The Effective Number of Shared Dimensions Between Paired Datasets

Hamza Giaffar*

University of California, San Diego
Dptmnt of Neurobiology &
Halıcıoğlu Data Science Institute

Camille Rullán Buxó*

New York University
Center for Neural Science

Mikio Aoi
maoi@ucsd.edu
University of California, San Diego
Dptmnt of Neurobiology &
Halıcıoğlu Data Science Institute

Abstract

A number of recent studies have sought to understand the behavior of both artificial and biological neural networks by comparing representations across layers, networks and brain areas. Increasingly prevalent, too, are comparisons across modalities of data, such as neural network activations and training data or behavioral data and neurophysiological recordings. One approach to such comparisons involves measuring the dimensionality of the space shared between the paired data matrices, where dimensionality serves as a proxy for computational or representational complexity. Established approaches, including CCA, can be used to measure the number of shared embedding dimensions, however they do not account for potentially unequal variance along shared dimensions and so cannot measure *effective* shared dimensionality. We present a candidate measure for shared dimensionality that we call the effective number of shared dimensions (ENSD). The ENSD is an interpretable and computationally efficient model-free measure of shared dimensionality that can be used to probe shared structure in a wide variety of data types. We demonstrate the relative robustness of the ENSD in cases where data is sparse or low rank and illustrate how the ENSD can be applied in a variety of analyses of representational similarities across layers in convolutional neural networks and between brain regions.

Proceedings of the 27[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238. Copyright 2024 by the author(s).

1 INTRODUCTION

Modern experiments across a range of scientific and technical fields are generating increasingly high dimensional datasets across multiple modalities. In neuroscience, for example, we can now combine information from various sources, including electrophysiological recordings from different brain regions, or from the same neurons under different conditions. Additionally, comparisons of such measured neural activity with different data modalities, from behavioural variables to neural network models, are increasing used to build and test theories of neural function.

One broad approach to unravelling the relationships contained in a given pair of multi-area and/or multimodal datasets involves decomposing the total variability in the two data matrices into independent and shared components. While both independent and shared components are typically informative, in the context of neuroscience, the subspace spanned by the shared variability between neural populations or modalities of data is of particular interest as it can often give useful insight into the computational processes at play (Semedo et al., 2019).

These subspaces can be individually characterized through their dimensionality, a measure which is often cast as a surrogate for computational or representational complexity. For example, low dimensional representations in neural networks are thought to make these robust to noise, improve their task performance, and offer interpretable and compact representations of data for downstream decoding. High dimensional representations, on the other hand, have a higher expressivity and more flexibly adapt to newer tasks (Sorscher et al., 2022). Thus, dimensionality is a way to understand the performance of a network and its ability to represent a wide range of data. This approach has been applied to biological and artificial neural network behaviour (Litwin-Kumar et al., 2017; Saxena

and Cunningham, 2019; Recanatesi et al., 2019; Ma et al., 2018), and an interesting vein of recent work has also used dimensionality to measure the accuracy of neural network models of biological systems (Elmoznino and Bonner, 2022; Canatar et al., 2023). The dimensionality of shared subspaces, however, has received comparatively little attention.

A number of recent studies have begun to explore the relationship between the dimensionality of neural responses and that of connected tasks or stimuli, finding that the two are related in theory (Gao et al., 2017) and in experiment (Semedo et al., 2019). To date, the study of shared dimensionality has relied on model based methods that may not generalize easily to other contexts and often require sequential hypothesis testing for each candidate shared dimension. Until now, a simple, model free measure has been lacking.

In this work, we present a candidate model-free measure of the dimensionality of the shared subspace between two datasets, which we call the *effective number of shared dimensions* (ENSD). The ENSD is a generalization of the participation ratio (PR), a widely used measure of the effective dimensionality for a single dataset (Del Giudice, 2021), to two datasets that share at least one dimension. We show that this measure can also be decomposed into interpretable components that describe geometric features of the shared subspace.

We start by exploring the properties of the ENSD, including an associated distance metric and measure of eigenvector alignment, using simple manipulations of synthetic data. We then contrast the ENSD with CKA and CCA, two frequently used techniques for analysing shared structure in paired datasets, finding that the ENSD provides a descriptive measure of shared dimensionality that is exceptional in its robustness to sparse data. Finally, we demonstrate how these tools can be used to probe both artificial and biological neural systems in different modalities: i) convolutional neural network activations, ii) neural recordings from the visual cortex and iii) sparse connectomic data from the fruit fly. We conclude with some discussion about areas of future development. Although the present work focuses on neuroscience and machine learning applications of the ENSD, interest in multimodal data analysis reaches well beyond these fields as integration of data from different sources becomes routine in applications across the sciences.

2 APPROACH

2.1 Background

The participation ratio The participation ratio (PR) is a measure of the dimensionality of a data matrix. Since its introduction in atomic spectroscopy, the PR has found use in disciplines from quantum and condensed matter physics to economics, sociology and machine learning. The PR has also found wide use in neuroscience (Altan et al., 2021; Mazzucato et al., 2016; Recanatesi et al., 2019, 2021; Litwin-Kumar et al., 2017), and is considered a reasonable measure of neural data dimensionality (Gao et al., 2017).

To estimate the PR, consider an $n \times p$ matrix $\mathbf{X}$, where the covariance of the rows of $\mathbf{X}$ are given by $\mathbf{C}_X = \frac{1}{p}\mathbb{E}[\mathbf{X}\mathbf{X}^\top]$. We assume $\mathbf{X}$ to be centered without loss of generality. We then denote the eigenvalues of $\mathbf{C}_X$ as $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. The PR is given by

$$\gamma_X = \frac{(\sum_{i=1}^{p} \lambda_i)^2}{\sum_{i=1}^{p} \lambda_i^2} = \frac{\text{trace}(\mathbf{C}_X)^2}{\text{trace}(\mathbf{C}_X^2)},$$

where we can estimate γ_X by noting that $\gamma_X \approx \frac{\text{trace}(\mathbf{X}\mathbf{X}^\top)^2}{\text{trace}(\mathbf{X}\mathbf{X}^\top\mathbf{X}\mathbf{X}^\top)}$. This measure quantifies the dispersion (inverse concentration) of the eigenvalue distribution, giving a more intuitive sense for the effective dimensionality of a data set than many of the existing alternatives (Gao et al., 2017; Recanatesi et al., 2020, 2021).

To see this, consider the set of n non-negative eigenvalues $\lambda_1, \ldots, \lambda_n$. By normalizing these values we can write the distribution of the variance over the eigenvectors: $\hat{\lambda}_i = \lambda_i / \sum_j \lambda_j$. A common measure of the concentration of a distribution is $\sum_{i=1}^{n} \hat{\lambda}_i^2$, which is constrained to the interval $[\frac{1}{n}, 1]$. Therefore, $(\sum_{i=1}^{n} \hat{\lambda}_i^2)^{-1}$ varies on $[1, n]$, where the lower limit corresponds to the most concentrated scenario, where all of the variance is described by one eigenvector, and n to the scenario where all of the normalized eigenvalues are equally distributed and equal to $1/n$. If λ_i are the eigenvalues of $\mathbf{C}_X$, then $\gamma_X = (\sum_{i=1}^{n} \hat{\lambda}_i^2)^{-1}$ is simply a measure of how evenly distributed the variance is among all eigenvectors. For several typical eigenspectra, the PR explains between 80-90% of the variance in a dataset (Gao et al., 2017).

2.2 Our Contribution: The effective number of shared dimensions

The PR can also be written as a matrix inner product. We show this by defining the scaled covariance matrix, $\hat{\mathbf{C}}_X = \frac{\mathbf{X}\mathbf{X}^\top}{\text{trace}(\mathbf{X}\mathbf{X}^\top)}$. Multiplying this quantity by the participation ratio of $\mathbf{X}$ we obtain the quantity $\mathbf{G}_X = \gamma_X \hat{\mathbf{C}}_X$. Then the PR can be expressed as

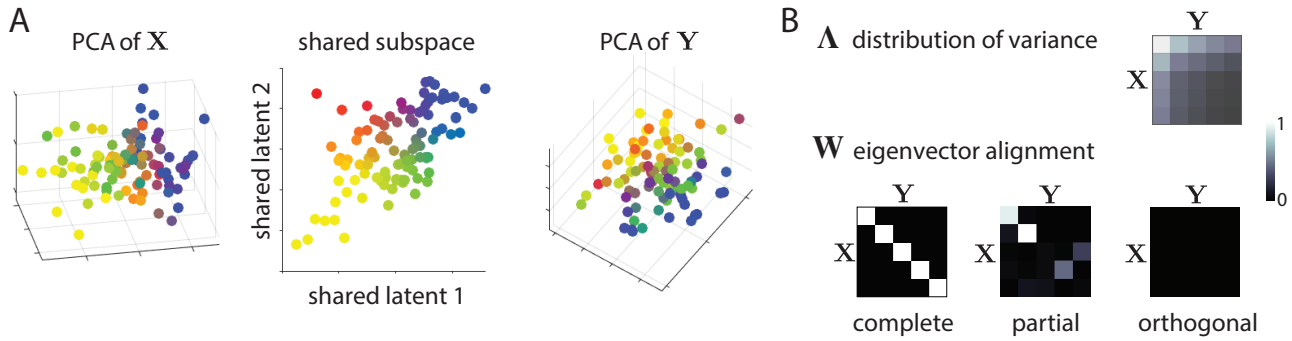


Figure 1: **A.** Schematic illustrating ENSD. **B.** Constituent parts of the similarity measure $\tau_{X,Y}$: $\mathbf{\Lambda}$, a rank-1 matrix of eigenvalue (outer) products (top), $\mathbf{W}$ a matrix of eigenvector overlaps (bottom). Examples of $\mathbf{W}$ are shown for datasets with i) fully aligned, ii) partially aligned and iii) orthogonal eigenvectors. $\mathbf{W}$s constructed from top 5 eigenvectors are shown, ordered by eigenvalue magnitude.

$$\gamma_X = \text{trace}(\mathbf{G}_X \mathbf{G}_X).$$

While the PR is a measure of the dimensionality of one dataset, the above representation for the PR suggests a simple generalization that can be applied to obtain the shared dimensionality of two datasets, $\mathbf{X}$ and $\mathbf{Y}$. The only requirement is a shared first dimension, n, so we define $\mathbf{Y}$ to be $n \times q$. We can now write the expression

$$\begin{aligned}\nu_{X,Y} &= \text{trace}(\mathbf{G}_X \mathbf{G}_Y) \\ &= \text{trace}(\gamma_X \hat{\mathbf{C}}_X \gamma_Y \hat{\mathbf{C}}_Y) \\ &= \gamma_X \gamma_Y \tau_{X,Y}\end{aligned}$$

where $\mathbf{G}_Y$ is defined as above, γ_Y is the participation ratio of $\mathbf{Y}$, and $\tau_{X,Y} = \text{trace}(\hat{\mathbf{C}}_X \hat{\mathbf{C}}_Y)$.

We call $\nu_{X,Y}$ the **effective number of shared dimensions** (ENSD). Just as the PR is a description of the dispersion of variance across eigenvectors, the ENSD is both a description of how much of the variability across datasets can be jointly explained and how dispersed that explanation is (Fig.1A). In the following sections, we explore the various properties of this shared dimensionality measure analytically and through a series of illustrative toy examples.

Properties of the ENSD The following are properties of the ENSD which can be proven analytically. The full derivations can be found in the supplementary material (SM1).

Equality with γ_X – The ENSD reduces to the PR when both matrices are equivalent up to orthonormal transformation. That is, $\nu_{X,X} = \gamma_X$.

Upper bound – Just as the PR is upper bounded by $\min\{n, p\}$, the ENSD also admits an upper bound: $\nu_{X,Y} \leq \sqrt{\gamma_X \gamma_Y}$. Note that the upper bound is not $\min(\gamma_X, \gamma_Y)$; the shared subspace can be of higher effective dimensionality than $\mathbf{X}$ or $\mathbf{Y}$. This occurs when dimensions of e.g. $\mathbf{X}$ which contribute little to γ_X align with high variance dimensions of $\mathbf{Y}$, amplifying the contribution of these low variance dimensions. The upper bound is achieved when $\mathbf{X} = \mathbf{YU}$, where $\mathbf{U}$ is an arbitrary orthonormal matrix, although there may be other conditions under which the upper bound is achieved.

Decomposition in terms of eigenvalues and eigenvectors – The ENSD may be rewritten in terms of the eigenvalues and eigenvectors of its constituent matrices. The scaled covariance matrix can be diagonalized as $\hat{\mathbf{C}}_X = \mathbf{U}_X L_X \mathbf{U}_X^\top$, where L_X is the diagonal matrix of normalized eigenvalues $\hat{\lambda}_X$ and $\mathbf{U}$ are the principal axes of the data. We may then write

$$\begin{aligned}\tau_{X,Y} &= \text{trace}(\mathbf{U}_X L_X \mathbf{U}_X^\top \mathbf{U}_Y L_Y \mathbf{U}_Y^\top) \\ &= \sum_i \sum_j \hat{\lambda}_{i,X} \hat{\lambda}_{j,Y} \text{trace}(\mathbf{u}_{i,X}^\top \mathbf{u}_{j,Y} \mathbf{u}_{j,Y}^\top \mathbf{u}_{i,X}) \\ &= \sum_i \sum_j \hat{\lambda}_{i,X} \hat{\lambda}_{j,Y} w_{i,j}^2, \qquad (1)\end{aligned}$$

where $\mathbf{u}_{i,X}$ is the i^{th} eigenvector of $\hat{\mathbf{C}}_X$, and $w_{i,j} = \mathbf{u}_{i,X}^\top \mathbf{u}_{j,Y}$ (for a full derivation, see SM1). This expression explicitly shows how $\nu_{X,Y}$ is a function of both the eigenvalues of $\hat{\mathbf{C}}_X$ and $\hat{\mathbf{C}}_Y$ and the inner products of their constituent eigenvectors. For $\mathbf{X} = \mathbf{Y}$, the entries of $w_{i,j}^2 = 1 \; \forall i = j$ and $w_{i,j}^2 = 0$ otherwise, and we therefore find $\gamma_X = \frac{1}{\sum_i \hat{\lambda}_{i,X}^2}$.

