Article

Rarely categorical, highly separable representations along the cortical hierarchy

https://doi.org/10.1038/s41586-026-10668-4

Received: 13 February 2025

Accepted: 15 May 2026

Published online: 15 July 2026

Open access

Check for updates

Lorenzo Posani[1,2,7]✉, Shuqi Wang[1,3,4,7], Samuel P. Muscinelli[1], Liam Paninski[1,5,6,8] & Stefano Fusi[1,6,8]✉

A long-standing debate in neuroscience concerns whether individual neurons are organized into functionally distinct populations that encode information differently (categorical representations[1–3]) and the implications for neural computation. Here we systematically analysed how cortical neurons encode cognitive, sensory and movement variables across 43 cortical regions during a complex task (14,000+ units from the International Brain Laboratory public Brainwide Map dataset[4]) and studied how these properties change across the sensory–cognitive cortical hierarchy[5]. We found that the structure of the neural code was scale dependent. At the whole-cortex scale, neural selectivity was categorical and organized across regions in a way that reflected their anatomical connectivity. However, within individual regions, categorical representations were rare and limited to primary sensory areas, and neuronal responses were instead very diverse. With theoretical arguments and empirical evidence, we demonstrate that the diversity of neural responses enables high-dimensional representations and therefore high separability, allowing linear readouts to separate experimental conditions in many arbitrary ways. Indeed, when accounting for information that is actually encoded in each area, all cortical regions exhibit maximal separability. Our results indicate that cortical circuits prioritize diversity over categorical structure, supporting a computational regime geared towards high-dimensional, highly separable neural representations.

Neural responses are often complex, depend on multiple variables (mixed selectivity[6]) and, in cognitive areas, are very diverse and seemingly disorganized[3,7,8]. One of the challenges in neuroscience is finding meaningful structure in seemingly disorganized data. This structure has traditionally been investigated in single-neuron responses by looking for easily interpretable tuning curves (for example, orientation selectivity in visual cortex or place cells in the hippocampus). When considering a large number of neurons, interesting structures are emerging, both in the statistics of the responses of individual neurons and at the level of population responses. Here we focused on both types of structure, how they are related to each other and what computational implications they entail. Consider the neural activity estimated in one particular time interval, recorded in different experimental conditions (for example, in response to different sensory stimuli)[3,9,10]: the responses of the different recorded neurons can be organized into a matrix (Fig. 1a) in which each row represents the responses of a neuron to all of the experimental conditions. This matrix defines two complementary spaces: one spanned by the rows, indexed by condition number, called the conditions space (Fig. 1a (red)), and the other spanned by its columns, indexed by neuron number, called the neural space (Fig. 1a (blue)).

In the conditions space, each point represents the response profile of a single neuron to the experimental conditions. If groups of neurons respond similarly, they will form clusters (functional groups) in this space, defining what was previously[1] called a categorical representation (Fig. 1a). Distinct clusters represent categories of neurons with specific functional properties. In non-categorical representations, neurons can exhibit diverse responses whose distributions do not cluster.

Categorical representations have been reported in the orbitofrontal cortex of rodents[2,11] and monkeys[12], whereas non-categorical representations were found in the rodent posterior parietal cortex[1] and in reward-sensitive frontostriatal brain areas in monkeys performing a variety of economic decision-making tasks[13]. In several other articles[6,14–17], the authors did not examine whether the representations are categorical, but the observation of diverse mixed-selectivity neurons and high-dimensional representations suggests non-categorical representations (Discussion). Whether, where and how neural populations are subdivided into functional clusters is still an open debate[3].

In the space spanned by the columns of the matrix, each point represents the population response to one experimental condition. The set of distances between all pairs of points defines the geometry of the neural representations (Fig. 1a). Analysing this geometry has been shown to provide insights into the brain's encoding strategies and their computational implications for learning and flexible behaviour[6,15–18]. For example, a high-dimensional neural geometry confers flexibility to a simple linear readout[6–8], as the same representation can support a large number of different output functions. High-dimensional

[1]Zuckerman Institute, Columbia University, New York, NY, USA. [2]ICM Paris Brain Institute, Sorbonne Université, CNRS, INSERM, Paris, France. [3]School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland. [4]School of Life Sciences, EPFL, Lausanne, Switzerland. [5]Department of Statistics, Columbia University, New York, NY, USA. [6]Kavli Institute for Brain Science, Columbia University, New York, NY, USA. [7]These authors contributed equally: Lorenzo Posani, Shuqi Wang. [8]These authors jointly supervised this work: Liam Paninski, Stefano Fusi. ✉e-mail: lorenzo.posani@icm-institute.org; sf2237@columbia.edu

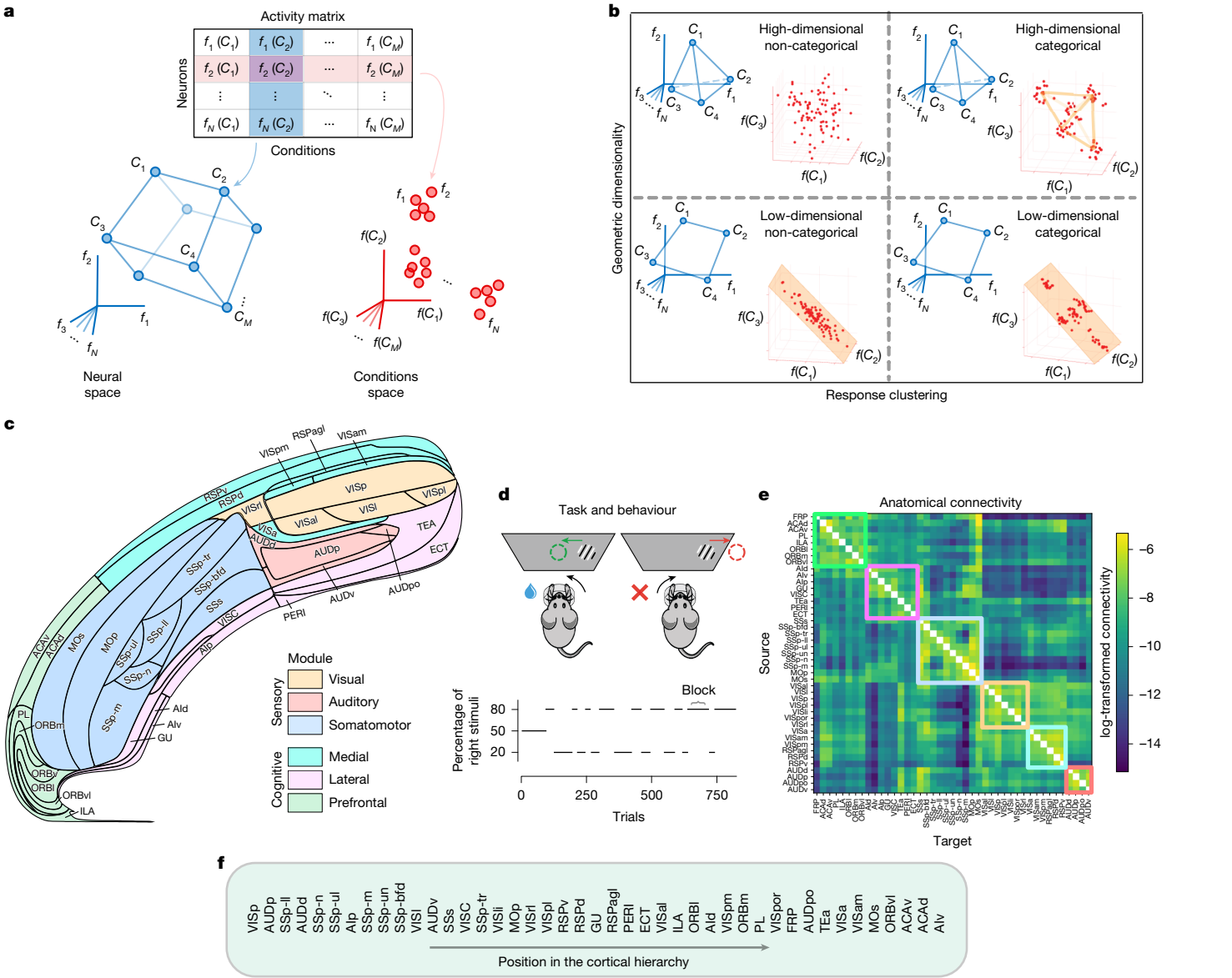


Fig. 1 | Conceptual framework and data structure. a, The matrix of neural responses of *N* neurons in *M* experimental conditions can be analysed in either its row space (conditions space, red) or column space (neural space, blue). The relative position of conditions in the neural space defines their representational geometry. If neurons are clustered in the conditions space, they define what is called a categorical representation[1–3]. **b**, The extent to which these two perspectives are related to each other is unclear, and neural populations could, in principle, occupy any portion of the clustering-dimensionality space. The image was adapted with permission from ref. 3, Elsevier. **c**, Swanson flat map of the 43 cortical regions that we analysed. Data were recorded by the IBL consortium (IBL Brainwide Map dataset) using Neuropixels probes. After selecting neurons in the cortex, we were left with around 14,000 neurons from around 180 recording sessions. **d**, In the IBL task, mice need to rotate a wheel to move a visual stimulus towards the centre of the screen. The stimulus appears left or right with an 80–20% biased probability in blocks of trials (bottom), adding prior contextual information to the task. The diagram was reproduced from refs. 4,50 under a CC BY 4.0 licence. **e**, Region-to-region anatomical connectivity in the cortex; data are from ref. 5. The coloured boxes highlight six anatomical modules with dense intramodule connectivity. **f**, Cortical hierarchy derived by the anatomical connectivity matrix of **e**. The order was derived previously[5], such that source regions (that is, regions whose connectivity is unbalanced outwards) are mostly placed low, and target regions (that is, regions whose connectivity is unbalanced inwards) are mostly placed high in the hierarchy. Additional details on how the hierarchy is computed and definitions of cortical region abbreviations can be found in ref. 5.

representations also maximize memory capacity[16], whereas low dimensionality is associated with abstract representations and generalization in novel situations[7,15,18]. How the dimensionality of population activity in the neural space is related to the clustering of response profiles in the conditions space is still an open question (Fig. 1b).

Here we systematically analysed the response profiles and the representational geometry of 14,000 neurons from 43 cortical regions (Fig. 1c) in mice performing a decision-making task (IBL public Brainwide Map dataset[4]; Fig. 1d) and related the functional properties of cortical areas with their anatomical properties as reported by the Allen Mouse Brain Connectivity Atlas[5] (Fig. 1e,f). As neuronal responses exhibit complex temporal profiles, we applied a reduced-rank regression (RRR) model to characterize both the temporal and selectivity components of individual neuron response profiles. We found that the selectivity profiles within individual areas are clustered (that is, categorical) only in primary sensory areas. When multiple brain areas are considered, we observe clustering that reflects the brain's large-scale organization. The response profiles of neurons in different brain regions differ enough that a decoder can determine the region to which a particular neuron belongs. Finally, we investigated

the computational implications of the observed representations by studying the separability of the different experimental conditions, a measure of how many different classifications in the activity space can be solved by a linear readout[6]. Brain areas with more diverse neural responses (less structured) exhibit higher separability.

Our results reveal a systematic relationship between clustering and representational geometry, providing a large-scale perspective on neural selectivity and population coding across the cortex.

Response profiles of individual neurons

To characterize the response profiles, we used a linear encoding model and, in particular, a RRR model that predicts the activity of single neurons in response to changes in a set of variables that characterize each experimental condition and the behaviour of the mouse (Fig. 2a). The core component of the RRR encoding model is a small set of temporal bases, learned from data and shared across neurons to describe their time-varying responses (see Extended Data Fig. 1 for a schematic of the RRR model and the Methods for its mathematical formulation and comparisons with previous regression models[4,19,20]). Sharing temporal bases substantially reduces the number of parameters and mitigates overfitting.

We used this model to fit the response of each neuron to eight variables that describe the IBL task (Fig. 2a). Mice are shown a visual stimulus on one side of a screen and must rotate a wheel to move it towards the centre. They are rewarded with water if they perform a correct wheel rotation. Stimuli are either shown on the left or right with a 50–50 balanced probability (50–50 block) or with an 80–20 imbalance (left or right block). In this analysis, we used only trials from unbalanced blocks to avoid confounding factors introduced by changes in neural signals over time, as the balanced blocks appeared only at the beginning of each session and units may weaken or drop out over time.

The variables used for the model range from cognitive (block) to sensory (stimulus side, contrast), motor (wheel velocity, whisking power, lick) and decision related (choice, outcome). For each trial, we considered a time window of −0.2 to 0.8 s relative to the stimulus onset. In the RRR model, each neuron is associated with a selectivity profile containing 40 coefficients (8 variables × 5 temporal bases). For each variable, we can then construct a time-varying coefficient by taking the weighted sum of the five temporal bases. These coefficients describe how sensitive the analysed neuron activity is to each variable in trial time (Fig. 2a). We then summed these coefficients over time to estimate the total effect on neural responses and obtained a selectivity profile (Fig. 2a; example neurons with large selectivity are shown in Extended Data Fig. 2). By normalizing the neuronal responses and input variables in the preprocessing steps (Methods), we ensured that the unit-free coefficients, $\beta_n^v(t)$, are not affected by the neuron's mean firing rate or the inherently different scales in different input variables and can be compared directly across neurons, input variables and time steps.

We first evaluated the predictive performance of our RRR model and compared it to that of the average firing rate per task condition (PSTH), finding a significant improvement across the whole population of neurons (RRR: mean $R^2 = 0.16$; PSTH: mean $R^2 = 0.09$, $P \approx 0$; Fig. 2b, see also Extended Data Fig. 1g,h for additional benchmarks and Extended Data Fig. 2a,b for example neuron fits). To prevent our results from being influenced by neurons that lack selectivity to any variable[4,20], we included only neurons for which the R^2 exceeded a minimum ΔR^2 threshold relative to a null model that does not incorporate variables as inputs (Methods; 4,617 neurons passed the threshold). Extended Data Fig. 6 shows that our results are robust to different values of this threshold.

Previous work has shown that neurons exhibit progressively longer autocorrelation timescales along the cortical hierarchy defined previously[5] from the Allen Institute Mouse Brain Connectivity Atlas[21–24]. Using the RRR-predicted responses, we likewise observed a significant positive correlation between a region's hierarchical position and its average response timescale (Methods and Extended Data Fig. 2d). Our analysis therefore complements previous work on intrinsic timescales of spontaneous activity using the IBL dataset[22].

Selectivity reflects anatomical organization

Consistent with the dataset's original paper[4], we found that many task-relevant variables are encoded across multiple brain regions. However, this does not mean that everything is everywhere and represented with the same strength. Indeed, we observed that the statistics of individual neurons' response profiles are distinct enough that we can guess the region that a neuron belongs to with greater-than-chance accuracy. The time-aggregated response profiles averaged across the neurons in each specific brain area are reported in Fig. 2c. We found that the cosine similarity of the average selectivity profiles for all pairs of regions is significantly correlated with the region–region anatomical connectivity reported previously[5] (Fig. 2d), indicating that anatomically distant regions have more distinct average response profiles.

We then labelled each neuron's selectivity profile with its region of origin and used a cross-validated multiclass decoding approach to assess whether the region could be decoded. The decoding accuracy is above chance both for time-varying selectivity profiles (Fig. 2e) and time-summed selectivity profiles (accuracy, 0.123). Similarly, we successfully decoded the anatomical position of individual neurons on the larger scale of brain modules (Fig. 2e), as defined in a previous article[5], which proposed these region groupings based on anatomical connectivity and known functional properties. We then considered pairs of regions and trained a decoder to report whether a neuron belongs to one or the other. The decoding performance inversely correlated with their anatomical connectivity (Fig. 2f). Note that, in all of these cases, we decoded the brain region from the selectivity profiles of individual neurons. The resulting decoding accuracy is surprisingly high, given the diversity of neuronal responses within each area, as discussed below.

Finally, we could observe systematic differences across the cortical hierarchy in the coding properties of neural populations for input variables. We performed an analysis to determine how many independent conditions (M_{IC}) a neural population encodes[25]. We defined a discrete set of conditions using combinations of four variables, chosen to span different categories (sensory, motor and cognitive), and so that each condition was well represented in the behaviour: whisking motion, block prior, stimulus contrast and stimulus side. Continuous variables (for example, whisking motion) were binarized to align with the other binary task variables (for example, block prior). By experimental design, variables like context and stimulus side are highly correlated in the biased blocks of trials. Thus, adding additional variables (such as choice or reward) was impractical due to the limited number of trials for certain variable combinations (for example, errors in context-side-matched trials). This choice of variables defines a maximum number of independent conditions $M_{IC}^{max} = 16$ (Fig. 2g).

We then estimated M_{IC} for each region using an iterative algorithm based on the cross-validated decoding performance of a linear classifier trained to distinguish pairs of conditions in a one-versus-one decoding test and grouping non-decodable conditions under a new single label. The algorithm iterates this procedure until all labelled conditions are pairwise decodable (Fig. 2h, Methods and Extended Data Fig. 3). The M_{IC} varied widely across cortical regions, ranging from a minimum of $M_{IC} = 5$ (SSp-n, the primary somatosensory area, nose) to $M_{IC} = M_{IC}^{max} = 16$ (MOs, the secondary motor area). The M_{IC} increased significantly along the cortical hierarchy (Fig. 2i), therefore encoding an increasing number of combinations of task variables. The analyses of single-cell response profiles and population decoding suggest that information is not encoded equally across cortical regions, but varies across the hierarchy in a way that reflects the large-scale anatomical organization of the cortex.

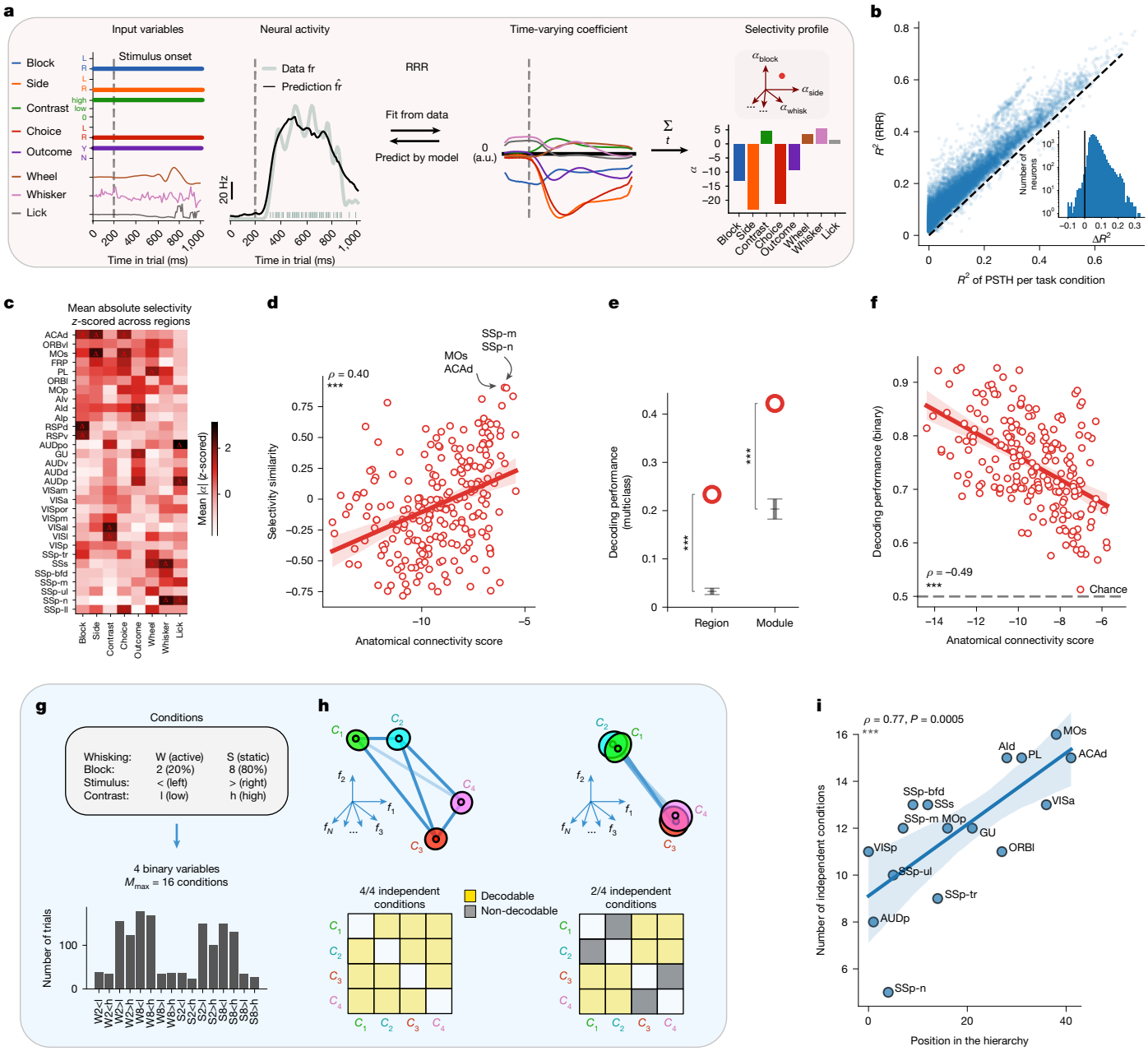


Fig. 2 | Large-scale functional organization of the cortex. a, Schematic of single-neuron selectivity estimation. A linear encoding model describes a neuron's temporal responses as a function of task and behavioural variables. Selectivity (α) is the sum of coefficients over time. Relevant variables, neural activity, estimated coefficients and selectivity are shown for an example trial. **b**, The goodness of fit (threefold cross-validated R^2) of our encoding model versus computing the PSTH per task condition. Inset: outperformance (ΔR^2). **c**, Average (absolute) selectivity profiles for neurons in the analysed cortical areas, normalized per input variable (columns). The triangles mark the top three selective areas per variable. **d**, Correlation between region-to-region functional similarity (cosine similarity between the average selectivity profiles in **c**) and their anatomical connectivity from Fig. 1e. Each dot is a pair of cortical areas (Spearman correlation, $\rho = 0.40$, $P = 1.8 \times 10^{-10}$). **e**, Multiclass decoding performance of the region or module labels from the time-varying response profiles of individual neurons (regions: accuracy = 0.233; null = 0.033 ± 0.003; $P \approx 0$; modules: accuracy = 0.422; null = 0.203 ± 0.010). Decoding performance is measured as the one-versus-one multiclass accuracy of a linear SVM. The error bars indicate the distribution of decoding performance after random label shuffling. Module labels are defined as shown in Fig. 1, following ref. 5. **f**, The binary region-to-region decoding performance is inversely correlated with the anatomical connectivity (from Fig. 1e) between the two regions involved (Spearman correlation $\rho = -0.49$, $P = 3.7 \times 10^{-14}$). **g**, Schematic of the independent conditions analysis. Conditions are defined by the largest set of combinations of motor, sensory and cognitive variables that are well represented in the behaviour. Bottom, an example of the number of trials per condition in a session. **h**, Two examples of geometries in the neural space with different numbers of independent conditions ($M_{IC} = 4$, left; $M_{IC} = 2$, right). **i**, The number of independent conditions increases along the hierarchy. The shaded areas in **d**, **f** and **i** indicate the bootstrap 95% confidence intervals of the fitted regression lines. *P* values for Spearman correlations were computed using SciPy's two-sided test; $^{***}P < 0.001$, $^{**}P < 0.01$.

