-carvajal-2026-reward-shapes-geometry)

Correspondence

John Jairo Madrid-Carvajal, Dimitra Maoutsa

Acknowledgements

This work was supported by the Impact Scholars Program. General-purpose large language models were used for grammar and writing corrections, and for limited code assistance (such as handling error messages and for parallelising and speeding up computations).

Data Availability

Published via Impact Scholars; original development repository.

A central challenge in systems neuroscience is that learning-related changes in cortical activity can arise either from unsupervised extraction of sensory regularities or from reward-driven assignment of behavioral relevance (Barlow, 1989; Botvinick *et al.*, 2020; Hebb, 2005). Passive sensory experience may reshape cortical circuits by repeatedly exposing them to the statistical structure of the environment (Bell & Sejnowski, 1997; Fiser *et al.*, 2010; Olshausen & Field, 1996; Simoncelli & Olshausen, 2001). Reward-driven learning, by contrast, does not merely expose the animal to stimulus regularities, but also assigns behavioral utility to particular sensory events (Poort *et al.*, 2015; Schultz, 2015; Schultz *et al.*, 1997). These two forms of learning are therefore computationally distinct: statistical learning extracts regularities from the environment, while rewarded learning uses those regularities to maximize reward (Ferrari *et al.*, 2026; Henschke *et al.*, 2020; Pakan *et al.*, 2018). Whether they produce similar or distinct changes in cortical population dynamics remain unresolved.

Zhong *et al.* (2025) showed that changes in visual cortical selectivity that accompany task learning are also reproduced in mice exposed to the same stimulus statistics without reward, suggesting that repeated sensory experience can account for a substantial fraction of learning-related plasticity. Henschke *et al.* (2020) found that reward-stimulus pairing also enhanced the reliability of the responses of selective neurons, improving the stimulus-specific structure of the representation. Yet, changes in single-neuron selectivity do not necessarily imply comparable changes in population dynamics. Rewarded and unrewarded learning could result in comparable numbers of selective neurons, while producing different representational geometries, with differences in how variance is distributed across dimensions, whether dominant activity modes align with stimulus-discriminating directions, or whether population structure becomes organized around other task-relevant variables ([Supplementary Figure 1](#)). This distinction is important because it changes how we interpret learning-related plasticity: as a better internal model of sensory structure, as a reward-driven reorganization of cortical activity around behaviorally relevant variables, or as an interaction between these processes. Here, we compare how rewarded and unrewarded learning alter cortical population geometry.

To determine how reward-driven learning and repeated sensory experience differentially shape visual cortical activity, we analysed the dataset from Zhong *et al.* (2025). In this experiment, mice ran through linear virtual-reality corridors displaying two types of naturalistic patterns (leaves or circles) presented in random order. Each leaf or circle trial consisted of a 4 m textured corridor followed by 2 m of uniform grey “inter-trial” space before the next corridor began. While the mouse traversed the corridor, a sound cue was delivered at a random position between 0.5 and 3.5 m from corridor entrance. In the rewarded cohort, water-restricted mice could obtain a 2.5 μ l water reward after the cue, but only in the rewarded (leaf) corridor and only if they licked the spout. In the unrewarded cohort, mice experienced the same corridors, trial structure, and sound cues without receiving any water reward ([Figure 1 A](#), top-left). We analysed calcium imaging recordings from the primary visual cortex and adjacent higher visual areas from both cohorts (3 mice each) ([Figure 1 A](#), top-center). Recordings were obtained during the first exposure session (“before learning”) and again after approximately two weeks of daily VR training or exposure (“after learning”), once rewarded mice had learned the task.

To separate temporal dynamics within a trial from learning dynamics across trials, we employed tensor component analysis (TCA) Williams *et al.* (2018). Since trials had unequal durations, we restricted the analysis to fixed windows of interest each of 3s length: one around the stimulus onset and one around a randomly timed sound cue that preceded reward delivery and was present in both cohorts. After regressing out mouse speed, we applied nonnegative TCA to the z-scored residual calcium activity in

each selected time window. The respective tensors (neurons \times time \times trials tensors; [Figure 1 A](#), top right) were constructed using either the full recorded population from each session ([Figure 1 B–D](#)) or fixed-size subsamples from each cortical area separately ($N_{sub} = 3000$; 5 subsamples per area) ([Figure 1 E](#) and [Supplementary Figure 2](#)). To estimate the rank of the decomposition of each tensor, we computed the reduction in mean reconstruction error between consecutive ranks, $\Delta(R) = err(R) - err(R + 1)$, and selected the smallest rank at which $\Delta(R)$ fell below half the maximum drop across the range ($\alpha = 0.5$); while requiring that all replicates had factor-similarity above 0.97. Because TCA components are non-orthogonal and arbitrarily ordered, we reordered them by greedy forward selection against the original data tensor, by sequentially adding the component that most reduced the residual error of the original tensor. This resulted in an incremental variance-explained order for comparison across mice, sessions, and cohorts.

Reward-driven learning results in a lower dimensional rank decomposition. Nonnegative TCA indicated that after learning, the variance in the rewarded cohort was consistently best captured by an estimated rank of 3 components at tunnel entrance before and after learning, while 4 components are required for the unrewarded cohort ([Figure 1 B](#)). At sound cue the decomposition after learning resulted in a decrease of one rank for the rewarded cohort. Across all 10 fitted ranks, error curves showed a decrease in reconstruction error for the rewarded cohort after learning, while it remained almost invariant for the unrewarded cohort. Averaged across all sessions and alignments, TCA rank estimated a ($\bar{R} = 3.00$) for the rewarded cohort and it was significantly lower than the unrewarded cohort ($\bar{R} = 4.17$; Mann–Whitney $U = 6.0$, $p = 0.028$, $r = 0.67$), whereas before learning the two groups did not differ significantly ($p = 0.34$). This reduction in rank was accompanied by a decrease in reconstruction error: total variance explained (VE) increased from 0.913 to 0.928 in the rewarded group after learning, while it remained stable in the unrewarded group (~ 0.926). Moreover, variance was more concentrated in the leading components for the rewarded cohort, with the first two components capturing 90.1% of total VE after learning compared to 78.6% in the unrewarded group ($U = 31$, $p = 0.041$, $r = -0.72$). Per-area TCA on size-matched neural populations showed that the increase in variance explained, together with the reduction in estimated rank, could not be explained by differences in the number of recorded neurons across sessions. This observation was supported by both a variance analysis conducted independently per-area and the consistent reduction in the participation ratio of the per-area neural activity after learning ([Supplementary Figure 2](#)). Together, these results indicate that rewarded learning in this visual paradigm yields a more compact, lower-dimensional neural representation in which fewer components suffice to reconstruct population activity, consistent with a consolidation of task-relevant dynamics into a smaller number of dominant modes. A quantification of shared and unique variance among component's further showed that the rewarded cohort's leading components were more independent than the unrewarded cohort's ([Supplementary Figure 3](#), [Supplementary Figure 4](#)).

