LEARNING BRAIN HYPERGRAPHS FROM PAIRWISE-IRREDUCIBLE INTERACTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Brain hypergraph and other higher-order models aim to capture information beyond pairwise functional connectivity. Current methods, however, assign higher-order status by construction, from group size, statistical order, or architecture, so neither multiregional coactivation nor improved accuracy shows that a model uses information that pairwise statistics cannot explain. We propose HyPI (**Hy**pergraph Learning from **P**airwise-**I**rreducible Interactions), which replaces structure-defined with evidence-defined higher-order modeling. For each group of three or four regions, HyPI measures how far the observed joint activity departs from a pairwise maximum-entropy reference that matches all univariate and pairwise activation statistics. Departures are calibrated against temporal surrogates and screened for split-half reproducibility, yielding a hypergraph with a shared topology and subject-specific interaction vectors, which together form a subject's interaction profile. These vectors gate a residual branch over a frozen pairwise predictor, so a group contributes nothing when its calibrated interaction is zero. To certify higher-order use, we compare held-out risk against the pairwise predictor and against a capacity-matched model trained on reassigned profiles, and we replace the profiles at a fixed checkpoint to test whether predictions depend on the correct subject-specific profile. Synthetic experiments show recovery of planted interactions and identification of label-relevant ones. On ABIDE, REST-meta-MDD, ADNI, and ADHD-200, HyPI attains the highest mean accuracy across eight cohort–parcellation settings, and the controls show that models with similar accuracy differ markedly in certified higher-order use.

1 INTRODUCTION

Resting-state functional magnetic resonance imaging (rs-fMRI) records spontaneous brain activity, commonly summarized by functional connectivity between regions of interest (ROIs) (Biswal et al., 1995; Smith et al., 2009). These pairwise representations support brain-based prediction, including disorder diagnosis, through connectome-specific neural networks, graph neural networks, and Transformer-based models (Kawahara et al., 2017; Li et al., 2021; Kan et al., 2022; Kim et al., 2021). However, pairwise correlations do not uniquely determine multiregional joint activity: distinct joint distributions can share the same connectivity (Amari, 2001; Schneidman et al., 2003). This motivates higher-order modeling of relationships among three or more regions (Luppi et al., 2022; Varley et al., 2023; Santoro et al., 2023).

Existing work follows two main directions. Explicit higher-order methods characterize multiregional dependence through information measures, co-fluctuation patterns, or learned structures (Qiu et al., 2024; Santoro et al., 2024; Zhang et al., 2025; Zhao et al., 2026). Brain hypergraph methods organize ROIs into hyperedges and learn over these relations (Jin et al., 2025; Han et al., 2026; Ma et al., 2026; Hu et al., 2026). Yet assigning higher-order status *by construction*, from group size, statistical order, or architecture, does not establish whether a relation contains structure beyond its pairwise explanation (Rosas et al., 2022; Amari, 2001). Pairwise maximum-entropy models can account for substantial collective neural and BOLD activity (Schneidman et al., 2006; Watanabe et al., 2013), so multiregional coactivation alone does not establish pairwise irreducibility. Likewise, a hypergraph deterministically derived from connectivity cannot add information absent from that input; related static-FC nulls can reproduce several edge-centric features and co-fluctuation events (Novelli & Razi, 2022; Ladwig et al., 2022). We therefore move from *structure-defined* to *evidence-defined*

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

higher-order modeling, in which interactions are defined relative to a pairwise reference and their predictive use is verified rather than assumed.