Since every term in (17) is indexed by i and j we can rewrite them as entries in the i, j^{th} position in a matrix. Let $\boldsymbol{\lambda}_X^\top = (\lambda_{1,X}, \lambda_{2,X}, \dots)$, and

$$\mathbf{W} = \begin{pmatrix} w_{1,1}^2 & w_{1,2}^2 & \cdots \\ w_{2,1}^2 & w_{2,2}^2 & \cdots \\ \vdots & & \ddots \end{pmatrix}$$

then $\tau_{X,Y} = \boldsymbol{\lambda}_X^\top \mathbf{W} \boldsymbol{\lambda}_Y$ and therefore the ENSD can be expressed in terms of eigenvalues of eigenvectors:

$$\nu_{X,Y} = \frac{\boldsymbol{\lambda}_X^\top \mathbf{W} \boldsymbol{\lambda}_Y}{(\boldsymbol{\lambda}_X^\top \boldsymbol{\lambda}_X)(\boldsymbol{\lambda}_Y^\top \boldsymbol{\lambda}_Y)}.$$

This way of expressing $\tau_{X,Y}$ allows for an intuitive interpretation of this factor: it captures the overlap between the two subspaces via $\mathbf{W}$, which is then scaled by the similarities of the eigenspectra, $\boldsymbol{\Lambda} = \boldsymbol{\lambda}_X \otimes \boldsymbol{\lambda}_Y$ (Fig.1B). This similarity measure is then weighted by the individual dimensionalities of the two subspaces in the equation for $\nu_{X,Y}$ to arrive at a final numerical estimate for the dimensionality of the shared subspace.

Corresponding distance metric – Williams et al. (Williams et al., 2021) noted that inner products can be converted to distances via an arccosine transformation. We can apply such a transformation such that the distance metric

$$d_{\nu_{X,Y}} = \frac{2}{\pi} \arccos \sqrt{\gamma_X \gamma_Y} \tau_{X,Y}$$

satisfies the equivalence and symmetry requirements as well as the triangle inequality and is therefore a proper distance metric. That we can define such a dissimilarity metric is interesting, first because we can demonstrate a formal relationship between the dimensionality of shared subspaces and a distance in a space defined by a set of equivalence relations. Second, because this intuition can be put to work in down-stream analyses such as k-nearest neighbors, hierarchical clustering, and a suite of other methods for representational similarity analysis.

3 RELATED WORK

The geometry of shared subspaces has been studied in a wide range of contexts in neuroscience under the rubric of representational similarity analysis (Schoonover et al., 2021; Zhuang et al., 2020; Mirza et al., 2018; Zhou et al., 2018; Freiwald and Tsao, 2010), and machine learning (Kornblith et al., 2019; Williams et al., 2021; Morcos et al., 2018; Wang et al., 2018), as well as in the intersection of the two domains (Khaligh-Razavi and Kriegeskorte, 2014; Yamins et al., 2014).

Of the methods developed and employed in this literature, the ENSD is closely related to the Hilbert-Schmidt Independence Criterion (HSIC) (Gretton et al., 2005) and Centered Kernel Alignment (CKA) measure (Kornblith et al., 2019). Specifically, the CKA is the ENSD normalized by its upper bound: $\text{CKA}(\mathbf{X}, \mathbf{Y}) = \frac{\nu_{X,Y}}{\sqrt{\gamma_X \gamma_Y}} = \sqrt{\gamma_X \gamma_Y} \tau_{X,Y}$. This relationship implies that the CKA score can be interpreted as the fraction of possible shared dimensionality that a system is using and allows us to draw a connection between similarity indices and dimensionality.

Nevertheless, to the best of our knowledge there are no other methods explicitly designed to compute the *effective* number of shared dimensions between two datasets in a manner analogous to dimensionality estimation for a single dataset via the PR. However, a number of methods can be employed to compute the *embedding* dimensionality of shared subspaces. These include CCA, where the number of shared dimensions between two datasets is estimated as the number of significant canonical correlations. Model based regression techniques such as reduced rank regression (RRR) have also been applied to measure the dimensionality of shared 'communication' subspaces between neural populations (Semedo et al., 2019). We compare these methods to the ENSD in the context of shared subspace dimensionality estimation below.

4 PROBING THE ENSD WITH SYNTHETIC DATA

In this section, we illustrate important properties of the ENSD via application to synthetic data. First, we clarify the difference between *effective* shared dimensionality, as estimated by the ENSD, and the *embedding* dimensionality as estimated via CCA. Second, we demonstrate how shared subspace dimensionality estimation via ENSD relates to two alternative methods, CCA and RRR, in the sparse and low sample number data regimes.

Effective vs. embedding dimensionality As shown above, the ENSD incorporates both the degree of overlap between the set of basis vectors and the spectral decay in both datasets, giving an estimate of *effective* shared dimensionality. We contrast this with the integer valued estimate of the *embedding* dimensionality obtained via CCA as the number of significant canonical correlations (with $p < 0.05$). Although many variants of CCA exist, here we focus on the original method as outlined in Hotelling (1936).

We compare estimated dimensionality via ENSD ($\nu_{X,Y}$) and CCA (D_{CCA}) for simple synthetic datasets $\mathbf{X}$ and $\mathbf{Y}$ (of sizes $n \times p$ and $n \times q$ respectively), with singular value decompositions given by $\mathbf{X} = (\mathbf{U}_1, \mathbf{U}_2)\mathbf{S}_X\mathbf{V}_X^\top$, where $\mathbf{S}_X = \text{diag}(\sigma_{X,1}, \ldots, \sigma_{X,p})$ and $\mathbf{Y} = (\mathbf{U}_1, \mathbf{U}_3)\mathbf{S}_Y\mathbf{V}_Y^\top$, where

Table 1: *Effective vs strict/embedding dimensionality*

	aligned basis		partially aligned basis			
	flat λ	decay λ	flat λ $\uparrow 5$	flat λ $\downarrow 5$	decay λ $\uparrow 5$	decay λ $\downarrow 5$
ENSD	10	3.4	5	5	3.1	0.1
CCA	10	10	5	5	5	5

Average estimated dimensionality over $n = 100$ synthetic datasets. $\uparrow$: highest variance dimensions are shared, $\downarrow$: lowest variance dimensions are shared.

$\mathbf{S}_Y = \text{diag}(\sigma_{Y,1}, \ldots, \sigma_{Y,q})$. The shared subspace is defined by the $n \times r$ matrix $\mathbf{U}_1$, while $\mathbf{U}_2$ is $n \times (p-r)$ and $\mathbf{U}_3$ is $n \times (q-r)$. We assume $\mathbf{U}_i^\top \mathbf{U}_j = 0$ for $i \neq j$. The $\nu_{X,Y}$ can then be written as

$$\nu_{X,Y} = \frac{(\sum_{i=1}^{r} \sigma_{X,i}^2 \sigma_{Y,i}^2)(\sum_{i=1}^{p} \sigma_{X,i}^2)(\sum_{i=1}^{q} \sigma_{Y,i}^2)}{(\sum_{i=1}^{p} \sigma_{X,i}^4)(\sum_{i=1}^{q} \sigma_{Y,i}^4)} \quad (2)$$

When all singular values are equal (flat spectra), we obtain

$$\nu_{X,Y} = \frac{(r\sigma_X^2 \sigma_Y^2)(p\sigma_X^2)(q\sigma_Y^2)}{(p\sigma_X^4)(q\sigma_Y^4)} = r \quad (3)$$

and $\nu_{X,Y}$ is precisely the number of matched eigenvectors as expected. When datasets $\mathbf{X}$ and $\mathbf{Y}$ contain low dimensional structure, their (ordered) singular value spectra will reflect this as a decay. We model such spectra as exponentials with decay rate β_X and β_Y. As β increases, the spectra decay more quickly and thus γ_X, γ_Y and $\nu_{X,Y}$ all decrease towards the lower limit of 1 (where all variance is concentrated on the first eigenvector).

To illustrate how $\nu_{X,Y}$ and D_{CCA} vary as functions of a) eigenvector alignment and b) spectral decay rates, we analyze simple synthetic data in which we independently modulate these components (details in SM2). For a), we consider i) $r = 10$ (complete alignment) and ii) $r = 5$ (partial alignment). For b) we consider both i) flat spectra ($\sigma_i = \frac{1}{n}, \forall i$) and ii) decaying spectra ($\beta_X = \beta_Y > 0$).

The results of this analysis are shown in Table 1. Both methods agree on the number of shared dimensions when spectra are flat: i) $\nu_{X,Y} = D_{CCA} = r = 10$ (col. 1) and ii) $\nu_{X,Y} = D_{CCA} = r = 5$ (cols. 3 and 4, respectively). When $\beta > 0$, the ENSD and CCA disagree, reflecting the fact that the ENSD incorporates information about the distribution of variance in the shared subspace. That is, when the subspaces for $\mathbf{X}$ and $\mathbf{Y}$ overlap but variance is unequally shared across these dimensions, the ENSD estimates a lower effective dimensionality, (e.g., $\nu_{X,Y} = 3.4$, vs. $D_{CCA} = 10$ col. 2)

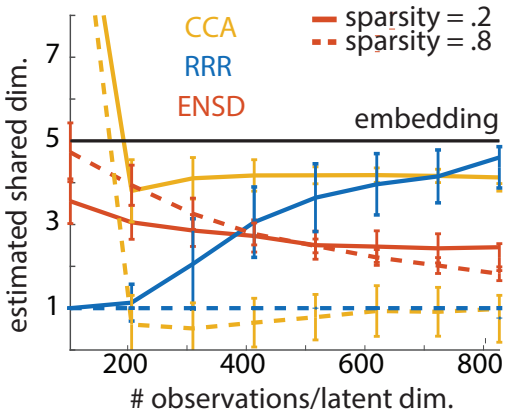


Figure 2: Comparison of dimensionality estimation accuracy for 3 methods in a simulated experiment, averaged over 50 experiments at different sample sizes and two levels of sparsity. Data were generated using the pCCA model (Bach and Jordan, 2005) with 5 shared latent variables. Details in SM3.

When only a subset of basis vectors are shared between the two datasets, the variance along those shared dimensions will determine the effective dimensionality. When basis vectors with the highest variance are shared , the estimated dimensionality is significantly larger than the case in which basis vectors with the lowest variance are shared ($\uparrow, D = 3.1$ vs $\downarrow, 0.1$, cols. 5 and 6, Table 1). In both cases, CCA correctly identifies the embedding dimensionality of the shared subspace, $D = 5$. In SM2, we also compare shared dimensionality estimates via ENSD and CCA as a function of spectral decay in the above synthetic data. CCA robustly identifies the embedding dimensionality even when spectral decay is significant, while ENSD smoothly decreases with the decay rate, accounting for the increasingly uneven distribution of variance with spectral decay (SM-Fig.6).

Estimating shared dimensionality in sparse and low sample regimes Here, we compare shared dimensionality estimation via ENSD to i) CCA (Hotelling, 1936) and ii) RRR (Izenman, 1975), for which we used the cross validation procedures outlined in (Semedo et al., 2019). We generated data using the probabilistic CCA model (Bach and Jordan, 2005; Browne, 1979) with two data sources each with observation dimension of size 30, 5 shared latent dimensions and private latent dimensions of 10 and 15 (details in SM3). In addition to varying the number of samples used, we randomly sparsified the observations at two

levels.

Two observations are apparent from our analysis (Fig.2). First, the ENSD is the most robust of the three methods to variations in both sample size and observation sparsity. Second, the ENSD is systematically biased downwards from the number of shared dimensions used to generate the data. This bias is consistent with the interpretation of ENSD as the *effective* number of shared dimensions. As we discussed in Section 2.2, any degree to which there is unequal variance of shared components would contribute to a deviation in the estimated number of dimensions. We therefore expect ENSD to routinely report shared dimensionality that is somewhat smaller than the integer number of dimensions needed to embed the shared structure.

5 APPLICATION TO ARTIFICIAL NEURAL NETWORKS

A recent line of work has sought to evaluate deep neural network (DNN) models of the visual system via their dimensionality. For example, Elmoznino and Bonner (2024) found that the ability of DNNs to predict neural activity was strongly positively correlated with the PR of their internal representations of natural images, irrespective of task optimization or architecture. The authors also introduce the notion of alignment pressure (AP), which describes how strongly aligned high variance dimensions are across representations. The ability of a DNN to predict neural responses, they reason, should be a combination of the breadth of its internal representation (the PR) and how well that representation describes the data (the AP). A DNN with a low PR, for example, can predict neural data reasonably well if its AP is high. The authors do not, however, provide a measure to quantify the AP of two datasets; the ENSD fills this gap, formalizing the empirical relationship between AP (through the $\mathbf{W}$ matrix) and PR.

Inspired by a recent study (Kornblith et al., 2019) which explored representational similarity within and between layers of convolutional neural networks (CNNs) using CKA, we applied the ENSD to measure the shared dimensionality between layers of a CNN. We trained a *TinyTen* CNN (Tatro et al., 2020) to classify the CIFAR-10 dataset and computed both the CKA and ENSD between all pairs of layers, repeating this across six random initializations of the same network. The average layer by layer ENSDs and CKAs across initializations are shown in (Fig.3A,B). This analysis reveals that the dimensionality of representations as measured by the PR (main diagonal, Fig.3A) first increases and then decreases with layer depth. The ENSD between subsequent layers (eg. along the

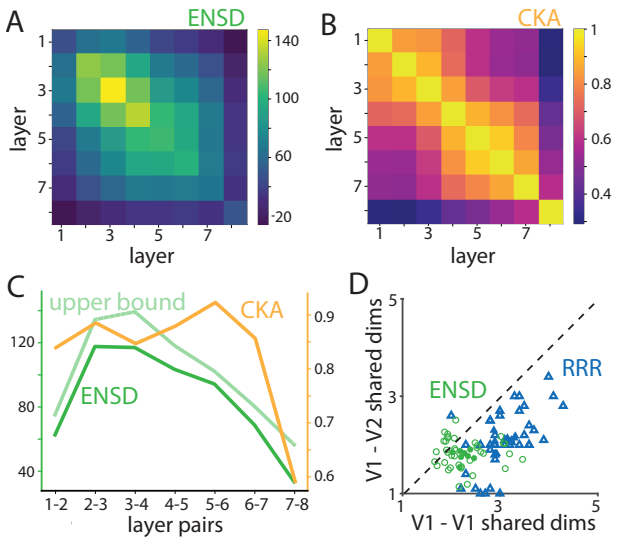


Figure 3: **A.** Across layer ENSD and **B.** CKA averaged over 6 networks. **C.** Average ENSD, upper bound and CKA across layers (off-diagonals of A and B). **D.** Comparison of shared dimensionality estimates between visual cortical areas using RRR and ENSD. Open markers represents a stimulus orientation for one session, closed markers are averages over sessions.