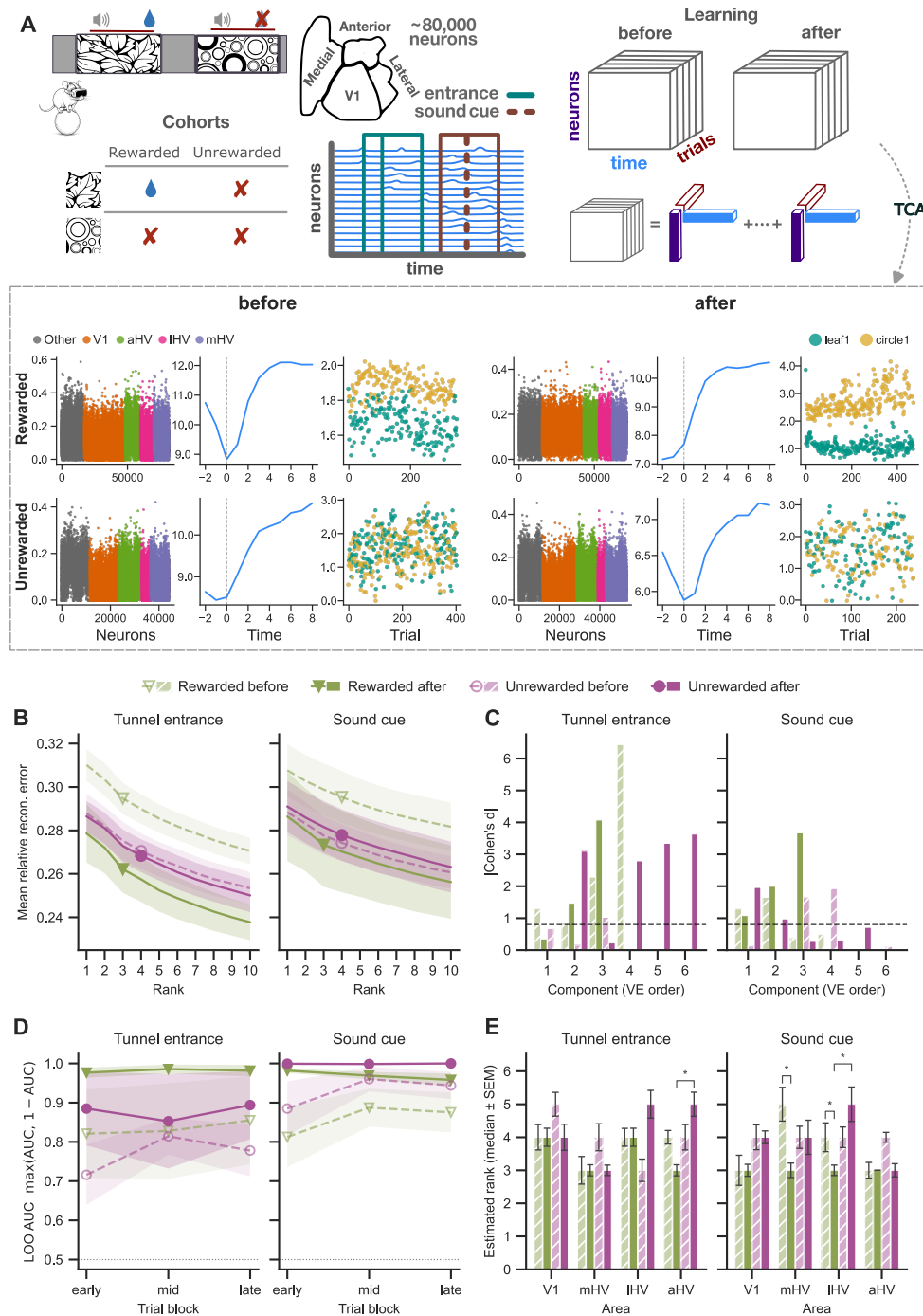


Figure 1: **A.** Schematic of experimental setup (*top-left*), neural recordings (*middle*), analysis method (TCA) (*top-right*), and obtained TCA factors for the first component for both learning paradigms in the before (*bottom-right*) and after sessions (*bottom-right*). **B.** Reconstruction error of neural activity versus fixed rank decomposition for before and after sessions and for the two windows of interest (tunnel entrance and sound cue). Triangles and dots indicate the estimated rank for reward-driven and unrewarded sessions respectively. For fixed rank reward-driven learning captures consistently more variance in the neural activity after learning indicated by lower reconstruction error, hinting to a lower rank representation. **C.** Mean trial factor's selectivity of each component, re-ordered

according to incremental variance explained. TCA components under reward-driven learning capturing most variance show stronger stimulus selectivity. **D.** Area under the curve (AUC) scores for stimulus discriminability in the low-dimensional representation comprising components that capture 80% of the variance in the neural data at early, middle and late trials blocks within a session. Reward-driven learning shows better stimulus discriminability already in the first session (before - dashed green line), as reflected by a modest increase from early to late trials. After learning both conditions increased discriminability, with a more pronounced effect for the reward-driven around tunnel entrance. **E.** Estimated rank for each recorded area. Rank was estimated over subsampled neural populations with fixed size $N_{sub} = 3000$ neurons and averaged over 5 independent realisations. Hatched, lighter bars = before learning; solid bars = after learning. Green = *rewarded*, (*water-reward*); magenta = *unrewarded*, (*no reward*). (*Left:*) trials aligned to tunnel entrance; (*Right:*) trials aligned to sound-cue onset. Significance brackets compare the four pairwise contrasts per area \times alignment: rewarded before vs after and unrewarded before vs after via exact paired sign-flip permutation ($2^{15} = 32\,768$ sign assignments); before and after (*rewarded vs unrewarded*) via Monte-Carlo independent permutation (10,000 label permutations). All p-values are Benjamini–Hochberg FDR-corrected across 4 contrasts \times 4 areas \times 2 alignments = 32 tests. * $p_{FDR} < 0.05$.