Making this shift requires addressing three gaps in establishing and evaluating higher-order relations (Figure 1): **(i)** *Absence of a pairwise reference.* Without a reference for hyperedge admission, multiregional structure cannot be distinguished from effects already explained by univariate and pairwise statistics (Amari, 2001; Schneidman et al., 2003). **(ii)** *Unreliable higher-order estimation.* Finite samples can bias information estimates, and short scans limit the reliability of functional summaries (Jiao et al., 2015; Birn et al., 2013). Together with the combinatorial candidate space, these issues motivate matched-null calibration and label-independent, selection-independent reproducibility assessment (Prichard & Theiler, 1994; Benjamini & Hochberg, 1995; Kriegeskorte et al., 2009). **(iii)** *Unverified higher-order use.* Reliable structure need not carry task information, and task information need not be used by a trained model (Schneidman et al., 2003; Santoro et al., 2026; Fisher et al., 2019). Accuracy gains can reflect extra capacity or pairwise inputs (Errica et al., 2019; Zhang & Xu, 2026), while hyperedge weights or attention scores alone do not establish dependence on the correct interaction profile (Jain & Wallace, 2019; Wiegreffe & Pinter, 2019).

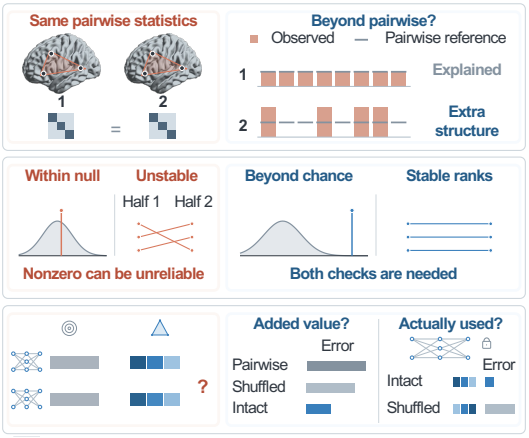

Figure 1: Gaps (i)–(iii), top to bottom. Left: structure-defined view; right: evidence-defined tests.

We instantiate this framework in HyPI (**Hy**pergraph Learning from **P**airwise-**I**rreducible Interactions). For variable-size regional groups, we quantify departures from a pairwise maximum-entropy reference matching binary activation and pairwise coactivation statistics, and retain their configuration through higher-order log-linear coefficients. Temporal-null calibration against second-order-matched surrogates and label-independent split-half screening produce a shared hypergraph topology with subject-specific interaction profiles. An interaction-gated residual branch corrects a frozen, strong pairwise predictor, so that a group sends no higher-order message when its calibrated interaction vector is zero. Held-out risk comparisons with the pairwise predictor and a capacity-matched model trained on reassigned profiles assess controlled predictive advantage; a same-checkpoint counterfactual assesses functional dependence on the correctly associated profile. Experiments on synthetic systems demonstrate recovery of planted interactions, and results on ABIDE, REST-meta-MDD, ADNI, and ADHD-200 show the highest mean accuracy across eight cohort–parcellation settings. The controls reveal markedly different certified higher-order use among models with similar accuracy.

Our contributions are threefold: **(i)** we formulate higher-order brain modeling through three distinct requirements: reliable pairwise irreducibility, controlled predictive advantage, and functional dependence on the correct interaction profile; **(ii)** we develop HyPI, which connects reference-based discovery of variable-size interactions to zero-preserving residual learning over a frozen pairwise predictor; **(iii)** we introduce a certification protocol for models with replaceable subject-specific interaction inputs, separating controlled predictive advantage from functional dependence at a fixed checkpoint.

2 Related Work

Higher-Order Interactions in fMRI. Higher-order fMRI studies characterize dependence among three or more regions through higher-order moments and multivariate information measures (Santoro et al., 2023; Rosas et al., 2019; Gatica et al., 2021; Varley et al., 2023), co-fluctuation analyses (Faskowitz et al., 2020; Santoro et al., 2023), and simplicial representations (Petri et al., 2014; Giusti et al., 2016). Predictive variants use signed or information-bottleneck formulations (Zhao et al., 2026; Zhang et al., 2025) or learn multiregional relations directly from BOLD signals (Qiu et al., 2024; Hu et al., 2026), with applications to task decoding, individual identification, and brain–behavior