Single regions are rarely categorical

We next focused on the selectivity properties of single neurons within each area and investigated whether neurons cluster into specialized subpopulations (categorical selectivity). Our RRR model associates a time-summed selectivity profile (eight-dimensional selectivity vector) with each neuron. Thus, a population of neurons within a region can be visualized as a cloud of points in an eight-dimensional

selectivity space. As a measure of clustering quality, we used the silhouette score[26] of clusters identified using *k*-means. For a given neuron, the silhouette score compares its mean distance to neurons within the same cluster with the distance to the nearest out-of-cluster points.

Figure 3a describes our pipeline for computing the clustering quality of a neural population (Methods): first, we used *k*-means to assign clustering labels. The *k* parameter is chosen to maximize the silhouette score. We then computed the silhouette score for these clusters, which we call SS_{data}. Importantly, across all clustering analyses, we considered only reproducible clusters, that is, those not dominated by neurons from a single experimental session. We then compared the silhouette score observed in the data with a null model sampled from a single Gaussian distribution (which is unimodal and thus non-clustered by design), while preserving the original data's mean, covariance structure and number of neurons. This null model is crucial for debiasing silhouette scores across varying dimensionalities (Extended Data Fig. 8b) and was inspired by the one used for the ePAIRS test previously[2]. By repeating many null model iterations, we could compare the SS_{data} value with a null model distribution of silhouette scores ($\{SS_{null}\}$). As a measure of clustering quality, we used the *z*-scored deviation of the datapoint from the null distribution.

We analysed the clustering structure of selectivity profiles in all of the recorded cortical regions. We included only regions with at least 50 neurons after applying our R^2 threshold. In Fig. 3b,c we show two examples of clustering results, for a clustered (VISp, primary visual area) and a non-clustered region (ACAd, anterior cingulate area, dorsal). To visualize the putative *k*-means clusters, we grouped neurons by cluster label and sorted them by their individual silhouette scores (highest at the top). As shown in Fig. 3b, the *k*-means algorithm clustered VISp neurons based on the block prior (c1, c2), whisking power (c3, c5) and a combination of side, contrast and choice (c4). All of the other cortical regions are shown in Extended Data Fig. 4a. Visualizing clusters in this way is often misleading because structured patterns appear whenever a few variables are more strongly encoded than others (Fig. 3b (middle, c1 and c3)). The presence of highly selective neurons for these variables does not necessarily imply that the clusters are well separated. Thus, we need to compare the silhouette score against a null model distribution to measure the clustering quality. For the VISp, the clusters do have a higher silhouette score than the null model (Fig. 3b (right)), indicating that neural selectivity is more categorical than a randomly distributed Gaussian null model. This was not the case for ACAd, a prefrontal area (Fig. 3c).

When applying this pipeline to all cortical areas, we found that only a few areas met the statistical threshold for significance ($P < 0.05$ with Bonferroni correction for multiple comparisons). The VISp and primary auditory area (AUDp), as well as the upper limb region of the primary somatosensory cortex (SSp-ul), showed a highly significant silhouette score, while the lower limb region (SSp-ll) and the gustatory cortex (GU) sit close to the threshold. To test which variables contributed to clustering, we removed from the analysis one variable at a time and measured the drop in silhouette score. As shown in Extended Data Fig. 4b, the most important variables for clustering varied across areas and were typically multimodal, spanning cognitive, movement and sensory variables (VISp: block, whisking, contrast and choice; AUDp: reward and whisking; SSp-ul: wheel and choice).

The clustering quality was significantly higher than that of the unstructured null model only in early sensory areas (Fig. 3d), suggesting that the single-neuron code becomes more diverse along the sensory-cognitive hierarchy. This negative trend was consistently found across several analysis choices such as varying the R^2 threshold for neuron inclusion (Extended Data Fig. 6b), changing the clustering algorithm (Leiden algorithm[27]; Extended Data Fig. 6c), and considering the specific temporal profiles of the time-varying selectivity coefficients on top of the sum across trial time (Extended Data Fig. 6d).

One potential weakness of the RRR-model-based analysis approach is that it depends on the assumption about the choice of the variables[2]. We therefore repeated the clustering analysis using the activity profiles in the conditions space, that is, the space of trial-average firing rates in different experimental conditions. We used the 16 conditions identified above for the analysis of independent conditions (Fig. 2g). We quantified the degree of clustering in the conditions space using a pipeline inspired by a previous study[2] (Methods and Extended Data Fig. 5): we preprocessed the mean firing rates (*z* score and dimensionality reduction), quantified the structural properties of neurons in the resulting preprocessed conditions space and compared them to a unimodal, non-clustered null model fitted to the data. Once again, we observed only a few categorical regions (VISp and AUDp; Extended Data Fig. 6f,h), but no correlation with the hierarchy. Finally, we used ePAIRS[2] to check whether there is any additional structure in the selectivity profiles that our pipeline might not capture: ePAIRS tests for deviations from a randomly mixed selectivity distribution without explicitly conducting clustering analysis. ePAIRS also yielded few categorical regions (VISp and MOs; Extended Data Fig. 6g,i).

Together, these results provide evidence that cortical areas are rarely categorical and that categorical areas are typically positioned at the lower end of the hierarchy.

Mesoscale clustering in cortical modules

The analyses of response profiles and independent conditions of Fig. 2 show that, at a larger scale, the brain is functionally organized. When considering multiple brain regions together, we therefore expect to observe more categorical representations, as each region contains one or more specialized neuronal populations. We applied our clustering pipeline to neural populations pooled across cortical modules defined previously[5]: visual, auditory, somatomotor, medial, lateral and prefrontal (Fig. 1c,e). As we excluded regions that were individually categorical (VISp, AUDp and SSp-ul), our mesoscale analysis focused on four modules: somatomotor, medial, lateral and prefrontal. To prevent over-representation by large regions, we selected the 100 best-encoded neurons from each area before pooling. The results were qualitatively similar when we considered all neurons.

This analysis revealed categorical representations in the somatomotor, medial and lateral modules (Fig. 3e), whereas the prefrontal module was not categorical (Fig. 3e). For the variables driving the module-level clustering, see Extended Data Fig. 7. We next computed the Rand index (RI), which quantifies the alignment between functional clusters and area labels (Methods). For the clustered modules (somatomotor and medial), the RI was significantly above chance (Fig. 3f and Extended Data Fig. 7), highlighting a partial alignment between clusters found by our algorithm in the selectivity space and anatomically defined regions. Finally, when we considered the entire cortex, representations were found to be categorical (Fig. 3e). In this case, nearly all variables contributed to cluster formation, and cluster labels significantly aligned with module labels (Extended Data Fig. 7b).

This analysis provides evidence that the brain is globally organized and that functional specialization relates to the anatomical organization of the cortex.

Functional implications of neural diversity

Figures 2i and 3d indicate that neural selectivity profiles are highly diverse, raising the question of what this diversity could be useful for. High diversity is important for achieving high-dimensional representations in the neural space, which, in turn, enables better separability (Fig. 4a). We first define a measure of diversity in neural response profiles that captures the structure observed in the data (α-diversity). Using empirical and theoretical arguments, we then link single-neuron response diversity to the embedding dimensionality of population

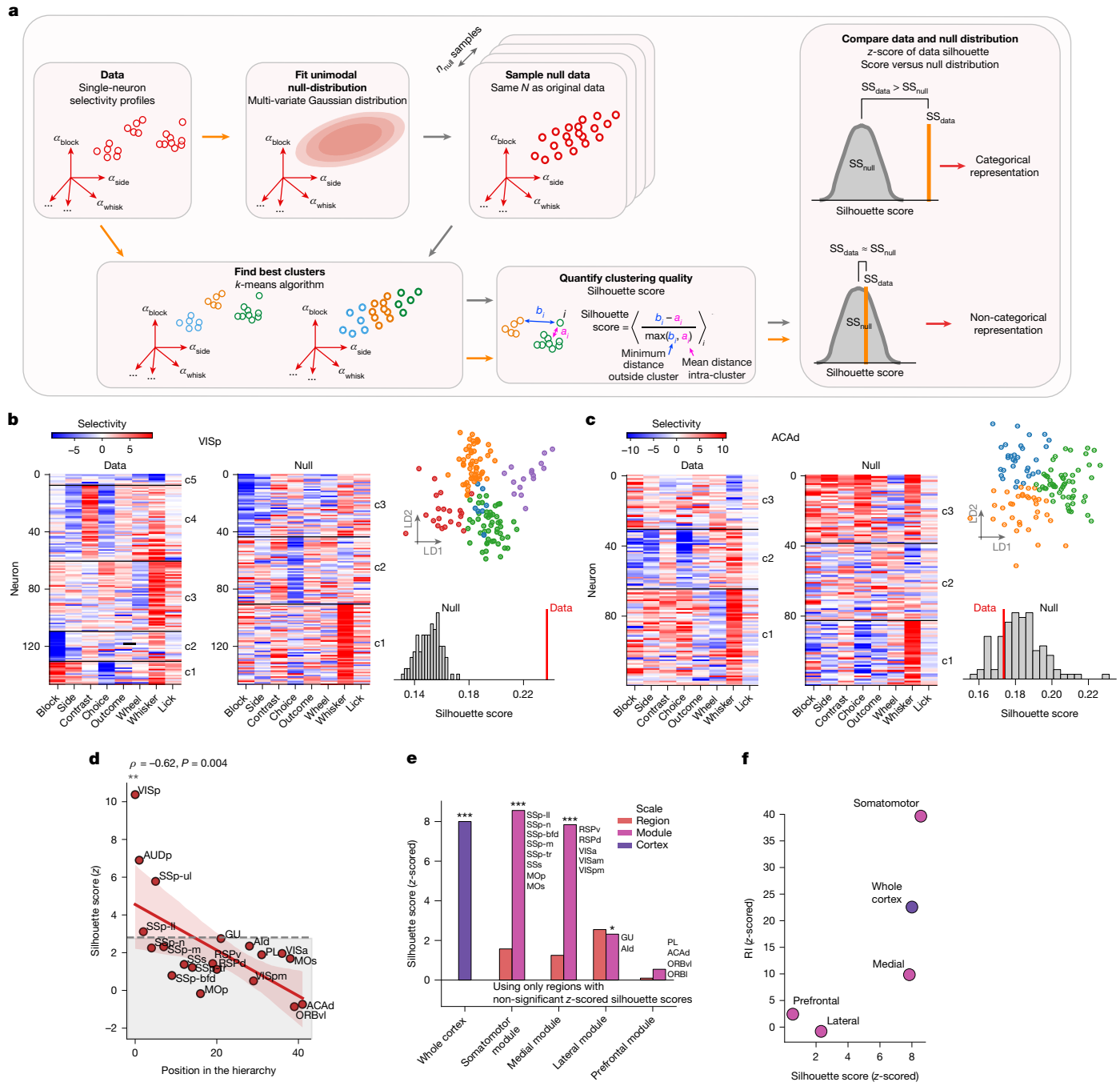


Fig. 3 | Clustering quality varies along the cortical hierarchy and across anatomical scales. a, Schematic of our pipeline to determine whether a representation is categorical. In categorical representations, neurons are clustered into multiple groups, resulting in a silhouette score that is significantly higher than that of the non-categorical null model (Gaussian distribution matched to the data, as proposed in refs. 1,2). **b**,**c**, Examples of categorical (**b**) versus non-categorical (**c**) areas. Left, selectivity matrix of individual neurons (neurons are sorted by the clustering labels and clusters are separated by black lines). Colour indicates the selectivity strength. Middle, selectivity matrix of an example null dataset. Right, data silhouette score and its null distribution (bottom) and a reduced-dimensionality visualization (top) obtained using linear discriminant (LD) analysis (points are neurons, and colours indicate cluster labels). **d**, Neuronal selectivity profiles are highly diverse for most areas, with clustering observed only in primary sensory areas. The horizontal dashed line is the threshold of significance ($P < 0.05$ with Bonferroni correction). The shaded area indicates the bootstrap 95% confidence interval of the fitted regression line. **e**, Clustering quality by pooling neurons at different scales (purple, whole cortex, $z = 8.0$, $P = 6 \times 10^{-16}$; pink, individual modules, somatosensory, $z = 8.56$, $P = 2.4 \times 10^{-17}$; medial, $z = 7.84$, $P = 8 \times 10^{-15}$; lateral, $z = 2.32$, $P = 0.04$; prefrontal, $z = 0.54$, not significant; modules were Bonferroni corrected). The orange bars show the mean clustering quality of individual regions (grouped data are from **d**). **f**, Cluster labels in well-clustered modules (medial, somatomotor) and across the whole cortex reflect the underlying anatomical organization. The similarity between anatomical labels and *k*-means clusters was quantified using the RI, *z*-scored against a shuffled null model; higher *z*-scored RI values indicate a better alignment between labels. For anatomical labels, neurons pooled within a module were assigned area labels, whereas neurons pooled across the whole cortex were assigned module labels. *P* values for Spearman correlations were computed using SciPy's two-sided test. *$P < 0.05$.

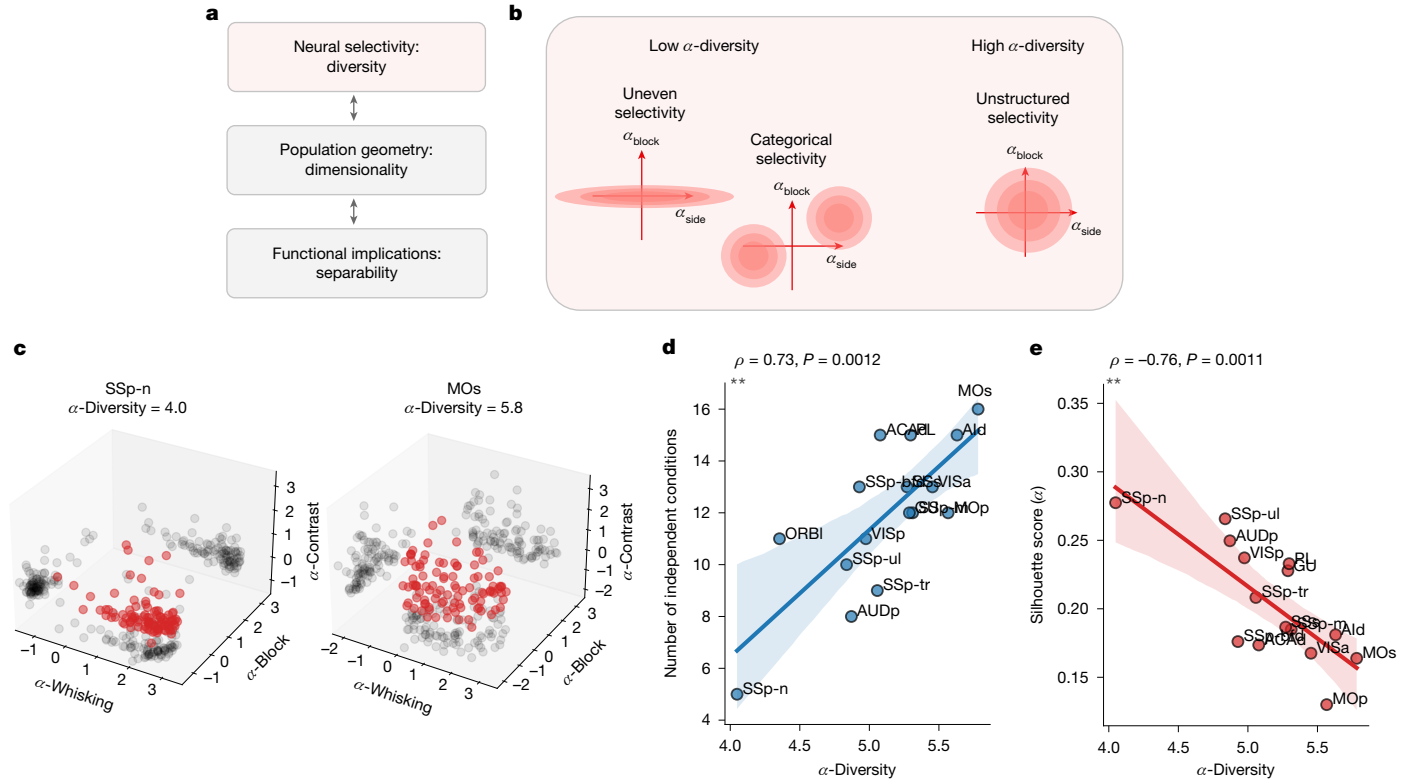


Fig. 4 | A unified measure of response profile diversity. a, Conceptual framework, integrating previous findings and current results, illustrating how diversity in neural selectivity can enhance linear separability by increasing the dimensionality of the population geometry. **b**, Schematic of the two main modes of structure observed in the data: uneven selectivity, whereby some variables or conditions are less encoded than others, and categorical selectivity, whereby neurons cluster into functionally distinct groups. **c**, Examples of response-profile distributions for two cortical areas (SSp-n, MOs). For visualization, only three variables (whisking, block, contrast) are shown, while α-diversity is defined as the participation ratio of response profiles in the full eight-dimensional space of regression coefficients. SSp-n exhibits an elongated response profile dominated by whisking encoding (uneven selectivity), and has a low α-diversity ($\simeq$4.0). By contrast, the MOs, a high-hierarchy region, displays a more uniform and diverse response distribution (α-diversity $\simeq$ 5.8). **d**, The relationship between α-diversity and M_{IC} across cortical regions. Higher α-diversity values were associated with a greater number of independent conditions (Spearman's $\rho = 0.73$, $P = 0.0012$). **e**, The relationship between α-diversity and the silhouette score in the α-diversity space. Regions with higher α-diversity showed lower clustering (Spearman's $\rho = -0.76$, $P = 0.0011$). The shaded areas in **d** and **e** indicate the bootstrap 95% confidence intervals of the fitted regression lines. P values for Spearman correlations were computed using SciPy's two-sided tests.

representations in the neural space (representation dimensionality). Finally, we show that neural response diversity is directly related to the number of dichotomies of conditions that can be decoded by a cross-validated linear classifier (separability), thereby connecting the structural properties of single-neuron selectivity to the functional properties of population-level representations.

Structure in the selectivity space

The results of Figs. 2 and 3 suggest that neural selectivity is structured in two ways (Fig. 4b): first, it is uneven as certain task variables are more strongly represented than others. This is revealed both by the diversity of the average linear regression coefficients across areas (Fig. 2c) and by the changes in the number of encoded conditions along the hierarchy (Fig. 2i). Second, it can be categorical (Fig. 3d), which is captured by the silhouette score. The silhouette score quantifies how well neurons cluster in functionally distinct subpopulations. To account for both forms of structure (or the lack of it), we introduce a new measure of diversity, which we call α-diversity. This is defined as the participation ratio in the space of linear regression coefficients (α; Fig. 2a), determined above using RRR. The participation ratio quantifies the number of dimensions spanned by a distribution of points[25,28]. Diverse responses are maximally unstructured when their distribution is an isotropic Gaussian. This distribution spans all dimensions of the α space uniformly, resulting in a high participation ratio. Structured selectivity profiles would lead to more correlated responses, a lower participation ratio and, therefore, lower α-diversity. Figure 4c shows the distribution of response profiles for two areas: the SSp-n, a low-hierarchy area, has an elongated distribution across one dominant variable, whisking, resulting in a highly uneven selectivity distribution and a low α-diversity (α-diversity $\simeq$ 4.0); whereas the MOs, a high-hierarchy area, shows a highly diverse unstructured distribution (α-diversity $\simeq$ 5.8). Both in simulations (Extended Data Fig. 8) and in the data α-diversity was directly correlated to the number of independent conditions (Fig. 4d) and negatively correlated with the silhouette score in the α space (Fig. 4e).

Thus, α-diversity provides a unified quantitative measure for characterizing neural selectivity, encompassing categorical and uneven forms of structure, as well as potentially more complex properties that fall outside these simplified models.

Diversity enables high dimensionality

High-dimensional representations have been shown to be important for separability, flexibility and memory capacity. Here we investigate how structure in single neuron response profiles affects the dimensionality of representational geometries. We consider the representation dimensionality of a set of neural activity vectors, each representing the average response of a population of N neurons to a different condition. This dimensionality can also be expressed as the participation ratio (the PR we considered in the previous section refers to a different space,

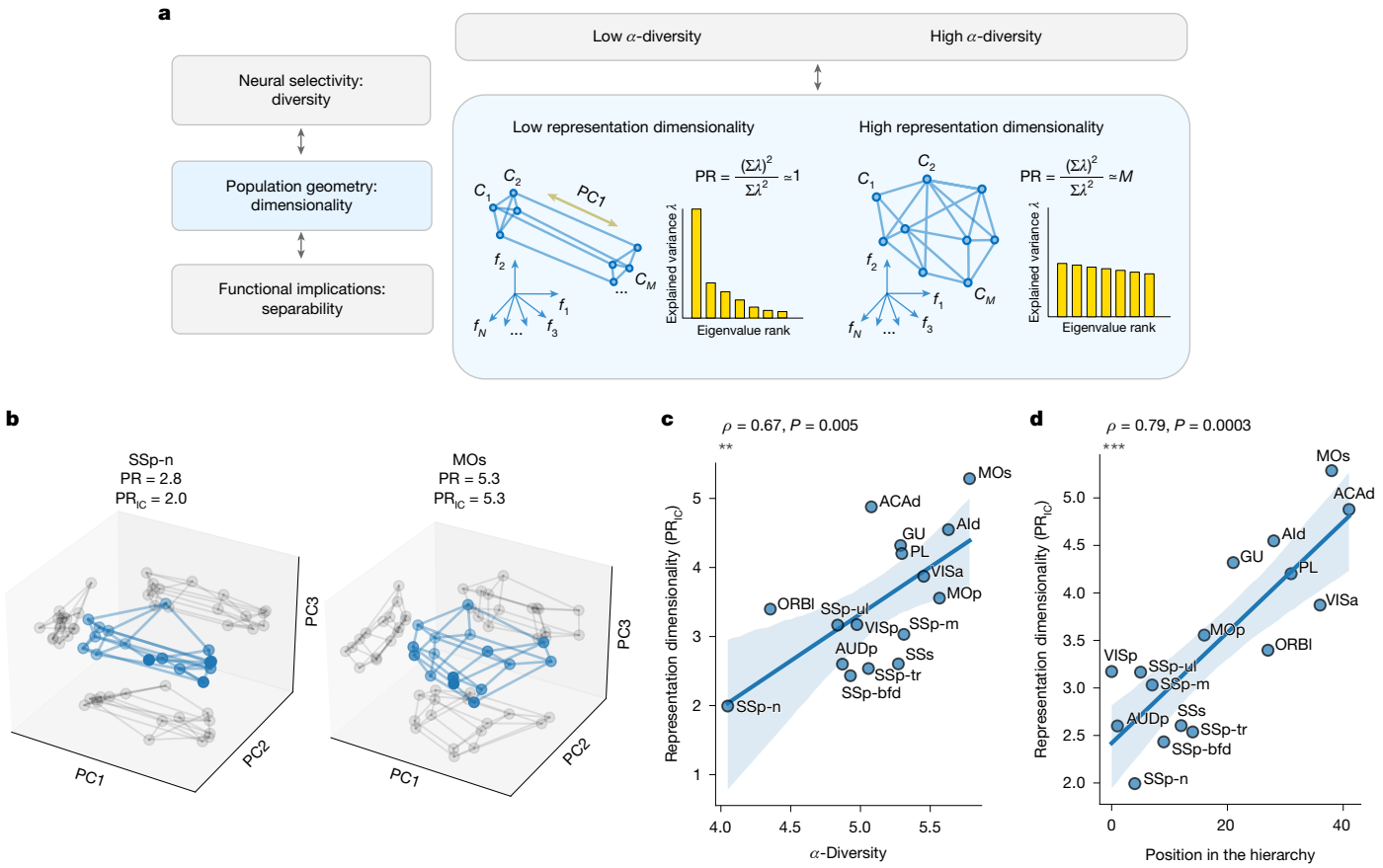


Fig. 5 | The relationship between structure in neural response profiles and embedding dimensionality of neural representations. a, Conceptual illustration of low- versus high-dimensional population geometries. When only a few eigenvalues dominate the covariance spectrum, the PR is low, indicating that most variance lies in a small number of dimensions (left). When eigenvalues are more evenly distributed, PR is high, corresponding to a more isotropic geometry (right). **b**, Example population geometries for two cortical regions (SSp-n, MOs) projected onto the first three principal components (PCs) of the $M = 16$ centroids. The low-hierarchy, low-diversity area, the SSp-n, exhibits reduced representation dimensionality compared with the MOs. **c**, Representation dimensionality is higher in areas with higher diversity of response profiles. **d**, Representation dimensionality increases systematically along the cortical hierarchy. The shaded areas indicate the 95% bootstrap confidence intervals for the regression lines. *P* values for Spearman correlations were computed using SciPy's two-sided tests.

the space of the α coefficients). We evaluated the PR using both the full set of $M = 16$ conditions defined in Fig. 2g (PR) and the subset of M_{IC} independent conditions (PR_{IC}) to account for differences in noise level and condition discriminability across conditions. The latter was taken as the definition of representation dimensionality. This choice does not qualitatively impact our results (Extended Data Fig. 10). In Fig. 5b, we show the first three principal components of the geometry of the SSp-n and MOs, already analysed in Fig. 4. The dimensionality is lower in the low-diversity area SSp-n. There is indeed a relationship between the α-diversity and dimensionality that was consistently observed across cortical regions (Fig. 5c).