Stimulus selectivity concentrates in the leading components under rewarded learning. To test whether the components carrying the most variance also carry the strongest stimulus discrimination, we quantified leaf vs. circle selectivity in the trial factor weights of each component using Cohen’s d , computed as the standardised mean difference between leaf and circle trial loadings divided by the pooled within-class standard deviation. We selected $|d| \geq 0.8$ as the threshold for a large effect (Cohen (2013)); at this magnitude the two trial-factor distributions overlap by less than $\sim 50\%$, which we interpret as strong stimulus selectivity (dashed lines in Figure 1 C). Mean d per cohort \times timepoint \times VE-ordered component revealed a consistent dissociation between the two cohorts in which components carried strong selectivity. In the rewarded cohort after learning, selectivity was concentrated in the leading VE positions: at both windows of interest, across the first 3 most VE components reached overpassed the $|d| = 0.8$ threshold. In contrast, the unrewarded cohort only expressed strong selectivity after learning in the late, less VE components. Together, these results indicate that rewarded learning consolidates stimulus discrimination into the same low-rank modes that capture the bulk of population variance, whereas unrewarded exposure leaves stimulus information distributed across higher-order, less variance-explaining components, a pattern that goes beyond the rank and VE-concentration results reported above by specifying also where in the variance spectrum stimulus information resides.

To complement the per-component Cohen’s d analysis with a multivariate readout, we asked whether the joint trial-factor representation of the leading VE components could discriminate between the two stimuli, and whether this discriminability evolved within sessions. For each session and alignment we ran leave-one-out (LOO) cross-validated logistic regression on the smallest set of greedy-VE-ordered components whose cumulative VE reached 80%, and computed the receiver-operating-characteristic (AUC) separately for the early, mid, and late trial blocks within a session. At the tunnel-entrance alignment, mean LOO AUC was highest in rewarded-after sessions, intermediate in unrewarded-after, and lower before learning in both cohorts (Figure 1 D). The mixed model confirmed a significant main effect of cohort on AUC ($\beta_{unrewarded} = -0.476$, $z = -2.15$, $p = 0.032$), indicating that, after learning, the unrewarded cohort’s leading components carried substantially less trial-level stimulus information at tunnel entrance than the rewarded cohort’s leading components. The main effect of timepoint trended in the expected direction

($\beta_{before} = -0.223$, $z = -1.72$, $p = 0.085$) and the cohort \times timepoint interaction was not significant ($p = 0.20$). At the sound-cue alignment, both cohorts reached near-ceiling AUC after learning, and even before learning AUC stayed above ≈ 0.80 in both cohorts. Together with the per-component selectivity results, these findings converge on a single interpretation: reward-driven learning rotates the dominant modes of population variance toward the task-relevant stimulus axis, producing a low-dimensional representation in which stimulus identity is both more prominent in the leading subspace and more consistently separated across trials.

Finally, to examine whether the reward-driven rank reduction was uniform across visual cortex, we applied TCA separately for V1, mHV, lHV, and aHV on size-matched neuron subsamples (3,000 neurons, 5 subsamples per session), and selected each model's rank via the marginal-gain rule. We compared per-subsample rank estimates across cohorts and timepoints with FDR-corrected permutation tests. At tunnel entrance, only aHV showed a significant between-cohort difference after learning ($p = 0.005$), with the rewarded cohort converging to a lower rank. (Figure 1 E). At the sound cue, the rewarded cohort significantly reduced its rank in mHV ($p = 0.023$) and lHV ($p = 0.031$), and the between-cohort gap at lHV after learning was the strongest area-level effect observed ($p = 0.003$). V1 dimensionality remained stable across cohorts and timepoints. These results indicate that the global rank reduction under reward was not driven by a single area in the visual cortex but reflected a reduction in the dimensionality in higher visual areas, while V1 remained comparatively unaffected. This reduction in dimensionality could be the result of reward-driven plasticity within each area with the observed effect, or due to changes in the interactions of those areas or the inputs they receive from higher-order areas in rewarded conditions.

We studied how reward and repeated sensory experience differentially shape population activity in the visual cortex. Reward-driven learning was associated with a consistent reduction in the dimensionality of cortical activity, indicating that population variance became concentrated in fewer components. This reduction was not merely a nonspecific compression of activity. In the rewarded cohort, the components explaining the most variance were also the components carrying the strongest stimulus discriminability, suggesting that learning reorganised the population embedding so that dominant modes of activity became aligned with the task-relevant stimulus-discriminant axis.

This co-localization of variance and selectivity constrains the interpretation of the dimensionality reduction. A uniform gain change could promote a high-variance mode without making it stimulus-selective, while a nonspecific reduction in trial-to-trial variability could suppress small variance components without aligning the leading subspace with stimulus identity. Neither account predicts the observed coincidence between variance ordering and selectivity ordering in rewarded animals. We therefore interpret the rank reduction as a **task-aligned reorganization of population geometry**. Consistently, the lower-rank rewarded representation comprised more functionally independent modes rather than a more redundant set of co-varying components (Supplementary Figure 3, Supplementary Figure 4).

This reorganization should be interpreted as a change in the embedding of stimulus information, not necessarily as a reduction in the intrinsic dimensionality of the circuit. The number of latent variables represented by visual cortical populations may remain unchanged, while reward changes how these variables project onto the dominant modes of activity. Distinguishing between these possibilities will require explicitly quantifying how non-task-related latent variables contribute to the dominant and non-dominant components and is the aim of future work. Moreover, we note that while our results characterize the geometric outcome of unrewarded exposure, the exact algorithm by which unrewarded mice extract

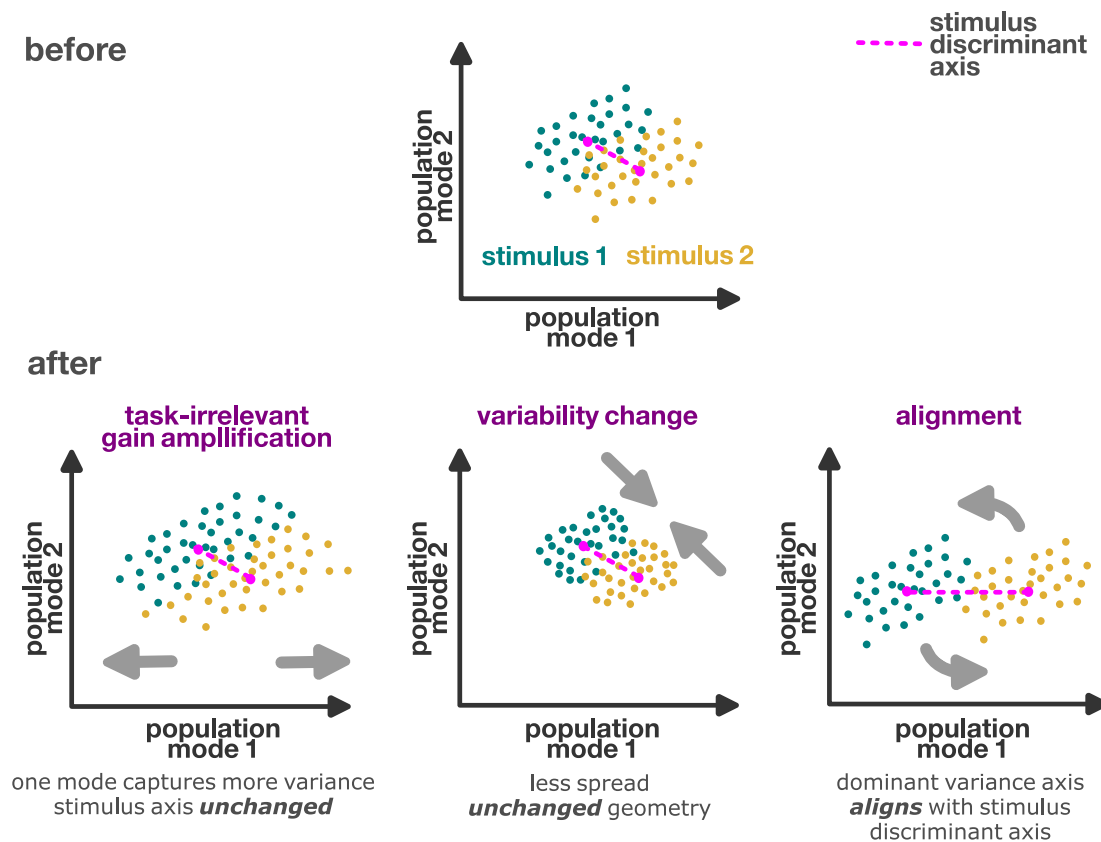
stimulus regularities remains an open question; we do not claim that the absence of reward implies nonreinforced learning.

