Testing Structured Dependence in Decodable Brain-State Representations: Higher-Order Information and Temporal Topology in fMRI

Valeriya Kirova [1] **Aleksandr Levin** [2] **Ivan Butakov** [3 4 5] **Alexander Semenenko** [3] **Alexey Frolov** [3]

Abstract

Decoding models can identify brain regions predictive of cognitive states, but they do not directly test how the information carried by these regions is organized. We formulate post-decoding analysis of task-fMRI representations as a set of structured hypothesis-testing problems. Building on an interpretable decoding pipeline, we test whether brain regions identified as important for classifying cognitive states differ from less classifier-relevant regions, and characterize how their information is organized using tools from information theory and topology.

Keywords fMRI, cognitive state decoding, mutual information, O-information, higher-order interactions, information theory, flow matching, temporal dynamics, functional connectivity, persistent homology, topological data analysis, Human Connectome Project.

1. Introduction

Functional magnetic resonance imaging (fMRI) provides a non-invasive way to study task-evoked brain activity through blood-oxygen-level-dependent (BOLD) signals (Ogawa et al., 1990; K., 2008). Task-fMRI data are naturally high-dimensional and spatiotemporal: each recording contains activity patterns distributed across many brain regions and evolving over time (Barch et al., 2013; Pereira et al., 2009). This makes fMRI a challenging but informative domain for machine learning methods aimed at decoding cognitive states and interpreting the neural structure underlying such predictions.

[1]National Research Nuclear University MEPhI, Moscow, Russia [2]Faculty of Mathematics, HSE University, Moscow, Russia [3]Applied AI Institute, Moscow, Russia [4]Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia [5]Moscow Independent Research Institute of Artificial Intelligence, Moscow, Russia. Correspondence to: Valeriya Kirova <lera.orlanovna@gmail.com>.

ICML 2026 Workshop on Hypothesis Testing, Seoul, South Korea, 2026. Copyright 2026 by the author(s).

Decoding models are widely used in neuroimaging to identify brain regions and activity patterns that are predictive of cognitive states (Haynes & Rees, 2006; Norman et al., 2006; Tong & Pratte, 2012). Such models can be used to derive candidate brain regions or activity patterns associated with a cognitive state, either through explicit model weights, feature-importance scores, attribution methods, or other interpretability tools, although the interpretation of such quantities requires care (Haufe et al., 2014; Kriegeskorte & Douglas, 2019; Vigotsky et al., 2024). However, identifying predictive regions is only the first step. A decoder can indicate whether information about a state is present in a form that the model can exploit, but it does not by itself explain how that information is organized across regions. In particular, it remains unclear whether state-specific regions contribute independently, redundantly, synergistically, or through coordinated temporal dynamics (Timme & Lapish, 2018). This motivates a complementary question that is naturally statistical rather than purely predictive: once a decoder has selected a set of brain regions, can we test whether these regions exhibit distinctive dependence or temporal structure relative to appropriate controls? In this view, interpretability is not only a matter of identifying predictive variables, but also of formulating and evaluating hypotheses about the structure that makes those variables informative.

In this work, we treat post-decoding analysis of task-fMRI representations as a structured hypothesis-testing problem. Given state-specific sets of informative regions produced by an interpretable decoding pipeline, we ask three questions. First, do classifier-selected regions exhibit stronger pairwise temporal dependence than matched low-weight control regions? Second, do compact and distributed cognitive-state representations differ in their higher-order multivariate dependence structure? Third, is the temporal topology of regional time series preserved under structure-preserving surrogates and altered under surrogates that destroy temporal ordering?

To address these questions, we combine information-theoretic and topological statistics with explicit control and surrogate comparisons. We estimate O-information (Rosas et al., 2019) on vectors of mean regional activations using an amortized flow-matching estimator (Butakov

et al., 2025) to assess whether selected region sets are redundancy- or synergy-dominated. For compact high-discriminability states, we compare the empirical distributions of pairwise mutual information (MI) between selected regions and matched low-weight regions using a two-sample Kolmogorov–Smirnov (KS) test. Finally, we apply persistent homology to delay-coordinate embeddings of regional time series and compare the resulting topological summaries against cyclic-shift and shuffle surrogate baselines.

Together, these analyses extend activation-based decoding into a hypothesis-testing framework for representation structure. Rather than only reporting which regions are predictive, the proposed framework evaluates whether the selected regions exhibit statistically distinguishable pairwise, higher-order, and temporal-geometric organization.

2. Related work

Prior work has studied fMRI representations through functional connectivity, graph-based models, and deep learning. In network approaches, brain regions are treated as nodes and statistical dependencies between regional signals as edges (Wang et al., 2010; Richiardi et al., 2011; Takerkart et al., 2014; Li et al., 2021; Saeidi et al., 2022; Bessadok et al., 2023). These methods characterize functional network organization, but often require substantial preprocessing, connectivity estimation, and additional modeling assumptions. Information-theoretic methods provide a complementary way to study neural representations. Entropy, MI, transfer entropy, and related conditional measures have been used to analyze population codes, decoding, and directed dependence in neural or complex time-series data (Quian Quiroga & Panzeri, 2009; Timme & Lapish, 2018; Schreiber, 2000; Runge et al., 2019). Closer to the present study, Hlinka et al. (Hlinka et al., 2011) compared MI with linear correlation for resting-state fMRI connectivity, Ince et al. (Ince et al., 2017) developed a Gaussian-copula framework for MI-based neuroimaging analysis, Gatica et al. (Gatica et al., 2021) used O-information to study redundancy- and synergy-dominated interdependencies in aging, and Varley et al. (Varley et al., 2023) identified synergistic cortical subsystems not visible from pairwise functional connectivity.

Topological methods have also been used to describe fMRI data. Persistent homology and graph filtrations have been applied to task-based and resting-state functional connectivity networks to obtain threshold-free multiscale descriptions of network organization (Kim et al., 2015; Wong et al., 2016; Anderson et al., 2018; Stolz et al., 2021; Ryu et al., 2023; Das et al., 2023), while related approaches have analyzed dynamic connectivity, robustness to acquisition variability, and disease- or aging-related changes (Talesh Jafadideh & Mohammadzadeh Asl, 2022; Xing et al., 2022; Kumar et al., 2023; Mondal et al., 2025). Saggar et al. (Saggar et al., 2018) used TDA to track task transitions in whole-brain activity maps, Ellis et al. (Ellis et al., 2019) studied persistent homology for event-related fMRI, and Cassidy et al. (Cassidy et al., 2015) showed that topological summaries can distinguish brain activity states. More recent dynamic TDA methods encode fMRI sequences as time-varying complexes, barcodes, or persistence diagrams for temporal tracking, clustering, and trajectory analysis (Songdechakraiwut & Chung, 2020; Rieck et al., 2020). Our analysis is complementary to these approaches.

Table 1. Two brain states for each cognitive task, between which classification is made.