1st diagonal) follows this same pattern of expansion and contraction. This kind of initial expanding of dimensionality may be helpful in expanding representations for ease of learning classification boundaries (Sorscher et al., 2022).

In contrast, the average between-layer CKA remains relatively constant across layers (Fig.3B). We note that the CKA can be interpreted as the *fraction* of the maximum possible ENSD that the measured system uses (see section 2.2). Since this proportion stays relatively constant and close to the maximum, early layers seem to be learning a representation of the data that remains relatively unchanged throughout the network. A decrease towards the readout layer may be a consequence of the network learning a lower dimensional representation that optimizes for the particular image classification task. That the ENSD includes dimensionality information therefore provides a more detailed assessment of the relationships between representations within a network.

6 APPLICATION TO NEURAL DATA

Re-analysis of visual electrophysiology data In a recent study, Semedo et al. (Semedo et al., 2019) examined correlated neural activity between primary and secondary visual cortices (i.e. V1, V2). Their analysis

demonstrated that correlations in trial-by-trial fluctuations were detectable both within and between regions and that these correlations were indicative of a shared ("communication") subspace that displayed dimensionality similar to the dimensionality of the stimulus used in the task. The authors used RRR to determine the dimensionality of this subspace. We propose that our measure can be equivalently used to conduct this analysis without model fitting.

We repeated their analysis, applying the ENSD to their data (Fig.3D) and obtaining qualitatively similar results for both target V1 and V2 populations from a size-matched source V1 population, although our results are somewhat downward-biased, as expected (see Fig.3D). Because we were able to conduct our analysis without model fitting, computation time was an order of magnitude smaller using ENSD as compared to RRR (see SM4), opening the door to analysis of much larger-scale datasets in the future.

Analysis of olfactory connectivity data The olfactory system faces a set of fundamentally difficult challenges, including the rapid identification and discrimination of odor stimuli drawn from the extremely high dimensional space of volatile chemicals (Mayhew et al., 2022). Odorants evoke responses in a large repertoire of receptors, which compress this high dimensional space via a poorly understood combinatorial code. In the fly, for example, 51 genetically defined receptors relay information, via antennal lobe projection neurons (AL PNs), to two downstream areas (Fig.4A): i) the mushroom body (MB), an area that has been shown to mediate adaptive olfactory learning (Aso et al., 2014) and ii) the lateral horn (LH), a relatively poorly characterized area thought to mediate innate olfactory behaviour (Das Chakraborty and Sachse, 2021). Understanding how the olfactory system makes consistent sense of the complex chemical space via the interaction of parallel processes is of importance beyond neuroscience, as the computational principles supporting olfaction can be applied to more general problems in AI.

Recent work has explored the relationship between the fruit fly olfactory network's connectivity, the effective dimensionality (PR) of its representations and performance in discrimination tasks, finding (similarly to the the DNNs described in Section 5) that high dimensional representations facilitate classification (Litwin-Kumar et al., 2017). Here, we use the ENSD to further probe structure-function relationships in the fly. We make use of a recently released connectomic dataset from a single fly to probe AL inputs to the two aforementioned downsteam areas (AL→LH and AL→MB connectivity data contained in highly sparse matrices $\mathbf{L}$ and $\mathbf{M}$ respectively, Fig.4B). Our results from Section 2.2 demonstrate that the ENSD is well suited to shared dimensionality estimation for such sparse datasets.

First, we asked to what extent the feedforward connectivity to these areas contains interpretable structure. We address this by computing the participation ratio of $\mathbf{L}$ and $\mathbf{M}$ and comparing these values to the corresponding null models. Our randomizing procedure removes correlations between channels but preserves the marginal statistics of the data matrices (Caron et al., 2013) (see SM5 for details). Deviations from randomness indicate the presence of a structured component. Our analysis reveals that i) both connectivity matrices contain a degree of nonrandom organization and ii) $\mathbf{L}$ is significantly more structured than $\mathbf{M}$ (Fig.4C). This is in line with results showing that input to the MB is largely unstructured, facilitating high dimensional representations for associative learning (Litwin-Kumar et al., 2017; Zheng et al., 2022), while inputs to the LH are more structured, with lower dimensional representations driving fast, innate behavioural programs.

To determine if the structure identified in the two pathways is related and significant, we compare the ENSD to that of shuffled data via $\Delta\nu = \frac{\nu(\mathbf{L},\tilde{\mathbf{M}})-\nu(\mathbf{L},\mathbf{M})}{\nu(\mathbf{L},\mathbf{M})}$ (the tilde indicates a shuffled matrix). The decrease in shared dimensionality (and increase in distance) after shuffling indicates that the LH shares a statistically significant representational subspace with the MB, emerging from similar input connectivity patterns (Fig.4D).

The $\mathbf{W}$ matrix allows us to probe this alignment further, as it directly describes the overlap between the eigenvectors odf the two matrices. We find a single statistically significant and strongly overlapping dimension: the first PC of $\mathbf{M}$ and the second PC of $\mathbf{L}$ (Fig.4E). We also found statistically significant but weak overlaps between two other pairs of eigenvectors. The strongly aligned subspace is therefore 1-dimensional. Comparing eigenvector loadings onto this shared subspace reveals that the structure is dominated by a set of channels that are tuned to food related odorants (specifically yeast and fermented fruit, Fig.4F). This is consistent with findings from (Choi et al., 2022), but our analysis also demonstrates that this food dominated shared subspace is statistically significant (i.e. not simply a byproduct of having more connections and therefore a higher connection probability amongst such channels).

Finally, we use a second dataset containing AL→MB connectivity from another fly (Zheng et al., 2022) to i) confirm that the shared input structure to LH and MB is consistent across individuals (see SM5) and ii) show

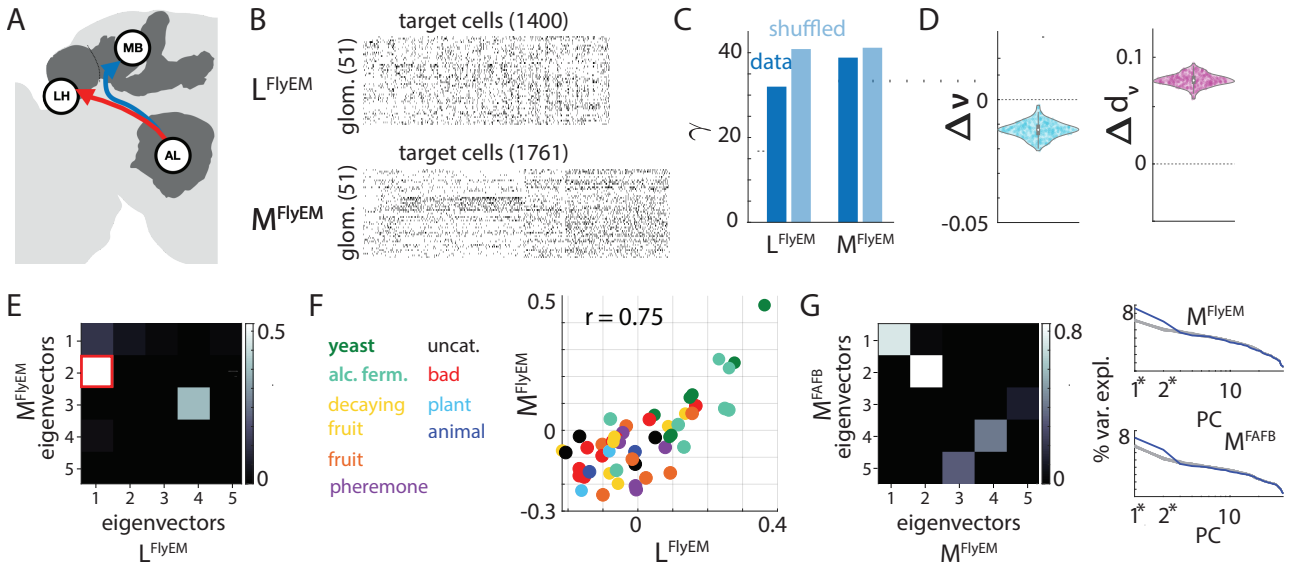


Figure 4: **A.** Schematic of the parallel pathways of the fly olfactory system. **B.** Binarized connectivity matrix between the 51 types of glomeruli in the AL and the principal neurons in the LH or MB. **C.** PRs of the connectivity matrices for the LH (left) and MB (right) and corresponding null models. **D.** Effect of shuffling on the ENSD (left) and the distance (right) between the LH and MB connectivity data. **E. W** matrix of eigenvector overlaps between the LH and MB. Eigenvector overlaps corresponding to the top 5 eigenvalues shown, all statistically insignificant overlaps ($p > 0.05$), are set to zero. **F.** Loadings of the different glomeruli onto the shared dimension (red border in E), color-coded by their odorant tuning (from (Choi et al., 2022)). **G. W** matrix showing eigenvector overlaps between the two MB datasets (left) and their corresponding eigenspectra (right) compared to null models (grey line). Star indicates statistical significance ($p < 0.05$).

that all of the structure in the MB inputs is shared across the two individuals (Fig.4G).

Overall these analyses reveal that i) part of the structure observed in inputs to the LH and MB is shared between the two areas, ii) this shared subspace relates to an ethologically relevant odor scene and iii) is consistent across individuals, and iv) all of the structure in the MB inputs is shared between individuals. The food-tuned shared subspace between LH and MB inputs revealed by our analysis may play an important role in the circuitry that permits olfactory learning in the MB pathway to alter innate food related behaviour mediated by the LH pathway, as recently observed in (Eschbach and et al, 2021; Lerner et al., 2020). SM5 contains additional ENSD based multimodal analyses of these connectivity matrices and paired odorant odorant evoked neural activity.

7 DISCUSSION

The capacity to characterize shared structure between diverse datasets will be increasingly important in many scientific fields as high dimensional and multimodal data collection becomes routine. Here we present the ENSD, a simple and interpretable tool to characterize shared linear subspaces in paired datasets. We use synthetic data to probe the behavior of the ENSD with respect to known structure, as well as comparing it to existing data analysis methods. We bring this comparison to the realm of real data in three examples: i) comparing CKA and ENSD analyses of CNN activations, showing that ENSD provides a richer insight into the structure of activity, ii) comparing ENSD to RRR, showing how the ENSD can be used to more easily arrive at existing results from electrophysiology experiments, and iii) showing how the ENSD can be used to generate novel analyses of fly olfactory connectivity data. We find that strategically probing the shared dimensionality of pairs of connectivity datasets reveals interpretable shared structure both within and between the inputs to two neuropils of the fly olfactory system (Fig.4). Furthermore, we find that these interpretable structures are consistent across individuals, implying a genetic origin. That we were able to recover all of this from application of a single measure to connectomic data serves to showcase the utility and interpretability of the technique.

We have shown that the ENSD has a number of advantages compared to existing, model-based methods for estimating shared variability in paired datasets:

robustness to sample size and observation sparsity, minimal computation time (just 5 matrix multiplications), and easily interpretable components. The ENSD also offers an interpretation of the recently introduced CKA score as a fraction of the possible shared dimensionality that a system is using. Because of this relationship, we expect that critiques of CKA also apply to the ENSD, including a lack sensitivity to low variance dimensions (Davari et al., 2022; Ding et al., 2021; Nguyen et al., 2021). Shared structure in low variance dimensions can still be detected via the empirical **W** matrix and an appropriate null model for the data.

Another important aspect to consider is the robustness of the ENSD to noise. A previous analysis of linear dimensionality estimation revealed that the PR tends to overestimate dimensionality when sufficient noise is added to synthetic data or when the underlying data manifold is nonlinear (Altan et al., 2021) and it is likely that the ENSD inherits these properties. However, amongst common shared dimensionality estimators, the ENSD changes the least in response to variations in sample size and random sparsening rate (Fig.3.A). Both of these features are indicative of a relative robustness to various sources of noise. Moreover, alternative techniques typically require costly multiple hypothesis testing to estimate shared dimensionality, a requirement that is troublesome for exploratory analyses.

While we focused on use cases in neuroscience and machine learning, we stress that this technique can be productively applied in other fields, from physics to sociology, to ask related questions about shared variance contained in data. Important areas for future development include i) extending the ENSD to incorporate scale-dependence, which can reveal differences in local and global structure that may be obscured by other dimensionality measures (Recanatesi et al., 2020), ii) extending the measure to non-linear dimensionality metrics via kernelization of the ENSD, and iii) extending the ENSD via a generalized notion of effective dimension (Del Giudice, 2021).

Acknowledgements

The authors are grateful to Stephen Keeley, Edoardo Balzani, Sergey Shuvaev and the reviewers for their many useful comments on the manuscript.

References

Ege Altan, Sara A. Solla, Lee E. Miller, and Eric J. Perreault. Estimating the dimensionality of the manifold underlying multi-electrode neural recordings. *PLOS Computational Biology*, 17(11):e1008591, 2021.

Yoshinori Aso, Daisuke Hattori, Yang Yu, Rebecca M. Johnston, Nirmala A. Iyer, Teri-T. B. Ngo, Heather Dionne, L. F. Abbott, Richard Axel, Hiromu Tanimoto, and Gerald M. Rubin. The neuronal architecture of the mushroom body provides a logic for associative learning. *eLife*, 3:e04577, 2014.

Francis R Bach and Michael I Jordan. A probabilistic interpretation of canonical correlation analysis. *Technical Report 688, Department of Statistics, University of California*, 2005.

L. Badel, K. Ohta, Y. Tsuchimoto, and H. Kazama. Decoding of context-dependent olfactory behavior in drosophila. *Neuron*, 91(1):155–67, 2016.