We interpreted the after-learning changes in rewarded-cohort population activity as primarily reflecting reward-guided task experience on visual cortical representations, supported by analyses of residual activity after regressing out running speed. However, running speed does not capture finer movements or internal state fluctuations such as attention, motivation, arousal, or fatigue. Whisking, sniffing, licking preparation, and anticipatory postural adjustments were not recorded. These may occur more often under reward anticipation and could modulate visual cortical activity (Musall *et al.*, 2018; Ramadan *et al.*, 2022), consistent with prior reports of increased whisking or sniffing before expected reward (Deschênes *et al.*, 2012; Dominiak *et al.*, 2019; Wesson *et al.*, 2008; Yoshimoto *et al.*, 2019). Some changes in the rewarded cohort could thus reflect anticipatory or motor-preparatory behavior rather than visual representations alone, particularly closer to the sound cue. The cohorts also differed in water-deprivation state, which could affect arousal, motivation, or engagement; since pupil diameter and other state variables were not recorded, these could not be regressed out. Thus, while our results are consistent with reward-dependent reorganization of visual cortical population geometry, unmeasured motor and state variables may also contribute to the cohort differences. Disambiguating these would require richer behavioral monitoring.

Together, our results suggest that reward-driven learning induces a task-aligned compression of visual cortical activity. Repeated sensory experience may be sufficient to generate stimulus selectivity, but reward determines how this information is embedded in the dominant population modes. More broadly, these findings show that population geometry may reveal which learning objective has shaped the resulting cortical representation.