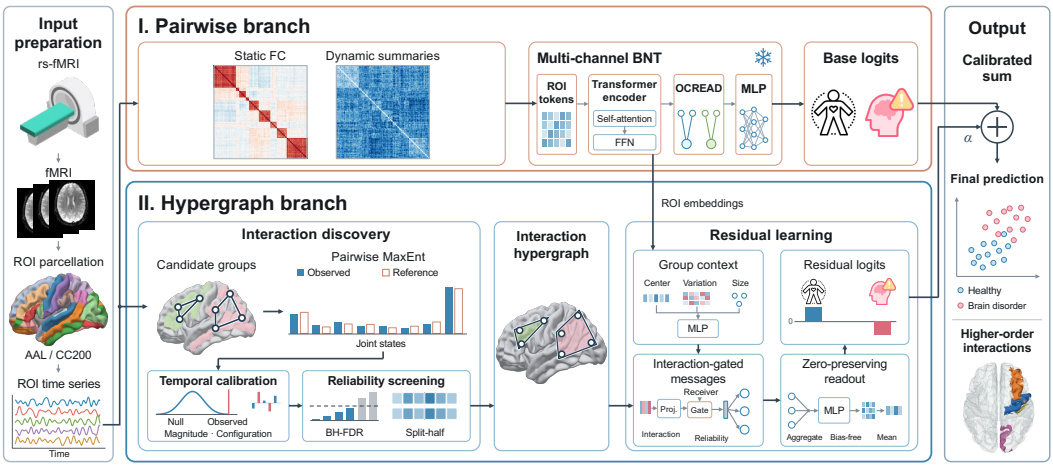

Figure 2: Overview of HyPI. Pairwise-reference comparison, temporal-null calibration, and reproducibility screening produce shared hyperedges with subject-specific interaction vectors. These vectors gate residual corrections to a frozen pairwise predictor. Capacity-matched and counterfactual controls evaluate predictive advantage and dependence on the correct profile.

analysis (Santoro et al., 2024; 2026). These methods describe multiregional dependence but do not separate it from what univariate and pairwise statistics already explain (Amari, 2001; Schneidman et al., 2003; Rosas et al., 2022). Pairwise maximum-entropy models account for much of the joint activity in neural populations and resting-state networks (Schneidman et al., 2006; Watanabe et al., 2013), but have so far served as descriptive baselines rather than as admission criteria for higher-order relations. HyPI uses the pairwise maximum-entropy model as a local reference: a group of regions becomes a hyperedge only if its departure from the reference survives temporal-null calibration and is reproducible across subjects.

Hypergraph Learning for fMRI. Hypergraph neural networks learn from multi-node relations (Feng et al., 2019; Chien et al., 2021). In fMRI, hyperedges have been built from connectivity similarity or sparse reconstruction (Jie et al., 2016; Zu et al., 2019), from predefined functional systems, temporal structure, or learned memberships, and extended with attention, self-supervised learning, and dynamic topology (Jin et al., 2025; Han et al., 2026; Ma et al., 2026; Hu et al., 2026). A hypergraph derived deterministically from the connectivity matrix adds no information beyond it (Novelli & Razi, 2022; Ladwig et al., 2022), and activity- or supervision-based constructions are higher-order by construction rather than by evidence of structure beyond a pairwise explanation. In both cases, predictive gains are attributed to the hyperedges without controlling for added capacity (Errica et al., 2019; Fisher et al., 2019) or testing whether the model depends on the specific hyperedge attributes it receives (Jain & Wallace, 2019; Wiegreffe & Pinter, 2019). HyPI differs on both counts: hyperedges are admitted by reference-based evidence (Section 3.3), and their contribution is isolated through a zero-preserving residual over a frozen pairwise predictor and verified against a capacity-matched null and a same-checkpoint counterfactual (Section 3.5).