Cognitive task	State 1	State 2
Working Memory	0-back	2-back
Gambling	Win	Loss
Motor Task	Left hand/foot	Right hand/foot
Language Processing	Story	Math
Social Cognition	Random motion	Mental interaction
Relational Processing	Relation	Similarity
Emotion Processing	Neutral	Fear

3. Data and problem setup

Kirova et al. (2026) proposed an interpretable classification pipeline for task-based fMRI data from the Human Connectome Project Elam et al. (2021). In their approach, each fMRI data unit was represented by a vector of mean regional activations, and linear classifiers were used to identify brain regions whose activity was most strongly associated with individual cognitive states. This analysis showed that several brain states can be decoded with high accuracy from compact and largely state-specific sets of regions, whereas other states require substantially larger and more distributed sets of contributing regions.

The fMRI data were obtained from the Human Connectome Project (HCP) 1200 Subjects Release (Human Connectome Project, 2017). The sample included 581 healthy, right-handed participants. Participants completed a series of task-based fMRI paradigms (Barch et al., 2013) designed to assess seven major cognitive domains, sampling the diversity of large-scale human brain networks: visual, motion, somatosensory, and motor systems; category-specific representations; working memory and cognitive control systems; language processing, including semantic and phonological processing; social cognition, including Theory of Mind; relational processing; and emotion processing. A detailed description of all tasks is provided in Appendix A. Each participant performed seven tasks with two conditions, resulting in 14 distinct brain states, summarized in Table 1, and 8134 fMRI data units in total.

Thus, each fMRI data unit is represented as an activity matrix $\mathbf{X}$ of size $r \times l$, where each line $\vec{x}_i = (x_{i1}, \ldots, x_{il})$ corresponds to the activity of the i-th brain region over a time interval of length l. Each element x_{ij} of the activity matrix $\mathbf{X}$ corresponds to the activity level of the i-th brain region at time point j. Brain parcellation was performed using the HCP Multi-Modal Parcellation (MMP) atlas, which defines 379 distinct brain regions.

Kirova et al. (2026) obtained, for each brain state class c, an informative set of significant brain regions

$$\mathcal{M}_c = \{\, i : i \text{ is the index of a significant brain region} \,\}.$$

These sets were identified by iteratively removing low-importance regions while preserving the classification performance of the corresponding one-vs-all classifier.

The authors represented each activity matrix $\mathbf{X}$ by the averaged activity values across brain regions: for each brain region $i = \{1, ..., r\}$ they calculated the mean value $\overline{\mathbf{x}}_i$ of the time series $\vec{x}_i$. By focusing on mean regional activity values, they aimed to demonstrate that each brain state is associated with a unique network of regional activations that can serve as a reliable signature for classification.

Their analysis showed that brain states with classification accuracy above 0.9 can be decoded from compact and largely state-specific sets $\mathcal{M}_c$. For these high-accuracy brain states, the number of selected regions $2 \leq |\mathcal{M}_c| \leq 13$, and the corresponding sets almost do not overlap. This indicates that each such brain state is associated with a small and unique set of brain regions whose mean activity provides a reliable classification signature. At the same time, authors also observed states for which the sets $\mathcal{M}_c$ are substantially larger and more similar in composition. In some cases, these sets approach broad coverage of the HCP MMP atlas. Such behavior indicates reduced specificity of feature selection: when many regions are retained, the distinctive contribution of individual regions becomes less discernible, making the functional interpretation of state-specific regions more difficult. A detailed breakdown of the number of identified regions is provided in Appendix A. We reproduced their region-selection procedure on the same preprocessed HCP dataset and obtained the same compact and distributed representation regimes reported in the original work. We therefore use these reproduced region sets as the starting point for the subsequent information-theoretic and topological analyses.

This naturally separates the brain states into two regimes. The first regime, which we refer to as the *compact high-discriminability regime*, contains states represented by relatively small sets of significant regions $|\mathcal{M}_c|$. This group consists of six high-accuracy states: *Math*, *Story*, *Random Motion*, *Mental Interaction*, *Left Hand or Foot*, and *Right Hand or Foot*. The second regime, which we refer to as the *distributed low-specificity regime*, contains states that require substantially larger sets of regions and therefore correspond to more distributed neural representations. This group includes *Similarity*, *0-back*, *2-back*, *Relation*, *Neutral*, and *Fear*.

For the subsequent analysis, we additionally define, for each brain-state class c, a complementary set $\mathcal{L}_c$ of low-importance regions such that $|\mathcal{L}_c| = |\mathcal{M}_c|$. To construct these sets, we use the ranked lists of brain regions, where regions are ordered according to their contribution to the corresponding one-vs-all classifier. While $\mathcal{M}_c$ consists of the most informative regions, the set $\mathcal{L}_c$ is formed by taking the same number of regions from the opposite end of the ranking, i.e., those with the smallest classifier weights.

Building on this separation, we perform a follow-up multiscale analysis of the identified brain-state representations. These comparisons define three post-decoding hypotheses. For pairwise temporal dependence, the null hypothesis is that the distribution of MI values within $\mathcal{M}_c$ is the same as within the matched control set $\mathcal{L}_c$. For higher-order dependence, the null hypothesis is that compact and distributed representation regimes do not differ systematically in their O-information structure. For temporal topology, the surrogate nulls test whether the observed H_1 summaries can be explained by marginal signal values alone or by cyclic re-indexing of the same temporal trajectory. These hypotheses allow us to evaluate whether decoder-selected regions are merely predictive, or whether they also carry distinctive statistical and geometric structure.

4. Methods

4.1. Mutual information

Mutual information (MI) provides a general information-theoretic measure of statistical dependence between random variables, quantifying how much information one variable contains about another (Cover & Thomas, 2005; Polyanskiy & Wu, 2025). Unlike the correlation coefficient, which captures only linear relationships, MI can reveal more complex dependencies. In this study, we leverage MI to demonstrate that the selected significant brain regions share meaningful interdependencies and perform a quantitative analysis of these relationships.

Definition 4.1 (Differential entropy)**.** Let X be a continuous random vector with probability density function $\pi(x)$. The differential entropy of X is defined as

$$\mathsf{h}(X) = -\,\mathbb{E}_X \log \pi(X).$$

Definition 4.2 (Mutual information)**.** Let X and Y be random vectors with joint distribution $\Pi_{X,Y}$ and marginal product distribution $\Pi_X \otimes \Pi_Y$. The MI between X and Y is defined as the Kullback–Leibler divergence between the joint

distribution and the product of marginals:

$$\mathsf{I}(X;Y) = D_{\mathrm{KL}}\left(\Pi_{X,Y} \,\|\, \Pi_X \otimes \Pi_Y\right).$$

Provided that the corresponding probability density functions exist, this can be written as

$$\mathsf{I}(X;Y) = \mathbb{E}\left[\log \frac{\pi(X,Y)}{\pi(X)\pi(Y)}\right] = \mathsf{h}(X)+\mathsf{h}(Y)-\mathsf{h}(X,Y).$$

MI equals zero if and only if the random vectors are independent. It is also *invariant* to bijective mappings: if g is one-to-one and measurable, then $\mathsf{I}(X;Y) = \mathsf{I}(X;g(Y))$.

In our problem, the random vectors X and Y correspond to the time series $\overline{x}_i$ and $\overline{x}_j$ of activity from two different brain regions $i, j \in \mathcal{M}_c$.

Thus, the MI $\mathsf{I}(X;Y)$ quantitatively reflects the degree of statistical dependence and the shared dynamics between the two brain regions.