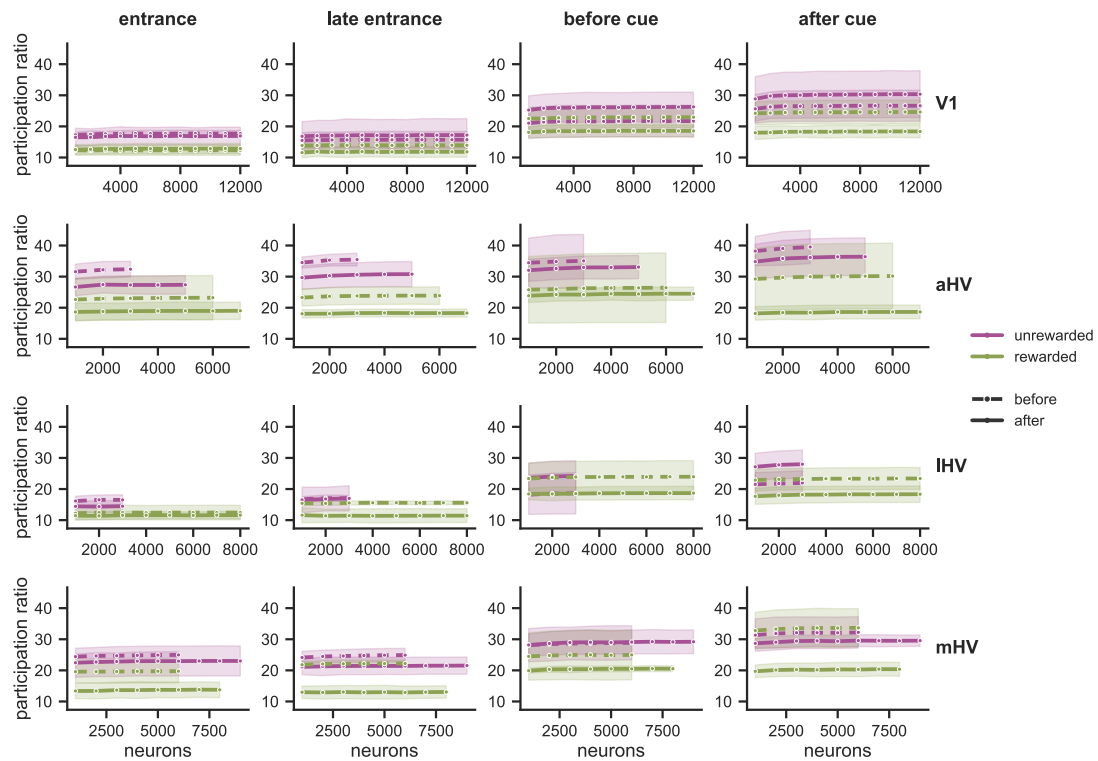
SUPPLEMENTARY MATERIAL

Population geometry changes



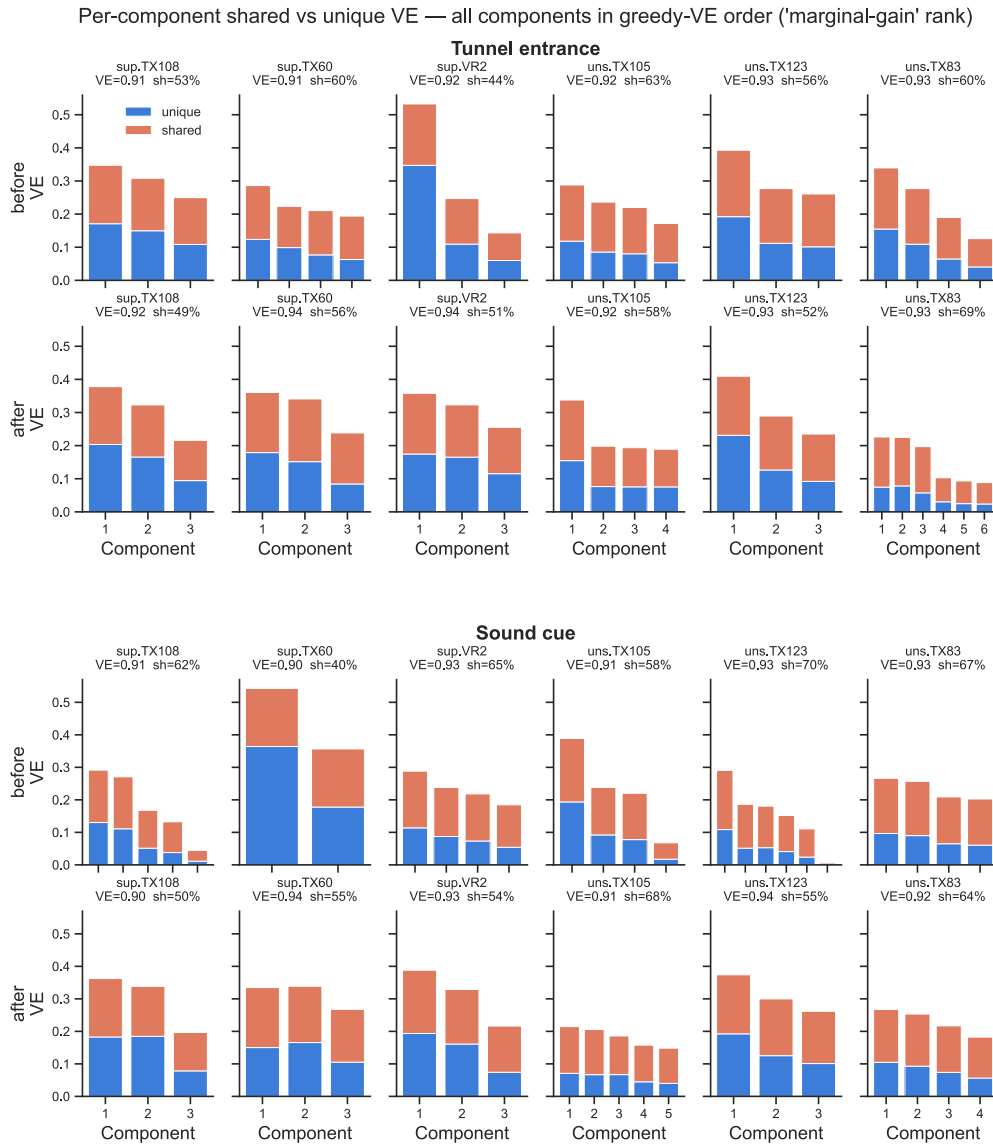
Supplementary 1: **Alternative hypotheses for population-geometry changes consistent with an increase in variance concentration.** Single-trial population activity projected onto the two leading modes (teal: *stimulus 1*; yellow: *stimulus 2*). The magenta line indicates the stimulus discriminant axis connecting the two condition means. (*Top*) baseline configuration before learning. (*Bottom*) three scenarios after learning that all produce a higher concentration of variance in the leading mode: (*left*) task-irrelevant gain amplification along the dominant axis, with the stimulus discriminant axis remaining unchanged; (*middle*) suppression of off-axis variability without change in the underlying geometry; (*right*) alignment of the dominant variance axis with the stimulus discriminant axis.

Participation ratio



Supplementary 2: **Participation ratio summary across cohorts.** Participation ratio computed as average over 20 subsamples from each mouse for fixed population sizes across areas and windows of interest. Participation ratio of activity after reward-driven learning results in lower dimensional activity than unrewarded learning for all examined areas.

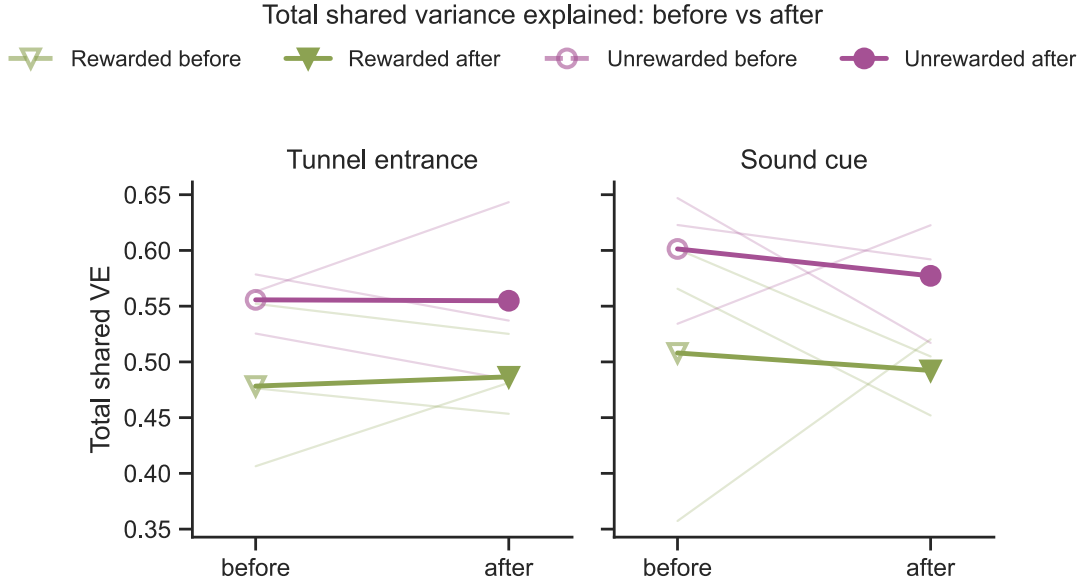
Shared vs. unique variance among TCA components.



Supplementary 3: Shared vs. unique variance distribution across TCA components.

For every session, the variance explained by the TCA model was partitioned, component by component, into a *unique part* (blue; the leave-one-out marginal contribution lost if only that component were removed from the model) and a *shared part* (orange; the overlap of that component with the remaining components, quantified as the pairwise cross-products between rank-1 component tensors). Components (x-axis) are shown in greedy incremental-VE order (C1 = highest variance-explaining). The two alignments (*top*: tunnel entrance; *bottom*: sound cue); rows are the before- and after-learning sessions and columns are the individual mice grouped by cohort (sup = rewarded; uns = unrewarded). Each panel is annotated with the model's total variance explained and the shared fraction, i.e. the proportion of the model-explained variance carried by inter-component overlap. Across mice, timepoints, and alignments, the rewarded cohort's components carry proportionally more unique variance (lower shared fraction) than the unrewarded cohort's, indicating that the lower-rank rewarded representation is assembled from more functionally independent modes.