To understand this relationship, we temporarily shift our focus from the space of regression coefficients (α space) to the condition space, defined by the average responses of single neurons to individual experimental conditions. Although these two spaces are linked in a non-trivial manner, structural organization in one is typically reflected in the other (for example, the VISp and AUDp are clustered both in the α space and in the conditions space; Fig. 3d and Extended Data Fig. 6f). The crucial intuition behind the relationship between structured selectivity and representation dimensionality is that the points representing conditions in the neural space (representational geometry) and the points that are the neurons in the conditions space have the same dimensionality as they are, respectively, the column and row space of the same activity matrix (Fig. 1a), and the column and row rank of a matrix is the same. Starting with this intuition, we can now say more about the quantitative relationship between uneven and categorical selectivity and dimensionality. Uneven selectivity means that certain conditions are more strongly encoded than others. This translates into distributions in the conditions space that are elongated along specific axes, resulting in a reduced participation ratio. In the data, we indeed observe a relationship between the number of independent conditions and dimensionality (Extended Data Fig. 10).

To gain an intuition about the relationship between categorical clustering and representation dimensionality, consider a scenario in which the neural response profiles are organized into $k < M$ small distinct clusters the centres of which are randomly positioned in the M-dimensional conditions space (Extended Data Fig. 9a (bottom left)). In the limit of small clusters, all of the neurons in each cluster are equivalent to a single neuron and the effective number of neurons (that is, the number of points in the conditions space) is therefore equal to the number of clusters. When the number of conditions is large enough, the dimensionality is bounded by the number of points in the space, and it therefore cannot be larger than the number of clusters ($PR \approx k < M$). Non-categorical representations allow for high-dimensional geometries (Extended Data Fig. 9a (right)). For intermediate scenarios, we computed analytically the dimensionality in the neural space as a function of properties of categorical clusters: their number k, their diversity δ and the number of conditions M (Methods and Extended Data Fig. 9b). The formula shows that more categorical representations (smaller δ) reduce the dimensionality, predicting an inverse relationship between

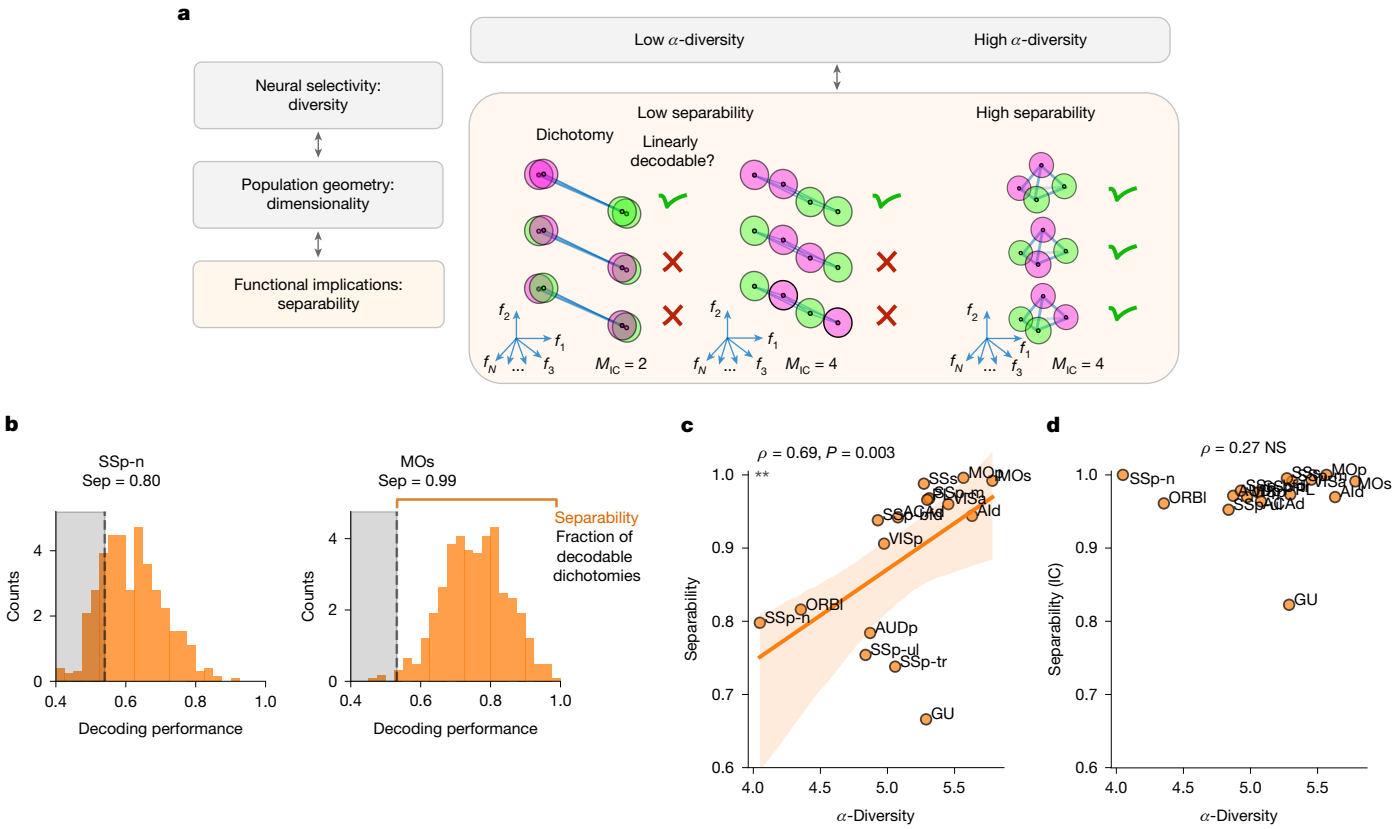


Fig. 6 | Diversity in neural response profiles is predictive of linear separability. a, Illustration of three geometries of $M = 4$ conditions with different degrees of separability, defined as the fraction of balanced dichotomies (represented by colouring schemes) that can be linearly separated in the activity space. Points represent conditions in the neural space, while clouds represent trial-to-trial noise. The geometry on the left has low separability (1/3 decodable dichotomies) owing to a lower number of independent conditions ($M_{IC} = 2$). The geometry in the middle has a maximal number of independent conditions ($M_{IC} = 4$), but still exhibits low separability (1/3 separable dichotomies) due to the collinearity of conditions in the neural space. The geometry on the right has the maximum dimensionality (three-dimensional tetrahedron) and M_{IC}, and supports maximal separability (3/3 decodable dichotomies). **b**, The distribution of cross-validated decoding performance of a linear classifier trained to decode $n = 200$ random balanced dichotomies of the $M = 16$ conditions defined in Fig. 2g for the SSp-n (left) and MOs (right). The dashed lines and shaded areas indicate the 99th percentiles of a shuffled null model for the decoding performance (Methods). Separability (sep) is defined as the fraction of dichotomies for which the linear decoding performance exceeds this threshold. **c**, Separability, computed across cortical regions, using the entire set of $P = 16$ conditions from Fig. 2g, is higher for regions with diverse neural responses. **d**, Conversely, separability of independent conditions (Fig. 2i) is always maximal (with one exception in the GU area) across all cortical regions. The shaded areas in **c** indicates the 95% bootstrap confidence interval for the regression line. *P* values for Spearman correlations were computed using SciPy's two-sided test. NS, not significant.

silhouette score and representation dimensionality (Extended Data Fig. 9b), which was observed in the data (Extended Data Fig. 9c). The theory also allows predicting the representation dimensionality from clustering properties and vice versa (Extended Data Fig. 9d,e).

As both categorical and uneven structure are more present in sensory than in cognitive areas (Figs. 2i and 3d), we expected and we observed that the representation dimensionality increases along the hierarchy (Fig. 5d), similar to what was observed in the visual hierarchy[29].

Diversity predicts linear separability

We have shown that diversity in selectivity profiles and the dimensionality of representational geometry are related. Next, we examined the functional implications of this relationship. Several studies have shown that representations characterized by a high embedding dimensionality[10] in the neural space are important for cognitive flexibility[6,7,30] or high memory capacity[16] as predicted by Cover's theorem[31], which states that a linear classifier can partition the points (corresponding here to different experimental conditions) into any two arbitrary groups (dichotomy). Here we define separability as the fraction of balanced dichotomies (different ways of dividing the conditions into two equally sized groups) that are linearly separable. A dichotomy is linearly separable if the cross-validated performance of a linear decoder is larger than that of a null model in which the labels are shuffled. As we consider a cross-validated performance, trial-by-trial variations are unlikely to make neural representations separable. In Fig. 6a, we show three schematic examples of how this notion of separability applies to different geometries. Each dot corresponds to a condition in the neural space, and the shaded area represents noise. In the first geometry (Fig. 6a (left)), only 2 out of 4 conditions are independent. This low M_{IC} implies a low dimensionality and has a direct impact on separability: only one-third of the dichotomies are linearly separable. However, separability can also be low when the number of independent conditions is maximal. This is the case of the collinear geometry of Fig. 6a (middle), in which the centroids lie along a line. There are clearly dichotomies that are not linearly separable (for example, conditions in the middle versus conditions at the borders), resulting in a low separability (1/3). Finally, the high-dimensional unstructured tetrahedron in Fig. 6a (right) has maximal separability (1). Simulations show that separability increases with dimensionality and saturates when the dimension is high enough (around $M/2$; Extended Data Fig. 8f,g). We also studied average decodability (AD), introduced previously[15] as shattering dimensionality and defined as the mean cross-validated decoding performance across all dichotomies (Extended Data Fig. 11a).

AD similarly increases with the representation dimensionality; however, in contrast to separability, it does not easily saturate (Extended Data Fig. 8g).

As diversity is important for dimensionality (see above), we expected α-diversity, separability and AD to be related. For both uneven and categorical selectivity structure (synthetic data), high neural diversity is related to high separability (Extended Data Fig. 8e). We next studied this relationship in real data. In Fig. 6b, we show the distribution of cross-validated linear decoding performance across dichotomies for the same two regions of Figs. 4 and 5, the SSp-n and MOs. For each region, we trained linear SVMs on 200 randomly sampled dichotomies of conditions and measured their cross-validated decoding performance: MOs achieves systematically higher decoding accuracy than SSp-n, resulting in higher separability. We then tested this relationship across all cortical regions. As predicted, both separability and AD increased systematically with α-diversity (α-diversity versus separability (Fig. 6c); α-diversity versus AD (Extended Data Fig. 11b)). We next examined how this relationship changes when dichotomies are defined over independent conditions only, that is, when the effects of uneven selectivity are minimized. Extended Data Fig. 8e shows using synthetic data that separability is severely limited in low-dimensional geometries, even when all pairs of conditions are separable (maximum independent conditions). Thus, by focusing on independent conditions, we can isolate the contribution of geometric dimensionality to separability. When analysing separability and AD for independent conditions in the data, we found that these quantities no longer correlated with α-diversity (Fig. 6d and Extended Data Fig. 11c). Separability was found to be near maximal ($\geq$0.95) across nearly all cortical areas, with the sole exception of the gustatory cortex (Fig. 6d and Extended Data Fig. 11e). Thus, once overlapping conditions are merged, cortical representations are not constrained by a low-dimensional geometry (Fig. 6a (middle)) but instead occupy high-dimensional, unstructured spaces that guarantee high separability (Fig. 6a (right)).

These results show that, although cortical regions differ in the number of encoded conditions and the degree of clustering, they represent those conditions with a dimensionality that is sufficiently high to achieve maximal linear separability.

Discussion

The brain has a clear anatomical organization and we observe its reflection in the organization of individual neurons' response profiles across different brain areas (functional organization). However, when one looks at the neurons within each brain area, only the responses of some primary sensory areas seem to be organized into functional clusters (categorical representations). The diversity of responses observed across brain areas has an important computational role by increasing the embedding dimensionality of representations, which, in turn, facilitates separability. Finally, all our analyses revealed that several aspects of the representational geometry and the statistics of the neural response profiles vary systematically along the cortical hierarchy: clustering decreases with position in the hierarchy, and the representation dimensionality increases, along with the number of independent conditions. All of these results show that the structure in the response profile space is related to the geometrical structure of the activity space. This relationship can help us to understand the computational implications of the neuronal response properties.

Categorical representations

In most individual brain areas, representations are non-categorical, whereas, when the entire brain is considered, they are categorical. It is important to briefly discuss the meaning of this result: according to our definition, a representation is categorical if its silhouette score is significantly different from that of a multivariate Gaussian distribution of neural response profiles. Any deviation from the null model distribution will be considered non-categorical, including distributions which do not contain well-separated clusters of neurons (for example, the continuous representations observed previously[32]). Our definition of non-categorical encompasses a relatively narrow class of unstructured distributions. Essentially, the primary structure observed in most individual brain areas is that the multivariate distribution is elongated along specific directions that differ across brain areas. Some investigators[33,34] examined representations that are categorical in a different sense: for them, 'categorical' refers to categories of stimuli, not to categories of response profiles, as we defined them here.

Modularity and cell types

In the brain, there are several different neuronal types[35], in particular for inhibitory neurons[36]. These different types of neurons are obvious candidates for potential clusters in the response profile space. One possible explanation for the lack of clustering in our data is that the discrete nature of neuronal types is not directly reflected in their functional organization. Another possibility is that the response profile reflects neuronal types that are not discrete or well separated, as suggested previously[37], in which the transcriptomic space looks only partially clustered, with some different neuronal types arranged in continuous filaments.

Clustering of responses in other species

A recent article[38] conducted a systematic analysis of structures in the space of neuronal response profiles, focusing on the visual and auditory systems of humans (functional magnetic resonance imaging) and monkeys (electrophysiology). Consistent with our results, when they examined mesoscale brain structures, they identified privileged neuronal axes that are preserved across individuals and reflect the brain's large-scale organization. Their result is compatible with our observation in Fig. 3e,f, which shows clear clustering in the case of mesoscale sensory brain structures. When they repeated the analysis at a more local scale, within category-selective regions of the high-level visual cortex, they did not find the same structure and reported a distribution of response profiles comparable to that of our null models.

Measures of separability

In all of the brain areas that we analysed, the representations are maximally separable when we consider only the independent conditions. In other words, when all pairs of conditions are separable, then all possible dichotomies are decodable. This serves as a warning for analysing neural data: when independent conditions are considered, all variables associated with potential dichotomies can be decoded better than chance. Consequently, the decodability of a specific variable often lacks significance, as all other variables are also likely to be decodable.

Separability is just one possible performance measure related to representation dimensionality. A single number cannot summarize all of the important aspects of the dependence of performance on geometry, which in turn is determined by the diversity of neuronal responses. This is why we also previously considered other measures of separability: the shattering dimensionality (here called 'average separability') introduced previously[15] considers not only whether a dichotomy is linearly decodable, but also the performance with which it is decodable. This is important as the chance level for decoding accuracy is often relatively low, and there is a wide range of accuracies above chance. This quantity is also clearly related to the diversity of the responses, both in simulations and in the data. A different approach adopted in another study[6] considered a dichotomy that was decodable only if the performance exceeded a high threshold (85%) in the limit of large N (number of neurons) and p (number of conditions). In this limit, the exact value of the threshold is actually not important, as predicted by Cover's theorem (see also the supplementary information of ref. 6). However, estimating this limit is laborious; we therefore decided to

use a different approach here, given the scale of the data and the higher noise level compared to monkey recordings.

Note that high-dimensional, unstructured representations are associated with maximal separability. However, high separability does not imply a complete absence of structure. Indeed, representations can possess the generalization properties of low-dimensional disentangled representations and still exhibit maximal separability[15].

Computational advantages of clustered response profiles

The brain is highly organized into functional and anatomical structures that can be considered to be modules. This large-scale organization is well known; it has computational implications, and its emergence can be explained using general computational principles, at least in the visual system[39–41]. However, within a particular brain area, the picture is less clear. Two main forms of modularity have been studied in theory and experiments. The first modularity (explicit) is observed when each relevant variable is represented by a segregated population of neurons (module) that forms a cluster in the selectivity space. This representation is efficient in terms of energy consumption and the number of required connections, under some assumptions[38,42]. A second type of modularity (implicit) would have the same geometry in the activity space as explicit modularity (it is simply rotated) but, in the response profile space, it is difficult to say whether clustering would be observed or not.

Both types of modularity have the same computational properties (enabling us to generalize more easily, learn new structures more rapidly[43] and even enable some form of compositionality, in which the dynamics of subcircuits can be reused in different tasks[44]) and have been characterized in artificial neural networks trained to perform multiple tasks[44,45] or to operate in different contexts[10,43,46].

Limitations of our study

Given the computational advantages of modular structures and the observation of categorical representations in some experiments[2,11,12], it is surprising that we observed significant clustering only in a few sensory areas. One possible explanation is that the assumptions of the theoretical studies on modularity are incorrect and that the biological brain operates in different regimes. For example, the metabolic advantage of modular representation could be too modest compared with the enormous baseline consumption[47].

However, there are other possible explanations. We looked for clustering using three ways of characterizing the response profile of each neuron (conditions space, regression coefficients of RRR and time-averaged regression coefficients), two different algorithms for finding clusters (*k*-means and Leiden algorithm) and two statistical tests (based on silhouette score and ePAIRS). It is possible that other approaches will reveal some form of clustering or entirely different interesting structures in the statistics of the response profiles[48]. The response profiles that we considered are based on discretized variables within the conditions space and a specific selection of variables for the continuous RRR coefficient space. Different assumptions about variable selection can yield different outcomes. We also did not consider the statistical properties of spontaneous activity. Recent work shows a relationship between spontaneous activity patterns and the intra-prefrontal cortex hierarchy[49], whereas our analysis of the response profiles did not reveal any categorical structure within the prefrontal module.

Finally, the IBL task is relatively simple, and the recorded animals are trained (often over-trained) to perform a single task. It is possible that repeating our analysis on a dataset involving multiple tasks would reveal more clustering and more specific modular structures. Indeed, a previous study[46] showed that, in RNNs, simple tasks do not require clusters, whereas complex tasks do. Similar considerations come from other theoretical studies[43,44]. Future studies on multiple complex tasks will reveal whether the organizational principles we identified are more general and valid in experiments closer to real-world situations.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-026-10668-4.

1. Raposo, D., Kaufman, M. T. & Churchland, A. K. A category-free neural population supports evolving demands during decision-making. *Nat. Neurosci.* **17**, 1784–1792 (2014).
2. Hirokawa, J., Vaughan, A., Masset, P., Ott, T. & Kepecs, A. Frontal cortex neuron types categorically encode single decision variables. *Nature* **576**, 446–451 (2019).
3. Kaufman, M. T. et al. The implications of categorical and category-free mixed selectivity on representational geometries. *Curr. Opin. Neurobiol.* **77**, 102644 (2022).
4. International Brain Laboratory et al. A brain-wide map of neural activity during complex behaviour. *Nature* **645**, 177–191 (2025).
5. Harris, J. A. et al. Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195–202 (2019).
6. Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
7. Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* **37**, 66–74 (2016).
8. Tye, K. M. et al. Mixed selectivity: cellular computations for complexity. *Neuron* **112**, 2289–2303 (2024).
9. Diedrichsen, J. & Kriegeskorte, N. Representational models: a common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Comput. Biol.* **13**, e1005508 (2017).
10. Ostojic, S. & Fusi, S. Computational role of structure in neural activity and connectivity. *Trends Cogn. Sci.* **28**, 677–690 (2024).
11. Hocker, D. L., Brody, C. D., Savin, C. & Constantinople, C. M. Subpopulations of neurons in lofc encode previous and current rewards at time of choice. *eLife* **10**, e70129 (2021).
12. Onken, A., Xie, J., Panzeri, S. & Padoa-Schioppa, C. Categorical encoding of decision variables in orbitofrontal cortex. *PLoS Comput. Biol.* **15**, e1006667 (2019).
13. Blanchard, T. C., Piantadosi, S. T. & Hayden, B. Y. Robust mixture modeling reveals category-free selectivity in reward region neuronal ensembles. *J. Neurophysiol.* **119**, 1305–1318 (2018).
14. Hardcastle, K., Maheswaranathan, N., Ganguli, S. & Giocomo, L. M. A multiplexed, heterogeneous, and adaptive code for navigation in medial entorhinal cortex. *Neuron* **94**, 375–387 (2017).
15. Bernardi, S. et al. The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell* **183**, 954–967 (2020).
16. Boyle, L. M., Posani, L., Irfan, S., Siegelbaum, S. A. & Fusi, S. Tuned geometries of hippocampal representations meet the computational demands of social memory. *Neuron* **112**, 1358–1371 (2024).
17. O'Neill, P.-K. et al. The representational geometry of emotional states in basolateral amygdala. *Nat. Neurosci.* https://doi.org/10.1038/s41593-026-02315-y (2026).
18. Courellis, H. S. et al. Abstract representations emerge in human hippocampal neurons during inference. *Nature* **632**, 841–849 (2024).
19. Pillow, J. W. et al. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**, 995–999 (2008).
20. Steinmetz, N. A., Zatka-Haas, P., Carandini, M. & Harris, K. D. Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576**, 266–273 (2019).
21. Murray, J. D. et al. A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* **17**, 1661–1663 (2014).
22. Zeraati, R., Shi, Y., The International Brain Laboratory, Levina, A. & Engel, T. A census of neural timescales across the mouse brain. In *Proc. Bernstein Conference 2024* (MPG. PuRe, 2024).
23. Song, M. et al. Hierarchical gradients of multiple timescales in the mammalian forebrain. *Proc. Natl Acad. Sci. USA* **121**, e2415695121 (2024).
24. Rudelt, L. et al. Signatures of hierarchical temporal processing in the mouse visual system. *PLOS Comput. Biol.* **20**, e1012355 (2024).
25. Gao, P. & Ganguli, S. On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* **32**, 148–155 (2015).
26. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
27. Traag, V. A., Waltman, L. & Van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
28. Litwin-Kumar, A., Harris, K. D., Axel, R., Sompolinsky, H. & Abbott, L. Optimal degrees of synaptic connectivity. *Neuron* **93**, 1153–1164 (2017).
29. Dahmen, D. et al. Strong coupling and local control of dimensionality across brain areas. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.02.365072 (2020).
30. Rigotti, M., Rubin, D. B. D., Wang, X.-J. & Fusi, S. Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Fron. Comput. Neurosci.* **4**, 24 (2010).
31. Cover, T. M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electron. Comput.* **EC-14**, 326–334 (2006).
32. Dyballa, L. et al. Population encoding of stimulus features along the visual hierarchy. *Proc. Natl Acad. Sci. USA* **121**, e2317773121 (2024).
33. Freedman, D. J. & Assad, J. A. Experience-dependent representation of visual categories in parietal cortex. *Nature* **443**, 85–88 (2006).

