Neural Manifold Capacity Captures Representation Geometry, Correlations, and Task-Efficiency Across Species and Behaviors

Chi-Ning Chou[1,✉], Luke Arend[2], Albert J. Wakhloo[3], Royoung Kim[4], Will Slatton[2], & SueYeon Chung[1,2,✉]

[1] *Center for Computational Neuroscience, Flatiron Institute, New York, NY, USA.*
[2] *Center for Neural Science, New York University, New York, NY, USA.*
[3] *Center for Theoretical Neuroscience, Columbia University, New York, NY, USA.*
[4] *Center for Neuroscience Imaging Research Institute for Basic Science, Sungkyunkwan University, Suwon, Republic of Korea*
✉{*cchou,schung*}*@flatironinstitute.org*

Abstract

The study of the brain encompasses multiple scales, including temporal, spatial, and functional aspects. To integrate understanding across these different levels and modalities, it requires developing quantification methods and frameworks. Here, we present effective Geometric measures from Correlated Manifold Capacity theory (GCMC) for probing the functional structure in neural representations. We utilize a statistical physics approach to establish analytical connections between neural co-variabilities and downstream read-out efficiency. These effective geometric measures capture both stimulus-driven and behavior-driven structures in neural population activities, while extracting computationally-relevant information from neural data into intuitive and interpretable analysis descriptors. We apply GCMC to a diverse collection of datasets with different recording methods, various model organisms, and multiple task modalities. Specifically, we demonstrate that GCMC enables a wide range of multi-scale data analysis. This includes quantifying the spatial progression of encoding efficiency across brain regions, revealing the temporal dynamics of task-relevant manifold geometry in information processing, and characterizing variances as well as invariances in neural representations throughout learning. Lastly, the effective manifold geometric measures may be viewed as order parameters for phases related to computational efficiency, facilitating data-driven hypothesis generation and latent embedding.

The neural population doctrine [1, 2] postulates that populations of neurons serve as both functional and structural units in brain computation. While there have been an explosive amount of experimental findings supporting this idea [3, 4], we need to develop interpretable and computationally relevant quantifications of neural population activity. This way, we can generalize across multiple scales of analysis [5]. For example, we have developed receptive fields and tuning properties in single-unit analysis. What analogous concepts at the population level can facilitate conceptual reasoning and scientific hypotheses?

Researchers have applied numerous methodologies to address the aforementioned question. The *structural alignment* approach, which seeks to compare low-order statistics of neural activities [6, 7] has faced criticism for its failure to establish connections with downstream computations and its lack of descriptors at the population level [8, 9]. Meanwhile, the *functional alignment* approach, which selects brain models based on their similarity to specific task performance or predictivity [10, 11], struggles with the challenge of multiple models exhibiting highly similar performance. Lastly, the *dimensionality reduction* approach [12, 13], while addressing some limitations of the aforementioned methods, has garnered concerns on a number of characteristics: its unreliability (e.g., horseshoe effect [9, 14, 15]), the absence of a standardized approach for quantifying neural signatures, and its limited applicability to situations where neural activities are low-dimensional. Further, dimensionality reduction fails to adequately account for data showing a wide spectrum of structure, including high-dimensional structure in certain regions [3, 16, 17].

Neural manifolds and manifold capacity theory

Neurons collectively represent task-relevant information in the brain [1, 2]. In neuroscience we informally call the collection of neural response vectors to some given task condition or input stimulus a "neural manifold" (see Fig. 1a for some examples) [18]. The neural manifold principle postulates that the geometrical, statistical, and structural properties of these manifolds are highly relevant to the study of the functional roles of population representations [18, 19, 20] (Fig 1.b).

The Manifold Capacity Theory (MCT) [28] quantifies the neural manifold's representational efficiency through the classification capacity [29, 30], which measures the amount of linearly decodable information per neuron. Using tools from statistical physics and machine learning theory, the MCT analytically characterizes the classification capacity as a function of the shape of a manifold. Such an analytical formula for capacity further suggests the definition of computationally relevant geometric terms such as effective dimension and effective radius of neural manifolds. These effective geometric measures lead to applications in analyz-

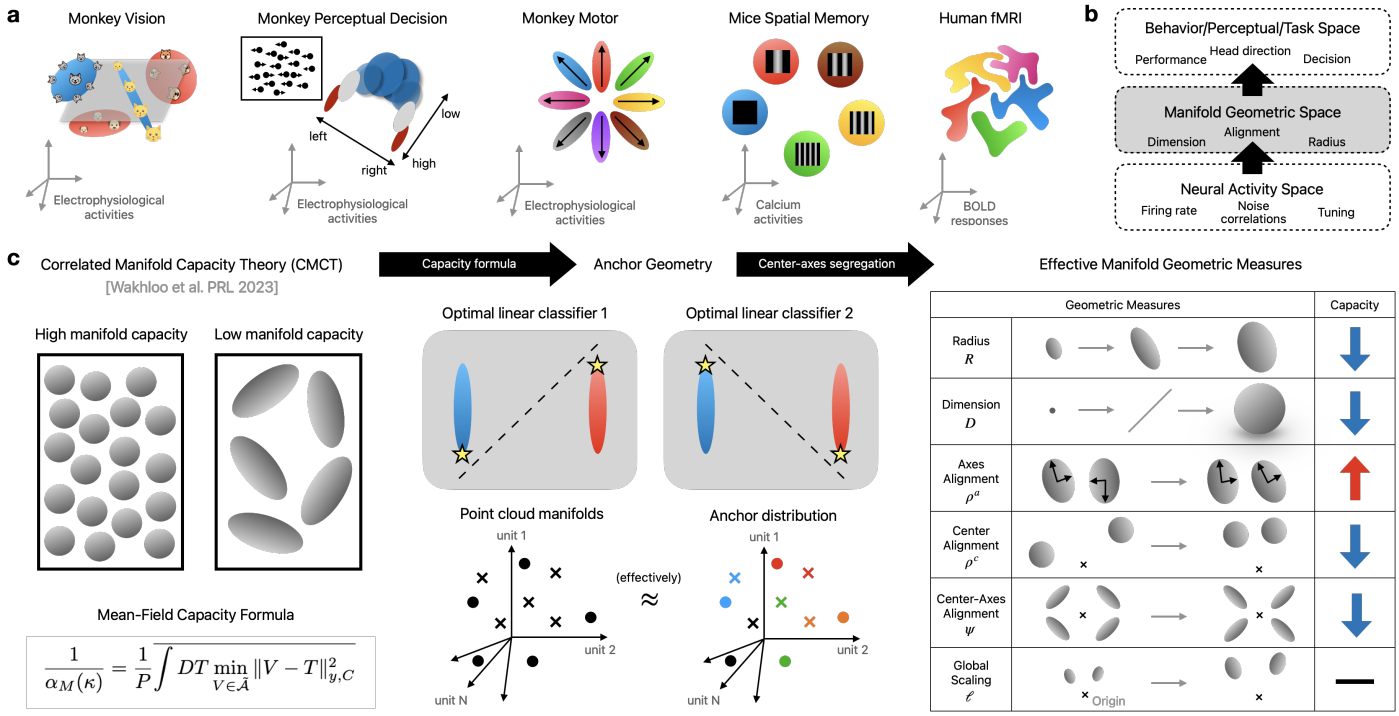

Figure 1: **Manifold geometry as intermediate descriptors for multi-scale neural data analysis.** **a**, Examples of using manifolds as analysis units. These manifolds lie in the neural state space with each coordinate being the (preprocessed) neural activity of a recording unit. Left to right: object recognition in monkeys' ventral visual pathway [21, 22], where a manifold is associated to stimuli from the same category; perceptual decision making in monkeys' prearcuate gyrus [23, 24], where a manifold corresponds to the stimulus's coherence level and the task outcome; a monkey delayed center-out reaching task, where a manifold corresponds to a target; a spatial memory task in mice hippocampus [25], where a manifold is associated to the reward location of an environment; human fMRI on THINGS dataset [26], where a manifold corresponds to a stimuli category and a region of interest (ROI). **b**, Effective manifold geometric measures serve as bridges between the neural activity space and the behavior/perceptual space. **c**, The effective Geometric measures from Correlated Manifold Capacity theory (GCMC). Left-top: manifold capacity captures the packing efficiency of neural manifolds in the neural state space. Left-bottom: previous theoretical work [27] derived an analytical formula for the capacity as a function of the manifold shapes and correlations in a mean-field model (Methods). Middle: the capacity formula induces an anchor geometry. Middle-top: the optimal linear classifier only sees the points (a.k.a., anchor points or support vectors) that are closest to the decision boundary. Middle-bottom: in the mean-field model, the quenched disorder (Methods) induces a distribution of anchor points, which give (non-uniform) probability weights to those points that are more important for linear classification. Points with the same color are anchor points that are simultaneously seen by an optimal linear classifier. Right: we define effective geometric measures on top of the anchor geometry. The table shows the qualitative relationship between each measure and the capacity. See Fig. SI1 and SI2 for intuitive and synthetic examples.

ing neural representations across biological datasets [31, 32, 33] and artificial neural networks [34, 35, 36].

However, the effective manifold geometric measures from MCT are ignorant to neural correlations, which play crucial roles in neural information processing [37]. Consequently, they have been limited to datasets with low correlations and sometimes yield inaccurate approximations to capacity [27]. This presents a fundamental research question: What effective geometric measures best capture capacity in the presence of neural correlations from biological datasets?

Results

In this work, we introduce effective Geometric manifold measures from Correlated Manifold Capacity theory (GCMC). Our contributions are three-fold: (1) GCMC incorporates the complex correlation structure via new effective geometric measures, which explicate how manifold geometry influences downstream computational efficiency. (2) GCMC connects noise correlations to manifold geometry, hence it unifies the concept of correlations at different system scales. (3) GCMC enables multi-scale data analysis such as quantifying the spatial progression of encoding efficiency across brain regions (Fig. 3), revealing task-relevant temporal dynamics (Fig. 4), and characterizing the variances and invariances in learning (Fig. 5). We demonstrate the power and the applicability of GCMC in a wide spectrum of datasets [21, 22, 26, 38, 23, 24, 25, 39] with various task modalities (e.g., vision, perceptual decision, motor, spatial memory), various model organisms (e.g., mice, monkeys, humans, artificial neural networks), and various recording methods (e.g., electrophysiology, calcium imaging, fMRI) (Fig. 1a). Finally, the effective manifold geometric measures can be conceptualized as order parameters for phases associated with computational efficiency, aiding in the generation of data-driven hypotheses and latent embedding.

Effective geometry for correlated neural manifolds

High-dimensional geometry can be deceptive, especially when it comes to correlations and computationally relevant geometry (see examples in Fig. SI1). To overcome this challenge, in GCMC, we derive effective manifold geometric measures that directly determine capacity based on a recent physics theory [27] for manifolds exhibiting correlations with each other (Fig. 1c, left). To be more specific, we derive from the theory the "anchor geometry space", which intuitively is the effective geometry space seen by the optimal linear classifier (Fig. 1c, middle). From there, we extract effective geometric measures through isolating terms in the analytic formula for manifold capacity.

Six effective geometric measures are highlighted here: the effective radius, the effective dimension, the effective center norm, and three distinct forms of effective manifold alignments (as depicted in Fig. 1c, right). The analytical connection to the capacity formula says that an increase in "effective radius" and/or "effective dimension" leads to a decrease in manifold capacity. Intuitively, the effective radius and dimension respectively serve as high-dimensional generalizations of the inverse signal-to-noise ratio's magnitude and degrees of freedom. The "effective center norm" captures the strength of mean neural responses whereas its effect on capacity has been factored into radius and dimension to ensure they are invariant to uniform scaling of the neural responses. A growth in "effective axes alignment" contributes to an increase in manifold capacity, as it allows for more efficient packing of manifolds oriented in the same direction within the neural state space. On the other hand, an increment in "effective center alignment" hurts manifold capacity, as it effectively reduces the available space for manifold packing by constraining manifolds to occupy a smaller portion of state space. Lastly, "effective center-axes alignment" provides insights into the extent of manifold variation distributed along the manifold center direction, which corresponds to the mean response. These effective alignment measures capture how the shared variability of a population of neurons affect downstream read-out efficiency. Note that these effective geometric measures are not independent to each other and they together orchestrate how manifold organization influences the capacity. More intuitions will be progressively developed through examples presented in the subsequent sections. In summary, effective radius, dimension, and center norm capture the geometric properties of individual manifolds while the three alignment measures delineate the effect of correlations between manifolds.

Understanding different types of neural correlations through effective manifold geometric measures

Correlated variability among neurons intricately connects to information encoding and decoding performance [37, 40, 41] of neural representations. There are multiple factors, including interactions between neurons and correlations among the input stimuli, posing a challenge for mechanistic understanding: How do neuron-wise correlations affect the structures of information encoding at the populational level?

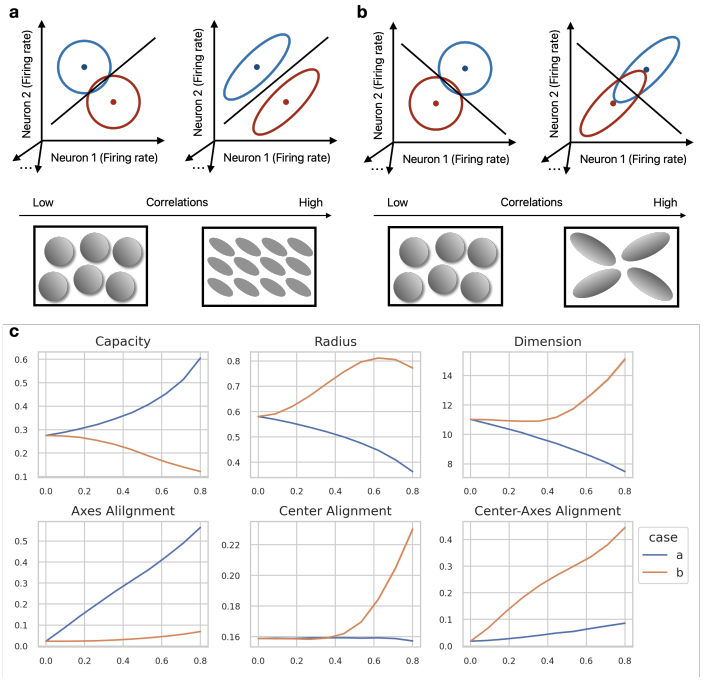

Figure 2: **Manifold geometry of noise correlations.** Classic examples of neural correlations from Averbeck et al. [37]. **a**, We depict the neural response distributions in terms of their mean response (i.e., the dot in the center) and the variation (i.e., the ellipse). Left: Uncorrelated, or shuffled, neural responses. Right: The ellipsoids stay in the "correct" side of the decision boundary. Bottom: The growth of neural correlations along the direction parallel to the decision boundary leads to the decrease of radius and dimension and leads to the increase of axes and center-axes alignments. As a joint effect, the manifold capacity increases. **b**, Left: Uncorrelated, or shuffled, neural responses. Right: The ellipsoids cross the decision boundary to the "wrong" side. Bottom: The growth of neural correlations toward the decision boundary leads to the growth of all manifold geometric measures, resulting in a decrease of manifold capacity. **c**, A quantitative characterization of the manifold geometry of noise correlations using GCMC (see Methods and Fig. SI3).

In GCMC, effective radius and dimension capture the inverse signal-to-noise ratio's variability strength and degrees of freedom respectively. Furthermore, the three manifold alignment measures in GCMC further differentiate the effects of correlations from signal (i.e., manifold center or mean response) and intrinsic variation (i.e., invariances or noises). Here, we present an illustrative examples on neural noises. More examples are provided in Fig. SI4.