Before versus after learning shared variance across components.



Supplementary 4: **Total shared variance explained across components, before vs. after learning.** Per-session shared variance ($VE_{\text{shared, total}}$, the proportion of the total data variance carried by overlap between the rank-1 TCA components, i.e. their summed pairwise cross-products) is shown for each mouse before and after learning. Faint lines are individual mice (one slope per animal connecting its before- and after-learning sessions); thick lines are cohort means. Left: trials aligned to tunnel entrance; Right: trials aligned to sound-cue onset. Green = rewarded, magenta = unrewarded; open markers = before learning, filled markers = after learning (triangles = rewarded, circles = unrewarded). Shared variance was consistently higher in the unrewarded than the rewarded cohort at both alignments, and changed only negligibly from before to after learning within each cohort, indicating that the greater inter-component overlap of the unrewarded representation is a stable cohort property rather than a learning-induced effect.

Shared vs. unique variance partition quantification.

Because non-negative TCA components are not orthogonal, the variance explained by the full model is not the sum of the variances explained by each component individually: the rank-1 components overlap, and a leading component's variance-explained silently includes variance it shares with the others. To separate these contributions we partitioned, for each session and alignment, the variance explained by the TCA model into a *unique* and a *shared* part per component.

Let \mathcal{X} be the data tensor, $\mathbf{T}_r = \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$ the r -th rank-1 component, and $\hat{\mathcal{X}} = \sum_{r=1}^R \mathbf{T}_r$ the full reconstruction. With $SST = \|\mathcal{X}\|_F^2$ (total variance in the data) and $SSE_{\text{full}} = \|\mathcal{X} - \hat{\mathcal{X}}\|_F^2$, the full-model variance explained is

$$VE_{\text{full}} = 1 - \frac{SSE_{\text{full}}}{SST}. \quad (1)$$

The **unique** variance of component r is its leave-one-out marginal contribution — the variance explained that is lost if only that component is removed from the otherwise complete model. Writing $SSE_{-r} = \|\mathcal{X} - (\hat{\mathcal{X}} - \mathbf{T}_r)\|_F^2$,

$$\text{VE}_{\text{unique}}(r) = \frac{\text{SSE}_{-r} - \text{SSE}_{\text{full}}}{\text{SST}}. \quad (2)$$

The **shared** variance of component r is its overlap with the remaining components, quantified by the inner products between rank-1 component tensors:

$$\text{VE}_{\text{shared}}(r) = \frac{\langle \mathbf{T}_r, \hat{\mathbf{X}} - \mathbf{T}_r \rangle}{\text{SST}} = \frac{\sum_{s \neq r} \langle \mathbf{T}_r, \mathbf{T}_s \rangle}{\text{SST}}, \quad (3)$$

where each component inner product factorises over the three modes, $\langle \mathbf{T}_r, \mathbf{T}_s \rangle = (\bar{\mathbf{u}}_r \mathbf{u}_s)(\bar{\mathbf{v}}_r \mathbf{v}_s)(\bar{\mathbf{w}}_r \mathbf{w}_s)$. For non-negative CP these cross-products are non-negative, so shared variance cannot cancel across component pairs.

These two terms partition the full-model variance exactly,

$$\text{VE}_{\text{full}} = \sum_{r=1}^R [\text{VE}_{\text{unique}}(r) + \text{VE}_{\text{shared}}(r)], \quad (4)$$

an identity we verified numerically (residual $< 10^{-8}$) for every session \times alignment. All quantities are computed against the original data tensor \mathbf{X} , with components taken in the greedy incremental-VE order.

Each session was summarised by its **shared fraction**, the proportion of the *model-explained* variance carried by inter-component overlap,

$$\text{shared fraction} = \frac{\sum_{r=1}^R \text{VE}_{\text{shared}}(r)}{\text{VE}_{\text{full}}} = \frac{2 \sum_{r < s} \langle \mathbf{T}_r, \mathbf{T}_s \rangle}{\text{SST} \cdot \text{VE}_{\text{full}}}, \quad (5)$$

which we report as mean \pm SD across the 3 mice in each cohort \times timepoint \times alignment cell. Note the denominator is VE_{full} (the variance the model explains), not the total data variance.

Experimental dataset

We analysed a publicly available dataset released in support of the experiments performed to investigate supervised and unsupervised learning in mice' primary visual cortex Zhong et al.,(Zhong et al. (2025)). In this study, neural activity was recorded from large populations of neurons across primary and higher visual cortical areas, while mice experienced a visual discrimination environment either **with** reward, in the task-trained or supervised condition, or **without** reward, in the unsupervised-exposure condition. This experimental contrast makes it possible to probe whether the reorganization of visual cortical representations arises from reward-guided task acquisition, or whether similar representational changes can emerge from plain exposure to the same sensory statistics of the stimuli alone.

Task structure and behavior

In the experiment, head-fixed mice were trained on a visual discrimination task within virtual reality (VR) corridors. Running speed on an air-floating ball was monitored across cohorts, whereas licking activity at a water spout was recorded only for mice in the task-trained, rewarded condition. The mice underwent daily training sessions (following a 1-hour-per-day regimen established during a running acclimation phase) for approximately two weeks. A typical session consisted of roughly 200-400 trials, during which mice were presented with two different naturalistic visual texture stimuli ("leaf" and "circle") in a pseudo-random order. Both cohorts were exposed to the same stimulus set and trial structure; however, only for the rewarded cohort only one stimulus had assigned behavioral significance through reward (in the analysed experiments here the "leaf" stimulus). During each individual trial, the mouse ran through a 4-

meter-long virtual corridor lined with one of the specific visual textures, separated from the next trial by 2 meters of gray space. As the mouse navigated the corridor, a sound cue played at a random spatial position (between 0.5 and 3.5 meters); in the rewarded trials, this cue signaled the availability of a 2.5 μ l water reward, which was delivered only if the mouse successfully licked the spout.

Neural recordings

Neural activity was recorded by the original authors using two-photon mesoscope calcium imaging while mice performed the task. The recordings targeted excitatory neurons, using mice that expressed the calcium indicator GCaMP6s in this population. Across sessions, the number of successfully tracked neurons ranged from 20,547 to 89,577 per animal. Recordings were obtained from the primary visual cortex (V1) and higher visual cortical areas (HVAs). We grouped neurons in HVAs into anterior (aHV: rostromedial area, RL; rostromedial lateral area, RLL), medial (mHV: retrosplenial cortex, RSP; posteromedial area, PM; anteromedial area, AM; mediomedial posterior area, MMP), and lateral (lHV: lateromedial area, LM; anterolateral area, AL) sub-areas.