34. Sarma, A., Masse, N. Y., Wang, X.-J. & Freedman, D. J. Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. *Nat. Neurosci.* **19**, 143–149 (2016).
35. Zhang, M. et al. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343–354 (2023).
36. Markram, H. et al. Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.* **5**, 793–807 (2004).
37. Bugeon, S. et al. A transcriptomic axis predicts state modulation of cortical interneurons. *Nature* **607**, 330–338 (2022).
38. Khosla, M., Srivastava, S., Williams, A. H., McDermott, J. & Kanwisher, N. Privileged representational axes in biological and artificial neural networks. *Nat. Hum. Behav.* https://www.doi.org/10.1038/s41562-026-02515-3 (2026).
39. Long, B., Yu, C.-P. & Konkle, T. Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proc. Natl Acad. Sci. USA* **115**, E9015–E9024 (2018).
40. Doshi, F. R. & Konkle, T. Cortical topographic motifs emerge in a self-organized map of object space. *Sci. Adv.* **9**, eade8187 (2023).
41. Prince, J. S., Alvarez, G. A. & Konkle, T. Contrastive learning explains the emergence and function of visual category-selective regions. *Sci. Adv.* **10**, eadl1776 (2024).
42. Whittington, J. C., Dorrell, W., Ganguli, S. & Behrens, T. E. Disentanglement with biological constraints: a theory of functional cell types. Preprint at https://arxiv.org/abs/2210.01768 (2022).
43. Johnston, W. J. & Fusi, S. Modular representations emerge in neural networks trained to perform context-dependent tasks. Preprint at *bioRxiv* https://doi.org/10.1101/2024.09.30.615925 (2024).
44. Driscoll, L. N., Shenoy, K. & Sussillo, D. Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nat. Neurosci.* **27**, 1349–1363 (2024).
45. Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T. & Wang, X.-J. Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.* **22**, 297–306 (2019).
46. Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F. & Ostojic, S. The role of population structure in computations through neural dynamics. *Nat. Neurosci.* **25**, 783–794 (2022).
47. Raichle, M. E. & Gusnard, D. A. Appraising the brain's energy budget. *Proc. Natl Acad. Sci. USA* **99**, 10237–10239 (2002).
48. Chen, Y. T. & Witten, D. M. Selective inference for *k*-means clustering. *J. Mach. Learn. Res.* **24**, 1–41 (2023).
49. Merre, P. L. et al. A prefrontal cortex map based on single neuron activity. *Nat. Neurosci.* **29**, 673–681 (2026).
50. The Inernational Brain Laboratory et al. Standardized and reproducible measurement of decision-making in mice. *eLife* **10**, e63711 (2021).

Methods

Data structure

We used the International Brain Laboratory (IBL) public data release[4]. For each experimental session, we collected time-series data on task, behaviour and electrophysiological recordings. These were segmented into trials based on key task events. The task recordings collected for each trial included information on the block prior as well as stimulus contrast and location. The behaviour recordings for each trial comprised the choice made, the outcome/reward received and the time-varying movement such as wheel movement velocity, whisker motion energy and licks. Other behavioural variables, such as paw movement, body motion energy and pupil diameter traces, could be potentially included. However, we did not include them because many sessions had missing values. The electrophysiological recordings for each trial contained time-varying spike trains of recorded neurons. All of these recordings can be accessed directly through the IBL's open API. The following section describes the steps for preprocessing these raw data into data matrices for the encoding model.

Criteria for session inclusion. We iterated over all cortical regions and downloaded the related sessions. Sessions were included only if all of the behaviour recordings (wheel velocity, whisker motion energy and licks) and electrophysiological data were in place. A maximum of 30 sessions was included per cortical area to encourage a more balanced coverage. Some analyses required additional inclusion criteria, such as a minimum number of trials per condition. These analysis-specific criteria are discussed in the relevant sections below.

Criteria for trial inclusion. All trials from the left or right unbalanced blocks were included except when the animals did not respond to the stimulus in time (the first movement time was longer than 0.8 s). Trials from the 50–50 balanced block were excluded from the analysis to avoid possible time artifacts arising from the fact that all of these trials were exclusively recorded in the first 90 trials of the session.

Criteria for neuron inclusion. All neurons were included in the downloaded data provided that their mean firing rate was higher than 0.5 Hz and lower than 50 Hz. For the selectivity and geometry analyses, we included only neurons whose activity was predicted accurately enough by the RRR model described below (above a minimal threshold of $\min\Delta R^2$ with respect to a simple model that assumes that the activity is equal to the average firing rate for all conditions). Unless specified differently, we used $\min\Delta R^2 = 0.015$. This threshold was necessary to avoid confounding effects from neurons that do not encode any relevant variable, including those recorded with a low signal-to-noise ratio. Although the main results of our article remain qualitatively the same for a broad range of values of ΔR^2 (Extended Data Fig. 6), it is important to avoid the extreme cases (that is, when no neuron is discarded, or when too few neurons are selected) when studying whether the representations are categorical or not. Indeed, as already noticed previously[13], when all neurons are considered, there is the risk that 'junk' neurons with very low selectivity to all variables are over-represented, leading to a peak in the distribution around zero selectivity. This distribution would be significantly different from our null distribution (multivariate Gaussian), but that does not mean the representation is actually categorical or that there is any interesting structure in the selectivity distribution. Indeed, in that previous study[13], they used as a null distribution the superposition of two Gaussians: one representing the distribution of the selective neurons, and the other, peaked around zero selectivity, to describe the junk neurons. In our case, we decided to discard the worst neurons by selecting only cells that have a large enough ΔR^2. Again, the exact value is not important, but keeping all neurons would lead to an extra peak around zero selectivity, and a misleading, inflated number of brain areas that would pass the criterion for being considered categorical. Again, this is not a real, interesting structure and, in general, when performing this kind of analysis, we recommend checking that the structure revealed by a statistical test is not just due to the over-representation of junk neurons. Similarly, if only very few neurons are selected, the centre of the selectivity distribution might be depleted, leading again to the misleading conclusion that the representation is categorical.

RRR encoding model

In this section, we describe the RRR model used to analyse the selectivity profiles of single neurons. We start by describing the input and target variables of the model, followed by a description of the model itself and its fitting procedure. Finally, we introduce a few quantities resulting from the fitted model that are key to the follow-up analysis. The notation that will be used is summarized in the 'Notations' section. The code for implementing and fitting the encoding model is available at GitHub (https://github.com/realwsq/brainwide-RRR-encoding-model).

Input and target variables

Target variables. The target variables (y in equation (1)) were the preprocessed neuronal responses. The preprocessing steps were applied as follows:

1. For each trial, we used the spike trains of the time window −0.2 to 0.8 s relative to stimulus onset, as the first movement time is typically less than 0.8 s (ref. 4). The activity of each neuron was first binned at 0.01 s, divided by the size of the time bin and then smoothed with a Gaussian filter with a s.d. of 0.02 s. We tried to apply the linear time warping technique[51] so that the stimulus onset time and first movement or response time aligned across trials, but the results did not differ substantially.
2. The resulting activity of neuron n, denoted as fr_n, was organized into a matrix of shape $K_n \times T$, where K_n is the number of trials and $T = 100$ is the number of time steps per trial. K_n depends on n as neurons may have different numbers of trials if they were recorded in different sessions.
3. Finally, for each neuron and each time step, we z-scored the activity fr across trials to obtain the target variable y as follows (Extended Data Fig. 1b,c,d):

$$y_n(k,t) = \frac{\mathrm{fr}_n(k,t) - \mu_n(t)}{\sigma_n(t)}, \text{ where } \mu_n(t) = \frac{\sum_k \mathrm{fr}_n(k,t)}{K_n} \text{ and } \sigma_n(t) = \sqrt{\frac{\sum_k (\mathrm{fr}_n(k,t) - \mu_n(t))^2}{K_n}}. \quad (1)$$

As the squared error between the preprocessed data and model predictions was used in the loss function for optimizing the model, the applied normalization prevented biases arising from inherent differences in activity scales and ensured that the predictions for all neurons and time steps were optimized equally. Notably, the z-score transformation is easily invertible, allowing the model's predictions to be mapped back to the original units of firing rate, therefore preserving interpretability (Extended Data Fig. 1b–d).

Examples of the processed neural activity are shown in Extended Data Fig. 1b.

Input variables. The input variables (x; Extended Data Fig. 1a) that we considered can be divided into two types: discrete task-based variables and continuous movement variables. Discrete task-based variables include task-related features, such as the block prior, stimulus contrast, stimulus side, choice and outcome. These are listed below:

- Block: the prior probability for the stimulus to appear on the left side is either p(left) = 0.2 (right block) or p(left) = 0.8 (left block). We used one input variable to encode the block prior: −1. representing p(left) = 0.2, and +1. representing p(left) = 0.8. As noted above, we excluded trials from the p(left) = 0.5 unbiased block.
- Contrast: the stimulus contrast is 0%, 6.25%, 12.5%, 25% or 100%. One variable was used to encode the stimulus contrast: 0, representing 0% contrast; 1, representing the low contrast (≤12.5%); and 4, representing the high contrast (>12.5%).
- Stimulus: the stimulus location is either on the left side (+1) or the right side (−1).
- Choice: the choice is indicated by the turning of the wheel: clockwise (+1) or counterclockwise (−1).
- Outcome: the outcome is either a water reward (+1) or negative feedback (−1).

As these values are static, all of the timepoints share the same values (Extended Data Fig. 1a).

Continuous movement variables included both instructed (for example, licking and wheel velocity) and uninstructed (for example, whisker motion energy) movement. These were as follows:

- Wheel: the velocity of the wheel movement (radian per second) per time bin.
- Whisking: the whisker motion energy per time bin is calculated as the motion energy for a square of the left/right camera roughly covering the whisker pad. The maximum value between the left and right whisker motion energy was used.
- Lick: the number of licks per time bin.

A few preprocessing steps were applied separately to each movement variable:

(1) For each trial, we first read out the continuous behaviour of the time window −0.2 to 0.8 s relative to stimulus onset and interpolated into 0.01 s time bins.
(2) Then, to account for the activity that was shifted in time, for each session, we computed the mean time-lagged correlation between the neuronal activity and the movement traces averaged across neurons and trials and shifted the movement traces so that the zero-lagged correlation was maximized.
(3) Last, significant differences were observed in the variance of input values across trials, both for different input variables and different time steps. To ensure optimal performance and clarity in interpretation, we z-scored the values for each input variable and time step across trials in the same way as in equation (1). The reasoning behind this normalization step is twofold. From a performance perspective, as an extra regularization term was incorporated to penalize the high-value coefficients, the scales of the coefficients and, therefore, the scales of the input values should be comparable for the regularization to work properly. From an interpretability standpoint, the inherent differences in scales must be carefully addressed to enable the comparison of coefficients across input variables and time steps.

Examples of the resulting input variables are shown in Extended Data Fig. 1a.

The model formulation

Linear encoding model. For each neuron n, we describe its temporal responses as a linear, time-dependent combination of input variables (a visual illustration is shown in Fig. 2a and Extended Data Fig. 1):

$$y_n(k,t) \approx \hat{y}_n(k,t) = \sum_{\upsilon} \beta_n^{\upsilon}(t) x^{\upsilon}(k,t), \quad \forall\, k,t,n, \tag{2}$$

where

- $y_n(k,t)$ is the preprocessed neuronal activity of the trial $k \in \{1, \ldots, K_n\}$ and time step $t \in \{1, \ldots, T\}$.
- $\hat{y}_n(k,t)$ is the corresponding model prediction, given by the value of the equation on the right side.
- υ represents the relevant input variables included in the model. $x^{\upsilon}(k,t)$ is the preprocessed value of the input variable υ for the trial k and time step t.
- $\beta_n^{\upsilon}(t)$ is the effect size of the input variable υ at time step t. It is further referred to as the regression coefficient.

Low-rank coefficient matrix. The time-varying coefficients $\boldsymbol{\beta}_n^{\upsilon} \in \mathbb{R}^T$ are the weighted sum of a set of temporal basis vectors shared across all of the neurons and input variables, that is

$$\boldsymbol{\beta}_n^{\upsilon} = \mathbf{U}_n^{\upsilon}\mathbf{V}, \quad \forall\, n, \upsilon. \tag{3}$$

Here, $\mathbf{U}_n^{\upsilon} \in \mathbb{R}^d$ is the neuron n and input variable υ-dependent loading of temporal basis vectors $\mathbf{V} \in \mathbb{R}^{d \times T}$. Specifically, we considered sharing a single set of temporal basis vectors across all of the neurons from all of the brain regions and across all of the input variables. We verified that this restriction did not compromise the goodness-of-fit. The rank d is generally a value much smaller than the number of time steps T. See Extended Data Fig. 1e for an example decomposition. Sharing the temporal bases across neurons and input variables significantly reduces the number of parameters (Extended Data Fig. 1f). Let N be the number of neurons, T be the number of time steps and $|\upsilon|$ be the number of input variables. An unconstrained full-rank coefficient matrix uses $N \times |\upsilon| \times T$ parameters, while a reduced-rank coefficient matrix of the same shape only needs $N \times |\upsilon| \times d + d \times T$ parameters. As N and T are typically much larger than d, the reduction in parameters is on the order of T, that is, around 100-fold.

Comparison to previous regression models

Well-established linear encoding models include generalized linear model (GLM)[4,19] and kernel regression model[20]. Below, we first describe these two models and then discuss how they compare to our RRR model. GLM is expressed as

$$y_n(k,t) \approx \hat{y}_n(k,t) = \sum_{\upsilon} \sum_{t'} \beta_n^{\upsilon}(t') x^{\upsilon}(k, t-t'), \forall\, k,t,n, \tag{4}$$

where the input filter vector, $\boldsymbol{\beta}_n^{\upsilon}$, composed of $\beta_n^{\upsilon}(t')$ over a neighbouring time window, is factorized as

$$\boldsymbol{\beta}_n^{\upsilon} = \mathbf{U}_n^{\upsilon}\boldsymbol{\kappa}^{\upsilon}, \forall\, n, \upsilon. \tag{5}$$

Here, $\boldsymbol{\kappa}^{\upsilon}$ represents pre-specified temporal basis vectors, which are not trained. Adapted from a previous study[4] (here we considered neural responses over a different time window from[4] and an enriched set of behaviour movements), we used the same set of raised cosine 'bump' functions in log space for Extended Data Fig. 1h. The kernel regression model uses the same linear formulation as the GLM (equation (4)). However, instead of relying on pre-defined temporal basis vectors, it allows these vectors to be trainable:

$$\boldsymbol{\beta}_n^{\upsilon} = \mathbf{U}_n^{\upsilon}\mathbf{V}^{\upsilon}, \forall\, n, \upsilon. \tag{6}$$

Both the kernel regression model and our RRR model fall into the category of RRR models. The primary distinction lies in the set of input features used by each approach.

In our RRR model, the influence of input variables on neural responses, $\beta_n^{\upsilon}(t)$ can vary over time. By contrast, both the GLM and the kernel regression model assume that the influence is time-independent. Therefore, for example, whether the mouse response time is early or late makes no difference and will modulate the neural responses the same way. Instead, these models provide a more descriptive account of input effects by allowing neural responses to depend on inputs from

neighbouring time steps. While this feature is absent in our current RRR model (equation (2)), it could be incorporated in future extensions. A performance comparison is shown in Extended Data Fig. 1h.

Estimation of the parameters

The parameters of the RRR model include a shared temporal bases matrix V of size $d \times T$ and loading vectors $\mathbf{U}_n^v$ of length d for each input variable v and neuron n. The approach we adopted to fit the parameters was to minimize the ridge-penalized mean square loss:

$$\mathcal{L}(\mathbf{V}, \{\mathbf{U}_n^v\}_{n,v}) = \sum_n \left(\sum_k \sum_t (y_n(k,t) - \hat{y}_n(k,t))^2 + \lambda \sum_v \sum_t \beta_n^v(t)^2 \right). \quad (7)$$

Minimizing this particular loss function is straightforward as a closed-form solution exists[52]. In practice, we chose to use the L-BFGS optimization algorithm to compute the optimum.

Moreover, to optimize the model hyperparameters, namely the rank d and the regularization penalty λ, we implemented a threefold cross-validation technique across trials. First, the dataset was stratified based on a composite target label that included the block prior and stimulus contrast to ensure that each fold was representative of the entire dataset. Then, for each combination of d and λ, the dataset was partitioned into three subsets by trials, using each subset in turn for testing the model while the remaining data served as the training set. Finally, the combination of d and λ that yielded the lowest average test error across all folds was selected. $d = 5$ turned out to be the optimal number of temporal bases.

Estimating the goodness of fit

We used the threefold cross-validated R^2 to measure the goodness-of-fit of single-trial predictions. For each session's data, we randomly sampled one-third of the trials as the test set held out during training. Once the model was trained—using the remaining two-thirds of trials—we computed the R^2 between the model predictions $\widehat{\mathrm{fr}}_n(k,t)$ ($\widehat{\mathrm{fr}}_n(k,t)$ is calculated by inversing the z-score transformation applied in the preprocessing step $\widehat{\mathrm{fr}}_n(k,t) = \sigma_n(t)\hat{y}_n(k,t) + \mu_n(t)$; Extended Data Fig. 1b–d) and the actual neuronal activity $\mathrm{fr}_n(k,t)$ in the test set. We repeated the whole split–train–test process three times and computed the mean of the three cross-validated R^2 as the measure of goodness-of-fit.

Null model and selectively modulated neurons. Conceptually, we distinguish two types of task modulation: the selective modulation and the non-selective modulation. Selective modulation, captured by $\hat{y}_n(k,t) = \sum_v \beta_n^v(t) x^v(k,t)$ is induced by the input variables and varied trial by trial. Non-selective modulation, captured by the mean time-varying response $\mu_n(t) = \frac{\sum_k \mathrm{fr}_n(k,t)}{K_n}$ is locked to the key events of the trial (stimulus onset in this case) and does not vary trial by trial. Both types have an important role in modulating neuronal responses. See Extended Data Fig. 2b for example selectively modulated (left) and non-selectively modulated (right) neurons. In this work, we focus mostly on the neuronal responses selectively modulated by the task. To distinguish the variation explained by the selective modulation from the non-selective modulation, we use the trial-average estimate as the null model that does not consider any effect of the input variables:

$$\hat{y}_n^{\mathrm{null}}(k,t) = 0, \ \forall k, t, n. \quad (8)$$

The outperformance, ΔR^2, defined as:

$$\Delta R^2(\mathrm{model}) = R^2(\mathrm{model}) - R^2(\mathrm{null}), \quad (9)$$

captures the overall selective modulation of all of the input variables combined.

Only selectively modulated neurons, identified as $\Delta R^2(\mathrm{RRR}) \geq 0.015$ (Extended Data Fig. 1g), are included in the further analysis.

Computing the selectivity profiles of single neurons to the input variable

To compute the selectivity profiles of single neurons to the individual variables, we used the estimated coefficient $\beta_n^v(t)$. For the clustering analysis of Fig. 3 and Extended Data Fig. 6, we took the sum of the coefficients across time as a measure of the total selectivity of neuron n to input variable v.

$$\alpha_n^v = \sum_t \beta_n^v(t) \quad (10)$$

Note that by normalizing the neuronal responses and input variables in the preprocessing steps, we ensured that the unit-free coefficients $\beta_n^v(t)$ are not affected by the neuron's mean firing rate or the inherently different scales in different input variables and can be compared directly across neurons, input variables and time steps. Thus, $\beta_n^v(t)$ can be interpreted as the expected change in normalized neuronal activity y_n per one s.d. change in the input variable x^v at time t, and α_n^v can therefore be thought as the expected total change across the whole trial.

The selectivity α_n^v captures whether individual neurons are selectively modulated by the given variable v or not (examples of strongly selective neurons are shown in Extended Data Fig. 2c). In the selectivity analysis (Fig. 2e,f), when the goal is to estimate the absolute modulation of an input variable, α_n^v is calculated as $\alpha_n^v = \sum_t |\beta_n^v(t)|$.

Computing the autocorrelation timescale of neural responses to task and behaviour variables

Given a matrix of single-neuron responses to task and behaviour variables $\hat{y}_n \in \mathbb{R}^{K \times T}$ with K being the number of trials and T the number of time steps per trial, we can compute the corresponding autocorrelation timescale.

To compute the timescale, we first calculate the time-lagged unnormalized autocorrelation sequence

$$c_n(i) = \frac{\sum_k \sum_t \hat{y}_n(k,t)\hat{y}_n(k,t+i)}{K}, \ i \geq 0.$$

We then linearly interpolate the autocorrelation sequence so that $c_n(i)$ is spaced at 1 ms resolution (original 10 ms). The timescale τ_n is approximated by the time the sequence first reaches half its peak value (that is, $c_n(0)$). The timescale of brain area a is further determined by averaging the values over all of the selectively modulated neurons within this area.

We observed a significant correlation between a region's hierarchical position and its estimated autocorrelation timescale (Extended Data Fig. 2d).