The direction of noise correlations with respect to the decision boundary (e.g., Fig. 2a,c) would affect the decoding accuracy and the amount of encoded information. While earlier research [37] quantifies the effects of noise correlations using information theory, GCMC is able to further unveil the diverse manners in which neurons exhibit correlations in terms of different manifold geometry (Fig. 2a-b). First, Fig. 2a and Fig. 2.b give two extreme examples of the manifold geometry of noise correlations. As real datasets are likely to encompass both types of noise correlations to certain degree, the effective geometric measures within GCMC offer a language to describe the computational consequences of neural noises at the population level. Second, GCMC can examine signals for multiple categories (instead of binary),

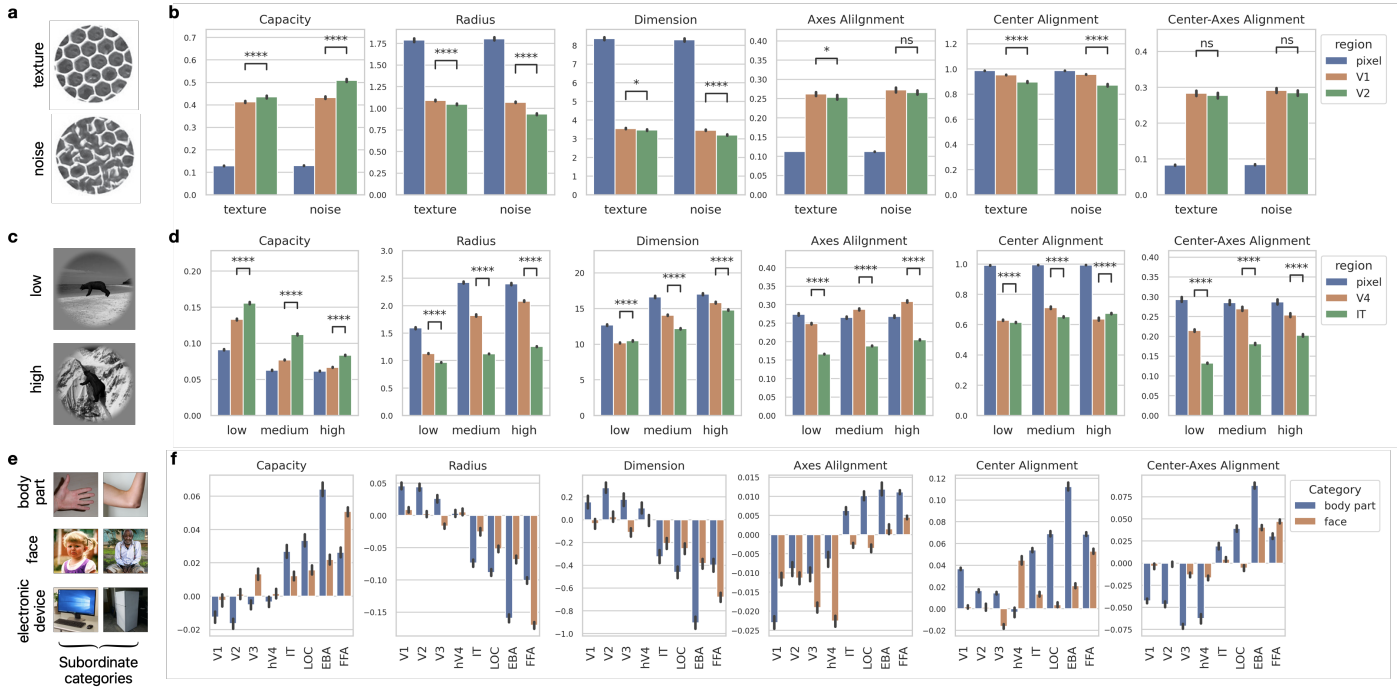

Figure 3: **Quantifying the spatial progression of encoding efficiency across brain regions.** **a**, A macaque monkey electrophysiological recording in V1 and V2 from Freeman et al. [21]. Monkeys were presented with images from two types. The first type contains 15 texture families (top) and the second type contains 15 noisy version of these texture families matching low order spectral statistics (bottom). Figure adapted from ref. [21], with permission. **b**, Effective geometric measures of texture manifolds and noise manifolds of an example monkey. Manifolds are defined as the collection of neural activities from the same brain area (and at pixel level), corresponding to the same image family. GCMC analysis was conducted on the 15 texture manifolds and the 15 noise manifolds respectively. Errorbar stands for ± 1 s.e.m. over 50 repetitions of subsampling 10 trials out of the 15 trials of each image family (Wilcoxon signed-rank test, $*p < 0.05$; $****p < 0.0001$; ns, not significant). See Fig. SI6 for additional results. **c**, A macaque monkey electrophysiological recordings in V4 and IT from Majaj et al. [22]. Monkeys gazed on a fixated dot and was presented with images composed of a foreground object over natural scenes. There are 64 objects each forming a sub-ordinate category. Top: Images with low variation. Bottom: images with high variation. **d**, Effective geometric measures of basic categories of an example monkey. There are 8 basic categories and each has 8 sub-ordinate categories. Manifolds are defined as the collection of neural activities from the same brain area, corresponding to the same category, and having the same variation. GCMC analysis was conducted on the 8 basic category manifolds. Errorbar stands for ± 1 s.e.m. over 50 repetitions of subsampling 35 trials from each category (Wilcoxon signed-rank test, $****p < 0.0001$). See Fig. SI7 for additional results on the subordinate identification. **e**, A human fMRI dataset from Hebart et al. [26]. We consider three image categories, each having 4 or 5 subordinate categories with 12 images. A manifold is defined as the preprocessed BOLD responses from the same region of interest (ROI) and corresponds to stimuli from the same subordinate category (Methods). **f**, Effective manifold geometric measures of an example human subject. GCMC analysis was conducted on all three categories and all ROIs. For each ROI, we subtract the resulting values of manifold geometric measures by that corresponding to the electronic device category. Errorbar stands for ± 1 s.e.m. over 50 repetitions of subsampling 10 images out of the 12 images from each subordinate category. See Fig. SI9 for details.

and hence opens up avenues for analyzing complex datasets (e.g., multi-class, hierarchical classes).

Quantifying the spatial progression of encoding efficiency across brain regions

The study of how neural representations progress across brain regions serves as one fundamental example. While neural tuning analysis has proven successful in early vision studies, higher visual areas present challenges due to the absence of clear single-neuron selectivity to higher-level concepts such as objects and faces. Consequently, prevailing analysis paradigms rely on data-driven frameworks like decoding analysis, comparison to representations from artificial neural network models, or visualization via dimensionality reduction. Although these methods excel in quantifying information in neural populations (e.g., decoding [42, 22]), ranking brain models [10], and providing intuitive insights (e.g., visualization [6, 43]), it remains a challenge to simultaneously capture the interaction between scales, e.g., read-out performance (top-down) and physiological structures (bottom-up).

GCMC excels at quantifying the stimulus-driven and/or task-driven structures in neural variabilities underlying behavioral invariance. Here, we utilize GCMC to give a unifying quantification of the spatial progression of encoding efficiency across various brain regions. These include electrophysiological recordings in monkey's early visual ventral stream (Fig. 3a-b) and late visual ventral stream (Fig. 3c-d), and a human fMRI experiment featuring multiple regions of interest (see Fig. 3e-f). In all examples, we see a clear improvement of manifold capacity along the visual hierarchy. Moreover, the changes of underlying geometry are different at different stages and different stimuli structures. For example, the growth of capacity from V1 to V2 is driven mainly by the shrinkage of radius and center alignment (Fig. 3b). On the other hand, the growth of capac-

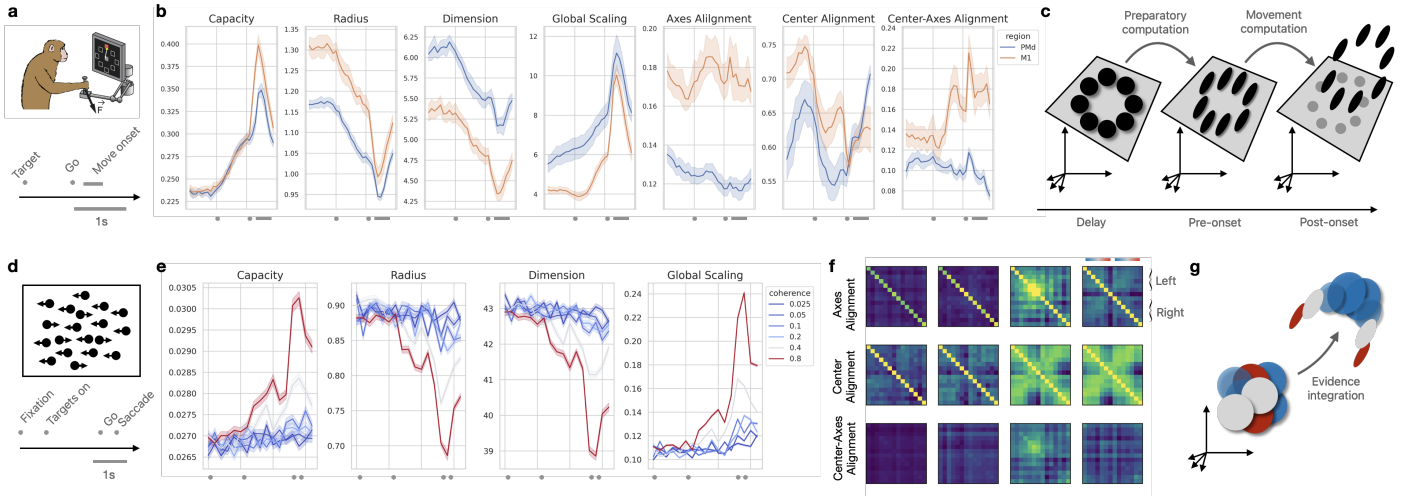

Figure 4: **Revealing the temporal dynamics of task-relevant manifold geometry.** **a**, Top: A monkey electrophysiological dataset on a delayed center-out reaching task with eight targets. Neural recordings are from premotor and primary motor cortices from Perich et al. [38]. Bottom: The trial structure. Figure adapted from ref. [38], with permission. **b**, The in-trial dynamics of the effective manifold geometry of an example monkey (see Methods and Fig. SI5). Errorband stands for 95% CIs over 10 sessions. **c**, The intuitive picture of the underlying computational strategies suggested by the manifold geometry. The preparatory computation compresses the manifolds along the signal direction while the movement computation springs out the manifolds. **d**, Top: A monkey electrophysiological dataset on a standard random-dot perceptual decision task from Kiani et al. [23, 24]. Bottom: The trial structure. **e**, The in-trial dynamics of the effective manifold geometry of an example monkey (Methods). For each coherence level of the random-dot experiment, we define two manifolds with respect to the two correct outcomes (left or right). The time bin size is 0.2 second. For each coherence level, we compute the manifold geometry of the two corresponding manifolds. Errorband stands for 95% CIs over 20 repetitions of subsampling 50 trials from each coherence level. **f**, The in-trial dynamics of the effective manifold alignments (see Methods). The time bin size is 0.2 seconds. We compute the manifold geometry of all the manifolds (i.e., two manifolds per coherence level) and plot the temporal dynamics of the alignment matrices with the manifold ordering being the top (resp. bottom) block containing manifolds corresponding to the left (resp. right) outcome with coherence level from small to large. **g**, The intuitive picture of the underlying computational strategies at the manifold level. The evidence integration compresses the manifolds with different extent according to how strong the evidence is.

ity from V4 to IT is simultaneously driven by the decrease of multiple factors, including radius, dimension, axes alignment, and center-axes alignment (Fig. 3d). Furthermore, the differences in effective manifold geometry across variations in stimuli reveal distinct organizational strategies of V4 and IT, warranting further study. In an fMRI dataset involving human subjects [26], we observed that the capacity difference between a target category (e.g., body parts or faces) and a control category (e.g., electronic devices) increases along the visual hierarchy. Notably, the greatest capacity gain occurs in brain regions known for their selectivity to the target category (e.g., the Extrastriate Body Area (EBA) exhibits the highest capacity gain for body part categories). While similar analyses using decoding methods (see Fig. SI9, Methods) did not reveal significant differences, our findings suggest that GCMC could be a valuable approach for investigating selectivity in brain regions. Further elaboration and additional details can be found in Methods and Fig. SI6, SI7, SI9.

Revealing the temporal dynamics of task-relevant manifold geometry

In recent years, there have been exciting advancements in the neural population dynamics framework, encompassing both computational modeling and theoretical analysis [44, 4, 45]. A growing body of experimental evidence highlights the pivotal role of temporal dynamics in areas such as motor control, decision-making, and working memory, among others. Here, we demonstrate how GCMC can elucidate the connection between population dynamics and the associated computations.

We first apply GCMC to a monkey dataset on a delayed center-out reaching task with electrophysiological recordings from premotor area (PMd) and primary motor area (M1) of macaque monkeys [38] (Fig. 4a). We divide each trial into time bins of length 0.2 seconds and define target-specific neural manifolds as collections of population responses associated with the same reaching target and measure the manifold geometry. GCMC unveils a wealth of manifold temporal dynamics (Fig. 4b). For example, we observe a gradual increase in capacity during the delay period, followed by a sudden burst immediately after the onset of movement. Meanwhile, the radius and dimension exhibit the opposite pattern, agreeing with the fact that capacity is inversely related to them (Fig. 1c). There are also quantitative differences between PMd and M1, e.g., the dimension of PMd is higher than M1, which agree to the previous findings [38] using a heuristic definition based on a dimensionality reduction method. Previously, it is known that the neural activities during preparatory and movement stage lie in subspaces that are orthogonal to each other. Here, the temporal dynamics of manifold geometry, particularly the sudden burst of center-axes alignments and center norm during the brief interval between the go cue and movement onset, paint out a more comprehensive picture: the preparatory computation compresses the target-specific manifold along the center di-