Michael W Browne. The maximum-likelihood solution in inter-battery factor analysis. *British Journal of Mathematical and Statistical Psychology*, 32(1):75–86, 1979.

Abdulkadir Canatar, Jenelle Feather, Albert Wakhloo, and SueYeon Chung. A spectral theory of neural prediction and alignment, 2023.

Sophie J. C. Caron, Vanessa Ruta, L. F. Abbott, and Richard Axel. Random convergence of olfactory inputs in the drosophila mushroom body. *Nature*, 497 (7447):113–117, 2013.

Kiri Choi, Won Kyu Kim, and Changbong Hyeon. Olfactory responses of *Drosophila* are encoded in the organization of projection neurons. *eLife*, 11, 2022.

Sudeshna Das Chakraborty and Silke Sachse. Olfactory processing in the lateral horn of drosophila. *Cell and Tissue Research*, 383(1):113–123, 2021.

Mohammad-Reza Davari, Stefan Horoi, Amine Natik, Guillaume Lajoie, Guy Wolf, and Eugene Belilovsky. Reliability of cka as a similarity measure in deep learning, 2022.

Marco Del Giudice. Effective dimensionality: A tutorial. *Multivariate Behav Res*, 56(3):527–542, 2021.

Frances Ding, Jean-Stanislas Denain, and Jacob Steinhardt. Grounding representation similarity with statistical testing, 2021.

Eric Elmoznino and Michael Bonner. Latent dimensionality scales with the performance of deep learning models of visual cortex. *Journal of Vision*, page 22(14):3704, 2022.

Eric Elmoznino and Michael F. Bonner. High-performing neural network models of visual cortex benefit from high latent dimensionality. *PLOS Computational Biology*, 20(1):1–23, 01 2024.

Claire Eschbach and et al. Circuits for integrating learned and innate valences in the insect brain. *eLife*, 10:e62567, 2021.

W. Freiwald and D. Tsao. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, 330(6005): 845–851, 2010.

Peiran Gao, Eric Trautmann, Byron Yu, Gopal Santhanam, Stephen Ryu, Krishna Shenoy, and Surya Ganguli. A theory of multineuronal dimensionality, dynamics and measurement. *BioRxiv*, page 214262, 2017.

A Gretton, O Bousquet, A Smola, and Schölkopf B. Measuring statistical dependence with hilbert-schmidt norms. 2005.

Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321, 1936.

Alan Julian Izenman. Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, 5(2):248–264, 1975.

Seyed-Mahdi Khaligh-Razavi and Nikolaus Kriegeskorte. Deep supervised, but not unsupervised, models may explain it cortical representation. *PLOS Computational Biology*, 10(11):e1003915, 2014.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 3519–3529. PMLR, 09–15 Jun 2019.

Hadas Lerner, Eyal Rozenfeld, Bar Rozenman, Wolf Huetteroth, and Moshe Parnas. Differential role for a defined lateral horn neuron subset in naïve odor valence in drosophila. *Scientific Reports*, 10(1):6147, 2020.

A. Litwin-Kumar, K. D. Harris, R. Axel, H. Sompolinsky, and L. F. Abbott. Optimal degrees of synaptic connectivity. *Neuron*, 93(5):1153–1164.e7, 2017.

Xingjun Ma, Bo Li, Yisen Wang, Sarah M. Erfani, Sudanthi Wijewickrema, Grant Schoenebeck, Dawn Song, Michael E. Houle, and James Bailey. Characterizing adversarial subspaces using local intrinsic dimensionality, 2018.

Emily J. Mayhew, Charles J. Arayata, Richard C. Gerkin, Brian K. Lee, Jonathan M. Magill, Lindsey L. Snyder, Kelsie A. Little, Chung Wen Yu, and Joel D. Mainland. Transport features predict if a molecule is odorous. *Proceedings of the National Academy of Sciences*, 119(15):e2116576119, 2022.

L. Mazzucato, A. Fontanini, and G. La Camera. Stimuli reduce the dimensionality of cortical activity. *Frontiers in Systems Neuroscience*, 10, 2016.

M. Mirza, R. Adams, C. Mathys, and K. Friston. Human visual exploration reduces uncertainty about the sensed world. *PLOS ONE*, 13(1):e0190429, 2018.

Ari S. Morcos, Maithra Raghu, and Samy Bengio. Insights on representational similarity in neural networks with canonical correlation. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, page 5732–5741, 2018.

Thao Nguyen, Maithra Raghu, and Simon Kornblith. Do wide and deep networks learn the same things? uncovering how neural network representations vary with width and depth. In *International Conference on Learning Representations*, 2021.

Stefano Recanatesi, Gabriel Koch Ocker, Michael A. Buice, and Eric Shea-Brown. Dimensionality in recurrent spiking networks: Global trends in activity and local origins in connectivity. *PLOS Computational Biology*, 15:1–29, 2019.

Stefano Recanatesi, Serena Bradde, Vijay Balasubramanian, Nicholas A Steinmetz, and Eric Shea-Brown. A scale-dependent measure of system dimensionality. *bioRxiv*, 2020.

Stefano Recanatesi, Matthew Farrell, Guillaume Lajoie, Sophie Deneve, Mattia Rigotti, and Eric Shea-Brown. Predictive learning as a network mechanism for extracting low-dimensional latent space representations. *Nature communications*, 12(1):1–13, 2021.

Shreya Saxena and John P Cunningham. Towards the neural population doctrine. *Current Opinion in Neurobiology*, 55:103–111, 2019. Machine Learning, Big Data, and Neuroscience.

Louis K et al. Scheffer. A connectome and analysis of the adult *Drosophila* central brain. *eLife*, 9, 2020.

Philipp Schlegel, Alexander Shakeel Bates, Tomke Stürner, Sridhar R. Jagannathan, Nikolas Drummond, Joseph Hsu, Laia Serratosa Capdevila, Alexandre Javier, Elizabeth C. Marin, Asa Barth-Maron, Imaan F. M. Tamimi, Feng Li, Gerald M. Rubin, Stephen M. Plaza, Marta Costa, and Gregory S. X. E. Jefferis. Information flow, cell types and

stereotypy in a full olfactory connectome. *eLife*, 10: e66018, 2021.

Carl E. Schoonover, Sarah N. Ohashi, Richard Axel, and Andrew J. P. Fink. Representational drift in primary olfactory cortex. *Nature*, 594(7864):541–546, 2021.

João D Semedo, Amin Zandvakili, Christian K Machens, M Yu Byron, and Adam Kohn. Cortical areas interact through a communication subspace. *Neuron*, 102(1):249–259, 2019.

Ben Sorscher, Surya Ganguli, and Haim Sompolinsky. Neural representational geometry underlies few-shot concept learning. *Proceedings of the National Academy of Sciences*, 119(43), October 2022.

Norman Tatro, Pin-Yu Chen, Payel Das, Igor Melnyk, Prasanna Sattigeri, and Rongjie Lai. Optimizing mode connectivity via neuron alignment. *Advances in Neural Information Processing Systems*, 33, 2020.

Liwei Wang, Lunjia Hu, Jiayuan Gu, Yue Wu, Zhiqiang Hu, Kun He, and John Hopcroft. Towards understanding learning representations: To what extent do different neural networks learn the same representation, 2018.

Alex H Williams, Erin Kunz, Simon Kornblith, and Scott Linderman. Generalized shape metrics on neural representations. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Daniel L. K. Yamins, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, Darren Seibert, and James J. DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111 (23):8619–8624, 2014.

Amin Zandvakili and Adam Kohn. Coordinated neuronal activity enhances corticocortical communication. *Neuron*, 87(4):827–839, 2015.

Amin Zandvakili and Adam Kohn. Paired v1-v2 neuronal spiking responses in anesthetized macaque monkey. *CRCNS.org*, 2019.

Zhihao Zheng, Feng Li, Corey Fisher, Iqbal J. Ali, Nadiya Sharifi, Steven Calle-Schuler, Joseph Hsu, Najla Masoodpanah, Lucia Kmecova, Tom Kazimiers, Eric Perlman, Matthew Nichols, Peter H. Li, Viren Jain, and Davi D. Bock. Structured sampling of olfactory input by the fly mushroom body. *Current Biology*, 32(15):3334–3349.e6, 2022.

Y. Zhou, B. Smith, and T. Sharpee. Hyperbolic geometry of the olfactory space. *Science Advances*, 2018.

Xiaowei Zhuang, Zhengshi Yang, and Dietmar Cordes. A technical review of canonical correlation analysis for neuroscience applications. *Human Brain Mapping*, 41(13):3807–3833, 2020.

Supplementary Material: The Effective Number of Shared Dimensions Between Paired Datasets

SM1 Properties of $\nu_{X,Y}$

In the following discussion, we consider the $n \times p$ matrix $\mathbf{X}$, and the $n \times q$ matrix $\mathbf{Y}$, where $n > p, q$.

Matrix expression for $\nu_{X,Y}$

$$\nu_{X,Y} = \frac{\text{trace}(\mathbf{Y}^\top \mathbf{X}\mathbf{X}^\top \mathbf{Y})\text{trace}(\mathbf{X}^\top \mathbf{X})\text{trace}(\mathbf{Y}^\top \mathbf{Y})}{\text{trace}(\mathbf{X}^\top \mathbf{X}\mathbf{X}^\top \mathbf{X})\text{trace}(\mathbf{Y}^\top \mathbf{Y}\mathbf{Y}^\top \mathbf{Y})}$$

Equality with $\gamma_{\mathbf{X}}$

$$\begin{aligned} \nu_{X,Y} &= \gamma_X \gamma_X \tau_{X,X} && (4)\\ &= \gamma_X \gamma_X \text{trace}[\hat{\mathbf{C}}_X \hat{\mathbf{C}}_X] && (5)\\ &= \gamma_X \gamma_X \frac{1}{\gamma_X} && (6)\\ &= \gamma_X && (7) \end{aligned}$$

Upper bound

Just as γ_X is upper bounded by p, the ENSD also admits an upper bound. To see this, first recall that the Cauchy-Schwartz inequality gives

$$(\mathbf{u}^\top \mathbf{v})^2 \leq (\mathbf{u}^\top \mathbf{u})(\mathbf{v}^\top \mathbf{v})$$

for any vectors $\mathbf{u}$ and $\mathbf{v}$. Noting that $\mathbf{u}$ and $\mathbf{v}$ in the above expression may be the vectorizations of matrices $\mathbf{U}$ and $\mathbf{V}$ we may equivalently write

$$\text{trace}[\mathbf{U}^\top \mathbf{V}]^2 \leq \text{trace}[\mathbf{U}^\top \mathbf{U}]\text{trace}[\mathbf{V}^\top \mathbf{V}].$$

Substituting $\mathbf{U} = \hat{\mathbf{C}}_X$ and $\mathbf{V} = \hat{\mathbf{C}}_Y$ gives

$$\text{trace}[\hat{\mathbf{C}}_X^\top \hat{\mathbf{C}}_Y]^2 \leq \text{trace}[\hat{\mathbf{C}}_X^\top \hat{\mathbf{C}}_X]\text{trace}[\hat{\mathbf{C}}_Y^\top \hat{\mathbf{C}}_Y], \tag{8}$$

and using the notation presented in Section 2 of the main paper we may write inequality (8) as

$$\tau_{X,Y}^2 \leq \frac{1}{\gamma_X \gamma_Y},$$

and therefore

$$\sqrt{\gamma_X \gamma_Y}\tau_{X,Y} \leq 1. \tag{9}$$

Finally, multiplying both sides of inequality (9) by $\sqrt{\gamma_X \gamma_Y}$ gives the upper bound for $\nu_{X,Y}$,

$$\nu_{X,Y} \leq \sqrt{\gamma_X \gamma_Y}.$$

and additionally, it is clear that $\tau_{X,Y}$ is upper bounded by the inverse of the geometric mean of the participation ratios, $\frac{1}{\sqrt{\gamma_X \gamma_Y}}$.

Decomposition in terms of eigenvalues and eigenvectors

The ENSD may be rewritten in terms of the eigenvalues and eigenvectors of it's constituent matrices, via the scaled diagonalized covariance matrix, $\text{trace}[\hat{\mathbf{C}}_X] = \mathbf{U}_X L_X \mathbf{U}_X$, where L_X is a matrix with eigenvalues $\hat{\lambda}_X$ on the diagonal and $\mathbf{U}$ are the principal axes of the data. We first rewrite the trace term,

$$\begin{aligned}
\tau_{X,Y} &= \text{trace}[\hat{\mathbf{C}}_X \hat{\mathbf{C}}_Y] && (10)\\
&= \text{trace}[\mathbf{U}_X L_X \mathbf{U}_X^\top \mathbf{U}_Y L_Y \mathbf{U}_Y^\top] && (11)\\
&= \text{trace}\left[\left(\sum_i \hat{\lambda}_{i,X} \mathbf{u}_{i,X} \mathbf{u}_{i,X}^\top\right)\left(\sum_j \hat{\lambda}_{j,Y} \mathbf{u}_{j,Y} \mathbf{u}_{j,Y}^\top\right)\right] && (12)\\
&= \text{trace}\left[\sum_i \sum_j \hat{\lambda}_{i,X} \hat{\lambda}_{j,Y} \mathbf{u}_{i,X} \mathbf{u}_{i,X}^\top \mathbf{u}_{j,Y} \mathbf{u}_{j,Y}^\top\right] && (13)\\
&= \sum_i \sum_j \hat{\lambda}_{i,X} \hat{\lambda}_{j,Y} \text{trace}\left[\mathbf{u}_{i,X} \mathbf{u}_{i,X}^\top \mathbf{u}_{j,Y} \mathbf{u}_{j,Y}^\top\right] && (14)\\
&= \sum_i \sum_j \hat{\lambda}_{i,X} \hat{\lambda}_{j,Y} \text{trace}[\mathbf{u}_{i,X}^\top \mathbf{u}_{j,Y} \mathbf{u}_{j,Y}^\top \mathbf{u}_{i,X}] && (15)\\
&= \sum_i \sum_j \hat{\lambda}_{i,X} \hat{\lambda}_{j,Y} \text{trace}[w_{i,j} w_{i,j}] && (16)\\
&= \sum_i \sum_j \hat{\lambda}_{i,X} \hat{\lambda}_{j,Y} w_{i,j}^2, && (17)
\end{aligned}$$

where $w_{i,j} \equiv \mathbf{u}_{i,X}^\top \mathbf{u}_{j,Y}$. Equation (17) explicitly shows how $\nu_{X,Y}$ is a function of both the eigenvalues of $\hat{\mathbf{C}}_X$ and $\hat{\mathbf{C}}_Y$ and also the inner products of their constituent eigenvectors. If we rewrite the participation ratio in the same way, we find that $w_{i,j} = 1 \; \forall i, j$, so

$$\gamma_X = \frac{1}{\sum_i \hat{\lambda}_{i,X}^2},$$

allowing us to rewrite the ENSD as

$$\nu_{X,Y} = \frac{\sum_i \sum_j \hat{\lambda}_{i,X} \hat{\lambda}_{j,Y} w_{i,j}^2}{(\sum_i \hat{\lambda}_{i,X}^2)(\sum_j \hat{\lambda}_{j,Y}^2)}.$$