In our problem, MI is used to quantify statistical dependence between activity patterns of pairs of brain regions. Thus, for two regions i and j, MI reflects the degree of shared information between their activity profiles and serves as a model-independent measure of interregional dependence.

We focus on the first regime, which we refer to as the *compact high-discriminability regime*. This regime includes the brain states that are decoded with high accuracy from compact sets of significant regions. For each brain state class c in this regime, we consider two region sets: the set of significant regions $\mathcal{M}_c$ identified by the decoding pipeline, and a control set $\mathcal{L}_c$ of low-weight regions with the same cardinality. For every pair of distinct regions within each set, we compute the mutual information $\mathsf{I}(\vec{x}_i;\vec{x}_j)$, where $\vec{x}_i$ and $\vec{x}_j$ denote the regional activity time series of regions i and j, respectively. Thus, for each class c, we compute $\mathsf{I}(\vec{x}_i;\vec{x}_j)$ for all distinct pairs $i, j \in \mathcal{M}_c$ and, separately, for all distinct pairs $i, j \in \mathcal{L}_c$.

This procedure produces two empirical distributions of pairwise mutual information values: $\mathcal{M}_c$ and one $\mathcal{L}_c$. Comparing these distributions allows us to test whether classifier-selected regions share stronger statistical dependence than regions that contribute minimally to the decoding task. To better understand the differences between these distributions, we apply the two-sample KS test.

4.2. O-information

Although MI serves as a fundamental measure for quantifying nonlinear dependence between two random variables, its applicability becomes limited when analyzing systems comprising more than two variables, which is a common scenario, for example, in neuroscience (Ganmor et al., 2011) and climatology (Runge et al., 2019). A natural multivariate extension of MI is interaction information, also known as co-information, which is defined via an inclusion–exclusion formula. While co-information can reflect the interplay between synergy and redundancy, its scalability and interpretability are limited for large systems.

A more computable and interpretable generalization is provided by O-information (Rosas et al., 2019; Bounoua et al., 2024).

Definition 4.3 (O-information)**.** Let $X = (X_1, \ldots, X_d)$ be a d-dimensional random vector. The O-information of X is defined as

$$\mathrm{OI}(X) = \mathrm{Redundancy}(X) - \mathrm{Synergy}(X),$$

where

$$\mathrm{Redundancy}(X) = \sum_{i=1}^{d} \mathsf{h}(X_i) - \mathsf{h}(X),$$

$$\mathrm{Synergy}(X) = \mathsf{h}(X) - \sum_{i=1}^{d} \mathsf{h}(X_i \mid X_{\setminus i}).$$

Here $X_{\setminus i}$ denotes the $(d-1)$-dimensional vector obtained by removing the i-th component from X, and $\mathsf{h}(\cdot)$ denotes differential entropy.

Intuitively, $\mathrm{OI}(X_c)$ measures both the overall strength and the dominant type of interactions in a multivariate system. A positive value of $\mathrm{OI}(X)$ indicates redundancy dominance, meaning that similar information is duplicated across different variables. Conversely, a negative value of $\mathrm{OI}(X)$ corresponds to synergy dominance, reflecting information that emerges only from the joint consideration of all variables and cannot be captured by lower-order or pairwise interactions alone.

Although less straightforward, $\mathrm{OI}(X_c)$ also admits a difference-of-entropies representation, which enables its estimation within our framework. To derive it, consider a $d \times d$ matrix $\mathbf{M}$ whose rows are i.i.d. samples from p_X. By construction,

$$\mathsf{h}(\operatorname{diag}\mathbf{M}) = \sum_{i=1}^{d} \mathsf{h}(X_i),$$

$$\mathsf{h}(\operatorname{diag}\mathbf{M} \mid \operatorname{offdiag}\mathbf{M}) = \sum_{i=1}^{d} \mathsf{h}(X_i \mid X_{\setminus i}).$$

Therefore,

$$\mathrm{Redundancy}(X) = \mathsf{h}(\operatorname{diag}\mathbf{M}) - \mathsf{h}(X),$$
$$\mathrm{Synergy}(X) = \mathsf{h}(X) - \mathsf{h}(\operatorname{diag}\mathbf{M} \mid \operatorname{offdiag}\mathbf{M}).$$

Following Butakov et al. (2025), we use an economical and amortized approach to approximate both terms.

Instead of learning two separate estimators, we employ a single flow matching-based estimator with two-dimensional time, one dimension for each term correspondingly. The standard flow-matching regression objective and affine interpolation paths are extended naturally to this multidimensional-time setting. This technique is also applicable to other information-theoretic quantities, such as transfer entropy (Galeano Munoz et al., 2025).

In this part of the analysis, we reproduce the $\mathrm{OI}(X_c)$ estimation protocol proposed by Butakov et al. (2025) and apply it to the fMRI representations considered in this study. We estimate $\mathrm{OI}(X_c)$ for the selected region set $\mathcal{M}_c$ of each brain-state class c from both regimes: the *compact high-discriminability regime* and the *distributed low-specificity regime*. Our goal is to compare whether compact states with a small number of significant regions and distributed states with substantially larger selected region sets differ in their higher-order dependence structure.

For every region $i \in \mathcal{M}_c$, we use the mean regional activation $\overline{\mathbf{x}}_i$ and collect its values across all 581 fMRI data units. This gives one empirical vector

$$X_i = \left(\overline{\mathbf{x}}_i^{(1)}, \ldots, \overline{\mathbf{x}}_i^{(581)}\right).$$

Thus, we obtain $|\mathcal{M}_c|$ such vectors, one for each selected region. Taken together, these vectors define an empirical joint distribution of mean activations over the selected regions,

$$X_c = \left(X_i, \ldots, X_{|\mathcal{M}_c|}\right).$$

We then estimate $\mathrm{OI}(X_c)$ to characterize higher-order dependence among the selected regions of brain state class c. Since $\mathrm{OI}(X_c)$ distinguishes redundancy-dominated from synergy-dominated multivariate structure, this analysis allows us to test whether compact and distributed cognitive states differ in the way information is shared among their selected regions.

Following the experimental setup of Butakov et al. (2025), for each setup from the fMRI dataset we perform 10 independent runs with different random seeds. In each run, the dataset of 581 samples is split into training and test sets with a 70%/30% ratio. Due to the limited dataset size, training employs full-batch updates and therefore consists of one gradient step per epoch for 10^4 epochs. For each run, we compute the median $\mathrm{OI}(X_c)$ over the last 100 training epochs on the test set. The seed-wise median values are then compared across brain states.

4.3. Topological analysis of temporal structure

To complement the information-theoretic analyses, we use topological data analysis to describe the geometry of individual temporal trajectories. Each regional time series $\vec{x}_i = (x_{i1}, \ldots, x_{il})$ is represented as a delay-coordinate point cloud, whose shape is summarized by persistent homology.