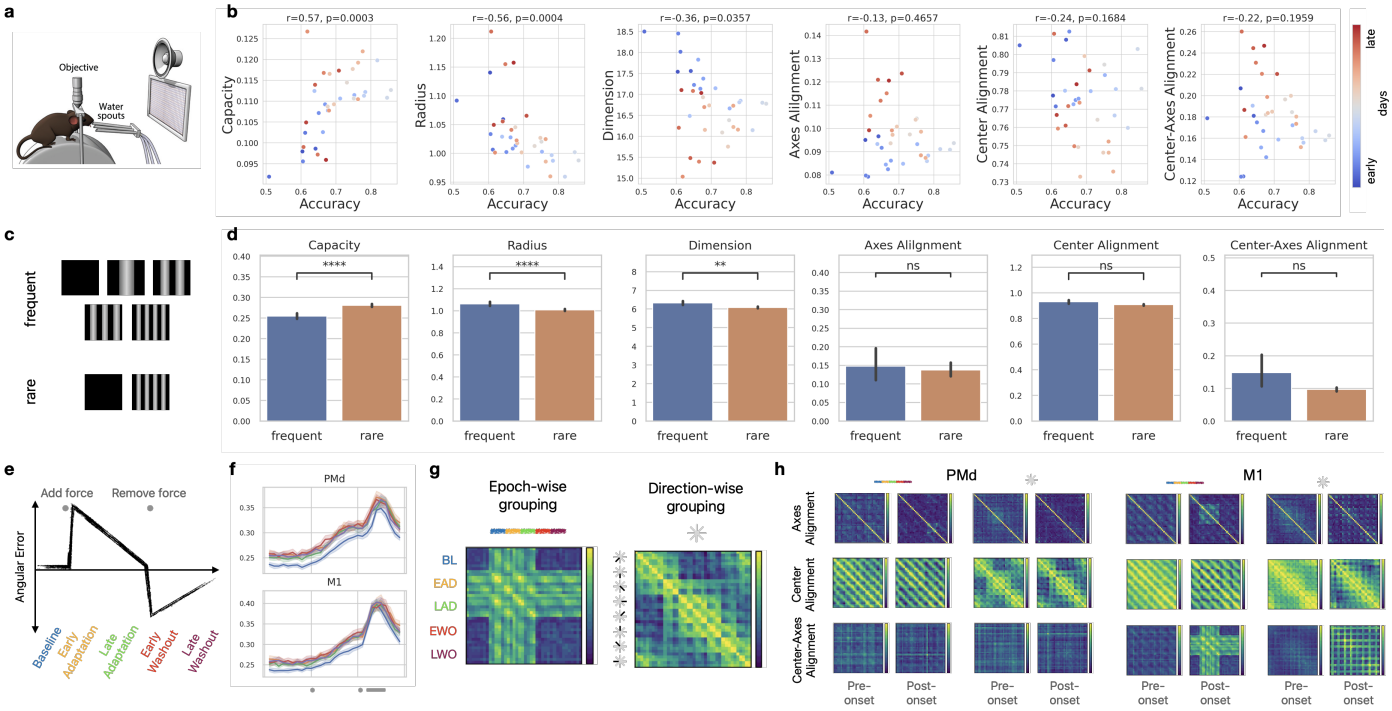

Figure 5: **Characterizing variances and invariances in learning.** **a**, A mice decision-making dataset. Calcium imaging recordings from posterior parietal cortex by Najafi et al. [39]. Behavioral setup: mice were trained to report whether the multisensory event rate is low or high. Figure adapted from ref. [39], with permission. **b**, Scatterplot of the effective manifold geometric measures versus behavioral performance of 35 training sessions of an example mouse. r, Pearson correlation coefficient. **c**, A mice spatial memory dataset. Calcium imaging recordings from hippocampus by Plitt et al. [25]. Behavioral setup: mice were trained to run in morphed virtual environment tracks. The frequent mice were exposed to all five morphed environments during training while the rare mice were only exposed to the two extreme morphed environments. **d**, The effective manifold geometric measures of the frequent and the rare condition. Errorbar stands for ± 1 s.e.m. over all sessions (12 for the frequent condition and 20 for the rare condition, see Methods). (Mann-Whitney-Wilcoxon two-sided test, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$; ns, not significant) **e**, A monkey rapid learning dataset. Electrophysiological recordings from premotor and primary motor cortices by Perich et al. [38]. Behavioral setup: monkey performed a center-out reaching task with eight targets. In each session, an external force or visuomotor rotation was added and divide the session into five epochs. **f**, Epoch-wise in-trial dynamics of manifold capacity in PMd and M1 are depicted. The color codes correspond to those specified in subfigure e. The capacity in the baseline epoch is noticeably smaller than the capacity of the other epochs, except for M1, right around movement onset. Errorband stands for 95% CIs over 10 sessions (Methods). **g**, Manifold alignments between the $40 = 5 \times 8$ (#epoch $\times$ #targets) manifolds. There are two ways to present the alignment matrices. Left: Grouping the manifolds epoch by epoch. Right: Grouping the manifolds according to the target direction. **h**, Effective manifold alignments across epochs in an example session. Both PMd and M1 exhibit ring-shaped manifold center alignment pre and post onset. Notably, the external perturbation only causes significant center-axis alignments in M1 during post onset. See Fig. SI5 for differences across sessions.

rection (as seen by the decrease of radius and dimension), whereas the movement computation pushes the manifolds outward along the center direction, effectively reducing the radius and dimension (Fig 4c).

Next, we investigate a monkey motion direction discrimination task regarding perceptual decision making (Fig. 4d). Specifically, we analyze an electrophysiological dataset with neural recordings from the prearcuate gyrus of macaque monkeys [23, 24]. Similarly, we divide trials into time bins of length 0.2 seconds and define neural manifolds as the collection of neural responses corresponding to the same outcome (left or right) and the same coherence level of random-dot motions. We first analyze the manifold geometry of two manifolds of the same coherence level. Consistent with task-performance, the manifold capacity is higher for the manifold pair with higher coherence level, and the capacity grows throughout the evidence-integrating period (Fig. 4e). We next analyze the manifold geometry of all the manifolds together. The three effective manifold alignments suggest the curve-shaped organization of decision manifolds (Fig. 4f), agreeing with previous findings in other brain areas related to decision making [46]. Furthermore, the difference in effective radius and dimension between manifolds with different coherence level indicates that there is a stronger compression when the monkey has more evidence (Fig. 4g).

Characterizing variances and invariances in learning

To study the underlying mechanism of learning at the population level, it is essential to establish quantitative measures that can characterize the variances and invariances in neural representations [47]. The effective geometric measures in GCMC naturally serve as tools for quantifying the computationally relevant information in the presence of learning.

In a learning task with clear supervisions or rewards, we expect the animals (or artificial neural networks) to improve their neural presentations in a way that facilitates performance after training. Manifold capacity can be used to quantify how the read-out efficiency of neural representations improves over time, and hence GCMC further de-

lineates the various "learning strategies" in terms of the changes of effective manifold geometry. For example, in a decision-making task in mice [39] (Fig. 5a), we find that the manifold capacity positively correlated with the behavioral performance and the improvement of capacity is mainly driven by the decrease of effective radius (Fig. 5b). Meanwhile, although not being strongly correlated with accuracy, there is a consistent decrease of effective dimension, center alignment, and center-axes alignment from early sessions to late sessions. On the other hand, for an object classification task trained on feedforward deep neural networks, we found that the increment of capacity across layers is mainly driven by the decrease of effective dimension (Fig. SI10).

When learning happens in a less straightforward or implicit manner, the effect on neural representations could be more complicated. For instance, the effective manifold geometry may exhibit unexpected trends or remain invariant, as demonstrated in the following two examples. Firstly, we analyze a hippocampal two-photon imaging dataset on mice spatial memory, where mice traversed linear tracks in morphed virtual reality environments [25] (Fig. 5c). In this study, mice experienced five different morph values with one group encountering all five environments throughout training (frequent morph), and the other group primarily encountering just two extreme morph environments (rare morph). The results were unanticipated: mice exposed to the rare morph condition exhibited higher capacity (across all five environments during the testing stage) compared to those exposed to the frequent morph condition (Fig. 5d). This potentially suggests that the rare morph condition facilitates formation of more compact hippocampal representations of the environment. In the second example, we investigated a rapid learning experiment involving a monkey reaching task [38], wherein monkeys received external perturbations after excelling at the task (see Fig. 5e). Our analysis reveals that the capacity of target-specific manifolds increased from the baseline after the introduction of external perturbations (Fig. 5f). Additionally, we observe variances as well as invariances in effective manifold alignment measures across various timescales (see Fig. 5h and Fig. SI5). Our results augment previous evidence suggesting that rapid motor adaptation is primarily due to a structural shift in the PMd population representation [38, 48].

Discussion

Advances in science are tightly linked to the development of experimental tools and analytical methods. In the era of neural population doctrine, it is of great importance to develop analytical frameworks for mechanistically describing how populations of neurons encode complex task information. Here, we present effective Geometric measures derived from Correlated Manifold Capacity theory (GCMC) as quantitative descriptors for neural manifolds. These measures analytically link the neural population activities to the corresponding functional properties such as downstream read-out efficiency. They play an analogous role to neuronal tuning and hence can serve as data-analysis tools for investigating the temporal and spatial structures of neural populational representations. By applying it to a wide range of datasets, we have shown how GCMC can robustly reveal task-relevant information in different modalities, recording methods, modal organisms, tasks, etc. In the following, we will further elaborate on potential avenues for future research, discussing the new types of biological questions that can be tackled by GCMC and how to develop mechanistic understandings in neural networks.

GCMC differs from dimensionality reduction methods in three ways: (i) there is no information loss and the high-dimensional neural activities are summarized into interpretable geometric measures (Fig. 1b); (ii) these effective geometric measures are quantitatively related to capacity through the formula (see Methods) and the qualitative connection is described in Fig. 1c (right); (iii) the segregation of manifold center and within-manifold variability allows rich investigation into how correlations at different levels affect geometry and computational efficiency. Finally, GCMC offers a complementary approach to decoding methods [37, 42, 22], by distinguishing among representations which exhibit high similarity in decoding performance (Fig. SI1a, SI9, Methods).

GCMC facilitates at least three types of analysis at the population level: (i) establishing connections between representational geometry and normative principles, (ii) describing and quantifying neural computational strategies based on the dynamics of effective manifold geometry, and (iii) providing a unifying framework for the study of stimulus-driven and task-driven neural correlations.

Normative principles refer to top-down functional or computational objectives for a neural processing system, such as the efficient coding principle and the redundancy reduction hypothesis [49]. To test the validity of normative principles at the neural population level, we need quantitative measures for neural representations that can reflect downstream read-out properties. The effective geometric measures in GCMC, connected to the capacity of neural manifolds, offer a foundation for testing normative principles. For example, the decrease in effective radius and dimension (Fig. 3) supports the untangling hypothesis [42] along the visual hierarchy.

Neural computational strategies entail a mechanistic understanding of a neural system, with a focus on the algorithmic aspects of the system. For instance, in the monkey reaching task, researchers investigate how motor areas organize their neural population dynamics for preparation and movement [50]. Using dimensionality reduction methods, researchers have been able to describe populational neural computation in terms of "neural signatures" [51] and interactions across different "neural modes" [52]. With GCMC, one can not only enhance these mechanistic understandings by providing finer characterizations of the changes in manifold geometry (e.g., Fig. 4c,g), but also can offer a more detailed description of the underlying computational strategy in terms of the dynamics of effective manifold geometry.

Neural correlations are ubiquitous in the brain and are challenging to study due to the complex interactions among at least three factors: the stimuli, the neurons, and the tasks. By defining neural manifolds according to a class of stimuli (e.g., object and category manifolds in Fig. 3), a task outcome (e.g., direction manifolds in Fig. 4a-c), or both (e.g., outcome-coherence manifolds in Fig. 4d-g), GCMC offers a unifying framework to quantify the computationally relevant correlations in terms of effective representational geometry. Therefore, in future studies, our quantification method is flexible enough to investigate normative principles and/or neural computational strategies in the presence of neural correlations.

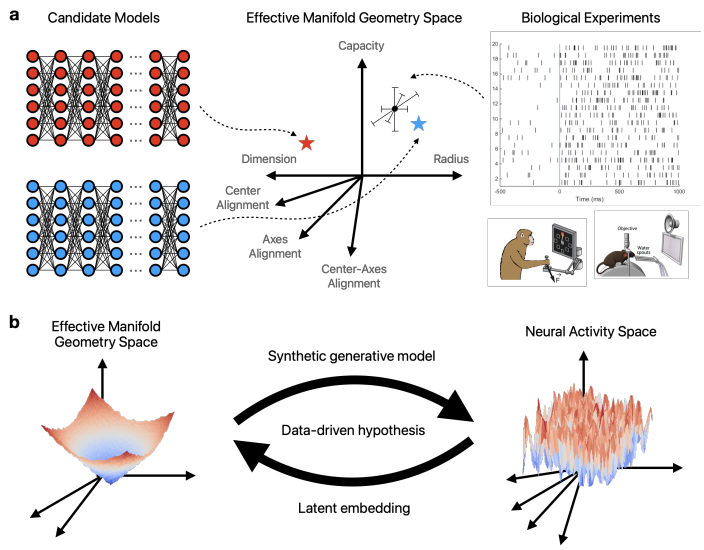

Figure 6: **Effective manifold geometric measures as computational order parameters.** **a**, To compare models (left, stars in different color) for a neural network (right, neural recordings), one can compare the distance in the manifold geometry space, where the coordinates are the effective manifold geometric measures. **b**, Manifold geometry space (left) contains the effective manifold geometric measures as coordinates and hence serve as an interpretable latent space for neural activities (right). Under the assumption of universality, one can further construct synthetic generative models (e.g., Gaussian processed) from these measures and study the synthetic neural activities. Different color refers to different task/stimulus condition. The global geometry (i.e., across conditions) is known to be complex in the neural activity space. Here we further hypothesize that the global geometry in the effective manifold geometry space would be much smoother.

In a broader context, we propose that we regard the effective geometric measures for neural manifolds as "computational order parameters" for constructing models so as to gain insights into both biological and artificial neural networks. As a fundamental concept in statistical physics, an order parameter provides a quantitative description of phase transitions in complex systems. Furthermore, it often exhibits "universality", meaning that different physical systems, regardless of their specific details and interactions, demonstrate similar critical behaviors near phase transitions and can be quantified using the same order parameter. Viewing effective manifold geometric measures as order parameters for a computational objective (i.e., capacity), GCMC allows the following two future directions (Fig. 6): (i) generating data-driven hypotheses and performing model selection via effective geometric measures, (ii) using effective manifold geometric measures as a latent embedding.

The formulation of scientific hypotheses relies on quantitative observables that can be extracted from the data and interpreted with respect to task behavior. For instance, neural tuning and selectivity represent fundamental quantification concepts in the context of the neuron doctrine. When it comes to the neural population doctrine, previous methodologies (e.g., similarity analysis, decoding performance, and geometric method) encounter at least one of the following two challenges: not formally connecting different levels of abstraction (e.g., structural and functional level) and being problem-specific. GCMC resolves these challenges by formally linking representational geometry to downstream read-out efficiency via the manifold capacity formula. By viewing effective manifold geometric measures as order parameters characterizing the computationally relevant properties of neural systems, GCMC provides an interpretable quantification for comparing neural systems. This enables a framework to perform hypothesis testing and model selection based on the closeness of effective manifold geometric measures [36].

A prominent approach in computational neuroscience and machine learning involves constructing latent models or embeddings [53] towards understanding a neural network. The concept is to embed neural representations into a low-dimensional and potentially interpretable latent space. Here, we propose using effective manifold geometry as a latent embedding. As the effective geometric measures from GCMC are order parameters for neural representations, using them as latent embedding offers several advantages. First, these measures quantify the computationally relevant properties of the system and are sample-efficient (due to the self-averaging property). Next, they are interpretable and hence can be used to develop conceptual understanding. If we further assume the universality of effective manifold geometric measures, then an advanced usage of GCMC would be building synthetic generative models and mapping them out in the effective manifold geometry space. We can then match a real neural network with a synthetic model via comparing their latent embeddings (i.e., distance in the manifold geometry space) (see Fig. 6a). It is worth noting that there are related approaches, such as approximate Bayesian computation [54] and simulation-based inference [55], which share the philosophy of inferring latent models through simulations or synthetic data. What sets our idea apart is the emphasis on making comparisons and selections at the "neural manifold" level through effective geometric measures of neural population activities. This provides a more interpretable and robust model comparison.

Lastly, it is an exciting direction for future research to apply GCMC to additional datasets or even design novel experimental setups suitable for neural manifold analysis. This approach opens up opportunities to explore new neuroscience questions at the neural population level.

Data Availability

Monkey early vision dataset: Freeman et al. [21]. Monkey late vision dataset: Majaj et al. [22]. Monkey motor dataset: Perich et al. [38]. Monkey perceptual decision dataset: Kiani et al. [23, 24]. Human fMRI dataset: Hebart et al. [26]. Mice spatial memory dataset: Plitt et al. [25]. Mice decision-making dataset: Najafi et al. [39].

Code Availability

Codes will be available public usage in the camera-ready version. Requests on accessing to the current version of the code should be made to the first and corresponding author.

References

[1] Rafael Yuste. "From the neuron doctrine to neural networks". In: *Nature reviews neuroscience* 16.8 (2015), pp. 487–497.

[2] Shreya Saxena and John P Cunningham. "Towards the neural population doctrine". In: *Current opinion in neurobiology* 55 (2019), pp. 103–111.

[3] Stefano Fusi, Earl K Miller, and Mattia Rigotti. "Why neurons mix: high dimensionality for higher cognition". In: *Current opinion in neurobiology* 37 (2016), pp. 66–74.