3 Methodology

3.1 Problem Setting and Requirements

Data and notation. Each subject in $\mathcal{D} = \{(\mathbf{X}_s, y_s)\}_{s=1}^{n}$ has a BOLD matrix $\mathbf{X}_s \in \mathbb{R}^{T_s \times N}$ over T_s time points and N shared ROIs $\mathcal{V} = \{1, \dots, N\}$, with label $y_s \in \{1, \dots, C\}$. A regional group $g \subseteq \mathcal{V}$ has size $m_g = |g| \in \mathcal{M} = \{3, 4\}$. Let $\mathbf{S}_s^{(2)} = \Psi_2(\mathbf{X}_s)$ denote a pairwise-only representation (e.g., functional connectivity), $P_{s,g}$ the joint binary-state distribution of group g, and $\mathbf{F}_s^{\mathrm{HO}} = \{\mathbf{r}_{s,g}^{\mathrm{HO}}\}_{g \in \mathcal{E}}$ the interaction profile over a hyperedge set $\mathcal{E}$. Sections 3.2–3.3 define the interaction vectors $\mathbf{r}_{s,g}^{\mathrm{HO}}$ and construct $\mathcal{E}$.

Data partition. Within each outer cross-validation fold, test subjects are held out before partitioning the remaining subjects into fitting, model-selection, and calibration sets $\mathcal{D}_{\mathrm{tr}}$, $\mathcal{D}_{\mathrm{val}}$, and $\mathcal{D}_{\mathrm{cal}}$,

respectively. Before supervised training, label-free discovery uses the subject-disjoint proposal and confirmation subsets $\mathcal{D}_{\mathrm{tr}} = \mathcal{D}_{\mathrm{prop}} \dot{\cup} \mathcal{D}_{\mathrm{conf}}$.

Three requirements. We assess three requirements separately:

(P1) *Reliable pairwise irreducibility.* Group g shows a departure from its pairwise maximum-entropy reference beyond the temporal null, with reproducible subject-level expression.

(P2) *Controlled predictive advantage.* The correctly associated interaction profile lowers held-out risk relative to both the pairwise predictor and a capacity-matched model trained on reassigned profiles.

(P3) *Functional dependence.* At a fixed trained checkpoint, replacing only the subject-specific interaction profile increases held-out risk while pairwise inputs and parameters are unchanged.

Section 3.5 operationalizes (P2)–(P3) through controlled held-out log-loss comparisons.

3.2 Pairwise-Irreducible Interaction Representation

Joint-state estimation. Let $Z_{s,i}(t)$ be the temporally standardized $X_{s,i}(t)$ and $B_{s,i}(t) = \mathbb{I}[Z_{s,i}(t) > 0]$ its binary state, with $\mathbb{I}$ the indicator function. Counting the scan occurrences of $\mathbf{b} = (B_{s,i}(t))_{i \in g} \in \{0,1\}^{m_g}$ gives $n_{s,g}(\mathbf{b})$ and the estimate

$$\widehat{P}_{s,g}(\mathbf{b}) = \frac{n_{s,g}(\mathbf{b}) + \tau_g \overline{P}_g(\mathbf{b})}{T_s + \tau_g}, \tag{1}$$

where the strictly positive population prior $\overline{P}_g$ is estimated on $\mathcal{D}_{\mathrm{prop}}$ and $\tau_g > 0$ is the shrinkage weight, ensuring positive state probabilities. Both are fixed before confirmation and reused for all subsequent subjects (Appendix B.1).

Pairwise reference. Following pairwise maximum-entropy analyses (Schneidman et al., 2006; Watanabe et al., 2013), the reference matches all univariate and pairwise binary-state moments of $\widehat{P}_{s,g}$:

$$\widehat{Q}^{(2)}_{s,g} = \arg\max_{Q} H(Q) \quad \text{s.t.} \quad \begin{aligned} \mathbb{E}_Q[B_i] &= \mathbb{E}_{\widehat{P}_{s,g}}[B_i], \qquad i \in g, \\ \mathbb{E}_Q[B_i B_j] &= \mathbb{E}_{\widehat{P}_{s,g}}[B_i B_j], \quad i < j \in g, \end{aligned} \tag{2}$$

where H is entropy; the reference has only univariate and pairwise log-linear terms and is fitted by convex moment matching (Appendix B.2). The information not accounted for by these statistics is

$$J^{\mathrm{HO}}_{s,g} = D_{\mathrm{KL}}\left(\widehat{P}_{s,g} \| \widehat{Q}^{(2)}_{s,g}\right). \tag{3}$$