Since every term in (17) is indexed by i and j we can rewrite each in terms of the i, j^{th} entries of a matrix. Let $\boldsymbol{\lambda}_X^\top = (\lambda_{1,X}, \lambda_{2,X}, \dots)$, and

$$\mathbf{W} = \begin{pmatrix} w_{1,1}^2 & w_{1,2}^2 & \dots \\ w_{2,1}^2 & w_{2,2}^2 & \dots \\ \vdots & & \ddots \end{pmatrix} \qquad \boldsymbol{\Lambda} = \boldsymbol{\lambda}_X \boldsymbol{\lambda}_Y^\top$$

then

$$\begin{aligned}
\tau_{X,Y} &= \mathbb{1}^\top (\boldsymbol{\Lambda} \odot \mathbf{W}) \mathbb{1} && (18)\\
&= \boldsymbol{\lambda}_X^\top \mathbf{W} \boldsymbol{\lambda}_Y && (19)
\end{aligned}$$

and

$$\nu_{X,Y} = \frac{\boldsymbol{\lambda}_X^\top \mathbf{W} \boldsymbol{\lambda}_Y}{(\boldsymbol{\lambda}_X^\top \boldsymbol{\lambda}_X)(\boldsymbol{\lambda}_Y^\top \boldsymbol{\lambda}_Y)}.$$

This way of writing out $\tau_{X,Y}$ allows for an intuitive interpretation of this term: it captures the overlap between two subspaces with $\mathbf{W}$, which is then scaled by the respective eigenspectra, $\mathbf{\Lambda}$. This similarity measure is then weighted by the individual dimensionalities of the two subspaces in the equation for $\nu_{X,Y}$ to arrive at a final numerical estimate for the dimensionality of the shared subspace.

The corresponding distance metric, according to Williams et al. (2021), may also be expressed in this way, as

$$d_{\nu_{X,Y}} = \frac{2}{\pi} \arccos \frac{\boldsymbol{\lambda}_X^\top \mathbf{W} \boldsymbol{\lambda}_Y}{(\boldsymbol{\lambda}_X^\top \boldsymbol{\lambda}_X)^{1/2} (\boldsymbol{\lambda}_Y^\top \boldsymbol{\lambda}_Y)^{1/2}} \tag{20}$$

$$= \frac{2}{\pi} \arccos \frac{\boldsymbol{\lambda}_X^\top \mathbf{W} \boldsymbol{\lambda}_Y}{\|\boldsymbol{\lambda}_X\|_2 \|\boldsymbol{\lambda}_Y\|_2} \tag{21}$$

SM2 Probing the ENSD using synthetic data

Partially shared subspaces

For the simplest toy example we construct a scenario where we define the number of shared dimensions between $\mathbf{X}$ and $\mathbf{Y}$ to be an integer r. Suppose the singular value decompositions of these matrices are given by the $n \times p$ matrix $\mathbf{X} = (\mathbf{U}_1, \mathbf{U}_2)\mathbf{S}_X\mathbf{V}_X^\top$, where $\mathbf{S}_X = \text{diag}(\sigma_{X,1}, \ldots, \sigma_{X,p})$ and the $n \times q$ matrix $\mathbf{Y} = (\mathbf{U}_1, \mathbf{U}_3)\mathbf{S}_Y\mathbf{V}_Y^\top$, where $\mathbf{S}_Y = \text{diag}(\sigma_{Y,1}, \ldots, \sigma_{Y,q})$, where the key feature we impose is that $\mathbf{U}_i \perp \mathbf{U}_j$ for $i \neq j$. The shared subspace is defined by the $n \times r$ matrix $\mathbf{U}_1$, while $\mathbf{U}_2$ is $n \times (p-r)$ and $\mathbf{U}_3$ is $n \times (q-r)$. A few relations follow from these assumptions:

$$\text{trace}(\mathbf{X}\mathbf{X}^\top) = \text{trace}\left((\mathbf{U}_1, \mathbf{U}_2)\, \mathbf{S}_X \mathbf{V}_X^\top \mathbf{V}_X \mathbf{S}_X^\top \begin{pmatrix} \mathbf{U}_1^\top \\ \mathbf{U}_2^\top \end{pmatrix} \right) \tag{22}$$

$$= \text{trace}\left((\mathbf{U}_1, \mathbf{U}_2)\, \mathbf{S}_X^2 \begin{pmatrix} \mathbf{U}_1^\top \\ \mathbf{U}_2^\top \end{pmatrix} \right) \tag{23}$$

$$= \text{trace}\left((\mathbf{U}_1, \mathbf{U}_2)\, \mathbf{S}_X \mathbf{V}_X^\top \mathbf{V}_X \mathbf{S}_X^\top \begin{pmatrix} \mathbf{U}_1^\top \\ \mathbf{U}_2^\top \end{pmatrix} \right) \tag{24}$$

$$= \text{trace}\left(\begin{pmatrix} \mathbf{U}_1^\top \\ \mathbf{U}_2^\top \end{pmatrix} (\mathbf{U}_1, \mathbf{U}_2)\, \mathbf{S}_X^2 \right) \tag{25}$$

$$= \text{trace}\left(\begin{pmatrix} \mathbf{U}_1^\top \mathbf{U}_1 & \mathbf{U}_1^\top \mathbf{U}_2 \\ \mathbf{U}_2^\top \mathbf{U}_1 & \mathbf{U}_2^\top \mathbf{U}_2 \end{pmatrix} \mathbf{S}_X^2 \right) \tag{26}$$

$$= \text{trace}\left(\begin{pmatrix} \mathbb{I}_r & \mathbf{0}_{r \times p-r} \\ \mathbf{0}_{p-r \times r} & \mathbb{I}_{p-r} \end{pmatrix} \mathbf{S}_X^2 \right) \tag{27}$$

$$= \text{trace}\left(\mathbf{S}_X^2 \right) \tag{28}$$

$$= \sum_{i=1}^{p} \sigma_{X,i}^2 \tag{29}$$

Following similar steps we find

$$\text{trace}(\mathbf{X}\mathbf{X}^\top\mathbf{X}\mathbf{X}^\top) = \text{trace}\left((\mathbf{U}_1, \mathbf{U}_2)\, \mathbf{S}_X^2 \begin{pmatrix} \mathbf{U}_1^\top \\ \mathbf{U}_2^\top \end{pmatrix} (\mathbf{U}_1, \mathbf{U}_2)\, \mathbf{S}_X^2 \begin{pmatrix} \mathbf{U}_1^\top \\ \mathbf{U}_2^\top \end{pmatrix} \right) \tag{30}$$

$$= \text{trace}\left((\mathbf{U}_1, \mathbf{U}_2)\, \mathbf{S}_X^2 \mathbf{S}_X^2 \begin{pmatrix} \mathbf{U}_1^\top \\ \mathbf{U}_2^\top \end{pmatrix} \right) \tag{31}$$

$$= \text{trace}\left(\mathbf{S}_X^2 \mathbf{S}_X^2 \right) \tag{32}$$

$$= \sum_{i=1}^{p} \sigma_{X,i}^4 \tag{33}$$

We therefore find that the PR for $\mathbf{X}$ can be written as

$$\gamma_X = \frac{\text{trace}(\mathbf{X}\mathbf{X}^\top)^2}{\text{trace}(\mathbf{X}\mathbf{X}^\top\mathbf{X}\mathbf{X}^\top)} \tag{34}$$

$$= \frac{(\sum_{i=1}^p \sigma_{X,i}^2)^2}{\sum_{i=1}^p \sigma_{X,i}^4} \tag{35}$$

Similarly,

$$\text{trace}(\mathbf{Y}\mathbf{Y}^\top) = \text{trace}\left((\mathbf{U}_1, \mathbf{U}_3)\mathbf{S}_Y\mathbf{V}_Y^\top\mathbf{V}_Y\mathbf{S}_Y^\top\begin{pmatrix}\mathbf{U}_1^\top \\ \mathbf{U}_3^\top\end{pmatrix}\right) \tag{36}$$

$$= \text{trace}\left((\mathbf{U}_1, \mathbf{U}_3)\mathbf{S}_Y^2\begin{pmatrix}\mathbf{U}_1^\top \\ \mathbf{U}_3^\top\end{pmatrix}\right) \tag{37}$$

$$= \text{trace}\left(\mathbf{S}_Y^2\right) \tag{38}$$

$$= \sum_{i=1}^p \sigma_{Y,i}^2 \tag{39}$$

Following equation (5) above for the formula for τ_{XY} we may write

$$\tau_{XY} = \frac{1}{\text{trace}(\mathbf{X}\mathbf{X}^\top)\text{trace}(\mathbf{Y}\mathbf{Y}^\top)}\text{trace}(\mathbf{X}\mathbf{X}^\top\mathbf{Y}\mathbf{Y}^\top) \tag{40}$$

$$= \text{trace}\left((\mathbf{U}_1, \mathbf{U}_2)\,\mathbf{S}_X^2\begin{pmatrix}\mathbf{U}_1^\top \\ \mathbf{U}_2^\top\end{pmatrix}(\mathbf{U}_1, \mathbf{U}_3)\,\mathbf{S}_Y^2\begin{pmatrix}\mathbf{U}_1^\top \\ \mathbf{U}_3^\top\end{pmatrix}\right) \tag{41}$$

$$= \text{trace}\left((\mathbf{U}_1, \mathbf{U}_2)\,\mathbf{S}_X^2\mathbf{S}_X^2\begin{pmatrix}\mathbf{U}_1^\top \\ \mathbf{U}_2^\top\end{pmatrix}\right) \tag{42}$$

$$= \text{trace}\left(\mathbf{S}_X^2\mathbf{S}_X^2\right) \tag{43}$$

$$= \sum_{i=1}^p \sigma_{X,i}^4 \tag{44}$$

and $\nu_{X,Y}$ can be written

$$\nu_{X,Y} = \frac{(\sum_{i=1}^r \sigma_{X,i}^2\sigma_{Y,i}^2)(\sum_{i=1}^p \sigma_{X,i}^2)(\sum_{i=1}^q \sigma_{Y,i}^2)}{(\sum_{i=1}^p \sigma_{X,i}^4)(\sum_{i=1}^q \sigma_{Y,i}^4)} \tag{45}$$

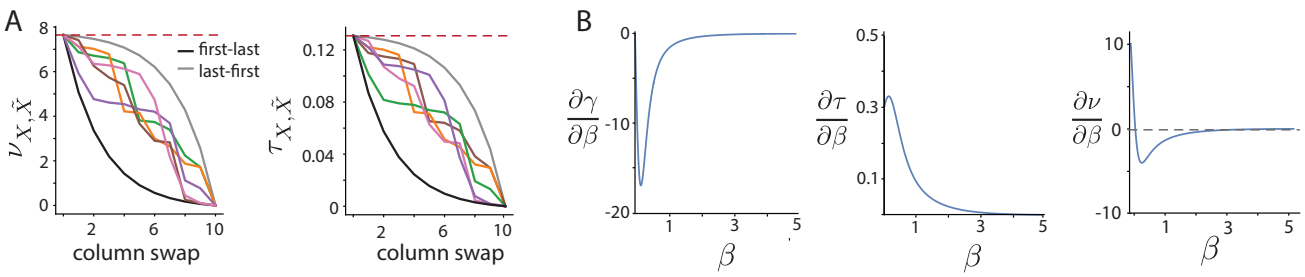
or, as described in the main text, if the dimensionality of both matrices are maximized (corresponding to flat eigenspectra),$\nu_{X,Y}$ reduces to exactly the number of shared eigenvectors.

Synthetic data in Section 4

We generate (zero-padded) data matrices by manipulating the singular value decompositions: $\mathbf{X} = (\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)_X\,\mathbf{S}_X\,(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3)_X^\top \in \mathbb{R}^{m\times p}$ and $\mathbf{Y} = (\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)_Y\,\mathbf{S}_Y\,(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3)_Y^\top \in \mathbb{R}^{m\times q}$. Matrices $\mathbf{U}_1 \in \mathbb{R}^{m\times a}$, $\mathbf{U}_2 \in \mathbb{R}^{m\times b}$ and $\mathbf{U}_3 \in \mathbb{R}^{m\times c}$, $a + b = c$ are a partition of a randomly generated orthonormal basis (eigenvectors are in columns). Matrix $\mathbf{S} = (\text{diag}(((\frac{\lambda_{1:c}}{\sum_i^c \lambda_i}), [0]_{1\times c})), [0]_{m\times d-m})$, where $d = p$ or q, $\lambda = e^{-\beta j}$ and j indexes singular values. We generate a set of common orthonormal bases $\mathbf{U}_A \in \mathbb{R}^{m\times a}$ and substitute these into columns of $\mathbf{U}_X$ and $\mathbf{U}_Y$ to construct the simple synthetic data considered in Section 4, with parameters $m = 2n,\ n = c = 10,\ a, b \in \{5, 10\}$, and $\beta_X = \beta_Y$.