Notations

- Indices:
  - n, N: index and number of neurons ($n \in \{1, \ldots, N\}$).
  - k, K: index and number of trials ($k \in \{1, \ldots, K\}$).
  - t, T: index and number of time steps ($t \in \{1, \ldots, T\}$).
  - $v, |v|$: index and number of input variables ($v \in$ {block prior, stimulus contrast, stimulus side, choice, outcome, wheel velocity, whisker motion energy, lick}).
  - i, d: index and rank of the RRR model ($i \in \{1, \ldots, d\}$).
- RRR model
  - $\beta_n^v(t)$: regression coefficient (arbitrary units (a.u.)) of neuron n, input variable v at time step t.
  - $\mathbf{U}_n^v \in \mathbb{R}^d$: neuron n and input variable v-dependent loading of temporal basis vectors (a.u.).
  - $\mathbf{V} \in \mathbb{R}^{d \times T}$: temporal basis vectors (a.u.).
- Random variables:
  - $x^v(k,t)$: preprocessed value (a.u.) of input variable v, trial k at time step t.

- $y_n(k,t)$: preprocessed neuronal response (a.u.) of neuron n, trial k at time step t.
- $\hat{y}_n(k,t)$: model prediction of preprocessed neuronal response (a.u.) of neuron n, trial k at time step t.
- $fr_n(k,t)$: smoothed, binned firing rate (Hz) of neuron n, trial k at time step t.
- $\widehat{\mathrm{fr}}_n(k,t)$: model prediction of smoothed, binned firing rate (Hz) of neuron n, trial k at time step t.
- $\mu_n(t)$: mean of smoothed, binned firing rate (Hz) of neuron n at time step t.
- $\sigma_n(t)$: s.d. of smoothed, binned firing rate (Hz) of neuron n at time step t.

Clustering analysis

To test for the presence of functional clusters, we followed the steps explained below. The required inputs include each relevant neuron's response profile and original session ID. Two types of response profile can be considered: the estimated selectivity to individual input variables (equation (10), referred to as clustering analysis in the variable selectivity space, used in Fig. 3 and Extended Data Fig. 6) or the average response in each experimental condition (referred to as clustering analysis in the conditions space, used in Extended Data Fig. 6). Performing clustering analysis in the selectivity space arguably has a few advantages: (1) It reduces the dimensionality in an interpretable and informed way. If we have $|v|$ variables, then there are at least $2^{|v|}$ conditions, assuming all of the variables are discrete and have more than one different value. (2) It mitigates the issue of unbalanced or even missing conditions. (3) It reduces the noise in the estimation of the response profile. As shown in Fig. 2b and Extended Data Fig. 2, neural responses are very noisy, and simple averaging may be non-satisfactory. By contrast, the selectivity estimated from the encoding model provides a more reliable account of the task-driven variance in the neural responses.

The code for the clustering analysis is available at GitHub (https://github.com/realwsq/clustering-analysis).

Clustering analysis in the variable selectivity space. We summarize our clustering pipeline as follows (a schematic is shown in Fig. 3a). (1) Check whether there are more than 50 neurons and only continue if so. (2) Given the selectivity profile of each neuron, run the k-means clustering algorithm (with 100 random initializations) with the number of clusters k varied from 3 to 20. (3) Then, select the optimal clustering result by maximizing the silhouette score. The silhouette score is defined as $<\frac{b_i-a_i}{\max(b_i,a_i)}>_i$ where i is the index of the neuron, $a_i=\frac{1}{|C_I|-1}\sum_{j\in C_I, i\neq j} d(i,j)$ is the mean Euclidean distance intracluster and $b_i=\min_{J\neq I}\frac{1}{|C_J|}\sum_{j\in C_J} d(i,j)$ is the minimum Euclidean distance outside cluster (Fig. 3a). (4) Iterate over the resulting clusters and check whether there is a cluster whose total silhouette score summed over all neurons is mainly contributed by neurons from one single session (>90%). If so, remove neurons from that cluster and session, and repeat steps 1–4. (5) Sample the same number of datapoints from the Gaussian distribution with the mean and covariance matrix matched to the data values and compute the sampled data's null silhouette score according to steps 2 and 3. (6) Repeat step 5 100 times and pool the null silhouette scores to form the null distribution. Finally, compute the z score of the data silhouette score with respect to the null distribution.

Clustering analysis in the conditions space. When clustering in the space of mean firing rate, two additional preprocessing steps are required. First, we normalized each neuron's mean firing rates separately across conditions to prevent clustering driven solely by overall firing rate differences between neurons. Second, as the number of conditions and the dimensionality of the activity profile is high, we reduced the dimensionality using principal component analysis. Moreover, we modified the null model by replacing the multivariate Gaussian distribution with a multivariate log-normal distribution to better capture the lower-bounded and heavy-tailed nature of mean firing rates. Neurons with zero activity were excluded from this analysis.

The clustering analysis in the conditions space follows these steps (a schematic is shown in Extended Data Fig. 5): (1) check whether there are more than 40 neurons and only continue if so. (2) z-Score the mean firing rates for each neuron. (3) Reduce the dimensionality using principal component analysis, retaining components that capture 90% of the total variance. (4) Run the k-means clustering algorithm (with 100 random initializations), varying the number of clusters k from 3 to 20. (5) Select the optimal clustering result by maximizing the silhouette score. (6) Iterate over the resulting clusters and check whether there is a cluster whose total silhouette score summed over all neurons is mainly contributed by neurons from one single session (>90%). If so, remove neurons from that cluster and session, and repeat steps 1–5. (7) Sample the same number of datapoints from the multivariate log-normal distribution with the mean and covariance matrix matched to the log data values, then compute the sampled data's null silhouette score following steps 2–5. (8) Repeat step 7 100 times and pool the null silhouette scores to form the null distribution. Finally, compute the z score of the data silhouette score with respect to the null distribution.

Measuring the similarity between cluster and area labels. The similarity between functional clusters and anatomical area labels was quantified using the RI, which measures the agreement between two labellings of the same dataset. For each pair of neurons, agreement occurs if both are assigned to the same cluster in both labellings, or to different clusters in both. The RI is defined as the fraction of agreeing pairs among all possible pairs. To assess statistical significance, we computed a z-scored RI by comparing the observed RI to a null distribution obtained from 10,000 random shufflings of the area labels. A significantly elevated z-scored RI indicates that functional clustering aligns closely with anatomical organization. To avoid biases towards areas or modules with larger neuron numbers, we considered the 100 best-encoded neurons from each module or area in the analysis of Fig. 3e,f and Extended Data Fig. 7.

Modified ePAIRS test

Given the average responses of single neurons in each experimental condition, we performed the ePAIRS test as follows (a schematic is shown in Extended Data Fig. 5): (1) z-score the mean firing rates for each neuron. (2) Reduce the dimensionality using principal component analysis, retaining components that capture 90% of the total variance. (3) Calculate the cosine distance to its nearest neighbour for each neuron. (4) Calculate the empirical median of these nearest-neighbour distances as the aggregated nearest-neighbour angle. (5) Sample the same number of datapoints from the multivariate log-normal distribution with the mean and covariance matrix matched to the log data values, then compute the null nearest-neighbour angle of the sampled data according to steps 1–4. (6) Repeat step 5 5,000 times and pool the null nearest-neighbour angles to form the null distribution. Finally, the z-score of the data aggregated nearest-neighbour angle with respect to the null distribution is computed.

α-Diversity

To measure α-diversity, we took the participation ratio of the $N\times V$ matrix of α coefficients resulting from the RRR analysis described above. The PR quantifies the effective dimensionality of a set of datapoints by measuring how evenly the variance is distributed across the eigenvalues of its principal component analysis decomposition[28]. The PR is defined as:

$$\mathrm{PR}=\frac{\left(\sum_j \lambda_j\right)^2}{\sum_j \lambda_j^2}, \tag{11}$$

where λ_j is the jth eigenvalue of the $N \times N$ covariance matrix. A higher PR indicates that the variance is more evenly spread across multiple dimensions, suggesting a higher effective dimensionality of the cloud of points in the high-dimensional space. Conversely, a lower PR implies that the variance is concentrated in fewer dimensions, indicating a lower effective dimensionality.

The number of neurons in individual areas was soft-equalized by taking a random subsample of $N_0 = 120$ neurons when $N > N_0$. In these cases, the participation ratio was computed over 100 random subsets, and the average was taken as the value of α-diversity.

Analysis of population representations

Data preparation. For the analysis of population neural representations, we used the same sessions as described above. For each trial within a session, we therefore have a collection of N-dimensional population activity vectors $\mathbf{f}^{t,k}$, where $k \in \{1, P\}$ is the trial index within the session, and $t \in \{1, T\}$ is the time-bin index within each trial. For the analysis below, we used data from 0 to 1,000 ms after the stimulus onset to capture a variety of sensory and behavioural variables. We then labelled each time bin according to the value of four binarized cognitive, sensory and movement variables:

- Block: left (20–80) versus right (80–20) prior block.
- Contrast: we binarized the contrast into low (0–0.125) versus high (0.25–1.0) values.
- Stimulus: left versus right side of the screen.
- Whisking: we binarized the whisking power using the distribution of whisking power values within each session. Time bins in which the mouse was whisking with a power larger than the 50th percentile across the distribution were annotated as high, while those below the 50th percentile were annotated as low.

These variables were chosen so that they span movement, cognitive and sensory variables while ensuring that all of the $M = 16$ conditions (combinations of the four variables) were well represented in the data. For example, we could not add Choice as a variable as mice are overtrained in the task and, as a consequence, make very few mistakes when block and stimulus are aligned (for example, choose 'left' when the block and the stimulus are both 'right').

For each condition $c \in \{1, M\}$ (for example, whisking = high, contrast = low, stimulus = left, block = right), we first identified those trials in which that specific combination of variables was present. We then defined a collection of 'conditioned trial' population activity vectors $\{\mathbf{f}^{k,c}\}$ as the mean firing rate of the population of neurons conditioned to the specific condition in each trial. Given a trial k and a neuron index i, the mean firing rate was computed as

$$f_i^{k,c} = \frac{\sum_t \delta^{t,k}(c) f_i^{t,k}}{\sum_t \delta^{t,k}(c)}, \quad (12)$$

where i indicates the neuron index and $\delta^{t,k}(c) = 1$ if the time bin t in trial k corresponds to the condition c, and 0 otherwise. These conditioned trial population vectors are the data samples that will be used for the dimensionality and decoding analyses below. Across all analyses, we considered only those recording sessions in which each condition was present in at least $M_{\min} = 5$ trials.

Representation dimensionality

To estimate the representation dimensionality of a neural geometry, we computed the PR of the centroids $\mathbf{f}_c$ of the M conditions, defined as the average activity pattern across all trials of the same condition:

$$\mathbf{f}^c = \langle \mathbf{f}^{k,c} \rangle_k, \quad (13)$$

To compute the PR for the set of centroids, we first calculated their covariance matrix, normalizing each neuron's mean activity vector by subtracting from each f_i^c the mean across conditions c and dividing them by their s.d. We then performed principal component analysis on this covariance matrix to obtain its eigenvalues, λ_j.

The number of neurons in individual areas was soft-equalized by taking a random subsample of $N_0 = 120$ neurons when $N > N_0$. In these cases, the participation ratio was computed over 100 random subsets, and the average was taken as the value of representation dimensionality. The PR was computed on the set of centroids to highlight signal dimensions and prevent noise from dominating the measure. When computed across all trial vectors (without condition averaging), the PR ranged from 10 to 350 and was highly correlated with that of the centroids (Spearman Correlation coefficient = 0.86).

Cross-validated decoding

We used Decodanda[53] (www.github.com/lposani/decodanda) to perform a cross-validated, class-balanced decoding analysis of different combination of condition labels from the neural activity within individual trials (condition trial vectors $\mathbf{f}^{k,c}$). See the individual sections below for additional details on the data input structure of our decoding analyses. As a decoder, we used a scikit-learn SVM classifier with linear kernel[54]. To ensure that results were comparable across regions, which might have a different number of recorded neurons, we created a pseudopopulation by resampling all of the recorded neurons within each region to a fixed number $N = 4{,}000$. Similarly, we resampled the same number of pseudopopulation for each analysis ($T = 100$ patterns per condition). Note that simultaneously recorded neurons were always kept together during resampling to keep the noise correlations intact within the pseudopopulation[53]. All cross-validated decoding analyses were performed using the following Decodanda parameters: training_fraction = 0.8, cross_validations = 100, ndata = 100.

Finding the independent conditions

To find the number of independent conditions encoded in the activity of a population of neurons, we developed an iterative algorithm based on linear decoding. The algorithm followed the steps below, and is shown in action on one example region in Extended Data Fig. 3.

(1) First, we performed a decoding analysis of the condition label c from trial population vectors $\mathbf{f}^{k,c}$ using a set of binary linear classifiers. For each pair of conditions (c_i, c_j), we estimated a cross-validated decoding performance $\varphi(c_i, c_j)$, resulting in an initial $M \times M$ condition–condition decoding matrix (C_0; Extended Data Fig. 3) defined as $C_0(ij) = \varphi(c_i, c_j)$.

(2) We then chose a decoding threshold $\varphi_{\min} = 0.666$; the pairs of conditions whose one-versus-one decoding performance was smaller than $\varphi_{\min}$ were defined as dependent. Using this threshold, we defined a binary dependency matrix D defined as $D_0(i, j) = 1$ if $\varphi(c_i, c_j) < \varphi_{\min}$, and $C_0(i, j) = 0$ otherwise.

(3) We then used the Bron–Kerbosch algorithm[55] to find all of the cliques, that is, subgroups of fully connected nodes, in the undirected graph defined by the dependency matrix D_0. This process enables us to identify whether there are groups of conditions that are all non-decodable from each other (dark squares in the sorted matrix in Extended Data Fig. 3).

(4) We next identified the largest clique and grouped together all of the trials of the conditions within that group into a new, merged condition (see the arrows and 'merge' conditions in Extended Data Fig. 3).

(5) We repeated steps 1 and 2 with the new reduced set of conditions, yielding a new C_t and a new D_t matrix of a different size M_t, where t denotes the iteration step.

(6) We then repeated steps 3 and 4, and iterated the whole process (1–4) until all of the merged and remaining conditions were found to be independent, that is, the dependency matrix $D_{\tilde{t}}$ was diagonal. The number of independent conditions was then defined as the size of the final dependency matrix: $M_{\text{IC}} := M_{\tilde{t}}$.

Separability and AD

Separability quantifies how many random dichotomies (equally sized groups) of experimental conditions can be decoded from neural activity using cross-validated linear classifiers. To estimate the separability of a neural population, we performed the following steps:

1. First, we randomly divided the set of M or M_{IC} independent conditions into two equally sized groups (dichotomy).
2. Given the dichotomy d, we then measured the cross-validated decoding performance φ_d of a linear classifier trained to report whether individual condition trial vectors $\mathbf{f}^{k,c}$ belonged to conditions within one or the other dichotomy groups. This decoding analysis was performed as described in the 'Cross-validated decoding' section above, resampling a fixed large number of neurons ($N = 4{,}000$) and a fixed number of trials ($T = 100$) per condition for all regions to ensure that performances could be compared across regions.
3. The random dichotomy assignment and decoding (steps 1 and 2) was then repeated $n = 200$ times to obtain a set of decoding performances $\{\varphi_d\}$.
4. The decoding analysis was then repeated $n = 200$ times with shuffled condition labels across the population vectors to obtain a distribution of null decoding performance values $\{\varphi^{\mathrm{null}}\}$.
5. Separability was then defined as the fraction of decodable dichotomies, that is, the fraction of φ_d larger than the 99 percentile of the null population $\{\varphi^{\mathrm{null}}\}$. AD was defined as the mean decoding performance across the $n = 200$ random dichotomies.

The distributions of decoding performances for all of the analysed regions are shown in Extended Data Fig. 11.

Synthetic data

Modelling uneven and categorical selectivity. We generated synthetic population responses for N neurons to all 2^V configurations of V binary variables $x_v \in \{-1, 1\}$. Each neuron's response combined linear and quadratic selectivity,

$$f_i(\vec{x}) = \sum_{v=1}^{V} \alpha_{iv} x_v + \gamma \sum_{u<v} \beta_{i,uv} x_u x_v + \epsilon, \tag{14}$$

where γ controls the strength of non-linear interactions and $\epsilon_i \sim \mathcal{N}(0, \sigma)$ introduces trial-to-trial variability. For each condition $\vec{x}$, we generated T independent trials and used the data in the decoding analyses as performed for real spiking data.

Sampling the selectivity structure. The coefficients $(\alpha_{iv}, \beta_{i,uv})$ were sampled in a feature space of dimension $D = V + \frac{V(V-1)}{2}$, which groups all linear and quadratic features on an equal footing. To generate either categorical or uneven selectivity, we first drew k cluster centroids in the D-dimensional feature space and repeated each centroid $k_N = N/k$ times, yielding N prototype vectors. These prototypes define the coarse structure of selectivity across neurons. To modulate within-cluster diversity, each prototype was perturbed by an additive diversity vector

$$\delta_i = \sigma_d \, \xi_i, \xi_i \sim \mathcal{N}(0, I_D),$$

where $\sigma_d = -\log(c)$ is set by a categorical specialization parameter $c \in (0, 1)$ (this is the x axis in Extended Data Fig. 8b). Larger c produces more tightly clustered selectivity (strong categorical structure), whereas smaller c yields more dispersed coefficients. To model uneven selectivity, we introduced anisotropy across the D feature dimensions by using $k = 1$ (no clustering) and scaling the diversity terms by a geometric decay profile:

$$\sigma_j = \sigma_d \, \log(1-r)^{j-1}, j = 1, \ldots, D,$$

where $r \in (0, 1)$ is the global specialization parameter. When $r \ll 1$, all dimensions contribute equally; when $r \approx 1$, variance is concentrated in a low-dimensional subspace, producing a strongly uneven selectivity spectrum. The final coefficients, including categorical and/or uneven structure, were

$$(\alpha_i, \beta_i) = \mathrm{centroid}_i + \sigma \odot \xi_i,$$

with the quadratic coefficients additionally scaled by γ to control non-linearity.

Trial generation. For each of the 2^V binary stimuli, we computed $f_i(\vec{x})$ through the linear-quadratic form above, and generated T noisy samples per condition,

$$r_{i,t}(\vec{x}) = f_i(\vec{x}) + \epsilon_{i,t}, \epsilon_{i,t} \sim \mathcal{N}(0, \sigma).$$

This procedure ensures that trial variability is independent across neurons and conditions, and that all structure in the population code arises exclusively from the geometry of the sampled coefficients.

Parameterizing categorical and uneven specialization. The two specialization parameters (c, r) therefore independently control: the categorical structure of selectivity (the number and separation of clusters, through c and k); and the unevenness of selectivity across feature dimensions (anisotropy of coefficient variances, through r).

By sweeping either c (categorical specialization) or r (global specialization) while keeping the other fixed, we isolated the effects of clustered versus uneven selectivity on representational dimensionality, separability and independent condition structure. For the analyses in Extended Data Fig. 8, we used $N = 100$ neurons, $V = 3$ variables (for a total of $P = 8$ conditions), $T = 20$ samples per condition and $\gamma = 0.25$.

Exploring the relationship between dimensionality and separability. To analyse how separability changes with the dimensionality of the geometry in the activity space, we performed a series of synthetic explorations shown in Extended Data Fig. 8. In these simulations, P centroids are randomly sampled from a Gaussian distribution spanning an L-dimensional subspace of the N-dimensional activity space. Each centroid vector is normalized. Trial-to-trial variability is then added to the centroids with a s.d. σ, scaled with $\sqrt{L}$ to keep the signal-to-noise level constant when pairwise distances between centroids increase with L. This obtains a $T \times N$ activity matrix for each of the P conditions. This synthetic activity is then analysed with the same pipeline used for the cortical data, yielding values of representation dimensionality, separability and AD shown in Extended Data Fig. 8g, in which we used $P = 16, N = 100$. For simulations in Extended Data Fig. 8i, we fixed $L = 14$ and multiplied the first dimension of each centroid by a factor γ to stretch the geometry along a single axis.

Theoretical considerations on the relationship between dimensionality and clustering

The conditions space and the neural space have the same dimensionality. As explained in Fig. 1a, response profiles of single neurons can be thought of as rows of a matrix X for which the columns define the geometry of conditions in the neural space. The PR in the conditions space is computed from the eigenvalue spectrum of the covariance matrix of the rows of X, that is, XX^T (assuming zero mean), while the PR in the activity space is computed from the eigenvalues of the covariance matrix of the columns of X, that is, X^TX. Given a matrix, the eigenvalues of its covariance matrix are the squared singular values of X. Let $X = USV^T$ be the SVD decomposition of X, where S is the diagonal matrix with singular values on the diagonal, then $X^T = VS^TU^T$. As $S = S^T$, the eigenvalue spectrum of X^TX and XX^T is the same. Thus, the participation ratio of the conditions space is the same as that in the neural space.

Mathematical derivation of the PR of Gaussian clusters. We consider a data model with N features (neurons) and M observations (conditions), in which observations are sampled from independent and identically distributed (i.i.d.) random variables as

$$\mathbf{x}^\mu = \mathbf{z}^\mu + \boldsymbol{\eta}^\mu, \tag{15}$$

where $\mathbf{z}^\mu$ and $\boldsymbol{\eta}^\mu$ are both vectors in $\mathbb{R}^N$ and represent the clustered and heterogeneous part of the data, respectively. More precisely, $\mathbf{z}^\mu$ is sampled from a normal distribution $\mathcal{N}(0, \mathbf{B})$ that has a clustered covariance matrix, that is, $B_{ij} = 1$ if i and j belong to the same cluster and $B_{ij} = 0$ otherwise. We call k the number of clusters and assume that all clusters have the same number of neurons $N_c = N/k$. By contrast, the heterogenous part $\boldsymbol{\eta}^\mu$ is sampled from $\mathcal{N}(0, \sigma^2\mathbf{I})$, where $\mathbf{I}$ is the identity matrix. Our goal is to compute the participation ratio (PR) of this representation, which we define as

$$\mathrm{PR} = \frac{\mathrm{Tr}(\mathbf{C})^2}{\mathrm{Tr}\mathbf{C}^2} = \frac{N\langle C_{ii}\rangle^2}{\langle C_{ii}^2\rangle + (N-1)\langle C_{ij}^2\rangle}, \tag{16}$$

where the averages are across neurons and the matrix C is the sample neuron-by-neuron covariance matrix, that is $C = \frac{1}{M}\sum_{\mu=1}^{M}\mathbf{x}^\mu(\mathbf{x}^\mu)^T$. We note that this definition assumes that the sample mean of both $\mathbf{z}$ and $\boldsymbol{\eta}$ are negligible or have been subtracted.