Definition 4.4 (Delay-coordinate embedding)**.** Let $d \in \mathbb{N}$ and $\tau \in \mathbb{N}$ satisfy $l - (d-1)\tau \geq 1$. The delay-coordinate embedding of $\vec{x}_i$ is the point cloud $Y_i(d,\tau) \subset \mathbb{R}^d$ formed by the vectors

$$\vec{y}_t^{\,i} = (x_{i,t+k\tau})_{k=0}^{d-1} \in \mathbb{R}^d, \qquad t = 1, \ldots, l-(d-1)\tau.$$

Each point $\vec{y}_t^{\,i}$ represents a short temporal fragment. Takens' theorem motivates this construction for smooth compact dynamical systems under generic observations (Takens, 1981), but it does not guarantee exact finite-sample reconstruction for noisy fMRI recordings. We use the resulting point cloud descriptively, as its shape reflects recurrent temporal patterns (Myers et al., 2019).

We then describe this point cloud quantitatively.

Definition 4.5 (Vietoris–Rips filtration)**.** Let $Y \subset \mathbb{R}^d$ be a finite point cloud. For $\varepsilon \geq 0$, the Vietoris–Rips complex $K_\varepsilon(Y)$ is the abstract simplicial complex whose simplices are the finite subsets $\sigma \subseteq Y$ such that $\|u - v\|_2 \leq \varepsilon$ for all $u, v \in \sigma$. The family $\{K_\varepsilon(Y)\}_{\varepsilon \geq 0}$ is a filtration, since $K_{\varepsilon_1}(Y) \subseteq K_{\varepsilon_2}(Y)$ whenever $\varepsilon_1 \leq \varepsilon_2$.

Thus nearby embedded points define edges, cliques are filled by simplices, and persistent homology tracks how topological features appear and disappear along the filtration (Zomorodian, 2010).

Definition 4.6 (Persistence diagram and summaries)**.** The p-dimensional persistence diagram D_p of a filtration is the multiset of birth–death pairs (b, δ) for p-dimensional homology classes. Let D_1^{fin} be the finite pairs in D_1. We use the feature count $N(D_1) = |D_1^{\mathrm{fin}}|$, total persistence $P(D_1) = \sum_{(b,\delta)\in D_1^{\mathrm{fin}}} (\delta - b)$, and maximal lifetime

$$M(D_1) = \begin{cases} \max_{(b,\delta)\in D_1^{\mathrm{fin}}} (\delta - b), & D_1^{\mathrm{fin}} \neq \emptyset, \\ 0, & D_1^{\mathrm{fin}} = \emptyset. \end{cases}$$

We use total persistence $P(D_1)$ as the primary summary for the paired subject-level analysis. The feature count $N(D_1)$ and maximal lifetime $M(D_1)$ are analyzed as secondary summaries.

In the present setting, H_0 describes connected components and H_1 describes one-dimensional loops; points farther from the diagonal $\delta = b$ correspond to longer-lived features.

We apply the topological analysis to the *compact high-discriminability regime*, comparing significant regions $\mathcal{M}_c$ with size-matched low-weight regions $\mathcal{L}_c$ for each state class c.

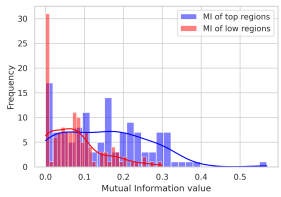


Figure 1. Empirical distributions of pairwise MI values computed for $\mathcal{M}_c$ and $\mathcal{L}_c$ in the *compact high-discriminability regime*.

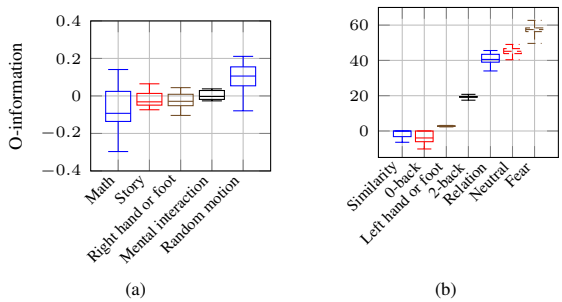


Figure 2. (a) $\mathrm{OI}(X_c)$ estimates for brain states in the *compact high-discriminability regime*. (b) $\mathrm{OI}(X_c)$ for brain states in the *distributed low-specificity regime*.

For every subject and every region from $\mathcal{M}_c$ or $\mathcal{L}_c$, we z-score the state-specific time series and construct a block-aware delay embedding with $d = 3$, $\tau = 2$, and unit stride. Using the original TR indices, we retain only delay vectors whose coordinates lie within the same contiguous acquisition block, pool the valid block-specific vectors into one point cloud, and compute Vietoris–Rips persistence diagrams up to dimension 1 with the Euclidean metric and coefficients in $\mathbb{F}_2$. These parameters keep the point clouds small while allowing nontrivial H_1 features; H_0 and H_1 are the most interpretable summaries for short noisy fMRI series.

Definition 4.7 (Cyclic-shift surrogate)**.** Let the state-specific time series of region i consist of contiguous blocks $\vec{x}_{i,r} = (x_{i,r,1}, \ldots, x_{i,r,n_r})$. For shifts $s_r \in \{0, \ldots, n_r - 1\}$, the blockwise cyclic-shift surrogate has entries

$$\widetilde{x}_{i,r,t}^{(s_r)} = x_{i,r,1+((t+s_r-1) \bmod n_r)}, \qquad t = 1, \ldots, n_r.$$

Definition 4.8 (Shuffle surrogate)**.** For a permutation π of $\{1, \ldots, l\}$, the shuffle surrogate of $\vec{x}_i$ is the time series $\widehat{\vec{x}}_i^{(\pi)}$ with entries

$$\widehat{x}_{it}^{(\pi)} = x_{i,\pi(t)}, \qquad t = 1, \ldots, l.$$

A blockwise cyclic shift preserves empirical values and within-block cyclic adjacency while changing the temporal origin. In our finite blockwise delay embedding, it can affect only vectors omitted at block boundaries, so we use it as an origin-sensitivity diagnostic rather than as an independent null for temporal ordering. Shuffle surrogates preserve the marginal distribution but destroy ordering.

For each regional signal, we generated five realizations of each surrogate with deterministic streams from base seed 42 and averaged their summaries. Cyclic shifts were sampled independently per block from nonzero shifts, and shuffles used uniform permutations of the complete state-specific series. Because five surrogates permit only descriptive per-series Monte Carlo assessment (minimum plus-one p-value $1/6$), formal inference is based on paired subject-level comparisons of the original recordings.

For $S \in \{\mathrm{cyc}, \mathrm{shuf}\}$, the reported regionwise contrast is $\Delta^{(S)}P(D_1) = P_{\mathrm{original}}(D_1) - 5^{-1}\sum_{r=1}^{5} P_{S,r}(D_1)$, averaged within each subject and region set. Lag-one Pearson autocorrelation was computed from within-block consecutive TR pairs and averaged similarly. We used two-sided paired Wilcoxon tests for $\mathcal{L}_c$ versus $\mathcal{M}_c$, and Spearman correlation between the subject-level autocorrelation difference and the shuffle response, defined as surrogate mean minus original $P(D_1)$. The autocorrelation and association analyses were Benjamini–Hochberg corrected separately across the six states.