The raw calcium imaging data had already been processed by Zhang et al, 2025 using the Suite2p software, which performed motion correction, neuropil correction, and spike deconvolution. All neural analyses we performed here were based on these deconvolved fluorescence traces provided by the authors. Neural recordings were obtained at the beginning (during the very first task exposure - “before learning”) and at the end of the training (“after learning”) for all mice. These sessions were for most mice approximately 11-14 days apart. The “after learning” neural dataset was acquired, once the mice demonstrated selective licking in the rewarded corridor in anticipation of reward delivery. From this dataset, we used 3 mice from the rewarded cohort (VR2, TX60, TX108) and (TX105, TX123, TX123) mice from non-rewarded cohort to obtain our results. No significant differences in behavioral results were noted in the original study between sexes.

Data preprocessing

An initial data preprocessing has been already performed on the provided imaging dataset by the authors of the original experiment. In particular, the authors have performed motion correction, segmentation and fluorescent extraction with Suite2p. In addition to the motion-corrected dataset, the authors provide a denoised version of the recordings obtained via principal component decomposition. Specifically, they performed a Principal Component Analysis (PCA) and discarded components corresponding to low-variance dimensions, under the assumption that these primarily reflect experimental noise. The retained principal components were then used to reconstruct the data back into the original neural activity space. Both the reconstructed (denoised) dataset and the original, unprocessed recordings are made available by the authors. For our analysis onwards we used this version of the dataset.

We used 3 task (“rewarded”) mice that learned a visual discrimination task with water reward contingent on a sound cue in one corridor, and 3 unrewarded mice that experienced the same visual stimuli and sound cues but without water rewards or water restriction. For each animal we analysed one recording *before learning* and one *after learning*, yielding one “before” and one “after” session per mouse in both the rewarded and unrewarded cohorts. Across trials mice were running with different running speeds. To account for the influence of running speed on neural activity, we applied a simple regression-based speed correction before tensor component analysis. For each recording, we first constructed a matrix of running speeds aligned to the event-centered analysis windows (see below). We then fit, independently for each neuron, a linear model of the form $\text{Activity} \sim 1 + \text{Speed}$ across all time-trial samples and used the residuals as speed-corrected activity. This procedure removes variance explained by moment-to-moment

speed while retaining other sources of variability, and the resulting residual activity was used to construct the tensors for TCA. After residualising the activity with respect to speed, we z-scored the data for each neuron with robust z-scoring. In particular, we rescaled the activity of each neuron using robust z-scoring to minimize the influence of extreme calcium transients

$$z_{i,t} = \frac{x_{i,t} - \text{median}(x_i)}{\text{MAD}(x_i) \times 1.4826} \quad (6)$$

Here x_i indicates the residual activity of neuron i and MAD is the Median Absolute Deviation.

Methods

Tensor Component Analysis (TCA): To uncover the latent structure of neural activity during visual learning, we utilized Tensor Component Analysis (TCA), a multi-way dimensionality reduction technique. Unlike standard matrix-based methods (e.g., PCA), TCA preserves the relational structure between neurons, time, and trials by treating the data as a third-order tensor $\mathcal{X} \in \mathbb{R}^{N \times T \times K}$, where N is the number of neurons, T is the number of time frames within a trial, and K is the number of trials.

The model approximates \mathcal{X} as a sum of R rank-1 components (Canonical Polyadic or CP decomposition):

$$\mathcal{X} \approx \hat{\mathcal{X}} = \sum_{r=1}^R \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r \quad (7)$$

where $\mathbf{u}_r \in \mathbb{R}^N$ represents the neuron factors (spatial ensembles), $\mathbf{v}_r \in \mathbb{R}^T$ represents the temporal factors (within-trial dynamics), and $\mathbf{w}_r \in \mathbb{R}^K$ represents the trial factors (across-trial evolution or gain).

Non-negative TCA and Reconstruction Error:

We employed a Non-negative TCA implementation using the Hierarchical Alternating Least Squares (HALS) algorithm. By constraining all factors (\mathbf{u} , \mathbf{v} , \mathbf{w}) to be non-negative, the model produces a parts-based representation. This is physiologically more interpretable for calcium imaging data, as it models neural activity as the additive sum of functional motifs rather than allowing for negative activity.

The model quality was assessed via the reconstruction error, defined as the squared Frobenius norm of the residual between the raw data and the model approximation:

$$E = \|\mathcal{X} - \hat{\mathcal{X}}\|_F^2 = \sum_{i=1}^N \sum_{j=1}^T \sum_{k=1}^K (x_{ijk} - \hat{x}_{ijk})^2 \quad (8)$$

The optimal rank R was determined by identifying the elbow of the reconstruction error curve using the Kneedle algorithm, ensuring a balance between model complexity and explanatory power.

Component Reordering and Variance Explained:

Because non-negative TCA components are not inherently orthogonal, the total variance explained (VE) is not the simple sum of individual component variances. To address this, we implemented a greedy incremental variance explained reordering (following the methodology of John Niebur).

We first calculated the $R \times R$ gram matrix G , where each entry G_{rs} represents the inner product between the r -th and s -th components:

$$G_{rs} = \langle \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r, \mathbf{u}_s \circ \mathbf{v}_s \circ \mathbf{w}_s \rangle = (\mathbf{u}_r^T \mathbf{u}_s)(\mathbf{v}_r^T \mathbf{v}_s)(\mathbf{w}_r^T \mathbf{w}_s) \quad (9)$$

Components were then reordered using a forward-selection greedy algorithm. The first component (C_1) was selected as the one explaining the most total variance. Each subsequent component was chosen to maximize the incremental variance added to the existing set. This ensured that components were ranked by their unique contribution to the neural data, allowing for consistent comparisons across different mice and sessions.

Stimulus Selectivity and other Statistical Analyses:

To evaluate the functional relevance of the identified latent components, we performed several downstream analyses:

- **Trial Factor Selectivity:** We quantified how well each component discriminated between visual stimuli (e.g., leaf vs. circle) by comparing the distributions of trial factors associated with each stimulus class.
- **Linear Classifier:** We trained a logistic regression decoder on the trial factors using Leave-One-Out - Cross-Validation (LOO CV). This allowed us to determine the accuracy and Area Under the ROC Curve (AUC) for predicting stimulus identity based on the latent neural state.
- **Learning Curves:** Trials were divided into chronological blocks (early, mid, late). By calculating the classification performance within each block, we tracked the temporal evolution of representational discriminability, allowing us to compare learning dynamics between the Rewarded (Supervised) and Unrewarded (Unsupervised) cohorts.
- **Cohen's d:** Cohen's d is an appropriate selectivity index because it is scale-invariant — making it directly comparable across components, sessions, and cohorts whose nonneg trial weights differ in absolute magnitude — and its sign encodes the direction of preference: $d > 0$ indicates higher loadings on leaf1 trials, while $d < 0$ indicates higher loadings on circle1 trials
- **Mixed effects model for discriminability:** Cohort \times timepoint \times block effects were tested in a linear mixed-effects model with mouse as a random intercept ($\text{auc} \sim C(\text{cohort}) * C(\text{timepoint}) * \text{block_num}$, $n = 36 \text{ block} \times \text{session observations per alignment, 6 mice}$).