Continuous partially shared subspaces

An illustrative extension of this simple example is a continuous relaxation of this model with the modification that $\mathbf{Y}(\alpha) = [\sqrt{\alpha}\mathbf{U}_1 + \sqrt{(1-\alpha)}\mathbf{U}_3, \quad \mathbf{U}_4]\mathbf{S}_Y\mathbf{V}_Y^\top$, where again all $\mathbf{U}_i$ are orthonormal and $\mathbf{U}_i \perp \mathbf{U}_j$ for $i \neq j$, and $\alpha \in [0, 1]$.



Supplementary Figure 5: **A.** Plot of $\nu_{X,\tilde{X}}$ and $\tau_{X,\tilde{X}}$ when swapping out n columns of U to generate $\tilde{U}$ in a particular order – the column corresponding to the first eigenvalue to last (black line), last to first (grey line) or in random order (colored lines). **B.** Plots of the derivatives of the PR (γ, left), the similarity term (τ, center) and the ENSD (ν, right) as a function of the eigenvalue decay of matrix $\mathbf{Y}$.

We then have

$$\text{trace}(\mathbf{Y}^\top\mathbf{Y}) = \sum_{i=1}^{q}\sigma_{Y,i}^2, \qquad \text{trace}(\mathbf{X}^\top\mathbf{X}) = \sum_{i=1}^{p}\sigma_{X,i}^2, \tag{46}$$

$$\text{trace}(\mathbf{X}^\top\mathbf{X}\mathbf{X}^\top\mathbf{X}) = \sum_{i=1}^{p}\sigma_{X,i}^4, \qquad \text{trace}(\mathbf{Y}^\top\mathbf{Y}\mathbf{Y}^\top\mathbf{Y}) = \sum_{i=1}^{q}\sigma_{Y,i}^4, \tag{47}$$

$$\text{trace}(\mathbf{X}^\top\mathbf{X}\mathbf{Y}^\top\mathbf{Y}) = \alpha\sum_{i=1}^{r}\sigma_{X,i}^2\sigma_{Y,i}^2 \tag{48}$$

Therefore,

$$\nu_{X,Y} \quad = \quad \alpha\frac{(\sum_{i=1}^{r}\sigma_{X,i}^2\sigma_{Y,i}^2)(\sum_{i=1}^{p}\sigma_{X,i}^2)(\sum_{i=1}^{q}\sigma_{Y,i}^2)}{(\sum_{i=1}^{p}\sigma_{X,i}^4)(\sum_{i=1}^{q}\sigma_{Y,i}^4)} \tag{49}$$

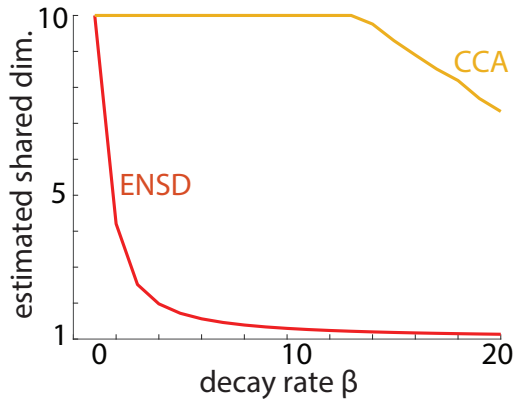
We can see that (49) is simply (45) multiplied by α. Therefore, in the case where the dimensionalities of $\mathbf{X}$ and $\mathbf{Y}$ are maximized (i.e. $\sigma_{X,i} = \sqrt{\mu_X}$ and $\sigma_{Y,i} = \sqrt{\mu_Y}, \forall i$) then we have, as expected, $\nu(\mathbf{X}, \mathbf{Y}(\alpha)) = \alpha r$. Additionally, since γ_X and γ_Y are constant, $\tau_{XY} \propto \alpha r$ as well and achieves the upper bound ($\frac{1}{\sqrt{\gamma_X\gamma_Y}}$) when $\alpha = 1$.

Dimensions explaining more variability contribute more to ν

Consider an $n \times p$ matrix $\mathbf{X} = \mathbf{U}_X\mathbf{S}_X\mathbf{V}_X^\top$, where $\mathbf{S}_X = \text{diag}(\sigma_1, \ldots, \sigma_p)$ and $\sigma_i > \sigma_{i+1}$, and a second basis $\mathbf{U}_{\tilde{X}}$, where again all columns of $\mathbf{U}_i$ are orthonormal and $\mathbf{U}_i \perp \mathbf{U}_j$ for $i \neq j$. We now generate a second matrix, $\tilde{\mathbf{X}}_k$, which is identical to $\mathbf{X}$ *except* that the kth column of $\mathbf{U}_X$ is replaced with the kth column of $\mathbf{U}_{\tilde{X}}$. Since all columns of $\mathbf{U}_X$ and $\mathbf{U}_{\tilde{X}}$ are mutually orthogonal, $\gamma_{\tilde{X}_k} = \gamma_X$ irrespective of which column of $\mathbf{U}_X$ is swapped. The only term in $\nu_{X,\tilde{X}}$ that is affected is the similarity term, $\tau_{X,\tilde{X}_k} = \mathbb{1}^\top(\mathbf{\Lambda} \odot \mathbf{W})\,\mathbb{1}$. Since the matrices are identical except for their eigenvectors, swapping columns of $\mathbf{U}_X$ only affects the matrix of eigenvector overlaps, $\mathbf{W}$.

For $i, j \neq k$, $W_{i,j} = 0$, as the associated eigenvectors are orthogonal, and for all $i = j \neq k, W_{i,j} = 1$, as $\mathbf{U}_{\tilde{X},j} = \mathbf{U}_{X,i}$. When $i, j = k$, now $W_{k,k} = 0$, whereas before the swap it was equal to 1. Since the kth term in the sum (17) is now zero, $\tau_{X,\tilde{X}_k}$ decreases by an amount $\hat{\lambda}_k^2$. The eigenvectors are strictly decreasing, so swapping out the kth eigenvector always results in a larger change to $\tau_{X,\tilde{X}}$, and therefore $\nu_{X,\tilde{X}}$, than swapping out the $(k+1)$th vector.

An example of this is shown in figure 5A, where we plot the ENSD and $\tau_{X,\tilde{X}}$ as a function of the number of columns swapped in $\tilde{X}$ for different swapping orders. In this case, both $\mathbf{X}$ and $\tilde{\mathbf{X}}$ are of dimensionality ~ 7.5. In the case where we successively swap the columns in $\mathbf{U}_X$ for columns in $\mathbf{U}_{\tilde{X}}$ from largest to smallest eigenvalues (black line), $\tau_{X,\tilde{X}}$ and $\nu_{X,\tilde{X}}$ decrease more quickly than if the columns were swapped from smallest to largest eigenvalue (grey line). Random column swaps (colored lines) all produce $\nu_{X,\tilde{X}}$ and $\tau_{X,\tilde{X}}$ within the envelope defined by these two modifications.



Supplementary Figure 6: Comparing estimated shared dimensionality via ENSD and CCA as a function of synthetic data spectral decay. Eigenvectors in both $\mathbf{X}$ and $\mathbf{Y}$ (10-dimensional datasets from Section 4 of main text), are aligned and spectral decay is defined by a common variable, β.

The ENSD captures amplification of shared variability

Finally, in section 3 of the main text we consider the case where the eigenspectra of the two matrices have different decay rates which produces some interesting phenomena. We consider two $n \times p$ matrices, $\mathbf{X} = \mathbf{U}\mathbf{S}_X\mathbf{V}^\top$ and $\mathbf{Y} = \mathbf{U}\mathbf{S}_Y\mathbf{V}^\top$, so $\mathbf{X}$ and $\mathbf{Y}$ are identical except for their rates of eigenvalue decay. The decrease in the PR is not linear, as shown by the plot of its derivative with respect to β (Fig.5B, left). Likewise, the derivative of $\tau_{X,Y}$ is strictly positive but decreasing over time as $\tau_{X,Y}$ approaches the upper bound (Fig.5B, center). In this limit, the shared dimensionality is contained entirely in the first dimension of $\mathbf{X}$ and $\mathbf{Y}$, and $\nu_{X,Y} \to 1$. The ENSD, however, shows some interesting behavior, first increasing then decreasing (see Fig.2C in the main text). We might expect that the ENSD would strictly decrease as the matrices become more different. However, its derivative goes from positive to negative in the brief period that $\frac{\partial \gamma}{\partial \beta} \leq \frac{\partial \tau}{\partial \beta}$ (Fig.5B, right), which causes the shared dimensionality to be greater than the participation ratio. After the inflection point, $\frac{\partial \gamma}{\partial \beta} > \frac{\partial \tau}{\partial \beta}$ the shared dimensionality is again below the participation ratio, and the decrease in ENSD is mainly due to the decrease in participation ratio of $\mathbf{Y}$.

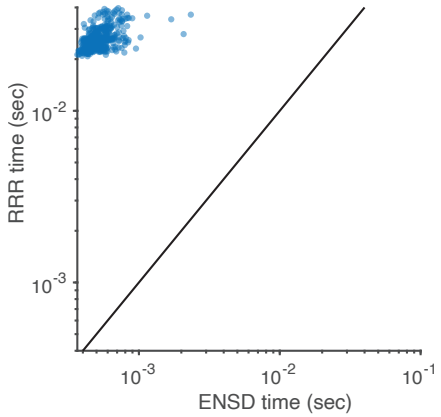
SM3 Comparison of shared dimensionality estimation techniques

We sought to evaluate the ENSD compared to alternative measures of shared dimensionality using synthetic data where ground truth is known. To our knowledge, only two alternatives exist: model-based estimation of shared dimensionality using reduced rank regression (RRR) and statistically-based estimation using canonical correlation analysis (CCA). For the RRR method we followed the procedure outlined in Semedo et al. (2019). Briefly, the RRR model is used to predict variability in one dataset from the variability of the other. Models are constructed with dimensionalities varying from 1 to k and each assessed via cross-validation. The model dimensionality with the best predictive CV score is selected as the shared dimensionality. For the CCA method, we conducted CCA using `canoncorr()` function in Matlab (version 9.13.0. Natick, Massachusetts: The MathWorks Inc., 2022.). Each canonical correlation with a p-value of <0.05 (corrected for false discovery rate) was considered significant. The significant canonical correlations were then summed to give the estimate of dimensionality.

Synthetic data was generated using the probabilistic CCA (pCCA) model Bach and Jordan (2005); Browne (1979). Briefly, the pCCA model is a structured factor analytic model where in observations of vector measurements for two data sources $(\mathbf{y}_A, \mathbf{y}_B)$ are determined by P latent variables whereby

$$\begin{pmatrix} \mathbf{y}_A \\ \mathbf{y}_B \end{pmatrix} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon} \tag{50}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{D})$, with $\mathbf{D}$ diagonal, and $\mathbf{x} \sim \mathcal{N}(0, \mathbb{I}_P)$. The shared subspace is structured through the partitioning of $\mathbf{x}$ such that $\mathbf{x} = \left(\mathbf{x}_A^\top, \mathbf{x}_S^\top, \mathbf{x}_B^\top\right)^\top$ where $\mathbf{x}_A$ is latent variability associated exclusively with measurements $\mathbf{y}_A$, $\mathbf{x}_B$ is latent variability associated exclusively with measurements $\mathbf{y}_B$, and $\mathbf{x}_S$ is latent variability associated



Supplementary Figure 7: Wallclock time of electrophysiology data analysis using ENSD and RRR. Each marker represents a single run of the analysis.

with both measurements $\mathbf{y}_A$ and $\mathbf{y}_B$. The latent dimensionality P is given by the sum of the dimensionality of all components; $P = P_A + P_S + P_B$. The mixing matrix is structured to maintain the partitioning of variability via the following block structure

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_A & \mathbf{A}_{SA} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{SB} & \mathbf{A}_B \end{pmatrix}. \tag{51}$$

In addition to the noise variable $\boldsymbol{\epsilon}$, we tested the robustness of these methods by randomly setting each observation to 0 with 2 different probabilities (0.2 and 0.8). We simulated data using this model with 31 observations of each $\mathbf{y}_A$ and $\mathbf{y}_B$, $P_A = 10,\ P_B = 15,\ P_S = 5$. Sample sizes ranged from 50 to 500 and each sample size was repeated across 50 different experiments. Nonzero elements of $\mathbf{A}$ were drawn iid from a standard normal distribution for each experiment. Means and standard deviations reported in Figure 3 are taken over all experiments.

SM4 Re-analysis of visual cortex electrophysiology data

For the electrophysiology data analysis in Section 5, we conducted estimation of dimensionliaty via RRR according to Semedo et al. (2019) using code retrieved from the authors' github repository (`https://github.com/joao-semedo/communication-subspace`). The data was retrieved from a publicly available CRCNS repository Zandvakili and Kohn (2019). Data collection was described in detail in Zandvakili and Kohn (2015).

Briefly, an anesthetized macaque was presented with a visual stimuli consisting of a series of oriented gratings with one of 8 different orientations. Extra-cellular electrophysiological recordings were conducted simultaneously in V1 (97 single units via Utah array) and V2 (31 single units via tetrodes) during stimulus presentation. Each stimulus orientation was analyzed separately. For each run of an analysis, the V1 sample sizes were matched to the V2 sample size by randomly sampling both a "target" and "source" subsample from V1. V1 dimensionality was taken to be the optimal RRR model dimensionality to predict the target V1 dataset from the source V1 dataset. Shared dimensionality was taken to be the optimal RRR model dimensionality to predict V2 neural responses using the source V1 dataset as regressors. The estimation procedure was repeated 10 times for each dataset and the results were averaged. We calculated ENSD values for each of these 10 resamplings for each stimulus orientation and results for each resampling were averaged to give 1 shared dimensionality value for each stimulus orientaion. Since the ENSD is not limited to matrices that share both dimensions, we were able to complete the analysis using the full V1 dataset. The results from this analysis reveal that size-matching may be causing an underestimation of the dimensionality of the communicaation subspace between the two areas.

Although absolute computation time for each method was fast, the wallclock computation time for ENSD was about two orders of magnitude faster than for RRR (Fig. 7), which has implications for analysis times for very large datasets.