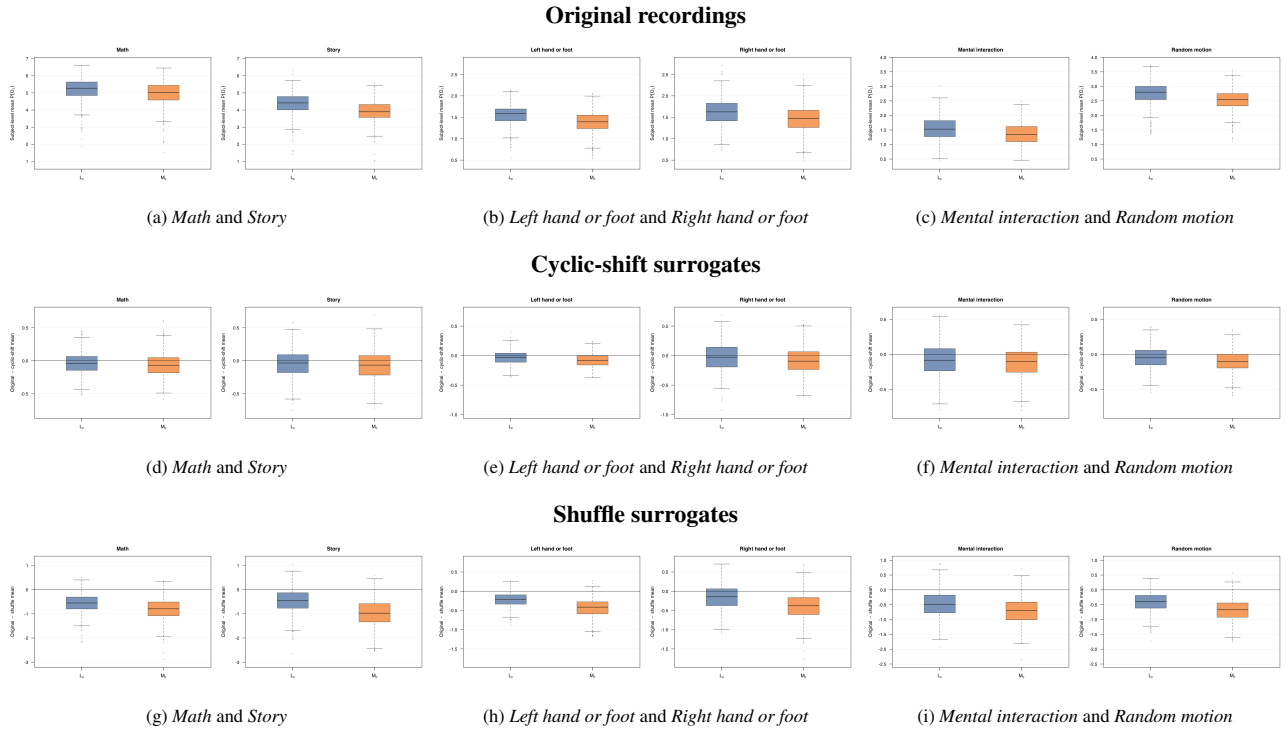
5. Results

5.1. Pairwise temporal dependence via MI

The empirical distributions of MI values for the top $\mathcal{M}_c$ and low-ranked regions $\mathcal{L}_c$ are visualized in Fig. 1.

For the top-ranked regions, the MI values had a mean of 0.153, a standard deviation of 0.109, a minimum of 0.000, a maximum of 0.569, and the 25th, 50th, and 75th percentiles were 0.059, 0.155, and 0.227, respectively. In contrast, for the low-ranked regions, the MI values had a mean of 0.074, a standard deviation of 0.068, a minimum of 0.000, a maximum of 0.299, and the 25th, 50th, and 75th percentiles were 0.023, 0.065, and 0.106.

To statistically evaluate the difference between these two empirical distributions, we applied the two-sample KS test, which formally assesses whether two samples originate from the same underlying distribution. The test yielded a KS statistic of 0.3906 with a p-value of 4.21×10^{-9}, indicating a substantial difference between the empirical distributions of MI values. Because MI values are not fully independent, the KS p-value should be interpreted descriptively rather than as an exact significance level. The result nevertheless indicates a clear distributional difference between classifier-selected and low-weight regions.



(a) *Math* and *Story* (b) *Left hand or foot* and *Right hand or foot* (c) *Mental interaction* and *Random motion*

(d) *Math* and *Story* (e) *Left hand or foot* and *Right hand or foot* (f) *Mental interaction* and *Random motion*

(g) *Math* and *Story* (h) *Left hand or foot* and *Right hand or foot* (i) *Mental interaction* and *Random motion*

Figure 3. Topological summaries for compact high-discriminability states. Top row: subject-level mean H_1 total persistence within $\mathcal{L}_c$ or $\mathcal{M}_c$. Middle and bottom rows: original-minus-mean-surrogate contrasts for blockwise cyclic shifts and complete shuffles. Each boxplot summarizes 581 participants; blue and orange denote $\mathcal{L}_c$ and $\mathcal{M}_c$.

5.2. Higher-order dependence analysis via O-inf.

Figure 2(a),(b) compares the O-information estimates $\mathrm{OI}(X_c)$ for the two regimes identified above. For classes c in the *compact high-discriminability regime*, where $|\mathcal{M}_c|$ is small, we observe $\mathrm{OI}(X_c) \approx 0$. These near-zero values, with small positive and negative deviations, indicate weak higher-order dependence among the selected regions in $\mathcal{M}_c$. Thus, the classifier appears to rely on a compact set of complementary regions with minimal redundancy or synergy.

In contrast, for classes c in the *distributed low-specificity regime*, where $|\mathcal{M}_c|$ is substantially larger, we observe $\mathrm{OI}(X_c) \gg 0$. This indicates redundancy-dominated higher-order dependence among the selected regions, suggesting that the corresponding cognitive states are represented by distributed and overlapping activity patterns across $\mathcal{M}_c$.

5.3. Temporal topology of $\mathcal{L}_c$ and $\mathcal{M}_c$.

As shown in Fig. 3, low-weight regions $\mathcal{L}_c$ had larger mean H_1 total persistence than significant regions $\mathcal{M}_c$ in all six state-specific settings; state-wise means ranged from 1.54 to 5.21 for $\mathcal{L}_c$ and from 1.36 to 4.98 for $\mathcal{M}_c$. For each subject, we averaged $P(D_{1,c,s,i})$ over regions in a set $\mathcal{A}$, writing this average as $\overline{P}_{c,s}(\mathcal{A})$, and tested $\Delta_{c,s}^{L-M} = \overline{P}_{c,s}(\mathcal{L}_c) - \overline{P}_{c,s}(\mathcal{M}_c)$, where $D_{1,c,s,i}$ is the H_1 diagram for region i of subject s in state c.

Table 2. Subject-level paired comparison of H_1 total persistence. Δ^{L-M} is the mean low-minus-selected difference over subjects; $L > M$ is the fraction with positive subject-level difference. p is the two-sided paired Wilcoxon value, and q is Benjamini–Hochberg adjusted across states.