[4] Saurabh Vyas, Matthew D Golub, David Sussillo, and Krishna V Shenoy. "Computation through neural population dynamics". In: *Annual review of neuroscience* 43 (2020), pp. 249–275.

[5] Anne E Urai, Brent Doiron, Andrew M Leifer, and Anne K Churchland. "Large-scale neural recordings call for new insights to link brain and behavior". In: *Nature neuroscience* 25.1 (2022), pp. 11–19.

[6] Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. "Representational similarity analysis-connecting the branches of systems neuroscience". In: *Frontiers in systems neuroscience* (2008), p. 4.

[7] Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. "Similarity of neural network representations revisited". In: *International conference on machine learning*. PMLR. 2019, pp. 3519–3529.

[8] Matthew R Whiteway and Daniel A Butts. "The quest for interpretable models of neural population activity". In: *Current opinion in neurobiology* 58 (2019), pp. 86–93.

[9] Mikhail A Lebedev, Alexei Ossadtchi, Nil Adell Mill, Núria Armengol Urpí, Maria R Cervera, and Miguel AL Nicolelis. "Analysis of neuronal ensemble activity reveals the pitfalls and shortcomings of rotation dynamics". In: *Scientific Reports* 9.1 (2019), p. 18978.

[10] Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J Majaj, Rishi Rajalingham, Elias B Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, et al. "Brain-score: Which artificial neural network for object recognition is most brain-like?" In: *BioRxiv* (2018), p. 407007.

[11] Blake A Richards, Timothy P Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, et al. "A deep learning framework for neuroscience". In: *Nature neuroscience* 22.11 (2019), pp. 1761–1770.

[12] John P Cunningham and Byron M Yu. "Dimensionality reduction for large-scale neural recordings". In: *Nature neuroscience* 17.11 (2014), pp. 1500–1509.

[13] Rich Pang, Benjamin J Lansdell, and Adrienne L Fairhall. "Dimensionality reduction in neuroscience". In: *Current Biology* 26.14 (2016), R656–R660.

[14] Maxwell Shinn. "Phantom oscillations in principal component analysis". In: *Proceedings of the National Academy of Sciences* 120.48 (2023), e2311420120. DOI: `10.1073/pnas.2311420120`.

[15] Eva L. Dyer and Konrad Kording. "Why the simplest explanation isn't always the best". In: *Proceedings of the National Academy of Sciences* 120.52 (2023), e2319169120. DOI: `10.1073/pnas.2319169120`. eprint: `https://www.pnas.org/doi/pdf/10.1073/pnas.2319169120`. URL: `https://www.pnas.org/doi/abs/10.1073/pnas.2319169120`.

[16] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Matteo Carandini, and Kenneth D Harris. "High-dimensional geometry of population responses in visual cortex". In: *Nature* 571.7765 (2019), pp. 361–365.

[17] Frederic Lanore, N Alex Cayco-Gajic, Harsha Gurnani, Diccon Coyle, and R Angus Silver. "Cerebellar granule cell axons support high-dimensional representations". In: *Nature neuroscience* 24.8 (2021), pp. 1142–1150.

[18] SueYeon Chung and LF Abbott. "Neural population geometry: An approach for understanding biological and artificial neural networks". In: *Current opinion in neurobiology* 70 (2021), pp. 137–144.

[19] Nikolaus Kriegeskorte and Rogier A Kievit. "Representational geometry: integrating cognition, computation, and the brain". In: *Trends in cognitive sciences* 17.8 (2013), pp. 401–412.

[20] Nikolaus Kriegeskorte and Xue-Xin Wei. "Neural tuning and representational geometry". In: *Nature Reviews Neuroscience* 22.11 (2021), pp. 703–718.

[21] Jeremy Freeman, Corey M Ziemba, David J Heeger, Eero P Simoncelli, and J Anthony Movshon. "A functional and perceptual signature of the second visual area in primates". In: *Nature neuroscience* 16.7 (2013), pp. 974–981.

[22] Najib J Majaj, Ha Hong, Ethan A Solomon, and James J DiCarlo. "Simple learned weighted sums of inferior temporal neuronal firing rates accurately predict human core object recognition performance". In: *Journal of Neuroscience* 35.39 (2015), pp. 13402–13418.

[23] Roozbeh Kiani, Christopher J Cueva, John B Reppas, and William T Newsome. "Dynamics of neural population responses in prefrontal cortex indicate changes of mind on single trials". In: *Current Biology* 24.13 (2014), pp. 1542–1547.

[24] Roozbeh Kiani, Christopher J Cueva, John B Reppas, Diogo Peixoto, Stephen I Ryu, and William T Newsome. "Natural grouping of neural responses reveals spatially segregated clusters in prearcuate cortex". In: *Neuron* 85.6 (2015), pp. 1359–1373.

[25] Mark H Plitt and Lisa M Giocomo. "Experience-dependent contextual codes in the hippocampus". In: *Nature neuroscience* 24.5 (2021), pp. 705–714.

[26] Martin N Hebart, Oliver Contier, Lina Teichmann, Adam H Rockter, Charles Y Zheng, Alexis Kidder, Anna Corriveau, Maryam Vaziri-Pashkam, and Chris I Baker. "THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior". In: *Elife* 12 (2023), e82580.

[27] Albert J Wakhloo, Tamara J Sussman, and SueYeon Chung. "Linear classification of neural manifolds with correlated variability". In: *Physical Review Letters* (2023).

[28] SueYeon Chung, Daniel D Lee, and Haim Sompolinsky. "Classification and geometry of general perceptual manifolds". In: *Physical Review X* (2018).

[29] Elizabeth Gardner. "The space of interactions in neural network models". In: *Journal of physics A: Mathematical and general* 21.1 (1988), p. 257.

[30] N Brunel, J-P Nadal, and G Toulouse. "Information capacity of a perceptron". In: *Journal of Physics A: Mathematical and General* 25.19 (1992), p. 5017.

[31] Justin D Yao, Klavdia O Zemlianova, David L Hocker, Cristina Savin, Christine M Constantinople, SueYeon Chung, and Dan H Sanes. "Transformation of acoustic information to sensory decision variables in the parietal cortex". In: *Proceedings of the National Academy of Sciences* 120.2 (2023), e2212120120.

[32] Nihaad Paraouty, Justin D Yao, Léo Varnet, Chi-Ning Chou, SueYeon Chung, and Dan H Sanes. "Sensory cortex plasticity supports auditory social learning". In: *Nature communications* 14.1 (2023), p. 5828.

[33] Emmanouil Froudarakis, Uri Cohen, Maria Diamantaki, Edgar Y Walker, Jacob Reimer, Philipp Berens, Haim Sompolinsky, and Andreas S Tolias. "Object manifold geometry across the mouse cortical visual hierarchy". In: *BioRxiv* (2020).

[34] Uri Cohen, SueYeon Chung, Daniel D Lee, and Haim Sompolinsky. "Separability and geometry of object manifolds in deep neural networks". In: *Nature communications* (2020).

[35] Joel Dapello, Jenelle Feather, Hang Le, Tiago Marques, David Cox, Josh McDermott, James J DiCarlo, and SueYeon Chung. "Neural population geometry reveals the role of stochasticity in robust perception". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 15595–15607.

[36] Michael Kuoch, Chi-Ning Chou, Nikhil Parthasarathy, Joel Dapello, James J DiCarlo, Haim Sompolinsky, and SueYeon Chung. "Probing Biological and Artificial Neural Networks with Task-dependent Neural Manifolds". In: *Conference on Parsimony and Learning (Proceedings Track)*. 2023.

[37] Bruno B Averbeck, Peter E Latham, and Alexandre Pouget. "Neural correlations, population coding and computation". In: *Nature reviews neuroscience* 7.5 (2006), pp. 358–366.

[38] Matthew G Perich, Juan A Gallego, and Lee E Miller. "A neural population mechanism for rapid learning". In: *Neuron* (2018).

[39] Farzaneh Najafi, Gamaleldin F Elsayed, Robin Cao, Eftychios Pnevmatikakis, Peter E Latham, John P Cunningham, and Anne K Churchland. "Excitatory and inhibitory subnetworks are equally selective during decision-making and emerge simultaneously during learning". In: *Neuron* 105.1 (2020), pp. 165–179.

[40] Brent Doiron, Ashok Litwin-Kumar, Robert Rosenbaum, Gabriel K Ocker, and Krešimir Josić. "The mechanics of state-dependent neural correlations". In: *Nature neuroscience* 19.3 (2016), pp. 383–393.

[41] Adam Kohn, Ruben Coen-Cagli, Ingmar Kanitscheider, and Alexandre Pouget. "Correlations and neuronal population information". In: *Annual review of neuroscience* 39 (2016), pp. 237–256.

[42] James J DiCarlo and David D Cox. "Untangling invariant object recognition". In: *Trends in cognitive sciences* 11.8 (2007), pp. 333–341.

[43] Elena A Allen, Erik B Erhardt, and Vince D Calhoun. "Data visualization in the neurosciences: overcoming the curse of dimensionality". In: *Neuron* 74.4 (2012), pp. 603–608.

[44] Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. "Context-dependent computation by recurrent dynamics in prefrontal cortex". In: *nature* 503.7474 (2013), pp. 78–84.

[45] Alexis Dubreuil, Adrian Valente, Manuel Beiran, Francesca Mastrogiuseppe, and Srdjan Ostojic. "The role of population structure in computations through neural dynamics". In: *Nature neuroscience* 25.6 (2022), pp. 783–794.

[46] Gouki Okazawa and Roozbeh Kiani. "Neural Mechanisms that Make Perceptual Decisions Flexible". In: *Annual Review of Physiology* 85 (2023), pp. 191–215.

[47] Claudia Clopath, Tobias Bonhoeffer, Mark Hübener, and Tobias Rose. "Variance and invariance of neuronal long-term representations". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1715 (2017), p. 20160161.

[48] Barbara Feulner and Claudia Clopath. "Neural manifold under plasticity in a goal driven learning behaviour". In: *PLoS computational biology* 17.2 (2021), e1008621.

[49] Horace B Barlow et al. "Possible principles underlying the transformation of sensory messages". In: *Sensory communication* 1.01 (1961), pp. 217–233.

[50] Juan A Gallego, Matthew G Perich, Lee E Miller, and Sara A Solla. "Neural manifolds for the control of movement". In: *Neuron* 94.5 (2017), pp. 978–984.

[51] Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. "Neural population dynamics during reaching". In: *Nature* 487.7405 (2012), pp. 51–56.

[52] Matthew T Kaufman, Mark M Churchland, Stephen I Ryu, and Krishna V Shenoy. "Cortical activity in the null space: permitting preparation without movement". In: *Nature neuroscience* 17.3 (2014), pp. 440–448.

[53] Steffen Schneider, Jin Hwa Lee, and Mackenzie Weygandt Mathis. "Learnable latent embeddings for joint behavioural and neural analysis". In: *Nature* (2023), pp. 1–9.

[54] Mark A Beaumont. "Approximate bayesian computation". In: *Annual review of statistics and its application* 6 (2019), pp. 379–403.

[55] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. "The frontier of simulation-based inference". In: *Proceedings of the National Academy of Sciences* 117.48 (2020), pp. 30055–30062.

[56] Martin N Hebart, Adam H Dickter, Alexis Kidder, Wan Y Kwok, Anna Corriveau, Caitlin Van Wicklin, and Chris I Baker. "THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images". In: *PloS one* 14.10 (2019), e0223792.

[57] Ha Hong, Daniel LK Yamins, Najib J Majaj, and James J DiCarlo. "Explicit information for category-orthogonal object properties increases along the ventral stream". In: *Nature neuroscience* 19.4 (2016), pp. 613–622.

Acknowledgements. We thank Matthew Perich for the insightful discussion on the monkey reaching dataset. We thank Roozbeh Kiani and Isabella Rischall for the helpful discussion related to monkey perceptual decision task. We also thank Billy Broderick, Jenelle Feather, Lada Kohoutová, Nikihil Parthasarathy, Won Mok Shim, and Choong-Wan Woo, for the stimulating conversation regarding many preliminary results. We thank Matthew Perich, Corey Ziemba, and Tony Movshon for share their datasets. This work was funded by the Center for Computational Neuroscience at the Flatiron Institute of the Simons Foundation and the Klingenstein-Simons Award to S.C. All experiments were performed on the Flatiron Institute high-performance computing cluster.

Author contributions. Conceptualization: CNC, AJW, SYC; Methodology: CNC, AJW, SYC; Software: CNC, LA, AJW; Data analysis and curation: CNC, LA, RK, WS; Writing-original draft: CNC; Writing-reviewing and editing: CNC, LA, AJW, RK, WS, SYC; Supervision: SYC; Funding acquisition: SYC.

Conflicts. The authors declare no additional conflicts of interest. The funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

Methods

Datasets

Monkey early vision task. We use the dataset from Freeman et al. [21]. 13 anesthetized macaques were presented with a set of 450 images of 15 naturalistic textures while neural responses were recorded from visual cortex.

Each stimulus is an intact or scrambled version of one of 225 grayscale images. Each intact image is one of 15 randomly synthesized examples of a texture from one of 15 texture families. The images were scrambled by randomizing phase in the 2D Fourier domain, preserving a naturalistic power spectrum while destroying the high-order statistics distinguishing texture.

The stimuli were presented multiple times in pseudorandom order. Presentation lasted 100 ms with a 100 ms gap between presentations. Responses were recorded from several neurons at a time using quartz-platinum tungsten microelectrodes (Thomas Recording) in cortical areas V1 and V2. Responses to 20 repeats of all stimuli were collected for 102 neurons in V1 and 103 neurons in V2.

Monkey late vision task. We use the dataset from Majaj et al. [22]. Two macaques were presented with images of objects rendered against naturalistic backgrounds while neural responses were recorded from visual cortex.

Each stimulus is one of 5760 grayscale images of a 3D object rendered against a randomly selected naturalistic background. The object is one of 64 3D models, where 8 example models belong to each of 8 object categories (animals, boats, cars, chairs, faces, fruits, planes, and tables). 90 stimuli are rendered from each 3D model by varying object pose, scale, location. Renderings are made at three levels of variation: none (10 images), medium (40 images), and high (40 images).

The stimuli were presented in series during passive fixation. Presentation lasted 100 ms with a 100 ms gap between presentations. Population responses were recorded using three 96-channel multi-electrode arrays in cortical areas V4 and IT. Responses to all stimuli were isolated for 189 sites in IT and 126 sites in V4 (a typical site captured spiking activity from 1-4 neurons).

Monkey reaching task. We use the dataset from Perich et al. [38]. Two macaques performed an eight-target center-out reaching task while neural responses were recorded from motor cortex.

A recording session consisted of about a hundred trials across three consecutive epochs: baseline, adaptation and washout. During baseline trials, the reaching task was not perturbed in any way. During adaptation trials, a motor perturbation was applied where either external forces were applied to the hand ("curl field") or a fixed rotation was applied to the visual feedback ("visual rotation").

Population responses were recorded using 96-channel multi-electrode arrays in cortical areas PMd and M1. Spike sorting was applied at the session level (9 curl-field and 7 visual-rotation sessions). Across all sessions, 137–256 PMd and 55–93 M1 neurons were isolated for Monkey C, and 66–121 PMd and 26–51 M1 neurons were isolated for Monkey M.

Monkey perceptual decision task. We use the dataset from Kiani et al. [23, 24]. Two macaques performed a

random-dot direction discrimination task while neural responses were recorded from prearcuate cortex.

In each task trial, two targets appeared on the screen after the monkey directed its gaze towards a central point. After a 250 ms delay, a random dot motion stimulus was presented for 800 ms, followed by a variable-duration delay period. Upon presentation of a Go cue, the monkeys were instructed to indicate their perceived motion direction by executing a saccade toward the corresponding target. Motion strengths were tuned for each monkey to provide a broad range of behavioral accuracies, ranging from chance to flawless performance.