SM5 Analysis of olfactory connectivity data

In section 5 of the main text, we explore shared structure in three olfactory datasets from *D. Melanogaster*. Each is a neural connectivity matrix obtained by electron microscopy Scheffer (2020); Zheng et al. (2022) describing connectivity from the fly Antennal Lobe (AL) to one of two downstream areas, the Lateral Horn (LH) and Mushroom Body (MB). Two such matrices were extracted from one individual via the Hemibrain dataset: $\mathbf{L}^{FlyEM}$ (dimensions 108×1400) and $\mathbf{M}^{FlyEM}$ (dimensions 108×1761) Schlegel et al. (2021). These data describe the number of synapses between individual AL and LH/MB cells. The final matrix was obtained from a second individual via the FAFB dataset, $\mathbf{M}^{FAFB}$ (dimensions 109×1344) and describes binary connectivity between individual AL and LH/MB celxls Zheng et al. (2022). We contract the first dimension (uniglomerular projection neuron (uPN)), by adding together uPNs associated to the same olfactory channel (originating in the same glomerulus). This results in a consistent first dimension of size $n_{channels} = 51$ across all three datasets. A threshold is then applied to binarize the FlyEM datasets, $\theta \in [1, 3, 5, 10]$. θ is set to 3 in the main text. Here, we also apply the ENSD to the synaptic count data.

As we are interested in investigating shared structure between these datasets, we compare each dataset to a null model in which correlations contained in the empirical data are removed via shuffling. Our shuffling procedure for binary connectivity matrices follows Caron et al. (2013); Zheng et al. (2022) and results in an ensemble of null model matrices, each of which has the same the marginal statistics (row and column sums) as the empirical data. We use an ensemble of size $x = 300$ for the $\theta = 3$ analyses and $x = 100$ for other θ values. To shuffle the synaptic data, we first apply a threshold to the data and then shuffle via the above procedure. Finally, synaptic counts are added back in randomly such that the row sums are held constant, i.e. synaptic counts from each channel/glomerulus are the same as the empirical data, but column sums are allowed to diverge. Rows in each matrix, empirical or null model, are mean centered.

It is important to note that in each analysis, only one matrix is shuffled (indicated by a grey bar in the figure header). We choose to shuffle the less structured matrix of the pair, i.e. $\mathbf{M}^{FlyEM}$ or $\mathbf{M}^{FAFB}$, rather than both matrices. Shuffling the matrix with more structure, i.e. $\mathbf{L}^{FlyEM}$, would result in a random matrix which may be (and in our case is) more similar to the less structured matrix. This similarity is, from our perspective, artificial, as we are interested in understanding how the structured components in two datasets relate to one another, not the random components.

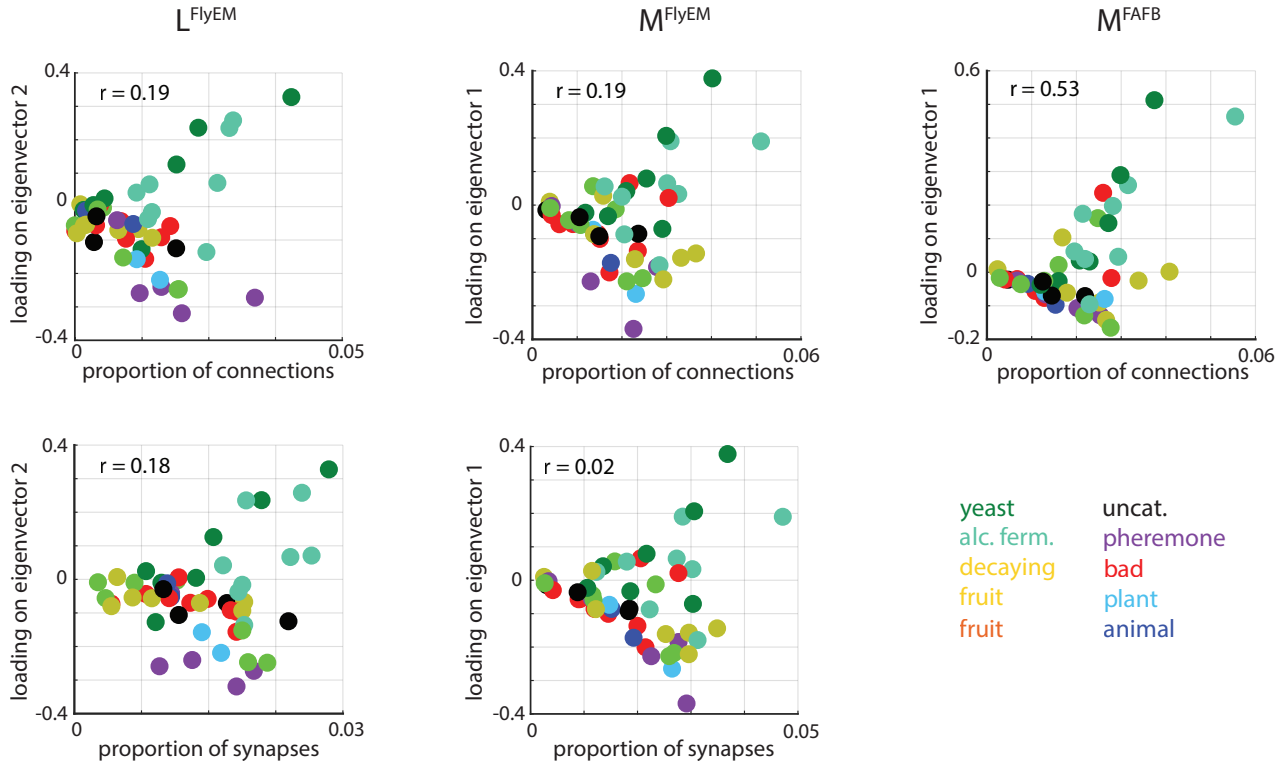
In Supplementary Figures 9-11, we present an extended analysis of the data from the main text. This includes pairwise analyses of the three connectivity matrices in SM-Fig.9A, with synaptic count data for the $\mathbf{L}^{FlyEM}$ and $\mathbf{M}^{FlyEM}$ replacing the binary data from the main text. In SM-Fig.9B, we show that the overlap between shared dimensions (eigenvector 1 in $\mathbf{L}^{FlyEM}$ and eigenvector 2 in $\mathbf{M}^{FlyEM}$ and $\mathbf{M}^{FAFB}$) is significantly increased in empirical data vs the null model. In SM-Fig.10, we confirm that the shared subspace is not an artefact of threshold choice, but a robust feature of the data by analyzing the changes in ν and d_ν for all synaptic and binary thresholding configurations. In In SM-Fig.8, we confirm that the shared structure (aligned $1D$ subspace) cannot be explained by the marginal statistics of the data alone as the loadings and proportions of synapses or connections for each information channel are relatively uncorrelated.

Because our shuffling procedure significantly affects the dimensionality of the data (shuffled matrices are higher dimensional), ν, which is sensitive to this dimensionality via the $\gamma_X \gamma_Y$ component, may be less indicative of shared structure than d_ν. For instance in the comparison between MB inputs across individuals (final column of SM-Fig.9), ν increases for the null model relative to the data. In this case and in others, the change in d_ν more clearly indicates shared structure that has been disrupted by shuffling. We expect that in studies where the procedure for generating samples from the null model doesn't significantly change dimensionality, ν would be equally useful for detecting shared structure, as well as giving an interpretable value for the size of the shared variability component.

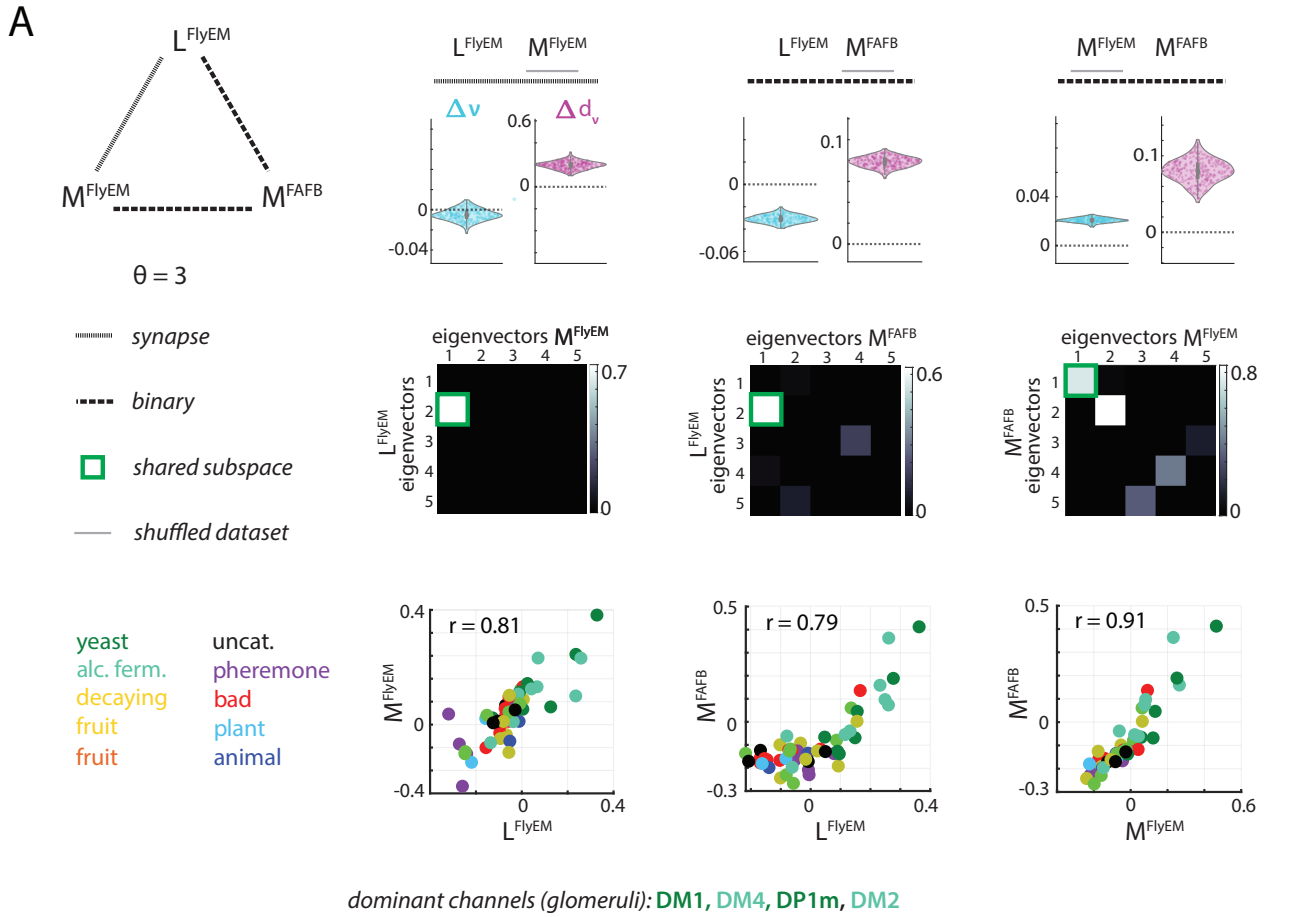
Finally, we show the connectivity data reordered by loading on the shared dimension in SM-Fig.11. The LH consists of two major cell types: output neurons (LHONs) and local neurons (LHLNs), as well as a few minor cell types (other). By reorganizing the components of $\mathbf{L}^{FlyEM}$ corresponding to these cell types, we can see that both the LHON and LHLN populations contain subpopulations of cells that are strongly tuned to the food related features of the shared dimension, while the other cell types do not contain any strongly tuned neurons. We also note that, while the output neurons contain subpopulations that are both positively and negatively tuned to the shared dimension, the local neuron population only contains a subpopulation that is positively tuned to these

features. This is an interesting circuit feature that we intend to study in future work. The MB contains a single primary cell type, the Kenyon Cell (KC), and is physically organized into a series of lobes. Reordering $\mathbf{M}^{FlyEM}$ by loading onto the shared dimension reveals that all three lobes, $\alpha/\beta, \alpha'/\beta'$ and γ, contain subpopulations of neurons that are both positively and negatively tuned.