State	Δ^{L-M}	$L > M$	p	q
Math	0.238	70.6%	3.85×10^{-33}	5.77×10^{-33}
Story	0.473	78.7%	5.63×10^{-51}	3.38×10^{-50}
Left hand/foot	0.174	76.2%	5.48×10^{-50}	1.64×10^{-49}
Right hand/foot	0.156	66.8%	1.60×10^{-18}	1.92×10^{-18}
Mental interaction	0.179	64.9%	3.36×10^{-16}	3.36×10^{-16}
Random motion	0.228	73.8%	2.60×10^{-38}	5.20×10^{-38}

The mean $\Delta_{c,s}^{L-M}$ was positive for all six compact states, and two-sided paired Wilcoxon tests remained significant after Benjamini–Hochberg correction (Table 2), indicating that the larger H_1 persistence of $\mathcal{L}_c$ is consistent across subjects.

Robustness checks were concordant: for $d \in \{2, 3, 4\}$ and $\tau \in \{1, 2, 3\}$, all 54 state–parameter combinations had positive mean $\Delta_{c,s}^{L-M}$ and remained significant after correction (maximum $q = 1.03\times10^{-4}$). The secondary summaries had the same direction, with low-minus-selected differences

ranging from 0.443 to 1.510 for $N(D_1)$ and from 0.0088 to 0.0221 for $M(D_1)$ (maximum corrected $q = 0.0106$).

Surrogate comparisons clarify which aspects of the delay geometry depend on temporal ordering. Blockwise cyclic shifts produced only small original-minus-surrogate contrasts ($\Delta^{(\mathrm{cyc})}P(D_1) \approx -0.116$ to -0.028), as expected for a transformation that preserves nearly all within-block delay neighborhoods apart from boundary-omitted vectors. This supports the stability of the TDA summary to the arbitrary temporal origin of each block. In contrast, complete shuffling produced much larger negative contrasts ($\Delta^{(\mathrm{shuf})}P(D_1) \approx -0.97$ to -0.16), indicating that randomizing the order, while preserving the marginal distribution of each regional signal, substantially changes its H_1 persistence. Thus the TDA summary is sensitive to sample ordering relative to complete shuffling.

Classifier-selected regions show a stronger response to temporal shuffling, consistent with the hypothesis that their delay-coordinate geometry is more closely related to temporal organization, although alternative explanations based on intrinsic local geometry cannot be excluded. Subject-level selected-minus-low autocorrelation differences were positively associated with shuffle-response differences in all states (Spearman $\rho = 0.216$–0.372, $q \leq 1.55\times10^{-7}$), indicating that part of this response may covary with ordinary serial dependence. Importantly, larger H_1 persistence should not be interpreted as evidence of greater task relevance. Persistent homology characterizes recurrence geometry within individual signals rather than discriminative information for cognitive-state decoding. Thus, larger persistence in low-weight regions may reflect richer local recurrence structure without implying a greater contribution to classification.

Code availability

https://github.com/dopeyhex/topofmri.

6. Conclusion

We formulated the analysis of decodable cognitive states in task-fMRI as a post-decoding hypothesis-testing problem. Rather than treating an interpretable decoder as the final output, we used it to define state-specific region sets and tested whether they exhibit distinctive statistical and temporal structure. This extends interpretability from variable selection to the validation of structural properties of learned representations.

Our results support three main conclusions. First, in the *compact high-discriminability regime*, classifier-selected regions exhibit stronger pairwise temporal dependence than matched low-weight controls. The distribution of MI values for $\mathcal{M}_c$ was consistently shifted toward larger values than that for $\mathcal{L}_c$, and a two-sample KS comparison revealed a clear distributional difference. Thus, selected regions are not only individually predictive but are also associated with stronger interregional dependence.

Second, flow-matching estimation of O-information revealed a clear distinction between compact and distributed representation regimes. States in the *compact high-discriminability regime* showed $\mathrm{OI}(X_c) \approx 0$, indicating weak higher-order dependence and relatively independent or complementary regional signatures. In contrast, states in the *distributed low-specificity regime* showed substantially positive $\mathrm{OI}(X_c)$, indicating redundancy-dominated higher-order structure and overlapping activity patterns across larger region sets. This provides a quantitative characterization of localized versus distributed cognitive-state representations.

Third, block-aware TDA showed larger H_1 persistence in low-weight regions, whereas selected regions responded more to complete shuffling. Small cyclic-shift effects and the autocorrelation association indicate ordering sensitivity relative to randomization, but not a specific nonlinear dynamical mechanism.

Taken together, these findings illustrate how hypothesis-testing methodology can serve as a post-decoding validation layer for high-dimensional scientific time-series data. After a predictive model identifies relevant variables, information-theoretic statistics, topological summaries, matched controls, and surrogate models can be used to test what makes those variables informative. In this sense, the proposed framework moves beyond activation-based decoding by testing not only which regions are predictive, but also how their information is statistically, geometrically, and temporally organized.

A limitation of this study is that the compact-versus-distributed dichotomy relies on the region-selection pipeline of (Kirova et al., 2026). Although we reproduced the original procedure and obtained the same representation regimes, their robustness to alternative selection strategies was not evaluated. Consequently, the reported information-theoretic and topological differences should be interpreted within the adopted selection framework. In addition, the same dataset was used for both region selection and downstream hypothesis testing. Thus, post-selection effects cannot be excluded, and some structural differences may reflect selection bias. A stronger confirmatory design would separate selection and testing using independent data splits or external validation data. Accordingly, the findings should be viewed as exploratory evidence for structured dependence rather than fully confirmatory statistical results. These limitations do not affect the descriptive patterns observed in the data, but should be considered when interpreting the inferential strength and generalizability of the findings.

References

Anderson, K. L., Anderson, J. S., Palande, S., and Wang, B. Topological data analysis of functional mri connectivity in time and space domains. In *Connectomics in NeuroImaging*, volume 11083, pp. 67–77. Springer, 2018. doi: 10.1007/978-3-030-00755-3_8.

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J. M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A. Z., and Van Essen, D. C. Function in the human connectome: Task-fmri and individual differences in behavior. *NeuroImage*, 80:169–189, 2013.

Bessadok, A., Mahjoub, M. A., and Rekik, I. Graph neural networks in network neuroscience. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5833–5848, 2023. doi: 10.1109/TPAMI.2022.3209686.

Bounoua, M., Franzese, G., and Michiardi, P. SΩi: Score-based o-information estimation. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.

Butakov, I., Semenenko, A., Kirova, V., Frolov, A., and Oseledets, I. Fmmi: Flow matching mutual information estimation, 2025. arXiv preprint arXiv:2511.08552.

Cassidy, B., Rae, C., and Solo, V. Brain activity: Conditional dissimilarity and persistent homology. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 1356–1359, 2015.

Cover, T. M. and Thomas, J. A. *Bibliography*. John Wiley & Sons, Ltd, 2005. ISBN 9780471748823. doi: 10.1002/047174882X.biblio. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/047174882X.biblio.

Das, S., Anand, D. V., and Chung, M. K. Topological data analysis of human brain networks through order statistics. *PLOS ONE*, 18(3), 2023.

Elam, J. S., Glasser, M. F., Harms, M. P., Sotiropoulos, S. N., Andersson, J. L. R., Burgess, G. C., Curtiss, S. W., Oostenveld, R., Larson-Prior, L. J., Schoffelen, J. M., Hodge, M. R., Cler, E. A., Marcus, D. M., Barch, D. M., Yacoub, E., Smith, S. M., Ugurbil, K., and Van Essen, D. C. The human connectome project: A retrospective. *NeuroImage*, 244:118543, 2021. doi: 10.1016/j.neuroimage.2021.118543.