Population responses were recorded using 96-channel multielectrode arrays in prearcuate gyrus. Spike sorting identified 100-250 units in each recording session (a unit could be an isolated single neuron or a multi-unit), for 15 sessions in total.

Mouse auditory decision-making. We use the dataset from Najafi et al. [39]. Four mice performed an auditory decision task while neural responses were imaged in posterior parietal cortex.

Trials consisted of simultaneous clicks and flashes, generated randomly (via a Poisson process) at rates of 5–27 Hz over 1,000 ms. Mice reported by licking a left or right waterspout whether event rates were high or low compared with a category boundary (16 Hz) learned from experience.

A two-channel, two-photon microscope was used to record the activity of neurons in layer 2/3 of posterior parietal cortex in transgenic mice. The images were processed by algorithm that simultaneously identified neurons and estimated *inferred spiking activity*, a number related to a neuron's spiking activity during one frame. Recordings were made for 135 sessions in total, with a typical session isolating activity from 400-600 cells.

Human fMRI dataset. We use the dataset presented in Hbart et al. [26]. Three human participants were presented with images of objects while fMRI data were collected. The images were taken from the THINGS object concept and image database [56].

Each stimulus is one of 8640 colored images of an object embedded in a natural background. The object belongs to one of 720 distinct concepts, with each concept consisting of 12 exemplars.

The stimuli were presented in series while maintaining central fixation. Presentation lasted 500 ms with a 4000 ms gap between presentations. To maintain engagement, participants were instructed to perform an oddball detection task, detecting occasional artificially generated images amidst the stimulus presentation series.

Whole-brain fMRI data were collected using a 3 Tesla Siemens Magnetom Prisma scanner and a 32-channel head coil with 2mm isotropic resolution. Functional data was acquired for the main experiment for 12 scanning sessions and functional localizers were acquired for additional 1-2 scanning sessions. .

Artificial neural networks. We use a pretrained ResNet-50 architecture trained on ImageNet with learning algorithm SimCLR.

Manifold capacity theory and GCMC

Mean-field model for correlated manifold capacity. Here we recall the mean-field model for correlated manifold capacity defined in [27]. Consider a collection of P manifolds M^μ lying in $\mathbb{R}^N$ (e.g., N-dimensional neural state space or feature space), where $\mu = 1, 2, \dots P$. These manifolds either come from the empirical data or a synthetic/mathematical model. As we focus on linear read-out, it suffices to consider the convex hull of each manifold. Thus, we describe manifold M^μ as $\{u_0^\mu + \sum_{i=1}^K s_i^\mu u_i^\mu \,:\, s_i \in \mathcal{S}^\mu\}$ where u_0^μ is the manifold center, $\{u_i^\mu\}_{i=1}^K$ forms the manifold intrinsic basis, and $\mathcal{S}^\mu \subset \mathbb{R}^K$ describes the shape of the manifold (see Fig. 2 in [28] or Fig. 1a in [27] for pictorial examples). See later section (Code implementations) on how to measure these parameters from a real dataset.

The mean-field model for correlated manifold capacity considers the following problem. Let $\kappa \geq 0$ be the margin. Let $y \in \{-1, 1\}^P$ be a random binary label vector, and let u_i^μ be random Gaussian vectors in $\mathbb{R}^N$ satisfying $\overline{\langle u_i^\mu, u_j^\nu \rangle} = C_{\nu,j}^{\mu i}$ for some positive definite covariance tensor C (either measured from data or specified mathematically). The correlated manifold capacity $\alpha_M(\kappa)$ is defined as supremum of the ratio P/N in the thermodynamic limit (i.e., $P, N \to \infty$) such that with probability 1 there exists a separating hyperplane $w \in \mathbb{R}^N$ such that $\min_{x \in M^\mu} y^\mu \langle w, x \rangle \geq \kappa$ for every μ. Concretely, this can also be phrased as solving a log partition function as described in Equation 1 in [27].

The main result of [27] shows in its Equation 5 that

$$\frac{1}{\alpha_M(\kappa)} = \frac{1}{P} \overline{\int D_{y,C} T \min_{V \in \mathcal{A}} \|V - T\|_{y,C}^2} \tag{1}$$

where $D_{y,C}$ is the zero-mean Gaussian measure with covariance tensor $y^\mu y^\nu C_{\nu,j}^{\mu,i}$, and the overline denotes the average over binary random label vector y. Next, $\mathcal{A}$ is a convex set reflecting the geometry of the manifold shapes:

$$\mathcal{A} := \left\{ V \in \mathbb{R}^{P \times (K+1)} \,:\, \min_{\mu, s^\mu \in \mathcal{S}^\mu} V_0^\mu + \sum_{i=1}^K V_i^\mu s_i^\mu \geq \kappa \right\} \tag{2}$$

where K is the number of intrinsic coordinates of a manifold (in data analysis, K will be the number of points per manifold and for convenience we usually assume every manifold have the same K). Note that in the above equation, an element of $\mathcal{A}$ is a matrix with rows and columns indexed by the superscript and the subscript of V respectively. Lastly, $\| \cdot \|_{y,C}$ denotes the Mahalanobis norm, i.e., $\|X\|_{y,C}^2 = \sum_{\mu,\nu,i,j} y^\mu y^\nu (C^{-1})_{\nu,j}^{\mu,i} X_i^\mu X_j^\nu$. Note that the above equation enables a simple algorithm to empirically

calculate the correlated manifold capacity given the covariance tensor C: simply sample a bunch of Gaussians T and random labels y and take the empirical average of the left hand side of Equation 1.

The emergence of correlated anchor geometry. To extract manifold geometry from the capacity formula of [27], i.e., Equation 1, we first perform a change of basis to rotate the correlated Gaussian measure (i.e., $D_{y,C}T$) to the isotropic Gaussian measure as follows.

$$\frac{1}{\alpha_M(\kappa)} = \frac{1}{P}\overline{\int DT \min_{V\in\tilde{\mathcal{A}}} \|V-T\|_{y,C}^2} \tag{3}$$

where $\tilde{\mathcal{A}}$ now also reflects the correlations between manifolds as follows.

$$\tilde{\mathcal{A}} := \left\{ V \in \mathbb{R}^{P\times(K+1)} : \min_{\mu,\tilde{s}\in\tilde{\mathcal{S}}^\mu} (y \odot \sqrt{C}V)_0^\mu + \sum_{j,\nu} V_j^\nu \tilde{s}_{\nu,j} \geq \kappa \right\}. \tag{4}$$

where $\odot$ denotes the entry-wise multiplication (i.e., $z = x \odot y$ where $z_i = x_i \cdot y_i$) and

$$\tilde{S}^\mu \equiv \left\{ \tilde{s} \in \mathbb{R}^{P\times(K+1)} : \exists s \in \mathcal{S}^\mu, \tilde{s}_{\nu,j} = \sum_i \sqrt{C}_{\mu,i}^{\nu,j} y^\mu s_i \right\} \tag{5}$$

Next, we follow the idea from [28] to rewrite the term with "min" in Equation 3 by the KKT condition. Concretely, the minimizer V^* of $\min_{V\in\tilde{\mathcal{A}}} \|V-T\|_{y,C}^2$ satisfies the following self-consistent equations:

$$\lambda^\mu \geq 0, \ \forall \mu \tag{6}$$

$$\min_{\tilde{s}\in\tilde{S}^\mu} (V^* \cdot \tilde{s}) - \kappa \geq 0, \ \forall \mu \tag{7}$$

$$\lambda^\mu \cdot \left[\min_{\tilde{s}\in\tilde{S}^\mu} (V^* \cdot \tilde{s}) - \kappa \right] = 0, \ \forall \mu\,. \tag{8}$$

As noticed in [28], the optimizer of the term $\min_{\tilde{s}\in\tilde{S}^\mu}(V^*\cdot\tilde{s})$ in Equation 8 is the anchor vector $\tilde{s}(y,T)$ depending on y and T. Note that there might be multiple anchor points as the feasible set of the optimization problem might be large. Finally, by putting Equation 3 and Equation 8 together, we have

$$\frac{1}{\alpha_R(\kappa)} = \frac{1}{P}\overline{\int DT (RT-\kappa\mathbf{1})^\top (RR^\top)^\dagger (RT-\kappa\mathbf{1})} \tag{9}$$

where the matrix R contains the anchor points $s(y,T)$ on its rows, $\dagger$ denotes pseudo-inverse, and $\mathbf{1}$ denotes the all one vector. Note that as $R^\top(RR^\top)^\dagger R$ is performing the orthogonal projection to the subspace spanned by the rows of R, the above equation is equivalent to Equation 1.

Effective manifold geometric measures from the capacity formula. Next, we again follow the intuition from [28] and decompose the anchor points $s(y,T)$ into the longitudinal component (i.e., manifold center) s_0 and the intrinsic component $s_1(y,T)$, i.e., $s(y,T) = s_0^\mu + s_1(y,T)$ where μ is the manifold this anchor point corresponds to. Concretely, s_0^μ is the mean of the anchor points coming from the μ-th manifold and $s_1(y,T) = s(t,T) - s_0^\mu$.

With the above decomposition at hand, we can also decompose the matrix $R = R_0 + R_1$ where R_0 contains s_0^μ on its rows and R_1 contains $s_1(y,T)$ on its rows. Thus, the gram matrix $RR^\top = G_1 + G_0$ where we define $G_1 = R_1R_1^\top$ and $G_0 = R_0R_0^\top + R_1R_0^\top + R_0R_1^\top$. Now, we are ready to define the effective manifold geometric measures for correlated manifolds. The key idea is using the Woodbury formula to rewrite $(G_1+G_0)^\dagger$ and aggregating terms based on a high-dimensional analogy of what was done in [28]. For example, in [28], it is shown that the capacity α can be approximated by $(1+R_M^{-2})/D_M$ where R_M and D_M are effective manifold radius and dimension respectively.

- Effective dimension:

$$D_M := \overline{\int DT (R_1T-\kappa\mathbf{1}) G_1^{-1} (R_1T-\kappa\mathbf{1})}\,. \tag{10}$$

- Effective radius:

$$R_M^2 := \overline{\int DT \frac{(R_1T-\kappa\mathbf{1})^\top (G_1+G_0)^\dagger (R_1T-\kappa\mathbf{1})}{(R_1T-\kappa\mathbf{1})(G_1+G_1G_0^\dagger G_1)^\dagger (R_1T-\kappa\mathbf{1})}} \tag{11}$$

- Effective center norm:

$$\ell_M^2 := \overline{\int DT \overline{\|s_0^\mu\|_2^2}} \tag{12}$$

  where the overline over the inner term is over the anchor points correspond to the same Gaussian sampling (same notation in the following).

- Effective axes alignment:

$$(\rho_M^a)_\nu^\mu := \overline{\int DT \overline{s_1(y,T)^\mu s_1(y,T)^\nu}} \tag{13}$$

  where the superscript for $s_1(y,T)^\mu$ indicates the index of the manifold that anchor point corresponds to.

- Effective center alignment:

$$(\rho_M^c)_\nu^\mu := \overline{\int DT \overline{s_0(y,T)^\mu s_0(y,T)^\nu}} \tag{14}$$

- Effective center-axes alignment:

$$(\psi_M)_\nu^\mu := \overline{\int DT \overline{s_1(y,T)^\mu s_0(y,T)^\nu}} \tag{15}$$

Finally, we remark that one can potentially design more effective manifold geometric measures through the anchor point distribution.

Code implementations and analysis pipeline. The Python implementation of our effective manifold geometric measures is written in NumPy, SciPy, and CVXOPT, and a GitHub repository will be available in the camera-ready version of this manuscript.

To run the analysis, one simply construct a list XtotT of 2D arrays with the rows being neurons/units and the columns being points in the manifold, and execute manifold_analysis_all(XtotT). The function will return a dictionary containing the resulting values of the effective manifold geometric measures. manifold_analysis_all() consists of several extra manifold analysis for comparison, one can use the argument mode to turn off those extra analyses (switching from 'expensive' to 'cheap'). The main correlated geometry analysis lies in the function manifold_analysis_corr(), which contains the following three steps:

1. prepare_manifold_data(): empirically estimate the correlation tensor C (using Cholesky decomposition from SciPy) and turn the input data into the manifold shape space, i.e., $\mathcal{S}^\mu$.
2. capacity_sampling(): conduct Gaussian sampling based on Equation 3 and compute anchor points for each sample based on Equation 8 (using CVXOPT for solving a quadratic programming problem).
3. summarize_results(): compute the correlated capacity based on Equation 1 and effective geometric measures based on Equation 10, 11, 12, 13, 14, 15.

Synthetic examples

Example on neural noise correlations. For examples in Fig. 2 and Fig. SI3, we use the following generative model:

- Parameters: $N \in \mathbb{N}$ (ambient dimension); $P \in \mathbb{N}$ (number of manifolds); $M \in \mathbb{N}$ (number of points per manifold); $\rho \in [0, 1)$ (positive noise correlations); $\psi \in [0, 1)$ (negative noise correlations).
- Procedures:
  1. Randomly sample P manifold centers $u_0^1, \ldots, u_0^P \in \mathbb{R}^N$ independently from the Gaussian distribution with mean 0 and variance $1/\sqrt{N}$ for each coordinate.
  2. For each $\mu = 1, \ldots, P$, randomly sample M vectors $u_1^\mu, \ldots, u_M^\mu$ from the Gaussian distribution with mean 0 and variance $1/\sqrt{N}$ for each coordinate.
  3. Let $A_\rho \in \mathbb{R}^{M\times M}$ be the square root of the matrix with diagonal being 1s and off-diagonal being ρ^2. Set the vectors $\hat{u}_1^\mu, \ldots, \hat{u}_M^\mu$ to be the columns of $A_\rho U$ where U contains $u_1^\mu, \ldots, u_M^\mu$ in its columns.
  4. Define the μ-th manifold to be the point cloud containing the following vectors: $u_0^\mu + \sqrt{1-\psi^2} \cdot \hat{u}_i^\mu + \psi \cdot (u_0^\mu - u_0)$ for each $i = 1, \ldots, M$ where $u_0 = \sum_\mu u_0^\mu / P$.

We pick $N = 2000$, $P = 40$, $M = 40$ in Fig. 2 and Fig. SI3 with 50 repetitions for each parameter setting. For case a in Fig. 2 and Fig. SI3, we set $\psi = 0$ and vary ρ from 0 to 0.8. For case b in Fig. 2 and Fig. SI3, we set $\rho = 0$ and vary ψ from 0 to 0.8.

Example on synthetic manifolds. For examples in Fig. SI2, we use the following generative model:

- Parameters: $N \in \mathbb{N}$ (ambient dimension); $P \in \mathbb{N}$ (number of manifolds); $D \in \mathbb{N}$ (manifold dimension); $M \in \mathbb{N}$ (number of points per manifold); $\rho_a \in [0, 1)$ (axes correlations); $\rho_c \in [0, 1)$ (center correlations); $\psi \in [0, 1)$ (center-axes correlations).
- Procedures:
  1. Randomly sample P manifold centers $u_0^1, \ldots, u_0^P \in \mathbb{R}^N$ independently from the Gaussian distribution with mean 0 and variance $1/\sqrt{N}$ for each coordinate.
  2. Let $A_{\rho_c} \in \mathbb{R}^{P\times P}$ be the square root of the matrix with diagonal being 1s and off-diagonal being ρ_c^2. Set the vectors $\hat{u}_0^1, \ldots, \hat{u}_0^P$ to be the columns of $A_{\rho_c} U$ where U contains $u_0^1, \ldots, u_0^P$ in its columns.
  3. For each $\mu = 1, \ldots, P$, randomly sample D vectors $u_1^\mu, \ldots, u_D^\mu$ from the Gaussian distribution with mean 0 and variance $1/\sqrt{N}$ for each coordinate.
  4. Let $A_{\rho_a} \in \mathbb{R}^{P\times P}$ be the square root of the matrix with diagonal being 1s and off-diagonal being ρ_a^2. For each $i = 1, \ldots, D$, set the vectors $\hat{u}_i^1, \ldots, \hat{u}_i^P$ to be the columns of $A_{\rho_a} U$ where U contains $u_i^1, \ldots, u_i^P$ in its columns.
  5. Define the μ-th manifold to be the point cloud containing the following vectors: $(1+a_j)\hat{u}_0^\mu + \sum_i \hat{u}_i^\mu b_j^i$ for each $j = 1, \ldots, M$ where a_j is independently sampled from the Gaussian distribution with mean 0 and variance ψ^2 and $(b_j^1, \ldots, b_j^D)$ is uniformly sampled from the unit sphere in $\mathbb{R}^D$.