We are interested in the regime in which $N \to \infty$ while M is allowed to be small, as it often happens in controlled experiments. Small M might cause the sample covariance matrix to differ substantially from the true covariance matrix. Defining C^{c}, C^{h}, and C^{ch} as the sample covariance matrices of $\mathbf{z}$, $\boldsymbol{\eta}$, and the cross-covariance between $\mathbf{z}$ and $\boldsymbol{\eta}$, respectively, we have that

$$\mathrm{PR} = N\frac{\langle C_{ii}^{\mathrm{h}} + C_{ii}^{\mathrm{c}} + 2C_{ii}^{\mathrm{ch}}\rangle^2}{\langle (C_{ii}^{\mathrm{h}} + C_{ii}^{\mathrm{c}} + 2C_{ii}^{\mathrm{ch}})^2\rangle + (N-1)\langle (C_{ij}^{\mathrm{h}} + C_{ij}^{\mathrm{c}} + 2C_{ij}^{\mathrm{ch}})^2\rangle}. \tag{17}$$

We therefore need to evaluate the first and second moments of both diagonal and off-diagonal elements of all covariance and cross-covariance matrices. The diagonal elements of these matrices have the following statistics:

$$\begin{aligned}
C_{ii}^{\mathrm{c}} &= \frac{1}{M}\sum_{\mu=1}^{M}(z_i^\mu)^2 \Rightarrow \langle C_{ii}^{\mathrm{c}}\rangle = 1\,, \quad \langle (C_{ii}^{\mathrm{c}})^2\rangle = \frac{M+2}{M} \\
C_{ii}^{\mathrm{h}} &= \frac{1}{M}\sum_{\mu=1}^{M}(\eta_i^\mu)^2 \Rightarrow \langle C_{ii}^{\mathrm{h}}\rangle = \sigma^2\,, \quad \langle (C_{ii}^{\mathrm{h}})^2\rangle = \frac{M+2}{M}\sigma^4 \\
C_{ii}^{\mathrm{ch}} &= \frac{1}{M}\sum_{\mu=1}^{M}z_i^\mu\eta_i^\mu \Rightarrow \langle C_{ii}^{\mathrm{ch}}\rangle = 0\,, \quad \langle (C_{ii}^{ch})^2\rangle = \frac{1}{M}\sigma^2,
\end{aligned} \tag{18}$$

and

$$\langle C_{ii}^{\mathrm{c}}C_{ii}^{\mathrm{h}}\rangle = \sigma^2,\ \langle C_{ii}^{\mathrm{c}}C_{ii}^{\mathrm{ch}}\rangle = 0,\ \langle C_{ii}^{\mathrm{h}}C_{ii}^{\mathrm{ch}}\rangle = 0. \tag{19}$$

For the off-diagonal elements, we have

$$\begin{aligned}
C_{ij}^{\mathrm{c}} &= \frac{1}{M}\sum_{\mu=1}^{M}z_i^\mu z_j^\mu \Rightarrow \langle C_{ij}^{\mathrm{c}}\rangle = \frac{1}{k}\,, \quad \langle (C_{ij}^{\mathrm{c}})^2\rangle = \frac{1}{M} + \frac{1}{kM} + \frac{1}{k} \\
C_{ij}^{\mathrm{h}} &= \frac{1}{M}\sum_{\mu=1}^{M}\eta_i^\mu\eta_j^\mu \Rightarrow \langle C_{ij}^{\mathrm{h}}\rangle = 0\,, \quad \langle (C_{ij}^{\mathrm{h}})^2\rangle = \frac{1}{M}\sigma^4 \\
C_{ij}^{\mathrm{ch}} &= \frac{1}{M}\sum_{\mu=1}^{M}z_i^\mu\eta_j^\mu \Rightarrow \langle C_{ij}^{\mathrm{ch}}\rangle = 0\,, \quad \langle (C_{ij}^{\mathrm{ch}})^2\rangle = \frac{1}{M}\sigma^2,
\end{aligned} \tag{20}$$

and

$$\langle C_{ij}^{\mathrm{c}}C_{ij}^{\mathrm{h}}\rangle = 0,\ \langle C_{ij}^{\mathrm{c}}C_{ij}^{\mathrm{ch}}\rangle = 0,\ \langle C_{ij}^{\mathrm{h}}C_{ij}^{\mathrm{ch}}\rangle = 0. \tag{21}$$

Most of the expressions above can be straightforwardly derived by writing down the definition of the sample covariance matrix for a zero-mean variable and then performing the average over neurons. To illustrate this procedure, let us consider one of the most involved terms:

$$\begin{aligned}
\langle (C_{ij}^{\mathrm{c}})^2\rangle &= \frac{1}{M^2}\sum_{\mu,\nu=1}^{M}\langle z_i^\mu z_j^\mu z_i^\nu z_j^\nu\rangle \\
&= \frac{1}{M^2}\sum_{\mu=1}^{M}\langle (z_i^\mu)^2(z_j^\mu)^2\rangle + \frac{1}{M^2}\langle z_i^\mu z_j^\mu\rangle\langle z_i^\nu z_j^\nu\rangle
\end{aligned} \tag{22}$$

The probability that z_i and z_j are part of the same cluster is given, for large N, by $\frac{1}{k}$. The expression for $\langle (C_{ij}^{c})^2\rangle$ then becomes

$$\begin{aligned}
\langle (C_{ij}^{c})^2\rangle &= \frac{1}{kM^2}\sum_{\mu=1}^{M}\langle (z_i^\mu)^4\rangle + \frac{1}{M^2}\left(1-\frac{1}{k}\right)\sum_{\mu=1}^{M}\langle (z_i^\mu)^2\rangle^2 \\
&\quad + \frac{1}{kM^2}\sum_{\mu\neq\nu}^{M}\langle (z_i^\mu)^2\rangle^2 \\
&= \frac{3}{kM} + \frac{1}{M} - \frac{1}{kM} + \frac{1}{kM}(M-1) \\
&= \frac{1}{M} + \frac{1}{kM} + \frac{1}{k}.
\end{aligned} \tag{23}$$

The other terms can be computed following the same steps.

Given that k, M are finite, we can approximate the PR for large N as:

$$\mathrm{PR} \simeq \frac{\langle C_{ii}^{\mathrm{h}} + C_{ii}^{\mathrm{c}} + 2C_{ii}^{\mathrm{ch}}\rangle^2}{\langle (C_{ij}^{\mathrm{h}} + C_{ij}^{\mathrm{c}} + 2C_{ij}^{\mathrm{ch}})^2\rangle}. \tag{24}$$

Expanding the square and using the results above for the first and second moments of the covariance matrices, we get to our final expression:

$$\mathrm{PR}\ \mathit{simeq}\ \mathit{frac}(1+\sigma^2)^2\frac{1}{k} + \frac{1}{kM} + \frac{1}{M}(1+\sigma^2)^2 = M\frac{k(1+\sigma^2)^2}{1+M+k(1+\sigma^2)^2} \tag{25}$$

From the mathematical expression in equation (25), we can see that, in the limit of perfect clusters ($\sigma \to 0$), the function is either limited by the number of rows-neurons (in this case, the k perfect clusters) or columns-conditions M, coherently with the intuition above:

$$\mathrm{PR} \underset{k\to\infty}{\to} M\mathrm{PR} \underset{M\to\infty}{\to} k\,, \tag{26}$$

However, things become more nuanced when k, M and σ are finite and non-zero. First, as shown in Fig. 5e, when k and M are kept fixed, the representation dimensionality decreases with the clustering quality (expressed as the average silhouette score of a population of Gaussian clusters with given k, M and σ). Second, if we fix the quality of clusters and the number of conditions (Fig. 5e (left)), we see that dimensionality increases with the number of clusters, with a magnitude that is larger for high silhouette scores (categorical representations). Finally, when fixing the number and quality of clusters, the dimensionality is determined by the number of conditions, with a magnitude that is larger for low silhouette scores (non-categorical representations; Fig. 5e (right)).

Different measures of dimensionality

Measuring dimensionality in the presence of noise and determining whether it is high or low can be quite challenging[7]. This is why, in neuroscience, multiple methods are used to assess dimensionality. Each approach is different and, as dimensionality is expressed as a single number, it inevitably emphasizes only specific aspects of the representational geometry.

In our Article, we always refer to the embedding dimensionality of the set of points representing different experimental conditions in the activity space[56]. These points represent patterns of activity recorded at the same time (not trajectories). To discount the dimensions due to the noise, we computed the representation dimensionality of the average positions of the conditions. An alternative approach is to consider a cross-validated measure of representation dimensionality[57]. Notice that separability and other measures related to it[6], which consider the computational consequences of high dimensionality, are typically insensitive to noise, as they are also cross-validated measures.

Other recent works have introduced dimensionality measures that depend on the spatial and temporal scales considered[58]. For large scales, they get an estimate of the embedding dimensionality and, for short scales, they get an estimate of the intrinsic dimensionality. While these quantities can reveal many other interesting aspects of the representational geometry, here we focused only on the embedding dimensionality because it is the one most relevant to the performance of a linear readout.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The original data used for this work are publicly available through the IBL and Allen Institute resources. The processed data and code used for the analysis are available at GitHub for clustering analysis (https://github.com/realwsq/clustering-analysis), the RRR model (https://github.com/realwsq/brainwide-RRR-encoding-model), and separability, decoding and dimensionality analyses (https://github.com/lposani/Separability).

51. Williams, A. H. et al. Discovering precise temporal patterns in large-scale neural recordings through robust and interpretable time warping. *Neuron* **105**, 246–259 (2020).
52. Izenman, A. J. Reduced-rank regression for the multivariate linear model. *J. Multivariate Anal.* **5**, 248–264 (1975).
53. Posani, L. Decodanda: a Python toolbox for best-practice decoding and geometric analysis of neural representations. Preprint at *bioRxiv* https://doi.org/10.64898/2026.03.16.711920 (2026).
54. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
55. Bron, C. & Kerbosch, J. Algorithm 457: finding all cliques of an undirected graph. *Commun. ACM* **16**, 575–577 (1973).
56. Jazayeri, M. & Ostojic, S. Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Curr. Opin. Neurobiol.* **70**, 113–120 (2021).
57. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D. High-dimensional geometry of population responses in visual cortex. *Nature* **571**, 361–365 (2019).
58. Recanatesi, S., Bradde, S., Balasubramanian, V., Steinmetz, N. A. & Shea-Brown, E. A scale-dependent measure of system dimensionality. *Patterns* **3**, 100555 (2022).
59. International Brain Laboratory et al. Reproducibility of in vivo electrophysiological measurements in mice. *eLife* **3**, RP100840 (2022).

Acknowledgements We thank F. Stefanini, M. Rigotti, M. Benna, R. Nogueira and G. Allen for discussions about categorical representations and their analysis; T. Engel for discussions about dimensionality and clustering; and Y. Zhang and J. Xia for a code review of the RRR encoding model. S. Wang thanks W. Gerstner for supporting her research visit to Columbia University. L. Posani thanks J. Johnston and J. Whittington for co-organizing the workshop 'Are Neurons Interpretable?' (COSYNE, 2023), which had a big role in inspiring this work.

Author contributions L. Posani and S.W. contributed equally to this work. L. Posani and S.F. conceived and designed the project. S.W. and L. Paninski designed and implemented the RRR approach. S.W. and L. Posani designed the clustering analysis pipeline. S.P.M. and L. Posani carried out the analytical computations. L. Posani and S.W. analysed the data. S.F. and L. Paninski supervised the project. All of the authors interpreted the results and wrote the manuscript.

Funding This work was supported by NIH RF1AG080818 and U19NS123716, the Simons Foundation, the Kavli Foundation, the Gatsby Foundation GAT3708, the Swartz Foundation and the NSF and by DoD OUSD (R&E) under cooperative agreement PHY-2229929 (the NSF AI Institute for Artificial and Natural Intelligence). L. Posani was supported by the NIH 1K99MH135166-01 grant.

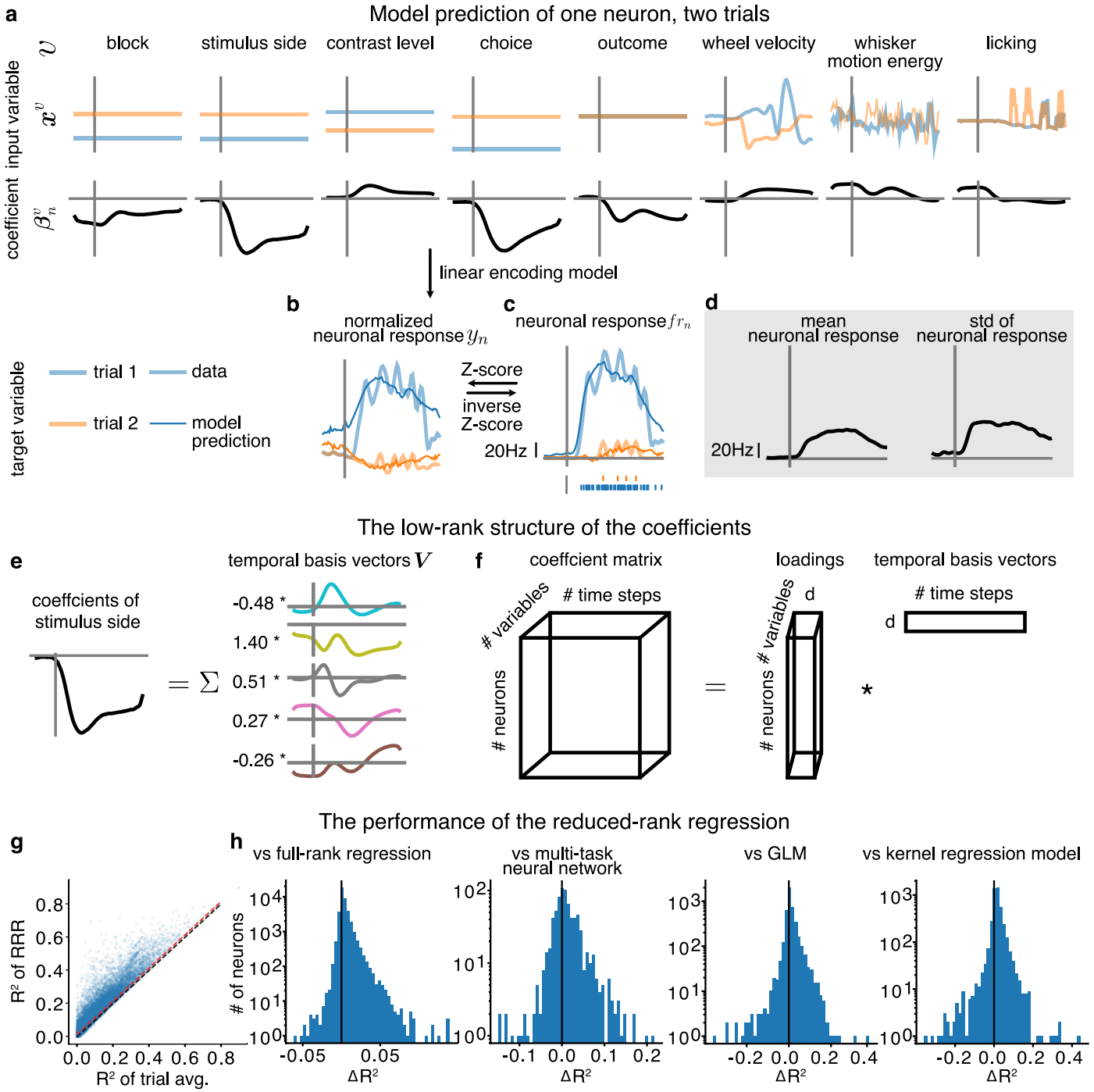
Competing interests The authors declare no competing interests.

Additional information

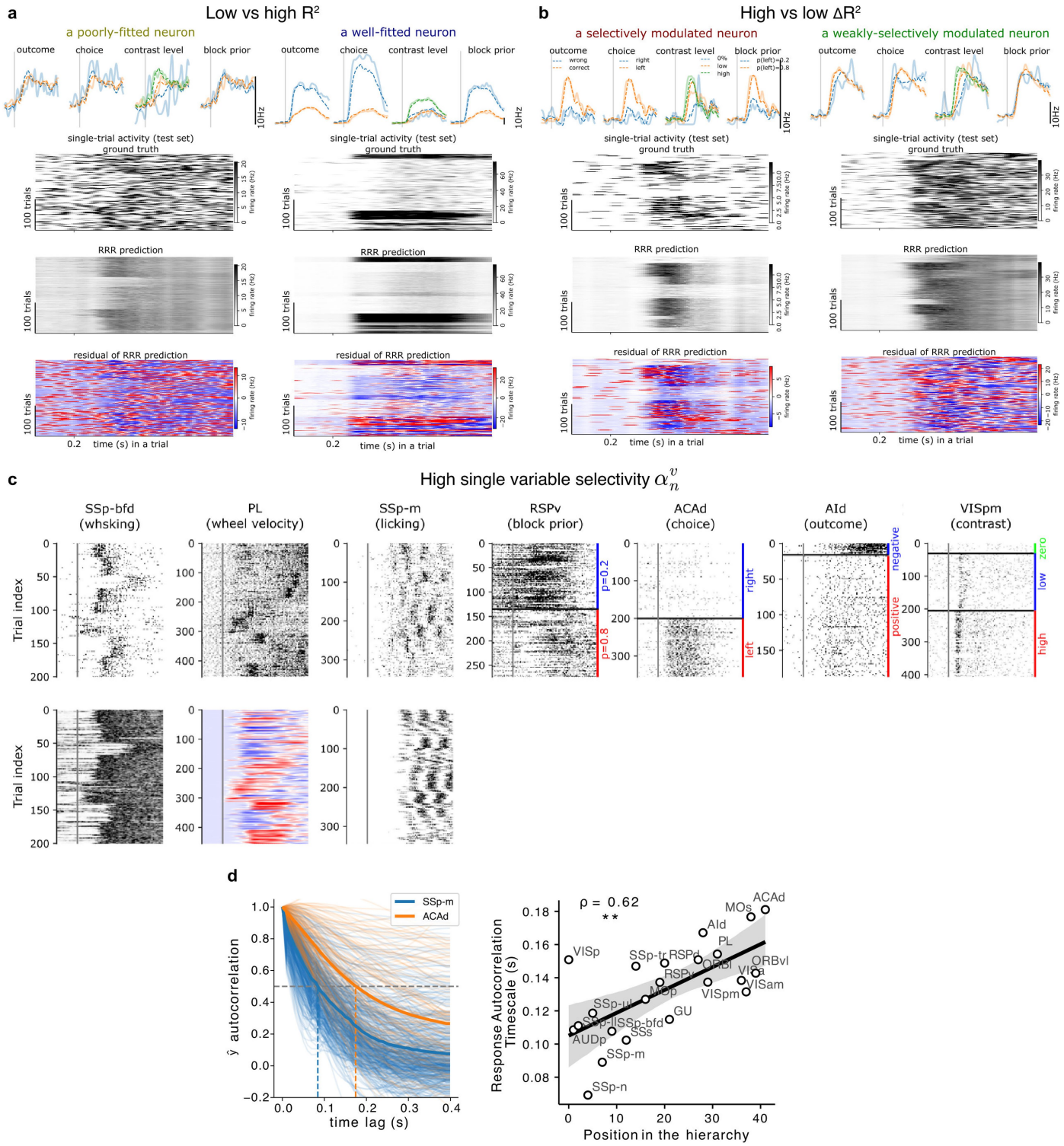
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-10668-4.

Correspondence and requests for materials should be addressed to Lorenzo Posani or Stefano Fusi.

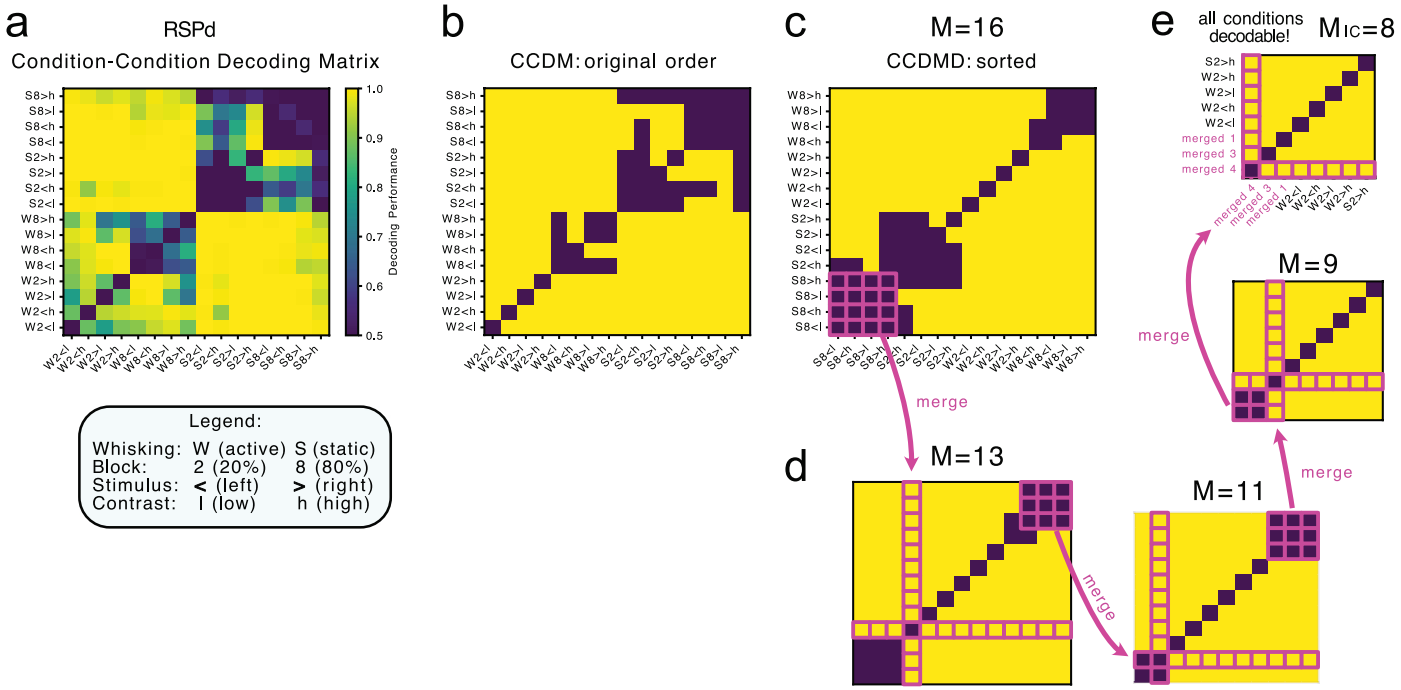
Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.