Ellis, C. T., Lesnick, M., Henselman-Petrusek, G., Keller, B., and Cohen, J. D. Feasibility of topological data analysis for event-related fmri. *Network Neuroscience*, 3(3):695–706, 2019. doi: doi.org/10.1162/netn_a_00095.

Galeano Munoz, S. P., Bounoua, M., Franzese, G., Michiardi, P., and Filippone, M. Tende: Transfer entropy neural diffusion estimation, 2025. Version 3, revised April 9, 2026.

Ganmor, E., Segev, R., and Schneidman, E. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *Proceedings of the National Academy of Sciences*, 108(23):9679–9684, 2011. doi: 10.1073/pnas.1019641108.

Gatica, M., Cofré, R., Mediano, P. A. M., Rosas, F. E., Orio, P., Diez, I., Swinnen, S. P., and Cortes, J. M. High-order interdependencies in the aging brain. *Brain Connectivity*, 11(9):734–744, 2021. doi: 10.1089/brain.2020.0982.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87:96–110, 2014.

Haynes, J.-D. and Rees, G. Decoding mental states from brain activity in humans. *Nature Reviews. Neuroscience*, 7(7):523–534, 2006.

Hlinka, J., Paluš, M., Vejmelka, M., Mantini, D., and Corbetta, M. Functional connectivity in resting-state fmri: Is linear correlation sufficient? *NeuroImage*, 54(3):2218–2225, 2011. doi: 10.1016/j.neuroimage.2010.08.042.

Human Connectome Project. 1200 subjects data release reference manual, 2017. URL https://www.humanconnectome.org/study/hcp-young-adult. Washington University School of Medicine and University of Minnesota.

Ince, R. A. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., and Schyns, P. G. A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human Brain Mapping*, 38(3):1541–1573, 2017. doi: 10.1002/hbm.23471.

K., L. N. What we can do and what we cannot do with fmri. *Nature*, 453(7197):869–878, 2008. doi: 10.1038/nature06976.

Kim, H., Hahm, J., Lee, H., Kang, E., Kang, H., and Lee, D. S. Brain networks engaged in audiovisual integration during speech perception revealed by persistent homology-based network filtration. *Brain Connectivity*, 5(4):245–258, 2015.

Kirova, V., Kadieva, D., Vlasenko, D., Blank, I. B., and Ratnikov, F. Decoding cognitive states from fmri using classical machine learning and temporal dynamics analysis: An interpretable approach using the human connectome project. 2026. doi: 10.64898/2026.05.29.728756. bioRxiv preprint: 2026.05.29.728756.

Kriegeskorte, N. and Douglas, P. K. Interpreting encoding and decoding models. *Current Opinion in Neurobiology*, 55:167–179, 2019.

Kumar, S., Shovon, A.-R., and Deshpande, G. The robustness of persistent homology of brain networks to data acquisition-related non-neural variability in resting state fmri. *Human Brain Mapping*, 44(13):4637–4651, 2023.

Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L. H., Ventola, P., and Duncan, J. S. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical Image Analysis*, 74:102233, 2021.

Mondal, M., Yadav, Y., Jost, J., and Samal, A. Node persistence from topological data analysis reveals changes in brain functional connectivity. *Patterns (New York, N.Y.)*, 7(3), 2025.

Myers, A., Munch, E., and Khasawneh, F. A. Persistent homology of complex networks for dynamic state detection. *Phys. Rev. E*, 100:022314, Aug 2019. doi: 10.1103/PhysRevE.100.022314. URL https://link.aps.org/doi/10.1103/PhysRevE.100.022314.

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. Beyond mind-reading: Multi-voxel pattern analysis of fmri data. *Trends in Cognitive Sciences*, 10(9):424–430, 2006.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24): 9868–9872, 1990. doi: 10.1073/pnas.87.24.9868.

Pereira, F., Mitchell, T., and Botvinick, M. Machine learning classifiers and fmri: A tutorial overview. *NeuroImage*, 45 (1 Suppl):S199–S209, 2009. doi: 10.1016/j.neuroimage.2008.11.007.

Polyanskiy, Y. and Wu, Y. Information theory: From coding to learning, 2025.

Quian Quiroga, R. and Panzeri, S. Extracting information from neuronal populations: information theory and decoding approaches. *Nature reviews. Neuroscience*, 10(3): 173–185, 2009. doi: 10.1038/nrn2578.

Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., and Van De Ville, D. Decoding brain states from fmri connectivity graphs. *NeuroImage*, 56(2):616–626, 2011.

Rieck, B., Yates, T., Bock, C., Borgwardt, K., Wolf, G., Turk-Browne, N., and Krishnaswamy, S. Uncovering the topology of time-varying fmri data using cubical persistence. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6900–6912. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4d771504ddcd28037b4199740df767e6-Paper.pdf.

Rosas, F. E., Mediano, P. A. M., Gastpar, M., and Jensen, H. J. Quantifying high-order interdependencies via multivariate extensions of the mutual information. *Phys. Rev. E*, 100:032305, Sep 2019. doi: 10.1103/PhysRevE.100.032305. URL https://link.aps.org/doi/10.1103/PhysRevE.100.032305.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11):eaau4996, 2019. doi: 10.1126/sciadv.aau4996. URL https://www.science.org/doi/abs/10.1126/sciadv.aau4996.

Ryu, H., Habeck, C., Stern, Y., and Lee, S. Persistent homology-based functional connectivity and its association with cognitive ability: Life-span study. *Human Brain Mapping*, 44(9):3669–3683, 2023.

Saeidi, M., Karwowski, W., Farahani, F. V., Lighthall, N. R., Sonmez, E. B., and Wylie, G. R. Decoding task-based fmri data with graph neural networks, considering individual differences. *Brain Sciences*, 12(8):1094, 2022.

Saggar, M., Sporns, O., Gonzalez-Castillo, J., Bandettini, P. A., Carlsson, G., Glover, G., and Reiss, A. L. Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature Communications*, 9(1):1399, 2018. doi: 10.1038/s41467-018-03664-4.

Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, Jul 2000. doi: 10.1103/PhysRevLett.85.461. URL https://link.aps.org/doi/10.1103/PhysRevLett.85.461.

Songdechakraiwut, T. and Chung, M. K. Dynamic topological data analysis for functional brain signals. In *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)*, pp. 1–4, 2020.

Stolz, B. J., Emerson, T., Nahkuri, S., Porter, M. A., and Harrington, H. A. Topological data analysis of task-based fmri data from experiments on schizophrenia. *Journal of Physics: Complexity*, 2(3):035006, 2021.

Takens, F. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, volume 898 of *Lecture Notes in Mathematics*, pp. 366–381. Springer Berlin Heidelberg, 1981.

Takerkart, S., Auzias, G., Thirion, B., and Ralaivola, L. Graph-based inter-subject pattern analysis of fmri data. *PLOS ONE*, 9(8):e104586, 2014.

Talesh Jafadideh, A. and Mohammadzadeh Asl, B. Topological analysis of brain dynamics in autism based on graph and persistent homology. *Computers in Biology and Medicine*, 150:106202, 2022.