We pick $N = 1000$, $P = 30$, $M = 30$, $D = 20$ in Fig. SI2 with 50 repetitions for each parameter setting. For each row in Fig. SI2a, we fix $\rho_a \in \{0.0, 0.9\}$, $\psi = 0$, and vary ρ_c from 0.0 to 0.9. For each row in Fig. SI2b, we fix $\rho_c \in \{0.0, 0.9\}$, $\psi = 0$, and vary ρ_a from 0.0 to 0.9. For each row in Fig. SI2c, we fix $\rho_a \in \{0.0, 0.75\}$, $\rho_c = 0$, and vary ψ from 0.0 to 0.9. For each row in Fig. SI2b, we fix $\rho_c \in \{0.0, 0.75\}$, $\rho_a = 0$, and vary ψ from 0.0 to 0.9.

Example on manifolds with a 2D ring-shaped latent organization. The implementations of dimensionality reduction methods are based on SciPy and scikit-learn. The implementations of representational similarity analysis are based on the python package rsatoolbox. Here, we give a brief overview on the three synthetic manifolds models in

Fig. 2. In all three cases, we generate 8 random spherical point clouds in a 200-dimensional space, each contains 20 points. Note that the choice of parameters is not fine-tuned and the results are very robust in a wide range of parameters. In Fig. 2a, we first generate a random Gaussian global center of expected length being 3.0. Next, we generate 2 random unit vectors and put 8 manifold centers on the unit circle spanned by these vectors with equal spacing. Finally, for each manifold, we generate 20 random Gaussian vectors of expected length 2.0 and add them to the manifold center to get the point cloud. In Fig. 2b, the generating process is exactly the same as that of the previous setting except in the final step we sample random points in a random 5-dimensional subspace (all the 8 manifolds use the same shared 5-dimensional subspace). In Fig. 2c, the generating process is exactly the same as that of the previous setting except the final step. For manifold 1,3,5,7, we generate 20 random Gaussian points of expected length 2.0 in the shared 5-dimensional subspace and a 20% bias toward the manifold center. For manifold 2,4,6,8, we generate 20 random Gaussian points of expected length 0.5. Lastly, in all the examples, we add very tiny random Gaussian noises (expected length being 0.01) to each point for numerical stability.

Analysis

Standard form. The raw datasets span a variety of formats. We compile each dataset into "standard form": a matrix whose columns are neural units and whose rows are population vectors sampled at varying task conditions and/or times throughout a trial. For vision datasets, we compile stimulus pixels into standard form as well.

Monkey early vision. The neural responses are a collection of population vectors corresponding to stimulus presentations. A population vector contains spike count in the 100 ms following response onset for each recorded neuron. In the pixel data, a population vector contains the intensity of each pixel in one image in the stimulus set.

We produce a set of manifolds for each population pixels, V1 and V2. For each population, we create 15 manifolds by grouping together neural responses to 15 different textures. Each manifold is a collection of presentation-averaged neural responses to a random subset of 10 images sampled from one texture. We randomly sample responses from a subset of 90 neurons in the ensemble. For pixel data, we use random projections to sample 90 dimensions from the 102400 pixel dimensions.

We assemble separate manifold sets from responses to scrambled and intact images. (We also compare against a random baseline where population responses are shuffled with respect to the stimuli before grouping into manifolds.)

We use GCMC to compute manifold capacity, radius, dimension, axes alignment, center alignment and center-axes alignment for each neural population. We show error bars across 10 random runs, and use a Mann-Whitney test to determine whether geometric measures differ significantly between neural populations.

Fig. SI6.

Monkey late vision. The neural responses are population vectors corresponding to stimulus presentations. A population vector contains the z-scored number of spikes from 50 to 150 ms after stimulus onset at one electrode site.

We produce a manifold set for each population pixels, V4 and IT by grouping neural responses to images of the same 3D object. Each manifold is composed of responses from a random subset of 100 units to 30 randomly selected images of the object. For pixel data, we use random projections to sample 100 dimensions from the 196608 pixel dimensions.

We assemble separate manifold sets from images at all 3 levels of variation. For a random baseline, we shuffle neural responses with respect to the stimuli before grouping into manifolds.

We use GCMC to compute manifold capacity, radius, dimension, axes alignment, center alignment and center-axes alignment for each neural population. We aggregate results across 10 random runs and apply a Mann-Whitney test for statistically significant differences between populations.

Monkey motor task. The neural responses are spike rates from each neuron at varying time throughout a reaching task. Spike rates are derived by binning spike trains in 0.1 s intervals and convolving with a Gaussian kernel of bandwidth 0.2 s.

We produce manifold sets for each motor cortex population by grouping together neural responses which occur during the same reach direction, under the same trial block, and at the same time with respect to the planning period of the trial.

We run two analyses on the manifold geometry, one for studying the timecourse of neural responses across a trial and one for studying the effects of adaptation to the motor perturbation throughout a session.

In the first analysis, we produce manifolds by grouping together neural responses which occur during the same reach direction and at the same timebin throughout the course of a trial. We use a timebin of width 0.2 s with step size 0.1 s to sample manifold sets at varying times throughout the trial. At each timestep, we assemble direction manifolds by sampling 10 neural responses within the timebin for each of the 8 reach directions. We use responses from a random subsample of 50 units for each manifold set.

In the first analysis we compute the six manifold capacity metrics for PMd and M1 separately, and show each measured value as a function of time averaged across 5 random runs.

In the second analysis we create manifolds at the session

level by sampling from neural responses occurring under the same reach direction and adaptation condition with respect to the motor perturbation. Correlations among these manifolds encode information about how the population representation changes structurally to adapt to perturbation of the task.

We delineate adaptation conditions by establishing "early" and 'late" trial blocks within the "adaptation" and "washout" epoch as well as a "baseline" block of trials before the perturbation is induced. We define early and late blocks respectively as trials before and after the median trial index throughout the epoch, restricting analysis to those trials ending in reward.

We then compute GCMC correlation matrices encoding geometric alignments among the 40 reach direction-trial block manifolds. We run the analysis separately on manifolds sampled from before and after the onset of the reaching motion. We plot correlation matrices two ways by different ordering of the columns: first by assigning to adjacent columns manifolds from the same adaptation condition, and then by assigning to adjacent columns manifolds associated with the same reaching direction.

The main text shows analysis of the curl-field perturbation session on date 2016-09-15.

Monkey perceptual decision dataset. The neural responses are spike rates from neural units at varying time throughout a reaching task. Spike rates are derived by binning spike trains in 0.1 s intervals and convolving with a Gaussian kernel of bandwidth 0.2 s.

We produce manifold sets by grouping together neural responses which occur under the same dot-motion conditions and at the same timebin throughout the course of a trial. We use a timebin of width 0.2 s with step size 0.2 s to sample manifold sets at varying times throughout the trial. At each timestep, we assemble dot-motion manifolds by sampling neural responses within the timebin from trials with the same motion coherence and direction. Each of the 12 dot-motion manifolds (6 for coherence levels and two opposing directions) consists of 20 randomly sampled population vectors from the same subset of 50 neural units. We restrict analysis to trials where the perceptual decision was correct.

We compute manifold capacity metrics for each timebin separately, and show measured values as a function of time averaged across 10 random runs. Time ranges from -1.0 s to 3.0 s where 0 is dot stimulus onset. We also show the correlation matrices for manifolds from four select timebins throughout the evidence integration period of the trial (0.0, 1.0, 2.0 and 2.2 s).

Mice spatial memory dataset. The neural responses are the inferred spiking activities during each frame of ROIs extracted from calcium images. Each ROI correspond to individual neurons identified during postprocessing. The population responses are sampled at 30.9 Hz.

We produce manifold sets by grouping together neural responses during trials with the same stimulus condition: high- or low-frequency multisensory stimulation. Each manifold is 50 randomly sampled population vectors from all recorded neurons. Manifolds are constructed from responses in the 100 ms following the go tone of each trial.

Recordings span multiple days of each individual learning the task. We run GCMC analysis on manifolds for each recording session and compare across days. For each session, we aggregate manifold metrics across 10 random runs.

Then, for one individual, we plot each measured manifold metric as a function of task performance where the points are colored by the day on which that measurement was taken throughout the course of learning. The metrics show various relationships with task accuracy which generally increases across days.

Human fMRI dataset. The neural responses are single-trial response estimates of the BOLD response amplitude to each object image from each voxel. Single-trial estimates are derived by fitting a single-trial general linear model on the fMRI time series. Responses from 19 visual ROIs are used. (V1, V2, V3, hV4, VO1, VO2, TO1, TO2, V3b, V3a, EBA, FFA, OFA, STS, PPA, RSC, TOS, LOC, and IT).

We select three higher-level categories (face, body part, and electronic device) and produce a set of manifolds for each higher-level category and for each ROI. For each ROI, we create 4-5 manifolds by grouping together neural responses to 4-5 different concepts belonging to the same higher-level category. Each manifold is a collection of neural responses to a random subset of 10 images sampled from one concept. We randomly sample responses from a subset of 400 voxels.

We assemble separate manifold sets from responses to three higher-level categories - face, body part, and electronic device.

We use GCMC to compute manifold capacity, radius, dimension, axes alignment, center alignment, and center-axes alignment for each ROI and each higher-level category. We show error bars across 50 replicates and do a two-sample t-test for statistically significant differences between higher-level categories within each ROI.

Artificial neural networks. The neural responses are extracted from a pretrained ResNet-50 architecture trained on ImageNet with the SimCLR learning algorithm. We focus on 16 layers in ResNet-50: x, conv1, layer1.0.relu, layer1.1.relu_2, layer2.0.relu_1, layer2.2.relu, layer2.3.relu_2, layer3.1.relu_1, layer3.3.relu, layer3.4.relu_2, layer4.0.relu_1, layer4.1.relu_1, layer4.2.relu, layer4.2.relu_1, layer4.2.relu_2, avgpool. For each random repetition, we fix a random projection matrix for each layer and project the neural activations to a 2000 dimensional subspace. In each repetition, we randomly select 50 categories from the 1000 ImageNet categories and randomly select 30 images from the top 10% accurate images of each category to for the manifolds (we follow the same protocol as in ref. [34, 27]). We run GCMC

analysis (as well as previous analysis methods for a comparison) on the manifolds of each layer and in total we conduct 5 random repetitions.

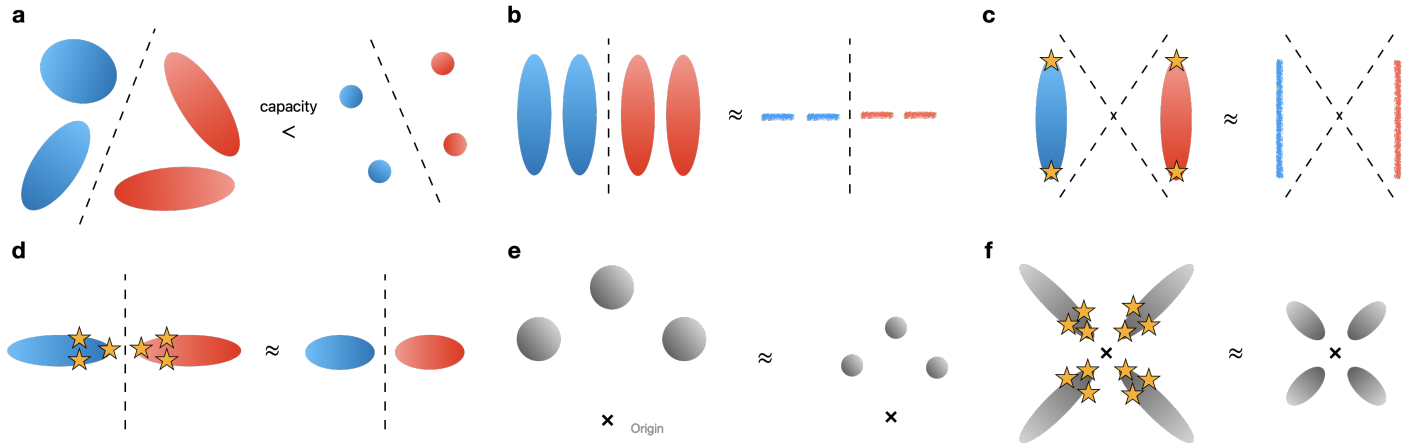

Figure SI1: Intuitive examples for effective manifold geometry. **a**, Manifolds in both left and right subfigure are perfectly separable. However, the ones on the left occupy more neural state space and are correlated in a way that deteriorate packing efficiency. Hence, the manifolds on the left exhibit higher manifold capacity. **b**, Left: When the intrinsic coordinates of the manifolds are well aligned, the optimal linear classifier would not utilize those shared subspaces. Right: Effectively, the dimension of the manifolds would not count those shared subspaces and thus is much lower than the raw dimensionality. See also Fig. SI2a for quantitative examples. **c**, Left: When the manifolds are far apart (i.e., low center correlations), the anchor points (i.e., support vectors, the starred points) of the optimal linear classifier are more concentrated to the directions with more variations. For example, in this case the anchor points will appear at the north and south pole of the ellipsoids. Right: Effectively, the dimension of the manifolds would only take those anchor points into account and hence will be lower than the raw dimensionality. See also Fig. SI2a for quantitative examples. **d**, Left: When the manifolds are more aligned to each other (i.e., high axes correlations), the anchor points will concentrate to this shared subspace. Right: Effectively, the radius of the manifolds would decrease as the anchor points concentrate to a smaller area. See also Fig. SI2a for quantitative examples. **e**, By globally scaling the neural state space by a multiplicative factor would not affect the manifold capacity. Thus, we factor out such a factor into the effective center norm so that the other effective manifold geometric measures would not change upon any multiplicative scaling. **f**, An example of the effective geometry inducing by center-axes correlations. When the center-axes correlations are high, the anchor points will concentrate toward the global origin. Effectively, this reduces the the radius and center norm, while the influence on capacity and dimension is further gated by the amount of center correlations and axes correlations in a non-monotonic manner.