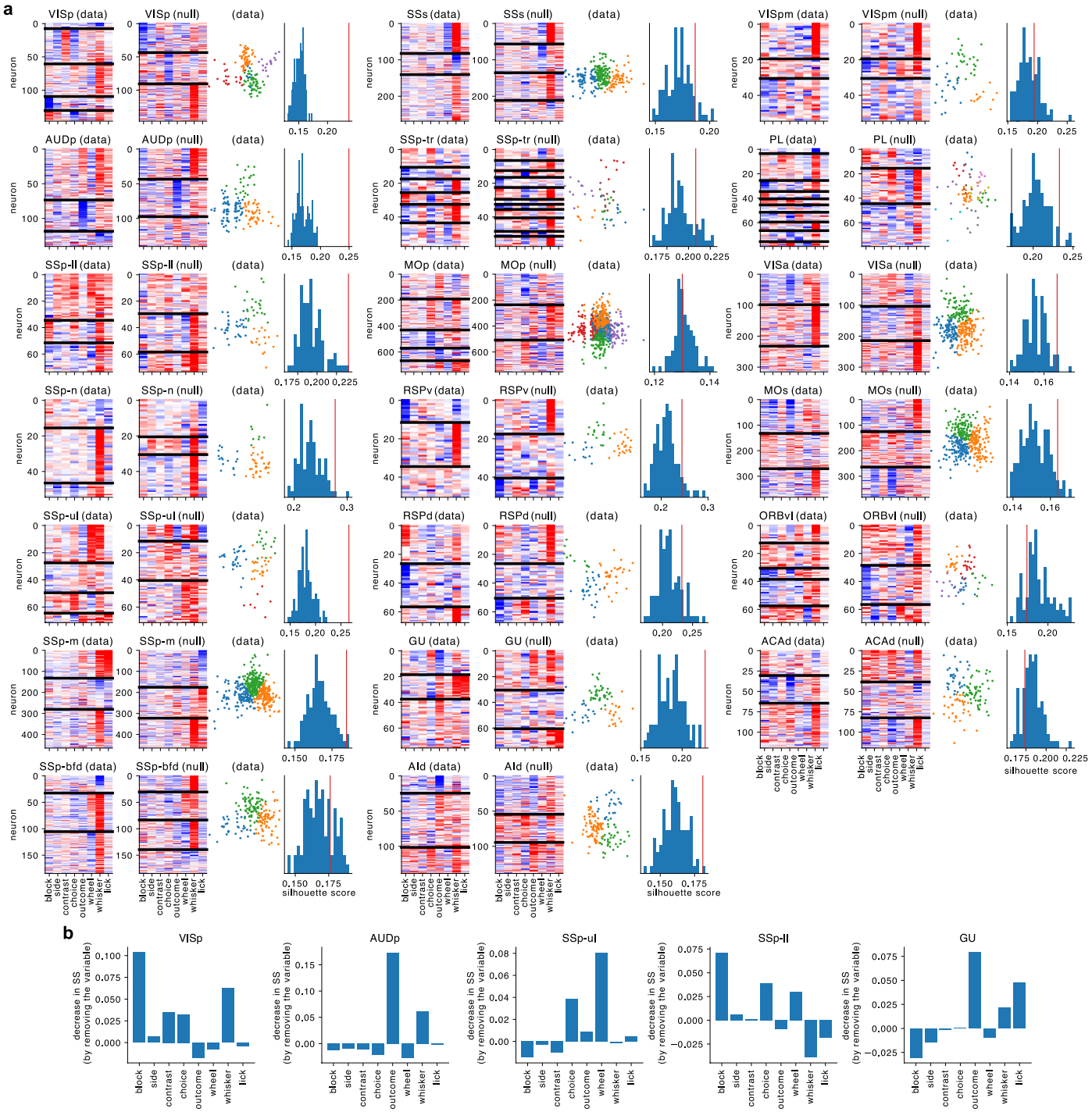
Extended Data Fig. 1 | Visual illustration and performance benchmark of reduced-rank regression (RRR) model. a-d Illustration of how the RRR encoding model predicts the neuronal responses using two example trials (blue and orange lines). The model weights the input variables x^v with the corresponding coefficients β_n^v (**a**) and sums the values over all the input variables to get the prediction $\hat{y}_n$ (**b**) (see Methods). Importantly, the model coefficients are time-dependent (**a**), and the inputs (x) and outputs (y) of the RRR encoding model are normalized across trials (**b**-**d**, see Methods). The prediction (thin and opaque lines) overlays the data (thick and transparent lines). The grey vertical lines indicate the onset of the stimulus (0.2 s), and the grey horizontal lines indicate the zero value. **e-f** Coefficients are enforced to be weighted sums of a small set of temporal basis vectors. **e** Example decomposition of the stimulus side coefficients in (**a**). **f** Schematic plot showing the low-rank structure of the coefficient matrix. **g** Goodness-of-fit analysis comparing the RRR encoding model to the trial-averaged neural activity (that is, the null model, Eq. (8)). Each point represents a neuron. The black line marks equal performance for both methods, while the red line indicates cases where RRR outperforms the trial-average estimate by $\Delta R^2 = 0.015$. **h** Performance comparison with other baseline models, measured as $\Delta R^2 = R^2(RRR) - R^2$ (baseline method). Baseline models include full-rank regression (which assumes a linear encoding model without rank constraints), a multi-task neural network (a nonlinear model with feed-forward and recurrent layers as in ref. 59), a generalized linear model (GLM, adapted from ref. 4), and a kernel regression model (adapted from ref. 20). See Methods for an explanation of the difference among RRR, GLM, and kernel regression model.



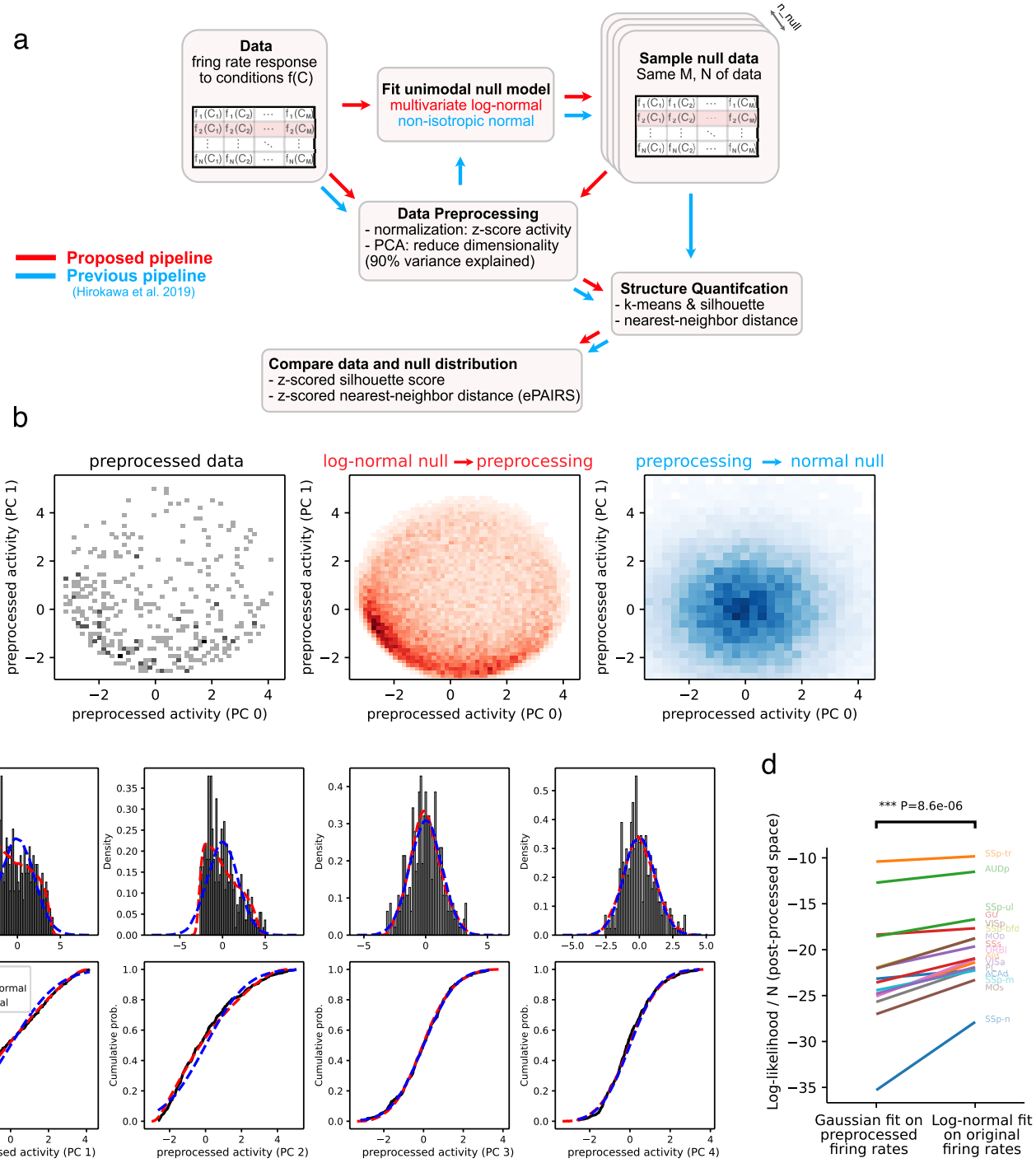
Extended Data Fig. 2 | Interpretation of reduced-rank regression (RRR) model. a The cross-validated R^2 reflects, overall, how well the model captures the task-related variance. Top: Peri-stimulus time histograms (PSTHs) for different task conditions (colour-coded). Middle: Single-trial activity (test set) alongside RRR predictions. Bottom: Residuals of the RRR predictions. **b** Conceptually, we distinguish two types of variance modulated by the task (see Methods): the *selective modulation* (left panel) and the *non-selective modulation* (right panel). While the selective modulation is induced by input variables and varies trial-by-trial, the non-selective modulation is locked to the key events of the trial (stimulus onset in this case) and does not vary trial-by-trial. The value of ΔR^2, defined as the difference between the R^2 of RRR and R^2 of the trial-average estimate (see Methods), is able to distinguish the two types of modulations. In this work, we care mostly about the selectively modulated neurons, identified as $\Delta R^2 \geq 0.015$. **c** Examples of neurons with high selectivity (that is, large RRR estimated coefficients, see Methods). The first row shows neuronal activity, while the second row illustrates the associated behavioural movements. **d** The analysis of the autocorrelation timescale reveals a gradient along the cortical hierarchy (Spearman correlation $\rho = 0.62$, $p = 0.002$). The autocorrelation functions were computed using RRR estimated responses to task and behaviour variables ($\hat{y}$, see Methods). Neurons from two example brain regions (SSp-m and ACAd) are shown in the left panel, along with their mean profiles (solid orange and blue lines) and estimated autocorrelation timescales (orange and blue dashed lines). In the right panel, shaded areas indicate the bootstrap 95% confidence interval of the fitted regression line. P values for Spearman correlations were computed using SciPy's two-sided test. $^{**}p < 0.01$.



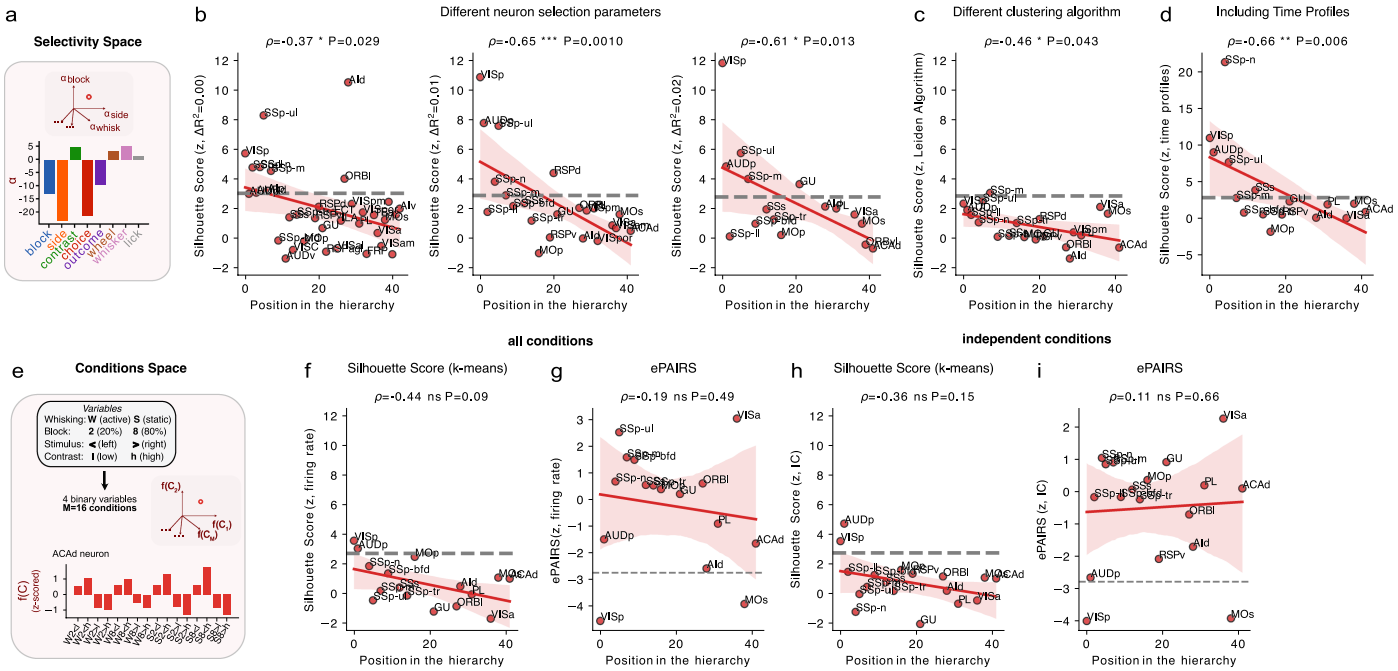
Extended Data Fig. 3 | Example of all the individual steps of the independent conditions algorithm applied to the population activity of an example region (RSPd). First, a condition-condition decoding matrix (CCDM) is computed using cross-validated decoding with a linear SVC (**a**), starting from the $M = 16$ conditions defined by the combinations of 4 binarized variables. The CCDM is then binarized into decodable and non-decodable pairs (**b**) by using a threshold (0.666). Then, the iterative algorithm finds the largest group of conditions that are non-decodable from each other, which corresponds to a dark block in the sorted matrix in (**c**). These conditions are grouped together and given a new single-condition label (**d**). The algorithm iterates until all the conditions are decodable from each other (**e**). The size of the resulting CCDM is the number of independent conditions M_{IC}.



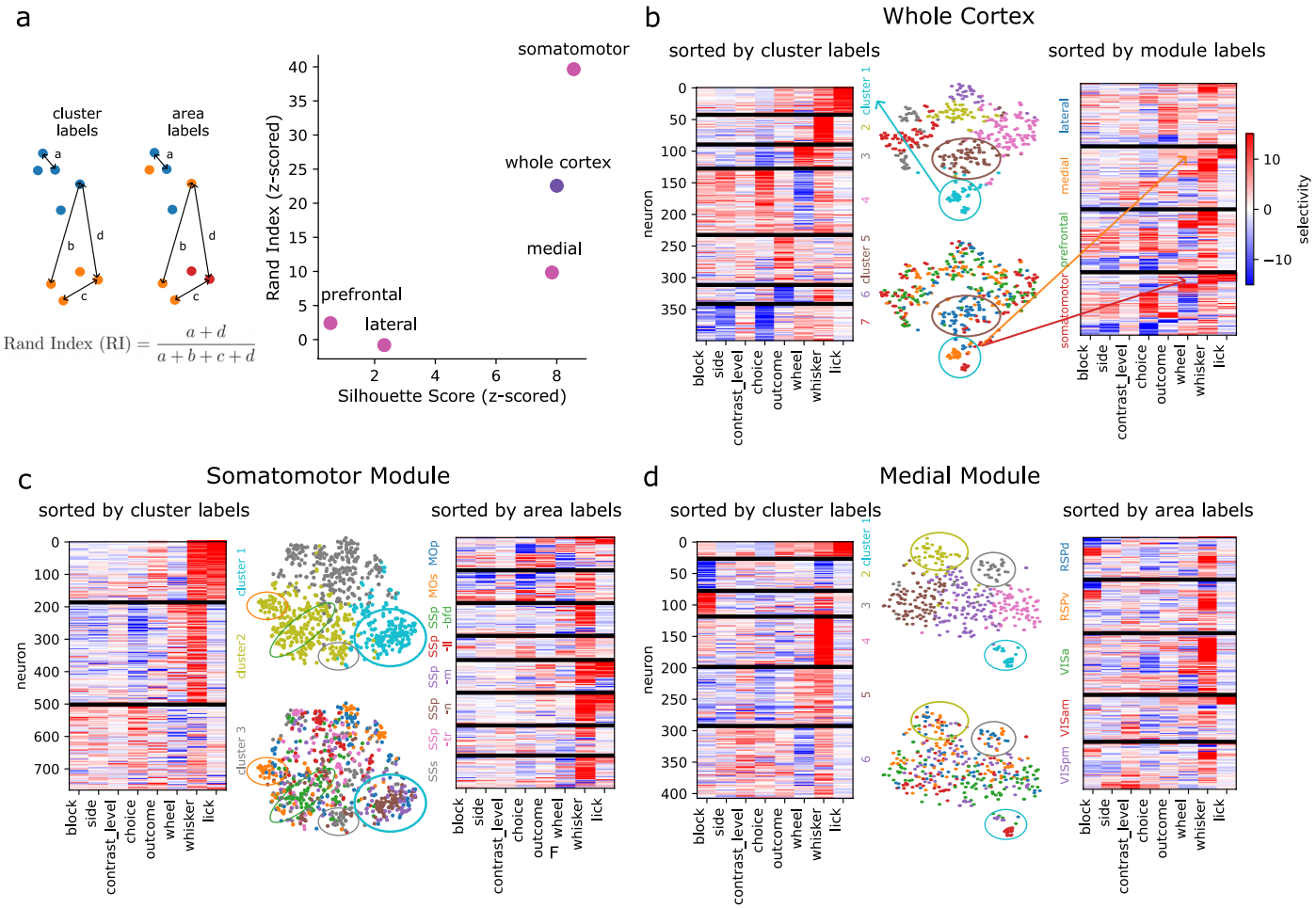
Extended Data Fig. 4 | **(a)** Clustering results of all cortical regions in the selectivity space. Subfigures are presented in the same format as Fig. 3b,c. (**b**) For the regions that passed the statistical test, we measured the contribution of individual variables to categorical clustering by computing the difference in silhouette score when removing each input variable.



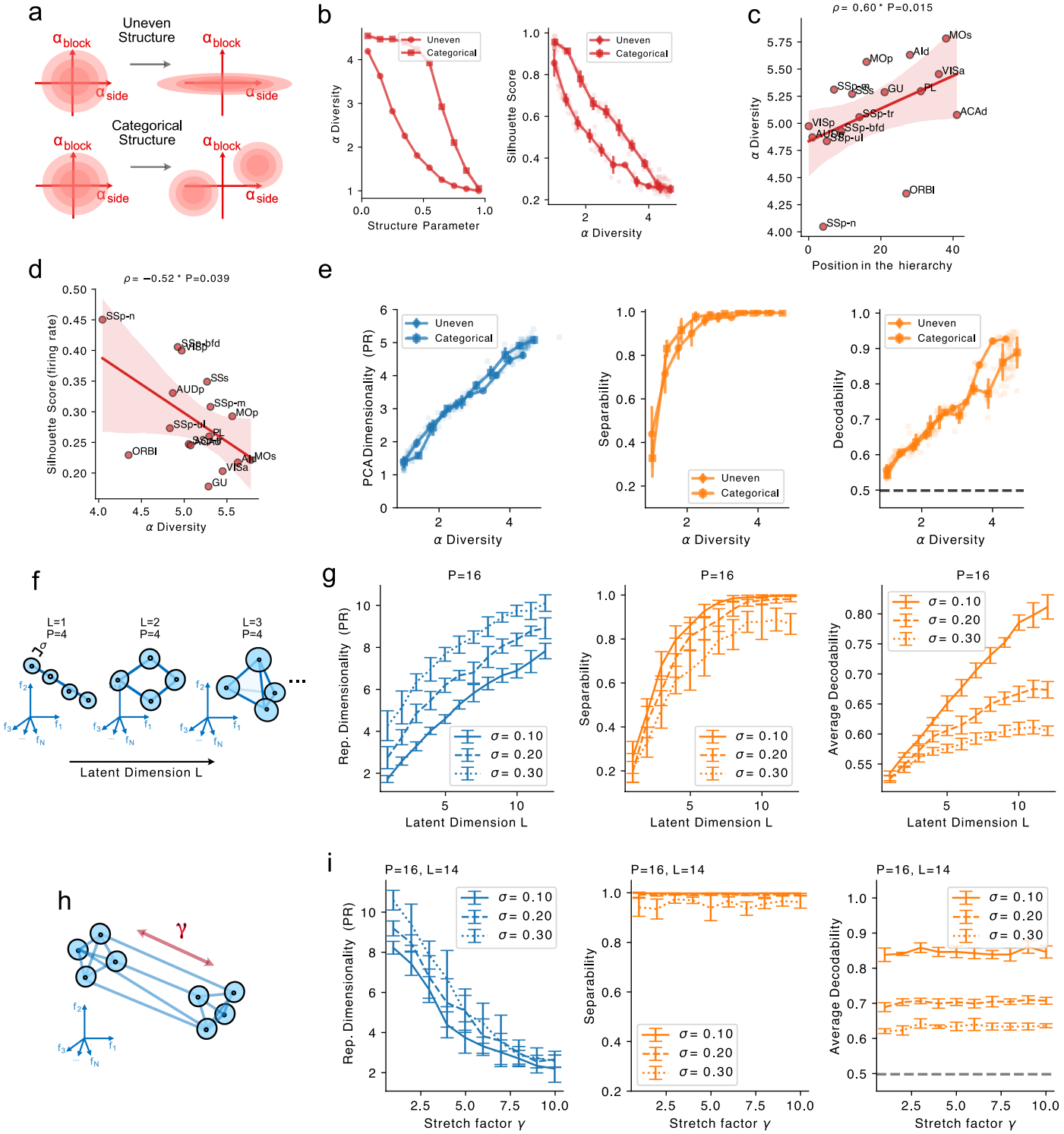
Extended Data Fig. 5 | Pipeline comparison for analysis of structure in the conditions space. (**a**) Flow charts for the analysis of structure in the conditions space defined by the mean firing rates of neurons in response to a given set of experimental conditions, comparing the proposed pipeline with the pipeline developed in[2] for the ePAIRS null model comparison (blue arrows). The main methodological difference is that, in our pipeline, the unimodal null model is fitted on the original data instead of the z-scored and dimensionality-reduced firing rate profiles. Note that in both cases, structural measures such as clustering quality or ePAIRS are computed in the post-processed, z-scored space. (**b**) Comparison of the joint distribution of the post-processed z-scored data (black) compared with the null model inferred using our pipeline (red: log-normal fitting in the original space, then z-scoring and dimensionality reduction) or using the pipeline as described in ref. 2 (blue: z-scoring and reducing the data dimensionality and *then* fitting a Gaussian null model), for an example region (MOs). (**c**) Marginal distributions of the data along the first 4 principal components of the reduced-dimensionality z-scored conditions space (that in which we compute structure). Top row: distribution of the data (black) compared with the null model distribution given by our pipeline (red) and the pipeline in ref. 2 (blue). Bottom row: cumulative distributions of the data in the top row (same colour scheme). Note that, in these data, the distributions of the data along the first dimensions of the reduced-dimensionality space (noted as PC0 and PC1) are typically asymmetric and non-Gaussian. The asymmetry is well-captured by a log-normal null model in the original space that undergoes the same preprocessing of the data (red curve), while it is less well-captured by a post-processed Gaussian fit (blue curve). In this example, PCs after the second are more Gaussian. (**d**) Comparison of the goodness of fit, quantified as the normalized log-likelihood of the data given the fitted null distribution, between the two pipelines. p-value computed via a paired t-test.