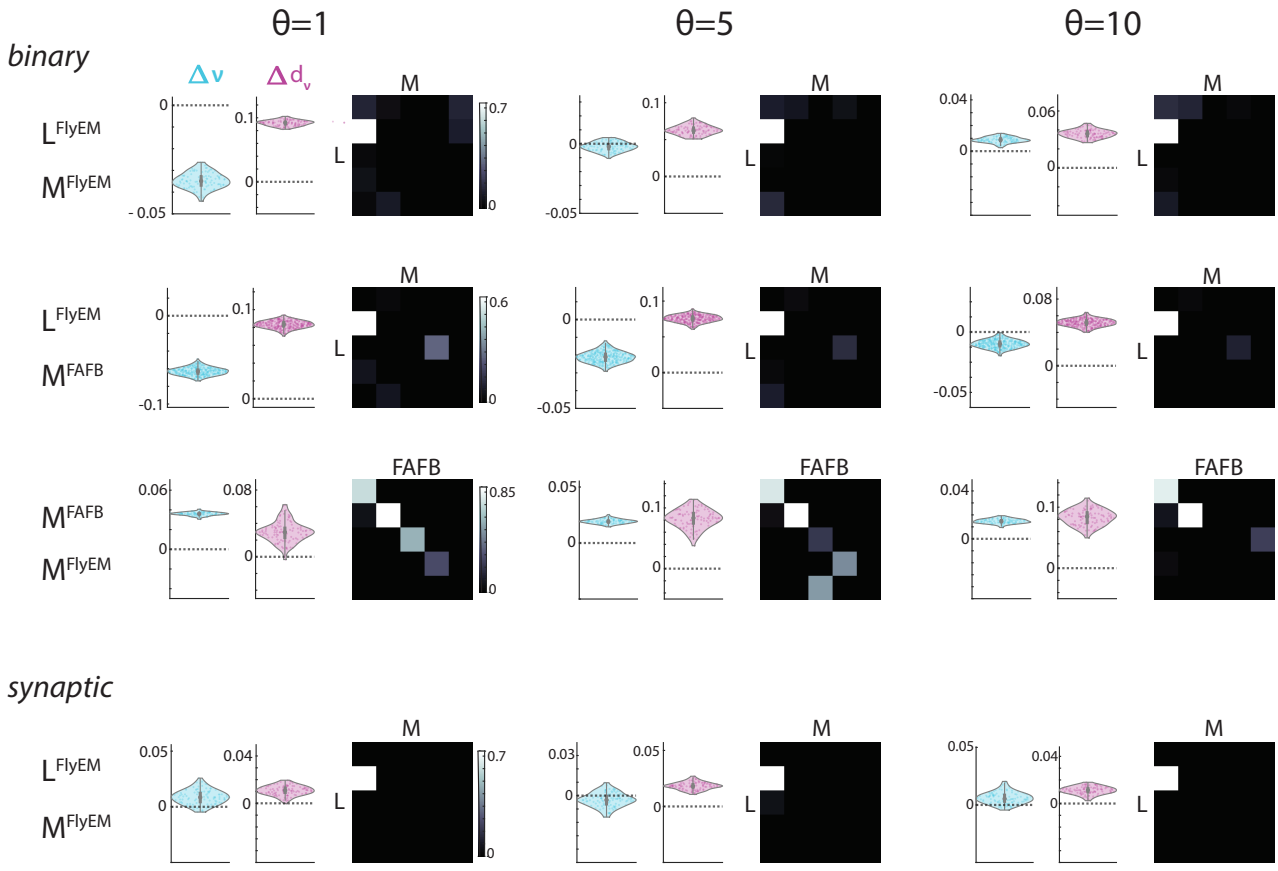
Additionally, we use the ENSD to probe variability shared between a subset of connectivity data from the Hemibrain dataset (*FlyEM*) and neural activity in a subset of channels obtained from Badel et al. (2016). The activity data describes the responses of 37/51 channels to a set of 84 odorous stimuli, including fermentation and other food related odorants/blends. The raw data is preprocessed to a matrix of Z-scored responses $\mathbf{R}$ shown in SM-Fig.12A. We find that input connectivity to the LH is significantly tuned to features of the neural responses (including but not limited to food related features. SM-Fig.12B, first column), relative to shuffled LH connectivity, whereas the MB input (second column) is more weakly tuned to the neural activity (comparison of Δd_ν plots in columns 1 and 2). We confirm that the subspace shared between LH and MB inputs is also present in the subset of connectivity data used for this analysis (last column). This difference in tuning between LH and MB inputs is consistent with the putative roles of these two neuropils in innate vs learned olfactory behaviour. This analysis demonstrates that the ENSD is a simple and useful tool for uncovering consistent relationships between datasets of different modalities.



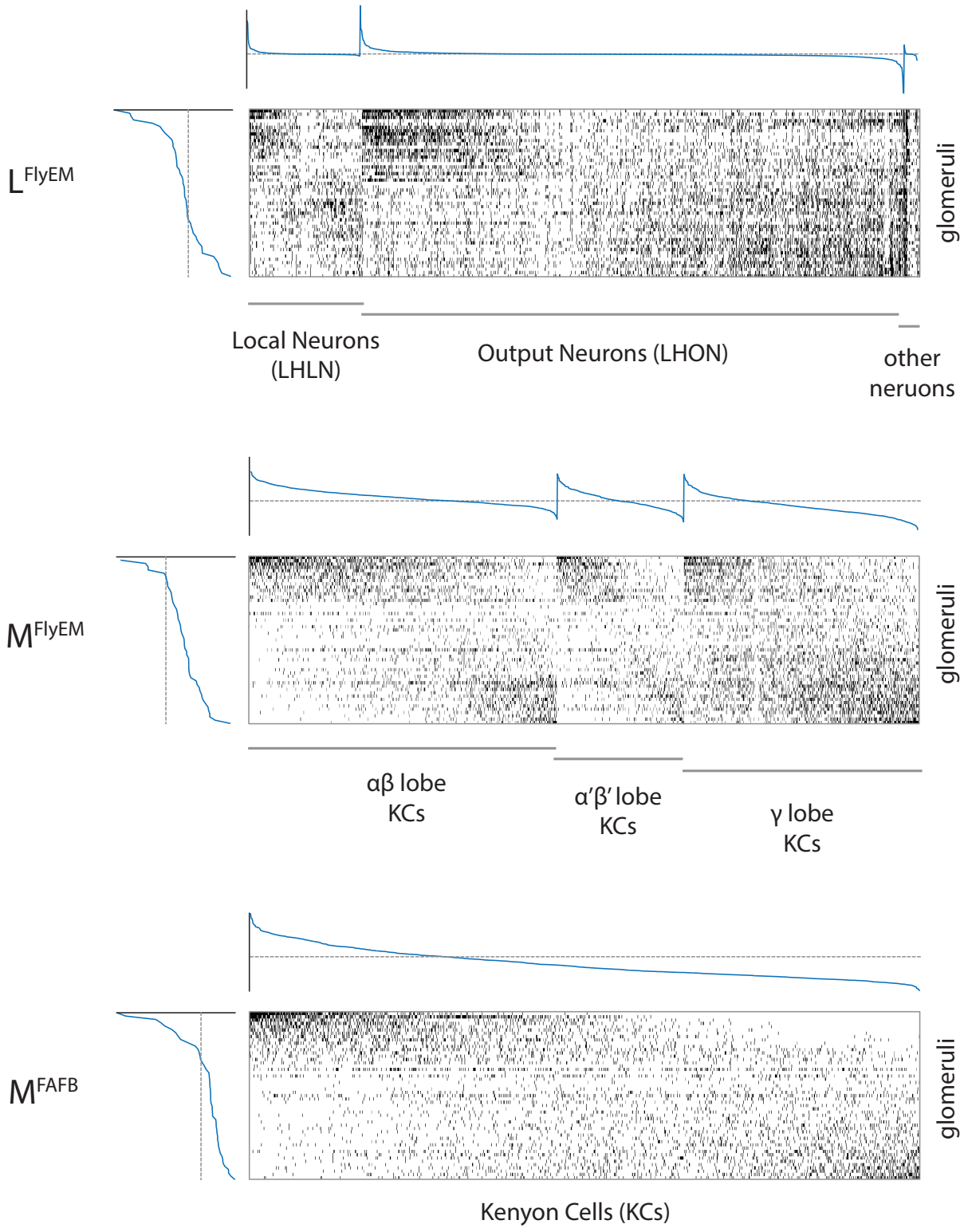
Supplementary Figure 8: Loadings on shared dimension from each dataset are compared against proportion of connections/synapses for each olfactory channel (glomerulus).



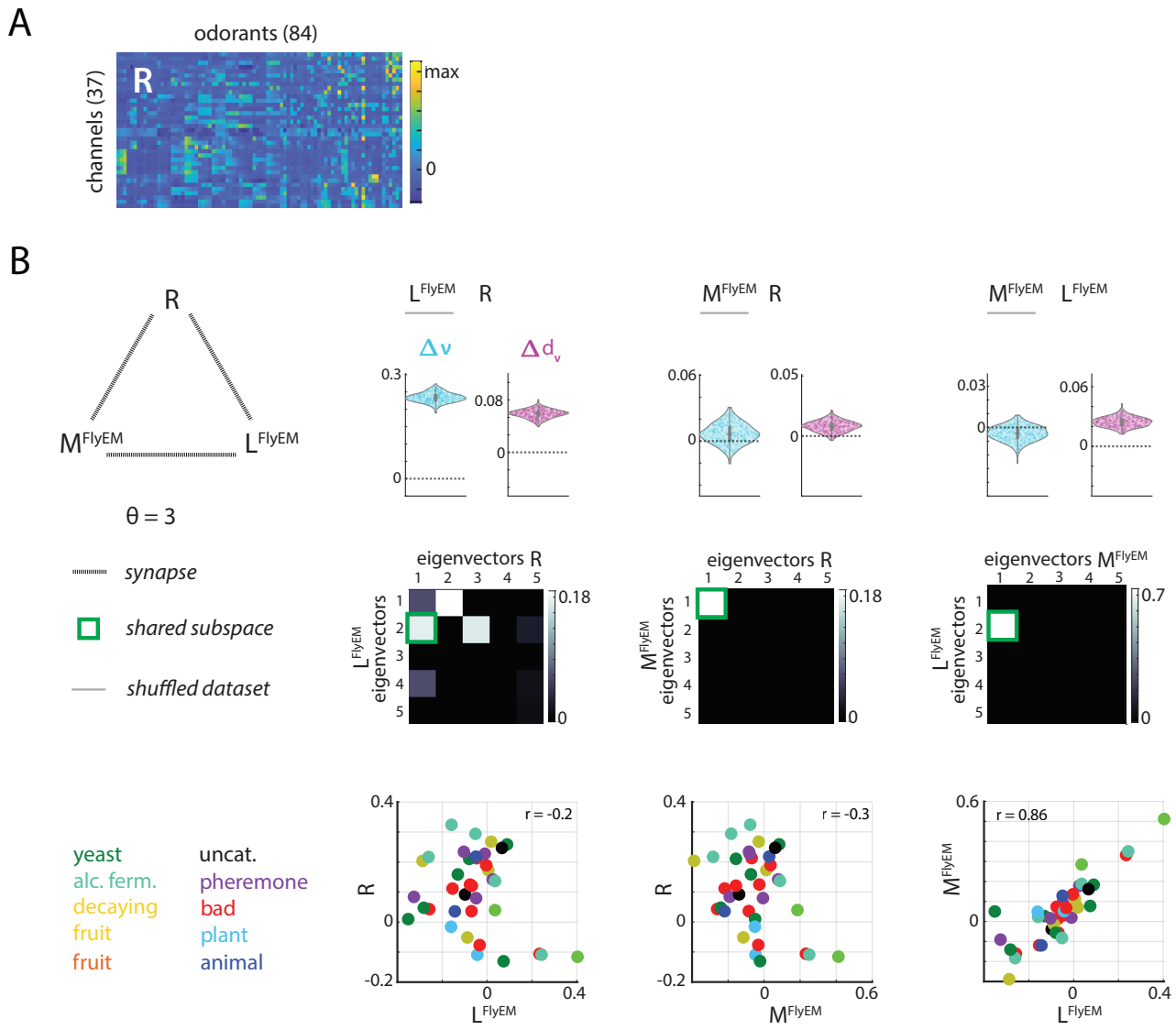
Supplementary Figure 9: **A.** Pairwise analysis of shared structure between connectivity matrices from FlyEM and FAFB datasets. *Top row*: changes in ENSD and distance metric when one dataset is shuffled (underlined in grey). *Second row*: **W** matrices containing eigenvector overlaps corresponding to the top 5 eigenvalues shown, all statistically insignificant overlaps ($p > 0.05$), are set to zero. *Bottom row*: Comparison of loadings onto shared dimension. **B.** The shared structure identified in **W** is not preserved when one dataset is shuffled.



Supplementary Figure 10: Shared subspace analysis from the main text is consistent across a range of synaptic threshold values.



Supplementary Figure 11: Binary connectivity matrices ordered in both dimensions by loading onto shared eigenvector (eigenvectors associated with the 2nd highest eigenvalue for $\mathbf{L}$ and the with the highest eigenvalue for both $\mathbf{M}$ matrices).



Supplementary Figure 12: **A.** The response of a subset of 37 olfactory channels to 84 odorants from four categories: monomolecular species, natural odorant blends, concentration series and binary mixtures (matrix **R**) Badel et al. (2016). **B.** Multimodal data analysis with the ENSD. Comparing synaptic count data from *FlyEM* (using the corresponding 37 channels) with odorant response data in **R**. *Top row*: changes in ENSD and distance metric when one dataset is shuffled (underlined in grey). *Second row*: **W** matrices containing eigenvector overlaps corresponding to the top 5 eigenvalues shown, all statistically insignificant overlaps ($p > 0.05$), are set to zero. *Bottom row*: Comparison of loadings onto shared dimension.

Code availability

Code to reproduce the main results of this work is available at https://github.com/hamzagiaffar/ensd.git.

References

[1] F. R. Bach and M. I. Jordan. A probabilistic interpretation of canonical correlation analysis. 2005.

[2] L. Badel, K. Ohta, Y. Tsuchimoto, and H. Kazama. Decoding of context-dependent olfactory behavior in drosophila. *Neuron*, 91(1):155–67, 2016.

[3] M. W. Browne. The maximum-likelihood solution in inter-battery factor analysis. *British Journal of Mathematical and Statistical Psychology*, 32(1):75–86, 1979.

[4] S. J. C. Caron, V. Ruta, L. F. Abbott, and R. Axel. Random convergence of olfactory inputs in the drosophila mushroom body. *Nature*, 497(7447):113–117, 2013.

[5] L. K. e. a. Scheffer. A connectome and analysis of the adult *Drosophila* central brain. *eLife*, 9, 2020.

[6] P. Schlegel, A. S. Bates, T. Stürner, S. R. Jagannathan, N. Drummond, J. Hsu, L. Serratosa Capdevila, A. Javier, E. C. Marin, A. Barth-Maron, I. F. M. Tamimi, F. Li, G. M. Rubin, S. M. Plaza, M. Costa, and G. S. X. E. Jefferis. Information flow, cell types and stereotypy in a full olfactory connectome. *eLife*, 10:e66018, 2021.

[7] J. D. Semedo, A. Zandvakili, C. K. Machens, M. Y. Byron, and A. Kohn. Cortical areas interact through a communication subspace. *Neuron*, 102(1):249–259, 2019.

[8] A. H. Williams, E. Kunz, S. Kornblith, and S. Linderman. Generalized shape metrics on neural representations. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 4738–4750. Curran Associates, Inc., 2021.

[9] A. Zandvakili and A. Kohn. Coordinated neuronal activity enhances corticocortical communication. *Neuron*, 87(4):827–839, 2015.

[10] A. Zandvakili and A. Kohn. Paired v1-v2 neuronal spiking responses in anesthetized macaque monkey. *CRCNS.org*, 2019.

[11] Z. Zheng, F. Li, C. Fisher, I. J. Ali, N. Sharifi, S. Calle-Schuler, J. Hsu, N. Masoodpanah, L. Kmecova, T. Kazimiers, E. Perlman, M. Nichols, P. H. Li, V. Jain, and D. D. Bock. Structured sampling of olfactory input by the fly mushroom body. *Current Biology*, 32(15):3334–3349.e6, 2022.