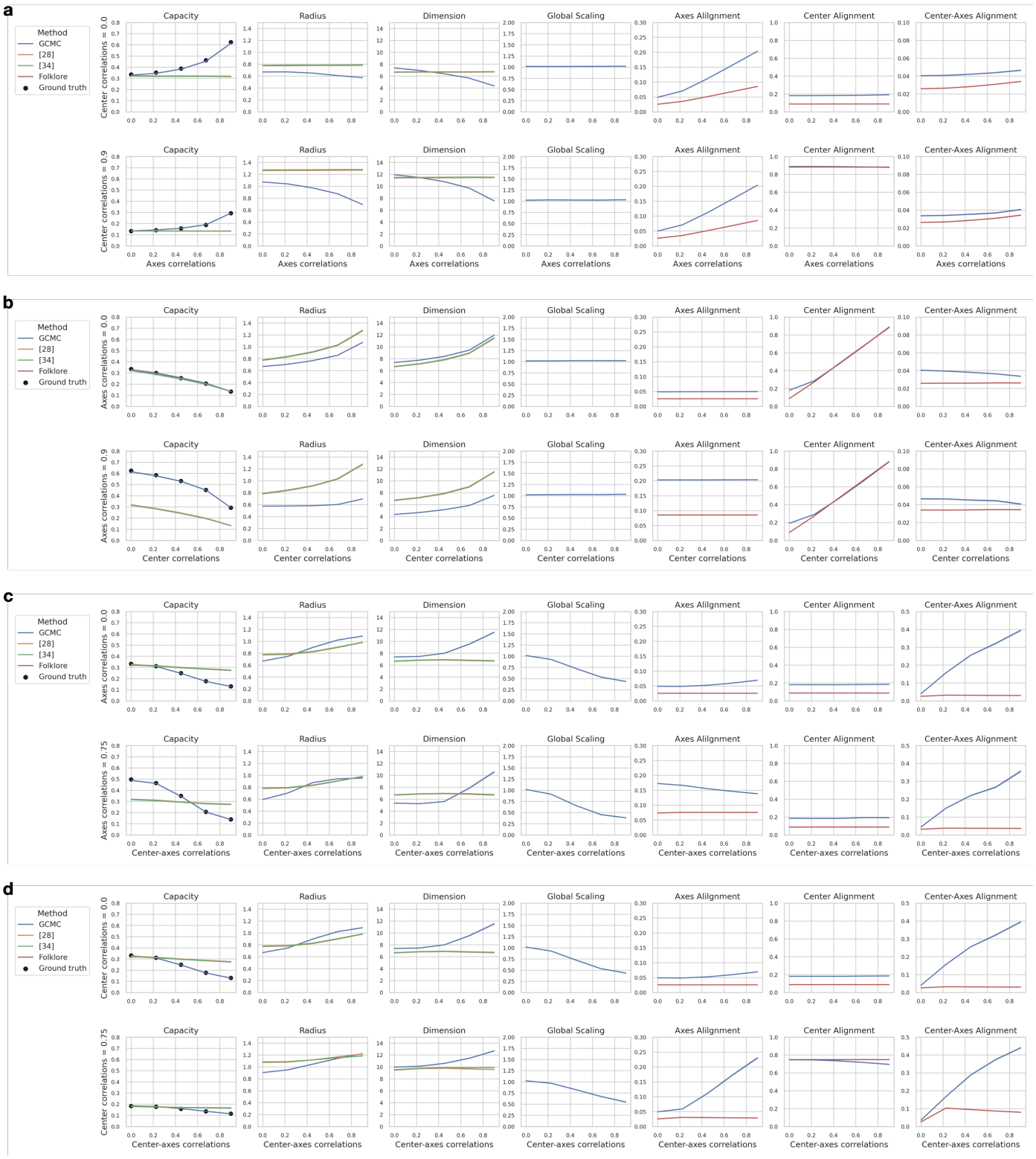

Figure SI2: Quantitative and synthetic examples for effective manifold geometry. See Methods for implementational details on the synthetic manifolds. The blue lines are the results from GCMC. The orange and green lines are the results from previous methods [28, 34] that do not take manifolds correlations into account. The black stars in the capacity plots are results from directly estimating the ground truth (a.k.a., simulated capacity) [28, 27]. **a**, Center-axis correlations is 0. For each row, we fix the axes correlations (0.0, 0.9) and varies the center correlations along the x-axis of the plots. **b**, Center-axis correlations is 0. For each row, we fix the center correlations (0.0, 0.9) and varies the axes correlations along the x-axis of the plots. **c**, For each row, we fix the axes correlations (0.0, 0.75) and varies the center-axes correlations along the x-axis of the plots. **d**, For each row, we fix the center correlations (0.0, 0.75) and varies the center-axes correlations along the x-axis of the plots.

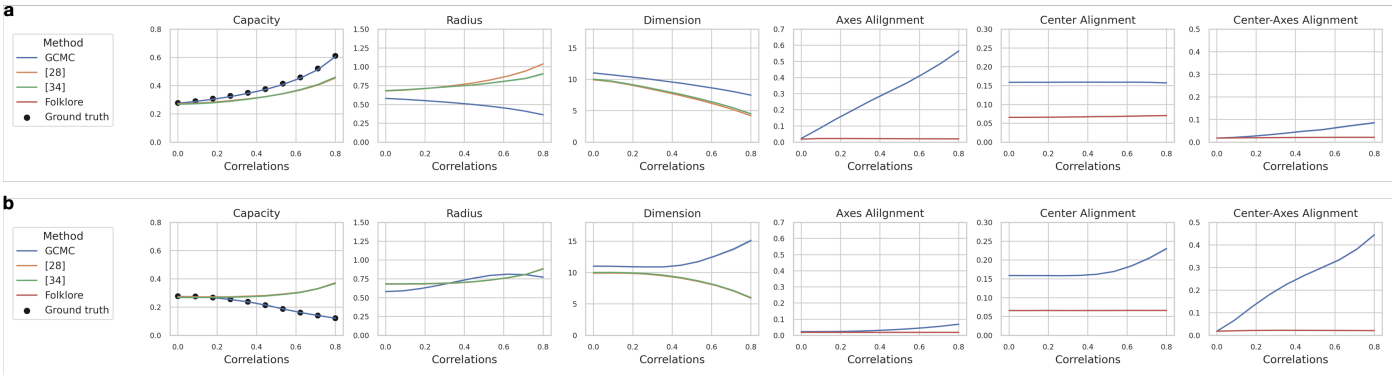

Figure SI3: Effective manifold geometry and traditional measures for noise correlations. This is an extended figure for Fig. 2. The legend for the plots is the same as that in Fig. SI2.

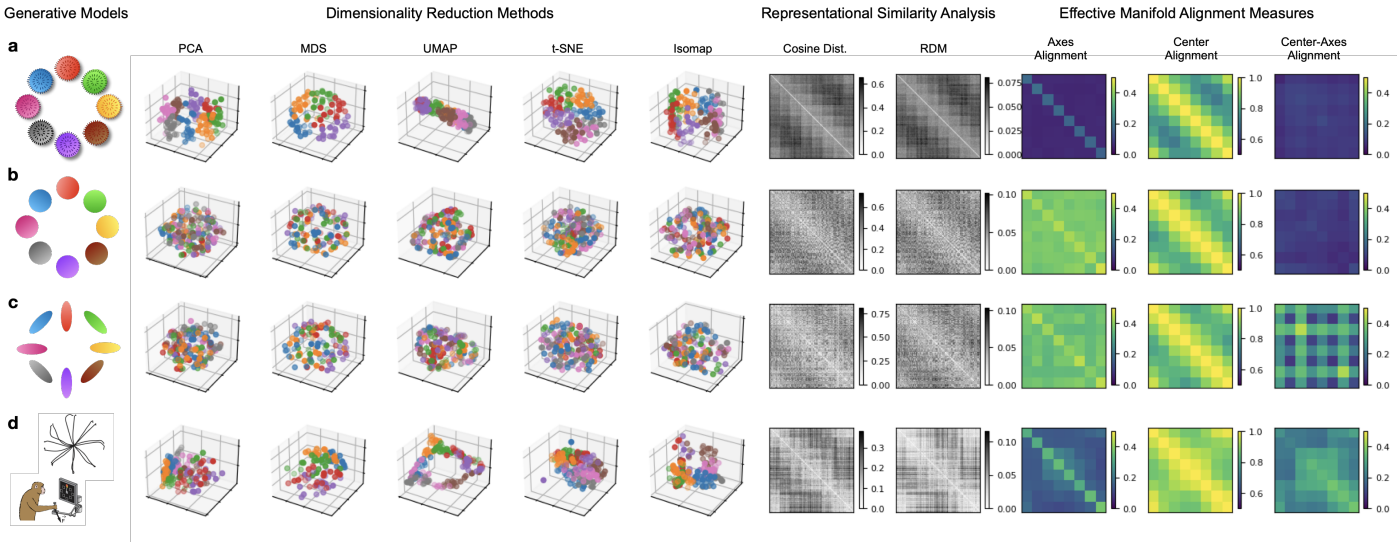

Figure SI4: **Extracting computationally relevant neural correlations into manifold alignments.** We consider three synthetic examples and one monkey center-out reaching dataset containing 8 manifolds with a 2D latent ring structure. We use common dimensionality reduction methods (PCA, MDS, UMAP, t-SNE, and Isomap), representational similarity analysis (RSA), and our effective manifold geometric measures to detect the underlying manifold structure. **a**, When the manifolds contain random independent white noises, every method can faithfully reveal the underlying ring structure. **b**, When the manifolds contain random noises residing in a shared low-dimensional subspace, dimensionality reduction methods fail horribly due to the confusion between intrinsic coordinates and manifold center directions. RSA barely captures the ring structure while manifold geometric measures successfully distill the correlations into uniform axes alignment and ring-shaped center alignment. **c**, When the manifolds have imbalanced center-axes correlations, both dimensionality reduction methods and RSA fail to detect the latent ring structure. Manifold geometric measures faithfully distill the correlations into uniform axes alignment, ring-shaped center alignment, and checkerboard-shaped center-axes alignment. **d**, In a center-out monkey reaching dataset [38], most dimensionality reduction methods could not demonstrate a clear manifold organization, RSA barely reveal a ring-shaped correlation pattern, while our manifold geometric measures distill out rich, quantitative, and interpretable manifold alignments.

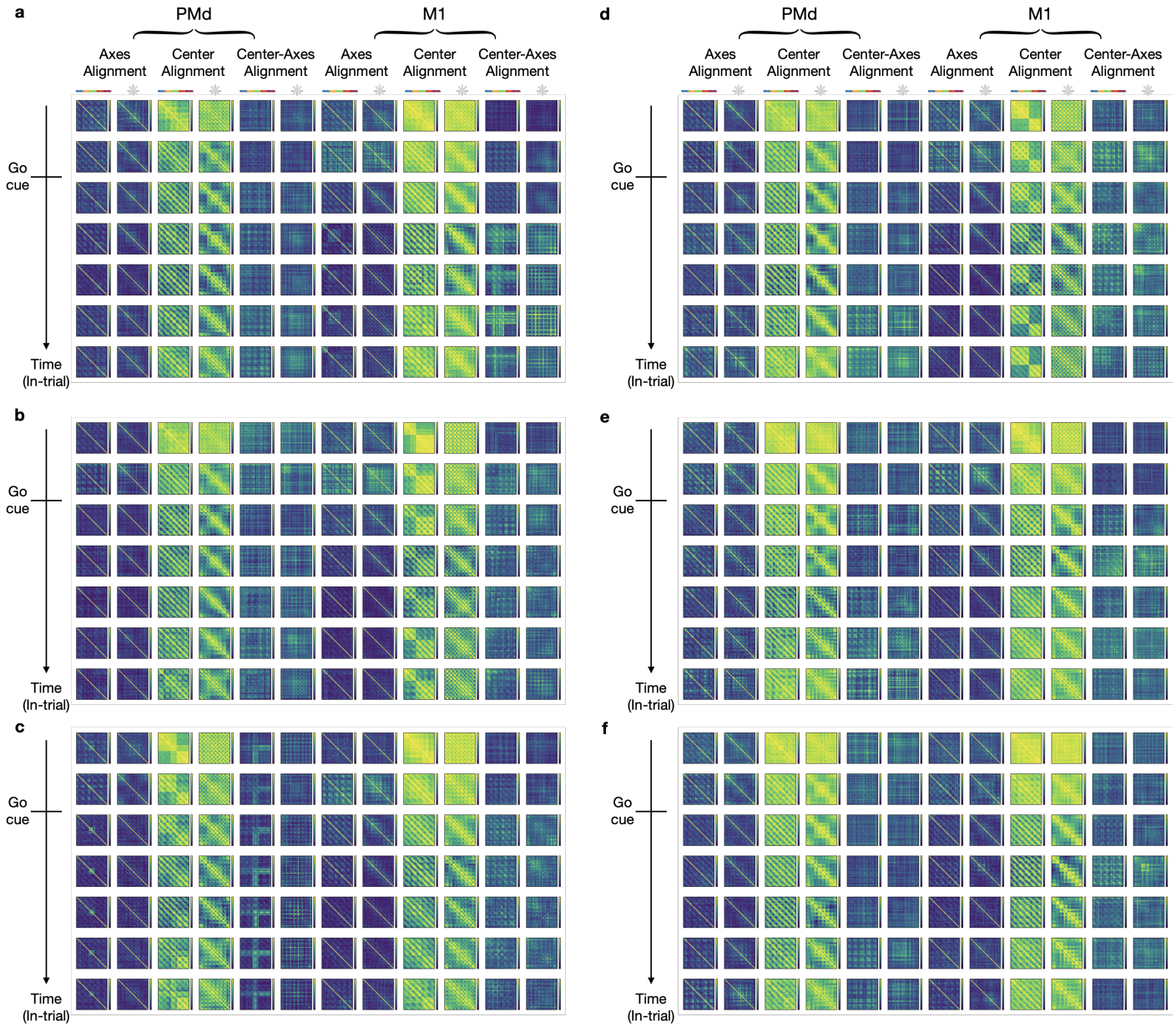

Figure SI5: Details for the analysis of the monkey reaching dataset [38]. For each subfigure, the 6 rows are time point -0.9, -0.4, 0.1, 0.3, 0.5, 0.7, 0.9. The first 6 columns are from PMd and the last 6 columns are from M1. Within the column group, the 1st column corresponds to axes alignment under epoch-wise grouping, the 2nd column corresponds to axes alignment under direction-wise grouping, the 3rd column corresponds to center alignment under epoch-wise grouping, the 4th column corresponds to center alignment under direction-wise grouping, the 5th column corresponds to center-axes alignment under epoch-wise grouping, the 6th column corresponds to center-axes alignment under direction-wise grouping. **a-c**, Curl-field task. Date: 2016-09-15, 2016-09-23, 2016-10-13. **d-e**, Visuomotor task. Date: 2016-09-12, 2016-09-29, 2016-1006.