Extended Data Fig. 6 | Additional analysis of structure in the selectivity and condition firing rate space. Comprehensive clustering analyses show consistent patterns in both selectivity space (**a**-**d**) and condition mean firing rate space (**e**-**i**) in agreement with the results in Fig. 3. **(a-d)** Analysis of structure in the selectivity space with different parameters and algorithms. (**a**) Visualization of an example neuronal selectivity profile in the selectivity space. (b-d) Same analysis as Fig. 3d (correlation of clustering quality in the selectivity space with the position of the hierarchy) reproduced with (**b**) different neuron selection ΔR^2 thresholds ($\Delta R^2 = 0.015$ used in the main text), (**c**) a different algorithm than k-means for finding the clusters (Leiden clustering algorithm), and (**d**) expanding the selectivity space by including the time-varying selectivity profiles (see Methods). **(e-i)** Results of structure analysis in the conditions mean firing rate space using the analysis pipeline explained in Extended Data Fig. 5. (**e**) Illustration of the conditions space framework incorporating four binary variables resulting in 16 experimental conditions. (**f**) Analysis of clustering quality in the 16-dimensional mean firing rate space of the 16 experimental conditions defined by the combinations of the four selected variables (see panel **e** and main text). (**g**) Structure quantified by the ePAIRS analysis (mean nearest-neighbour distance compared to a null model through z-scoring, see Methods) in the same conditions mean firing rate space of panel **f**. **(h, i)** Same analyses as (**f**, **g**) in the space defined by the mean firing rate of independent conditions of each individual area. Shaded areas indicate the bootstrap 95% confidence interval of the fitted regression line. P values for Spearman correlations were computed using SciPy's two-sided test. n.s. non-significant; *p < 0.05; ** p < 0.01; *** p < 0.001.

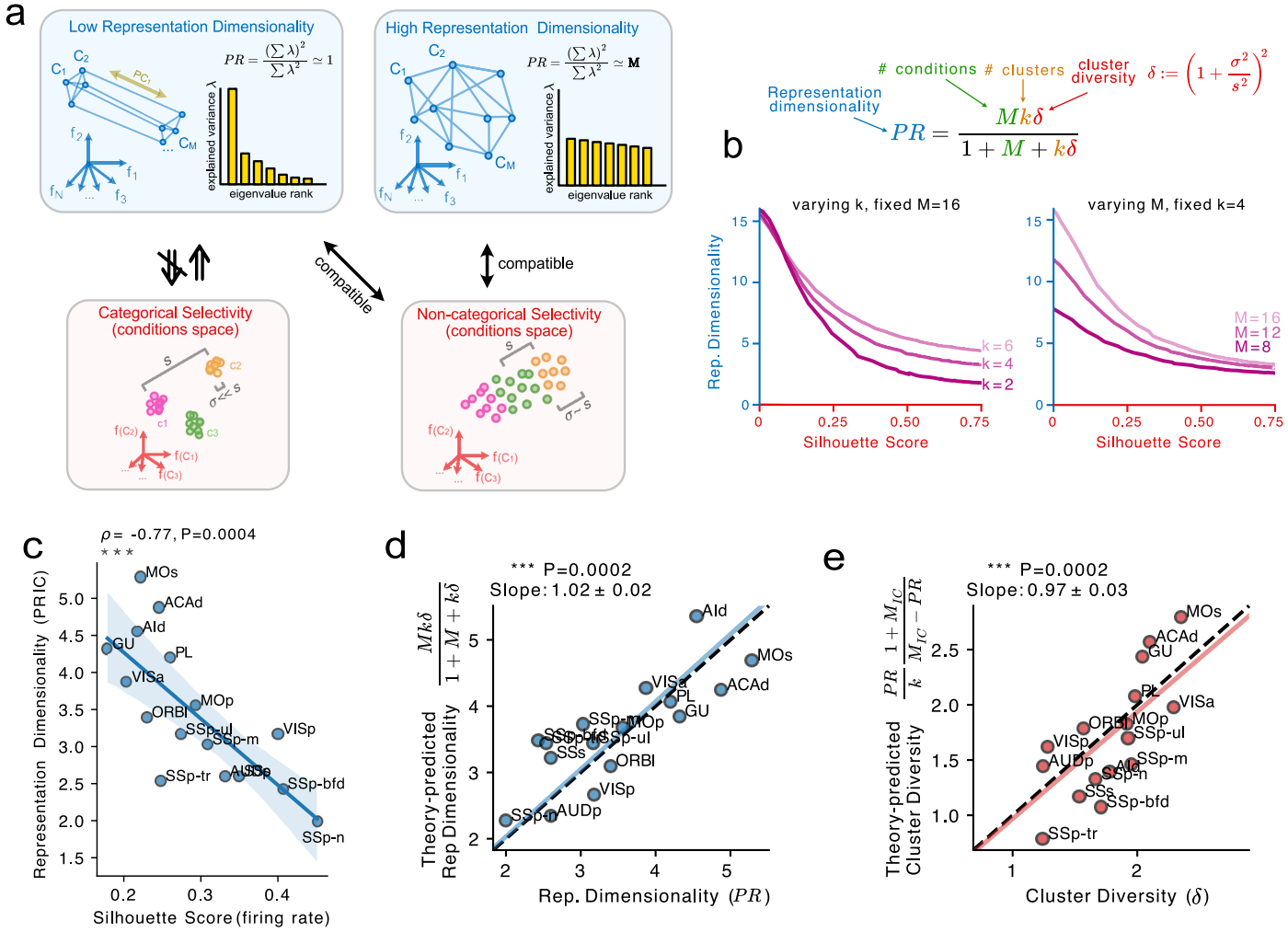


Extended Data Fig. 7 | A globally organized brain, with functional clustering closely reflecting anatomical structure. a We propose a measure of how well neurons are clustered by anatomical structure using Rand index (z-scored) (see Methods). Left: Rand index (RI) quantifies the similarity between two clusterings. We compare the clustering based on neuron area labels to the optimal clustering solution. Right: Neurons in clustered modules (somatomotor and medial) and across the whole cortex are significantly organized by anatomical structure labels. **b** Neuron selectivity matrix with neurons pooled across the cortex, sorted either by cluster labels (left) or module labels (right), noted aside and distinguished with colours. Middle: t-SNE visualizations of the same matrix, coloured by cluster (top) or module (bottom). As arrows illustrate, the full selectivity profile can be traced through the colours, while ellipses highlight examples of clustering by anatomical areas. Cluster 1, with strong selectivity for licking, for example, is primarily from somatomotor and medial modules. When analysing a single population spanning the entire cortex, we observed categorical representations ($SS\_z \simeq 8$), with key variables contributing to cluster formation encompassing nearly all variables considered. **c-d** Somatomotor (**c**) and medial (**d**) modules show clustering by areas. In the somatomotor module, clustering reflects differences in selectivity for whisking power and licking. Cluster 1 (high selectivity for both variables) is enriched in neurons from nose (SSp-n) and mouth (SSp-m) regions of primary somatosensory cortex. Cluster 2 (high whisking selectivity, low licking selectivity) contains neurons mainly from barrel field (SSp-bfd), secondary somatosensory cortex (SSs), and secondary motor cortex (MOs). Within this cluster, neurons from the same area are more similar to each other than to neurons from different areas. In the medial module, clustering is shaped by selectivity to block prior, whisking power, and licking. Cluster 1 (strong licking selectivity) is enriched in VISam, while clusters with strong block-prior selectivity are dominated by RSPd and RSPv neurons.

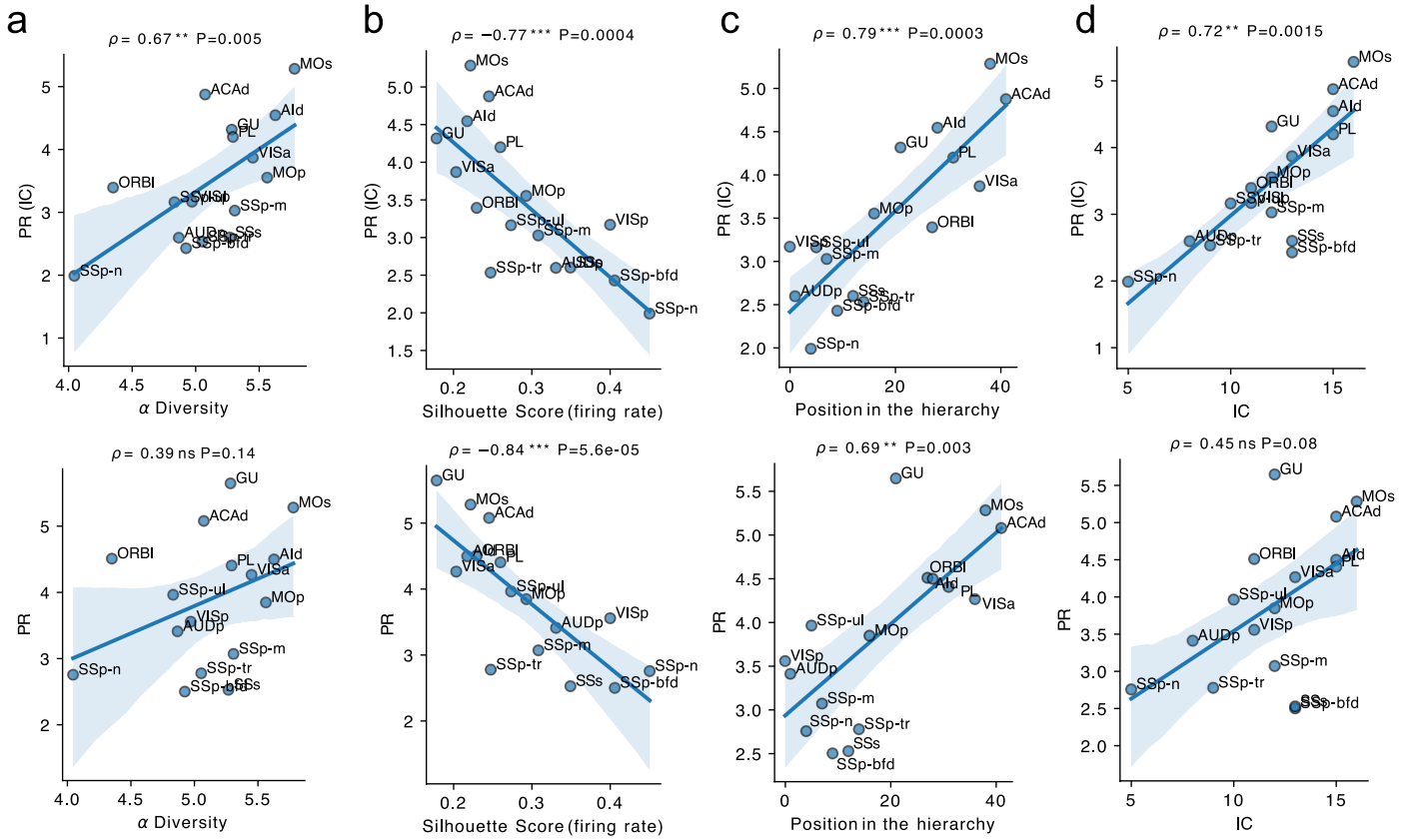


Extended Data Fig. 8 | See next page for caption.

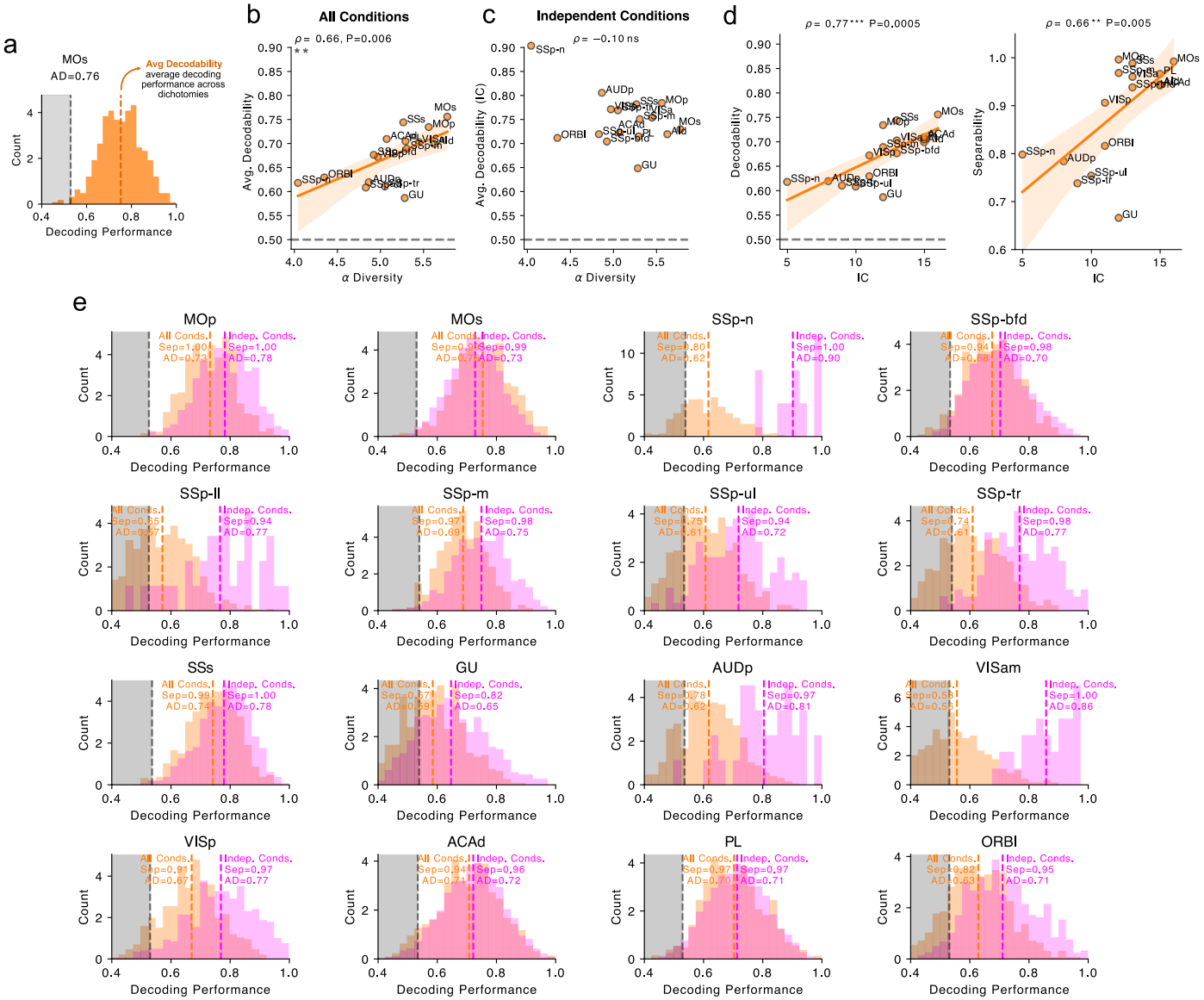
Extended Data Fig. 8 | Synthetic exploration and additional analysis of α-diversity. (**a**) Conceptual model of how uneven and categorical structures shape selectivity profiles. In the simulations, structure is imposed on the space of regression coefficients α for the neuronal response to a set of V binary variables, see Methods. Simulations are parametrized by a structure parameter that controls how strong these structures are. When this parameter is zero, alpha coefficients are maximally unstructured (random Gaussian). When the parameter is close to 1, alphas are maximally structured (categorical: perfect clusters; uneven: just one variable dominating the distribution). See Methods. (**b**) Simulations showing how Silhouette Score changes with α-diversity for the two models. Points show individual simulations, while error bars show mean and standard deviation over a running window of 10% of the x-range. We used $\gamma = 0.25$, $k = 2$, $V = 4$, $N = 100$ (see Methods). (**c**) α-diversity increases along the cortical hierarchy. (**d**) α-diversity is inversely correlated to the silhouette score in the conditions firing rate space. (**e**) Simulations showing how Representation Dimensionality, Separability, and Average Decodability change with α-diversity for the two models. Points show individual simulations, while error bars show mean and standard deviation over a running window of 10% of the x-range. We used $\gamma = 0.25$, $k = 2$, $V = 4$, $N = 100$ (see Methods). (**f**) Schematic of the subspace dimensionality analysis: in these simulations, P centroids are randomly sampled from an L-dimensional subspace of the N-dimensional activity space (see Methods). Trial-to-trial variability is then added to the centroids with a standard deviation σ (scaled with $\sqrt{L}$ to keep signal-to-noise invariant with L) in the full N-dimensional space. (**g**) Representation Dimensionality, Separability, and Average Decodability as a function of L for different values of the trial-to-trial noise. Low-dimensional ($L \lesssim M/2 = 8$) geometries limit the Separability, which saturates at larger dimensionalities. This behaviour is observed at all different levels of noise. Decodability, on the other hand, increases almost linearly with L for small noise levels, and sub-linearly with medium and high noise. Each error bar shows mean and standard deviation across 10 simulations with fixed level of noise and latent dimensionality L. These simulations were performed using N = 100, P = 16, T = 20. (**h**) While Representation dimensionality and Separability are generally linked, this relationship is not 1-1. It is indeed possible to construct a geometry that has maximal separability *and* low Representation dimensionality, by starting from a high-dimensional geometry and stretching one particular dimension while keeping the noise level unchanged. (**i**) This operation will decrease the Representation dimensionality while keeping separability intact. Decodability is also largely invariant upon this transformation (and is mostly determined by the noise level).



Extended Data Fig. 9 | Theoretical Investigation of the relation between clustering of neural response profiles and embedding dimensionality of neural representations. (**a**) Schematic of how categorical versus non-categorical selectivity in the conditions space shapes representational dimensionality. Categorical clusters in the conditions space imply a low-dimensional geometry, while non-categorical responses allow for a high-dimensional geometry. Note that low-dimensional geometries are compatible with non-categorical representations (see[3]), as long as the embedding dimensionality is the same across the two spaces. The intra-clustering diversity δ is quantified as the ratio of the dispersion of points within each cluster (denoted as σ) to the distance between different clusters (denoted as s). (**b**) Top: analytical relation between the features of a set of clusters in the conditions space (number of clusters k, number of conditions M, intra-cluster diversity δ) and their Representation dimensionality (PR) in the neural space. Bottom: visualization of how the Representation dimensionality (PR) is limited by clustering quality and how this limit varies with the number of clusters k (left panel) and the number of conditions M (right panel). For these visualizations, intra-clustering diversity δ was converted into a silhouette score to aid comparison with the data (see Methods). (**c**) Coherently with the theoretical model, Representation dimensionality decreases with increasing Silhouette Score in the conditions space (Spearman's $\rho = -0.77$, $P = 0.0004$). (**d**) Correlation between the value for PR predicted by the theory, using k, δ and M_{IC} (y-axis) and that computed from the centroids of the M_{IC} conditions in the neural space. The blue line is the best fit zero-intercept slope, while the dashed black line is the identity function $y = x$. (**e**) Similar to **d**, the intra-cluster diversity δ in the conditions space could also be quantitatively predicted from M_{IC}, PR, and k using an inversion of the equation in **b**.



Extended Data Fig. 10 | a) Representation dimensionality computed on independent conditions increases with α – diversity (top). PR of all conditions exhibit a non-significant trend in the same direction (bottom). **b)** Both Representation dimensionality computed on independent conditions (top, same data as Fig. 5c) and on all conditions (bottom) are inversely related to the Silhouette Score. **c)** Both measures increase along the cortical hierarchy (top: same data as Fig. 5d). **d)** Representation dimensionality of independent conditions increases with the number of independent conditions (top); a similar trend (non-significant) is observed for Representation dimensionality of all conditions.



Extended Data Fig. 11 | (a) Distribution of linear decoding performance across $n = 200$ random dichotomies of the $M = 16$ conditions in MOs (same data as Fig. 6b). Average Decodability (AD) is defined as the mean of this distribution. (**b**) AD increases with α-diversity when computed on all conditions. (**c**) This relationship is not present when computing AD for independent conditions only. (**d**) Across cortical areas, both AD and Separability are positively correlated with the number of independent conditions. (**e**) Distribution of decoding performance of $n = 200$ random dichotomies of all conditions (orange) and independent conditions (magenta) for all the regions used in Main Figs. 5 and 6. Each performance was obtained by a cross-validated analysis of the decoding performance of a random balanced binary classification of the independent conditions found for each region. The cross-validation was performed on mean firing rate vectors corresponding to individual trials. Within each analysis, individual conditions were balanced within each side of the dichotomy to avoid grouped or large conditions dominating the analysis (see Methods). Separability was then computed as the fraction of dichotomy decoding performances that was larger than a null distribution found by shuffling the labels of individual trials, while Average Decodablility (AD) was computed as the mean decoding performance across the sampled random dichotomies.

nature portfolio

Corresponding author(s): Lorenzo Posani; Stefano Fusi

Last updated by author(s): Apr 6, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection, as we did not collect any data.
Data analysis	The following python packages were used for the analysis: decodanda 0.7.3; scikit-learn 1.7.0; numpy 2.2.0; scipy 1.16.0; seaborn 0.13.2; matplotlib 3.10.3; pandas 2.2.3; torch 2.6.0+cu124; one 3.0.0; iblatlas 0.5.6; ibllib 3.4.1; the custom code used for the clustering analysis is available on GitHub: https://github.com/realwsq/brainwide-RRR-encoding-model.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

No new datasets were generated in this study. The electrophysiological data analysed here are publicly available from the International Brain Laboratory (IBL) Brain-Wide Map resource and were accessed through the IBL ONE interface/API (International Brain Laboratory et al., Nature, 2025). Cortical connectivity data used for

the hierarchy analyses were obtained from the public Allen Institute MouseBrainHierarchy resource associated with Harris et al. (Nature, 2019). Custom code used to retrieve, preprocess and analyse these data is available as described in the Code Availability section.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	this study does not involve human participants
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. This study is a secondary analysis of publicly available datasets, and sample sizes were determined by the number of sessions, trials and neurons available in the International Brain Laboratory Brain-Wide Map dataset that satisfied predefined inclusion criteria. We analyzed approximately 14,000 cortical units from approximately 180 recording sessions spanning 43 cortical regions. Additional analysis-specific inclusion criteria were applied where appropriate, including minimum trial counts per condition, minimum predictive performance thresholds for neuron inclusion in selectivity analyses, and minimum numbers of neurons per region for clustering analyses, as described in the Methods. These sample sizes were sufficient to support the statistical analyses reported here, and the main results were robust across reasonable variations in analysis thresholds.
Data exclusions	As clarified in the main text and Methods, neurons were included only if their firing activity was selective in any way to at least one of the variables analyzed. This was measured by taking the difference between the predictive power (measured with R2) of a multi-variable linear regression model and that of a model that is agnostic to variables (only time in trial). In the main text, we use delta R2=0.015. In the Supplementary material, we show that results are robust across different choices of this threshold.
Replication	No experiments were performed specifically for this study. This work is a secondary analysis of previously collected public datasets, and no independently repeated experiments were generated by the authors. Analytical robustness was assessed through cross-validation and through supplementary analyses testing the stability of the results across different inclusion thresholds.
Randomization	There was no group allocation in this study.
Blinding	There was no group allocation in this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	The study involved previously collected public mouse datasets. No new animal experiments were performed by the authors.
Wild animals	No wild animals were involved.
Reporting on sex	Both male and female mice were included in the source dataset. Sex was not a variable analyzed in the present study; further details are provided in the original dataset publication.
Field-collected samples	No field-collected samples were involved.
Ethics oversight	No new ethical approval was required for this study because it involved only secondary analysis of previously collected public datasets. Ethical approval and animal oversight for the original experiments are described in the source publication for the International Brain Laboratory dataset.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	This study does not involve plants
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*