Timme, N. M. and Lapish, C. A tutorial for information theory in neuroscience. *eNeuro*, 5(3):ENEURO.0052–18.2018, 2018. doi: 10.1523/ENEURO.0052-18.2018.

Tong, F. and Pratte, M. S. Decoding patterns of human brain activity. *Annual Review of Psychology*, 63:483–509, 2012.

Varley, T. F., Pope, M., Faskowitz, J., and Sporns, O. Multivariate information theory uncovers synergistic subsystems of the human cerebral cortex. *Communications Biology*, 6(1):451, 2023. doi: 10.1038/s42003-023-04843-w.

Vigotsky, A. D., Iannetti, G. D., and Apkarian, A. V. Mental state decoders: Game-changers or wishful thinking? *Trends in Cognitive Sciences*, 28(10):884–895, 2024.

Wang, J., Zuo, X.-N., and He, Y. Graph-based network analysis of resting-state functional mri. *Frontiers in Systems Neuroscience*, 4:16, 2010.

Wong, E., Palande, S., Wang, B., Zielinski, B., Anderson, J., and Fletcher, P. T. Kernel partial least squares regression for relating functional brain network topology to clinical measures of behavior. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1303–1306, 2016.

Xing, J., Jia, J., Wu, X., and Kuang, L. A spatiotemporal brain network analysis of alzheimer's disease based on persistent homology. *Frontiers in Aging Neuroscience*, 14:788571, 2022.

Zomorodian, A. Fast construction of the vietoris-rips complex. *Computers & Graphics*, 34(3):263–271, 2010. ISSN 0097-8493. doi: 10.1016/j.cag.2010.03.007. URL https://www.sciencedirect.com/science/article/pii/S0097849310000464. Shape Modelling International (SMI) Conference 2010.

A. fMRI Tasks

Description of tasks completed in the fMRI scanner.

Working Memory. This task was designed to study short-term memory and information retention processes. Participants were shown blocks of images depicting places, tools, faces, and body parts. Each run included blocks from four types of stimuli, with half of the blocks performed with a 2-back task requiring memory retention, and the other half with a 0-back task for comparison. The 2-back task required participants to remember the sequence of images and determine whether the current image matched the one shown two steps earlier. The 0-back task only required participants to identify whether the current image matched a given target. Task blocks alternated with fixation blocks, where participants looked at a cross on the screen, allowing researchers to track brain activity changes in the absence of cognitive load.

Gambling Task: This task simulated decision-making processes under uncertainty and risk. Participants played a card game in which they had to guess whether the number on a hidden card was higher or lower than 5. Depending on their answer, they could win (green upward arrow with "1"), lose (red downward arrow with "0.50"), or receive no reward (neutral outcome). The task was divided into blocks dominated by either winning or losing outcomes, enabling researchers to examine how the brain responds to reward anticipation and loss.

Motor Task. This task was aimed at mapping motor areas of the brain. Participants were presented with visual cues instructing them to perform movements such as tapping their fingers on the left or right hand, squeezing their toes, or moving their tongue. Movement blocks lasted 12 seconds, and each of the two runs contained 13 blocks, including tongue, hand, and foot movements, as well as fixation blocks. These tasks help identify motor cortex areas activated in response to different types of movements.

Language Processing. This task aimed to study language comprehension and arithmetic processing. In the story blocks, participants listened to short narratives, such as adaptations of Aesop's fables, and then answered questions about the story content. In the math blocks, participants solved verbal arithmetic problems. The tasks alternated, and the block durations were adjusted to ensure equal completion times, allowing comparison of brain activity during language processing and numerical operations.

Social Cognition (Theory of Mind). Participants were shown short video clips where geometric shapes (squares, circles, triangles) either interacted with each other or moved randomly. After each video, participants evaluated whether the shapes had intentions and were interacting or whether their movements were random. This task allowed researchers to study brain areas associated with understanding social interactions and recognizing intentions.

Relational Processing. This task involved analyzing relationships between objects on the screen. In one condition, participants determined which characteristic (shape or texture) differentiated object pairs and assessed whether this distinction applied to another pair. In the control condition, they simply identified whether the lower object matched one of the upper objects based on a given criterion. This task helped researchers study the brain's ability to perform comparative analysis and identify relationships between objects.

Emotion Processing. Participants were shown faces displaying fear or anger, as well as neutral figures, and had to choose one of two faces or figures that matched the presented stimulus. Task blocks alternated with fixation blocks, allowing researchers to assess brain activity during emotional stimulus processing compared to neutral conditions.

B. The HCP indices and anatomical labels of significant brain regions

Index	Label
Language Processing: Story	
80	IFJa_L, IFJa, Fr, Inferior Frontal
117	AIP_L, AIP, Par, Superior Parietal
150	PGi_L, PGi, Par, Inferior Parietal
Language Processing: Math	
44	6ma_L, 6ma, Fr, Paracentral Lobular and Mid Cingulate
128	STSda_L, STSda, Temp, Auditory Association
135	TF_L, TF, Temp, Medial Temporal
137	PHT_L, PHT, Temp, Lateral Temporal
146	IP0_L, IP1, Par, Inferior Parietal
186	V4_R, V4, Occ, Early Visual
313	TE1p_R, TE1p, Temp, Lateral Temporal
337	FST_R, FST, Occ, MT+Complex and Neighboring Visual Areas
367	LP_L, Thalamus: Lateral Posterior
Social Cognition: Mental Interaction	
4	V2_L, V2, Occ, Dorsal Stream Visual
338	V3CD_R, V3CD, Occ, MT+Complex and Neighboring Visual Areas
Social Cognition: Random Motion	
13	V3A_L, V3A, Occ, Dorsal Stream Visual
128	STSda_R, STSda, Temp, Auditory Association
146	IP0_L, IP0, Par, Inferior Parietal
185	V3_R, V3, Occ, Early Visual
203	MT_R, MT, Occ, MT+Complex and Neighboring Visual Areas
309	STSdp_R, STSdp, Temp, Auditory Association
Motor Task: Right Hand or Foot	
39	5L_L, 5L, Par, Paracentral Lobular and Mid Cingulate
55	6mp_L, 6mp, Fr, Paracentral Lobular and Mid Cingulate
231	1_R, 1, Par, Somatosensory and Motor
Motor Task: Left Hand or Foot	
8	4_L, 4, Fr, Somatosensory and Motor
36	5m_L, 5m, Par, Paracentral Lobular and Mid Cingulate
39	5L_L, 5L, Par, Paracentral Lobular and Mid Cingulate
54	6d_L, 6d, Fr, Prem, otor
55	6mp_L, 6mp, Fr, Paracentral Lobular and Mid Cingulate
105	PFcm_L, PFcm, Par, Early Auditory
148	PF_L, PF, Par, Inferior Parietal,
220	24dd_R, 24dd, Fr, Paracentral Lobular and Mid Cingulate
231	1_R, 1, Par, Somatosensory and Motor
232	2_R, 2, Par, Somatosensory and Motor
281	OP1_R, OP1, Par, Posterior Opercular
285	PFcm_R, PFcm, Par, Early Auditory
296	PFt_R, PFt, Par, Inferior Parietal