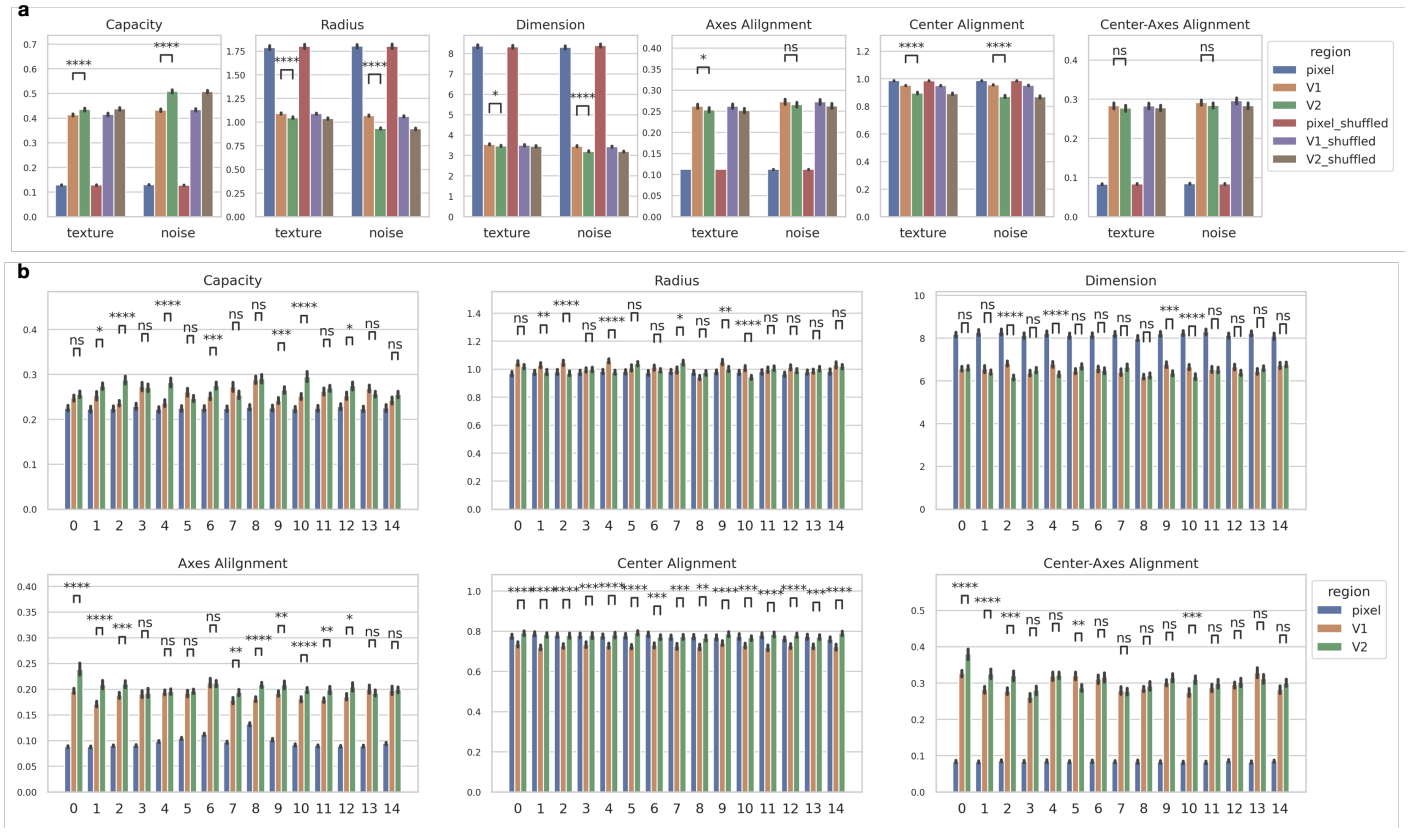

Figure SI6: Details for the analysis of monkey early vision dataset [21]. **a**, Same analysis as Fig. 3b with additional results on shuffled manifolds. Here shuffled manifolds refer to relabeling each trial to a random texture (resp. noise) family. Note that there is no significant difference between unshuffled and shuffled manifold geometry. This suggests that, in early vision, the global representational geometry dominates the local representational geometry. Errorbar stands for the standard error over the 50 repetitions of subsampling 10 trials out of the 15 trials of each image family (Wilcoxon signed-rank test, $*p < 0.05$; $**p < 0.01$; ns, not significant). **b**, GCMC analysis on texture vs. noise manifold. For each texture family (recall that there are 15 of them), we conduct GCMC analysis on the associated texture and noise manifold. This setting corresponds to the analysis done in [21]. Errorbar stands for the standard error over the 50 repetitions of subsampling 10 trials out of the 15 trials of each image family (Wilcoxon signed-rank test, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$; ns, not significant).

Figure SI7: Details for the analysis of monkey late vision dataset [22, 57]. GCMC analysis on subordinate classification. For each category family (e.g., Animals), there are 8 subordinate categories and we conduct GCMC analysis on the 8 manifolds associated to each subordinate categories. This setting corresponds to some of the analysis done in [57]. Errorbar stands for the standard error over the 50 repetitions of subsampling 35 trials out of the 40 trials of each subordinate category with a fixed variation strength (either medium or high) (Wilcoxon signed-rank test, $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$; ns, not significant).

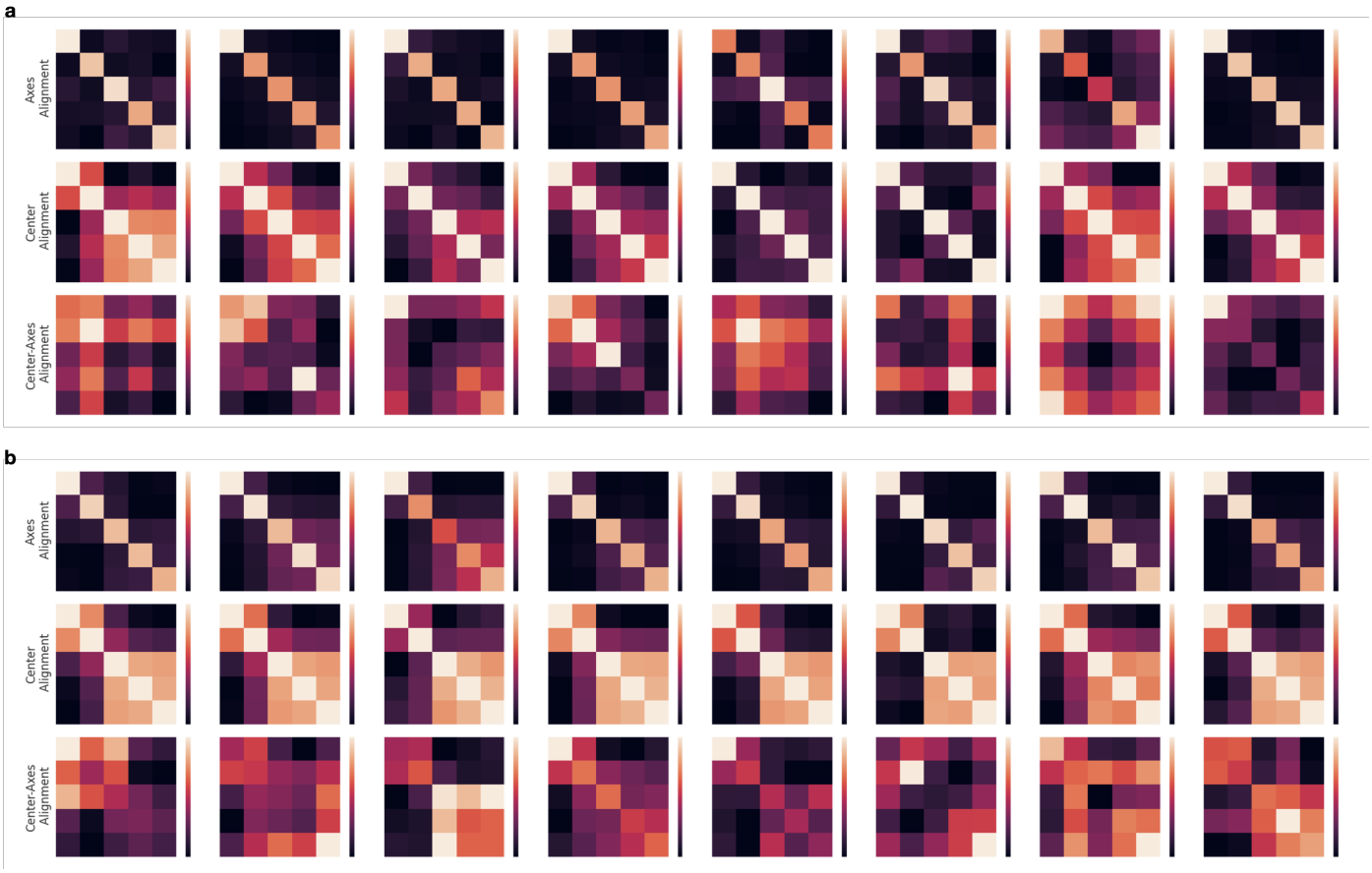

Figure SI8: Details for the analysis of mice spatial memory dataset [25]. **a**, Effective alignment measures for the morph manifolds under the frequent condition. Each column corresponds to a session and is randomly selected for presentation. **b**, Effective alignment measures for the morph manifolds under the rare condition. Each column corresponds to a session and is randomly selected for presentation. There is stronger block structure in both center and center-axes alignment matrix.

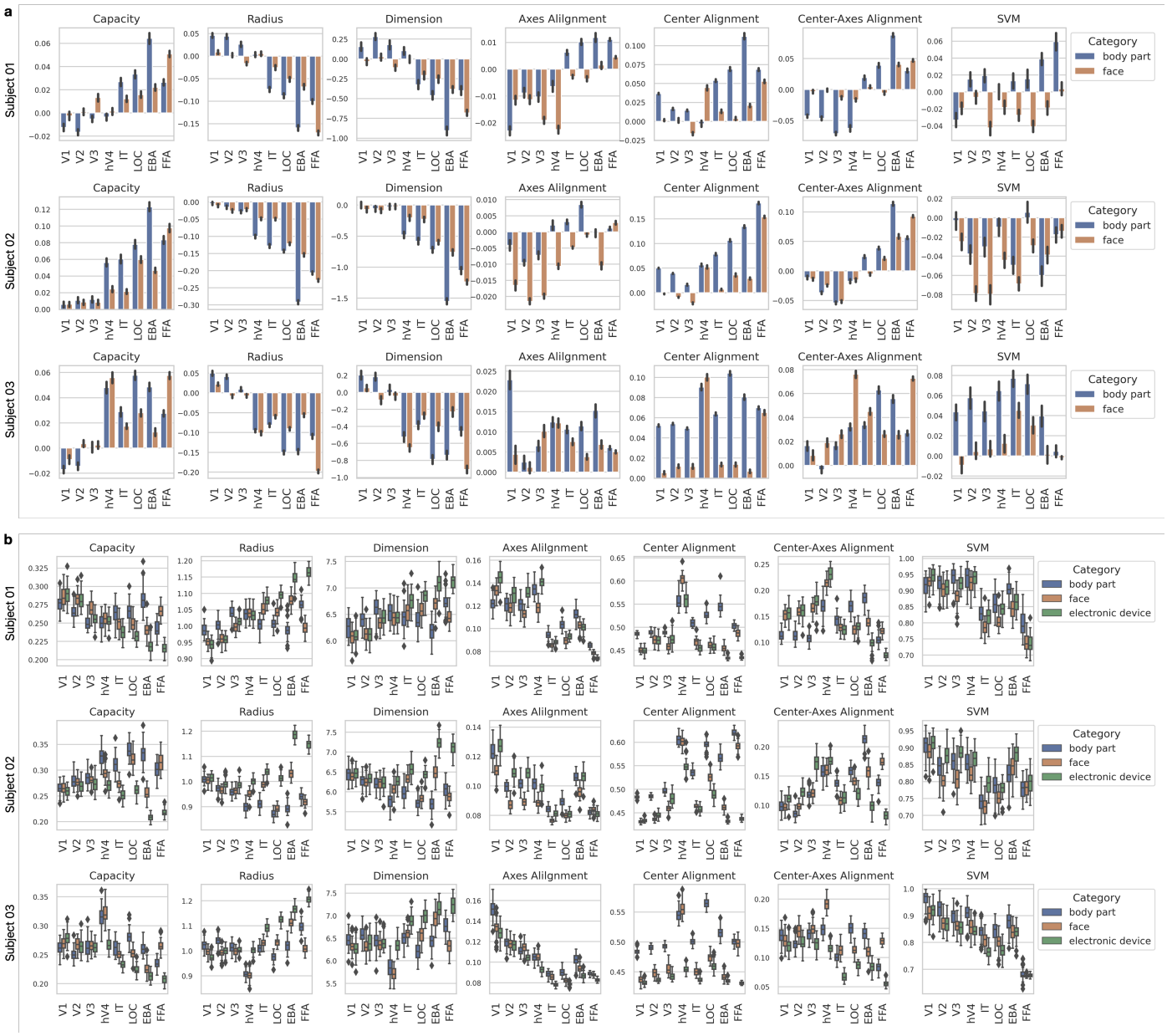

Figure SI9: Details for the analysis of fMRI dataset [26]. GCMC on the BOLD responses from all three human subjects. The GCMC analysis setting is the same as Fig. 3e with an additional result on decoding method via SVM (Methods). Notably, the SVM testing accuracy does not reveal the selectivity of EBA and FFA.

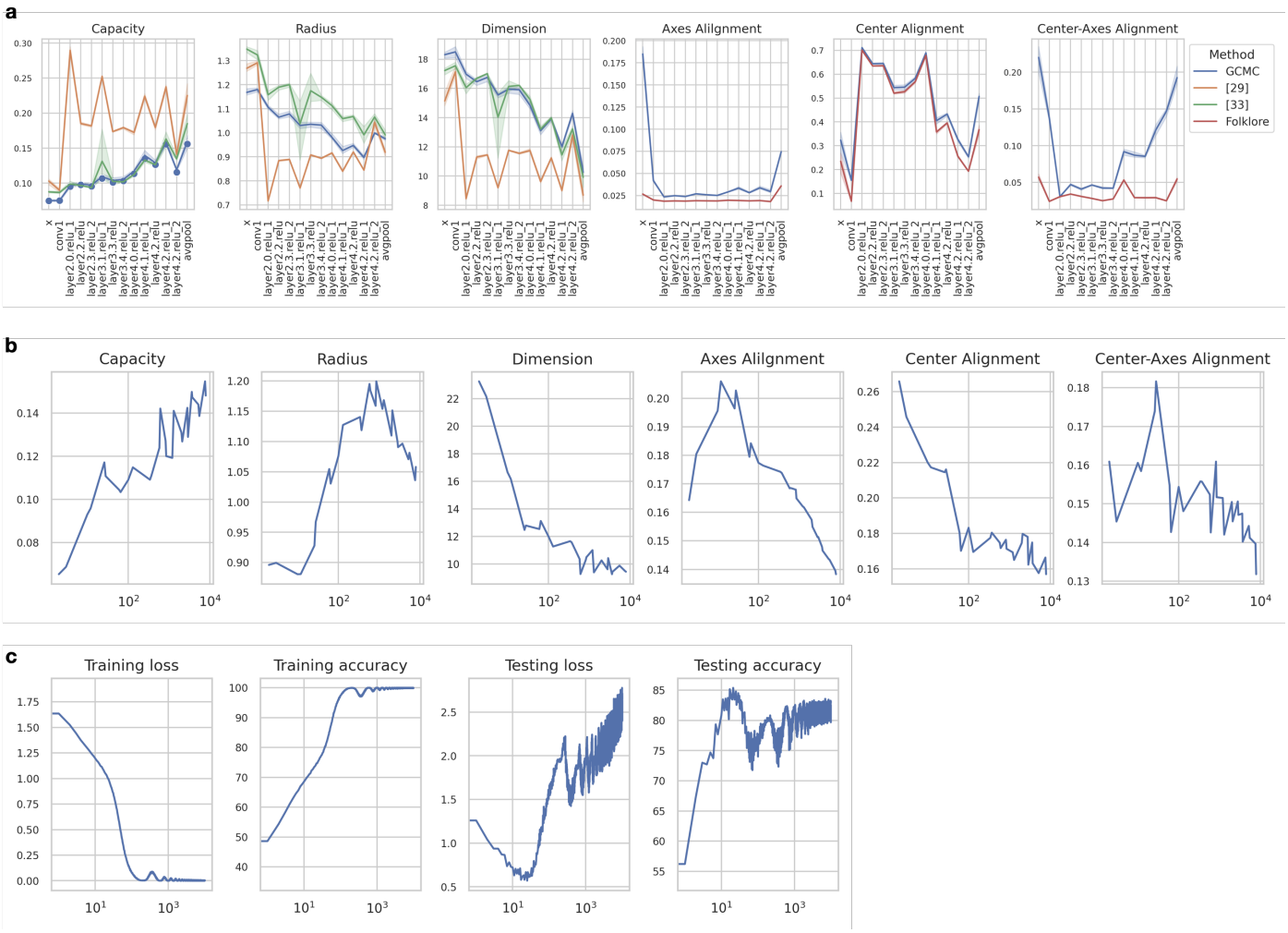

Figure SI10: Details for the analysis of artificial neural networks. **a**, Effective geometric measures of pretrained ANNs. **b**, Effective geometric measures of ResNet-18 trained on CIFAR10. **c**, Training and testing loss/accuracy. Notice that the both types of accuracy saturate very fast (whereas testing accuracy undergoes a double descent) while the effective geometric measures in **b** reveal finer details of the changes of representations.