Participation ratio analysis:

To quantify the dimensionality of population activity while controlling for differences in recorded neuron number across sessions and cortical areas, we computed the participation ratio of trial-averaged residual population activity for repeated neuron subsampling. For each session, area, and analysis epoch, we formed a trial-by-neuron matrix ($X \in \mathbb{R}^{T \times N}$) of neural activity, employed speed correction and performed z-scoring across trials.

For each trial-by-neuron matrix (X), we centered the activity across trials and computed the empirical covariance matrix

$$C = \frac{1}{T-1} X_c^\top X_c, \quad (10)$$

where $X_c = X - \frac{1}{T} \mathbf{1} \mathbf{1}^\top X$. If we denote by $\lambda_1, \dots, \lambda_N$ the positive eigenvalues of C , the participation ratio of the neural activity X is computed by

$$\text{PR} = \frac{\left(\sum_i \lambda_i \right)^2}{\sum_i \lambda_i^2}. \quad (11)$$

The participation ratio quantifies the effective number of covariance dimensions occupied by the population activity. It has a value close to one when most of the variance is concentrated along a single dominant axis, and increases when variance is distributed more evenly across multiple dimensions.

Because the number of recorded neurons varied substantially across cortical areas and sessions, full-population participation ratios would confound population dimensionality with sampling density. We therefore estimated participation ratio scaling curves by repeatedly subsampling neurons within each area and session. For each subsample size, we drew 20 independent random subsets of neurons, computed the participation ratio for each subset, and repeated this procedure across increasing subsample sizes. $N_{\text{sub}} \in 1000, 2000, \dots, 12000$. For each value of (N_{sub}) , if at least (N_{sub}) neurons were available in the corresponding area, we sampled (N_{sub}) neurons uniformly without replacement and computed the participation ratio on the resulting reduced activity matrix.

This subsampling procedure served two purposes. First, it allowed comparisons across cortical areas and sessions with different numbers of recorded neurons. Since the participation ratio can increase with increasing neuron number, comparing full-population values would confound population dimensionality with sampling density. Second, the scaling of participation ratio with neuron number provides information about the structure of the underlying population code. If activity is dominated by a small number of shared latent modes, the participation ratio should saturate as we add more neurons. In contrast, if additional neurons contribute independent or weakly correlated activity dimensions, the participation ratio should continue to increase with population size.

REFERENCES

- Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, 1(3), 295–311.
- Bell, A. J., & Sejnowski, T. J. (1997). The “independent components” of natural scenes are edge filters. *Vision Research*, 37(23), 3327–3338.
- Botvinick, M., Wang, J. X., Dabney, W., Miller, K. J., & Kurth-Nelson, Z. (2020). Deep reinforcement learning and its neuroscientific implications. *Neuron*, 107(4), 603–616.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. routledge.
- Deschênes, M., Moore, J., & Kleinfeld, D. (2012). Sniffing and whisking in rodents. *Current Opinion in Neurobiology*, 22(2), 243–250.
- Dominiak, S. E., Nashaat, M. A., Sehara, K., Oraby, H., Larkum, M. E., & Sachdev, R. N. (2019). Whisking asymmetry signals motor preparation and the behavioral state of mice. *Journal of Neuroscience*, 39(49), 9818–9830.
- Ferrari, A., Lange, F. P. de, & Akrami, A. (2026). Where learning paths meet: Convergence and divergence of statistical and reinforcement learning. *Current Opinion in Neurobiology*, 98, 103181.
- Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, 14(3), 119–130.
- Hebb, D. O. (2005). *The organization of behavior: A neuropsychological theory*. Psychology press.
- Henschke, J. U., Dylida, E., Katsanevaki, D., Dupuy, N., Currie, S. P., Amvrosiadis, T., Pakan, J. M., & Rochefort, N. L. (2020). Reward association enhances stimulus-specific representations in primary visual cortex. *Current Biology*, 30(10), 1866–1880.
- Musall, S., Kaufman, M. T., Gluf, S., & Churchland, A. K. (2018). Movement-related activity dominates cortex during sensory-guided decision making. *Biorxiv*, 308288.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583), 607–609.
- Pakan, J. M., Currie, S. P., Fischer, L., & Rochefort, N. L. (2018). The impact of visual cues, reward, and motor feedback on the representation of behaviorally relevant spatial locations in primary visual cortex. *Cell Reports*, 24(10), 2521–2528.
- Poort, J., Khan, A. G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., Bauza, M., Sahani, M., Keller, G. B., Mrsic-Flogel, T. D., & others. (2015). Learning enhances sensory and multiple non-sensory representations in primary visual cortex. *Neuron*, 86(6), 1478–1490.

- Ramadan, M., Lee, E. K., De Vries, S., Caldejon, S., Kato, I., Roll, K., Griffin, F., Nguyen, T. V., Larkin, J., Rhoads, P., & others. (2022). A standardized nonvisual behavioral event is broadcasted homogeneously across cortical visual areas without modulating visual responses. *Eneuro*, 9(5), ENEURO-0491.
- Schultz, W. (2015). Neuronal reward and decision signals: from theories to data. *Physiological Reviews*, 95(3), 853–951.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593–1599.
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1), 1193–1216.
- Wesson, D. W., Donahou, T. N., Johnson, M. O., & Wachowiak, M. (2008). Sniffing behavior of mice during performance in odor-guided tasks. *Chemical Senses*, 33(7), 581–596.
- Williams, A. H., Kim, T. H., Wang, F., Vyas, S., Ryu, S. I., Shenoy, K. V., Schnitzer, M., Kolda, T. G., & Ganguli, S. (2018). Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron*, 98(6), 1099–1115.
- Yoshimoto, J., Ozaki, J., Mizutani, K., Nakano, T., Ikeda, K., & Yamashita, T. (2019). Statistical analysis on characteristic whisker movements observed in reward processing. *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 723–726.
- Zhong, L., Baptista, S., Gattoni, R., Arnold, J., Flickinger, D., Stringer, C., & Pachitariu, M. (2025). Unsupervised pretraining in biological neural networks. *Nature*, 644(8077), 741–748.