It aggregates orders three through m_g and vanishes exactly when $\widehat{P}_{s,g} = \widehat{Q}^{(2)}_{s,g}$.

Configuration. To distinguish joint patterns with equal divergence, we expand $\log \widehat{P}_{s,g}$ in the saturated log-linear basis and retain coefficients of order three and higher:

$$\log \widehat{P}_{s,g}(\mathbf{b}) = \sum_{A \subseteq g} \theta_{s,g,A} \prod_{i \in A} b_i, \qquad \boldsymbol{\theta}^{\mathrm{HO}}_{s,g} = (\theta_{s,g,A})_{A \subseteq g,\ |A| \geq 3} \in \mathbb{R}^{d_{m_g}}, \tag{4}$$

where the empty product is one, $d_m = 2^m - 1 - m - \binom{m}{2}$, and subsets follow a fixed coordinate ordering (Appendix B.3). Triplets have $d_3 = 1$; quadruplets have $d_4 = 5$, comprising four three-way and one four-way coefficient from the same four-region distribution. By Lemma 1, a positive triplet coefficient means that the all-active state occurs more often than the pairwise reference predicts.

Temporal-null calibration. Finite scan length, temporal autocorrelation, binarization, and estimation produce positive divergences even for a stationary Gaussian process. We generate M_{cal} phase-randomized surrogate scans (Prichard & Theiler, 1994) preserving the observed scan length and auto- and cross-spectra, hence its pairwise temporal structure. Repeating Eqs. (1)–(3) on each yields the null mean $\mu^{\mathrm{null}}_{s,g}$ and standard deviation $\sigma^{\mathrm{null}}_{s,g}$ (Appendix B.4). The calibrated interaction vector is

$$\mathbf{r}^{\mathrm{HO}}_{s,g} = \left[\frac{J^{\mathrm{HO}}_{s,g} - \mu^{\mathrm{null}}_{s,g}}{\sigma^{\mathrm{null}}_{s,g} + \epsilon}\right]_+ \mathcal{N}_0\left(\boldsymbol{\theta}^{\mathrm{HO}}_{s,g}\right), \tag{5}$$

where $[x]_+ = \max(x, 0)$, $\epsilon > 0$ is a small constant, and $\mathcal{N}_0(\boldsymbol{\theta}) = \boldsymbol{\theta}/\|\boldsymbol{\theta}\|_2$ with $\mathcal{N}_0(\mathbf{0}) = \mathbf{0}$. Its norm is the calibrated magnitude in null-standard-deviation units, and its direction retains the higher-order configuration. In particular, $\mathbf{r}^{\mathrm{HO}}_{s,g} = \mathbf{0}$ when $J^{\mathrm{HO}}_{s,g} \leq \mu^{\mathrm{null}}_{s,g}$; triplets yield signed scalars.

3.3 Reliable Interaction Hypergraph Construction

To establish (P1) over $\mathcal{U}_m = \{g \subseteq \mathcal{V} : |g| = m\}$, with $|\mathcal{U}_m| = \binom{N}{m}$, we rank candidates inexpensively on $\mathcal{D}_{\text{prop}}$ separately for each size and confirm the shortlist on $\mathcal{D}_{\text{conf}}$.

Size-wise proposal. On $\mathcal{D}_{\text{prop}}$, each scan is divided into two non-overlapping segments a and b. For each $g \in \mathcal{U}_m$, segment distributions use a symmetric pseudocount $\lambda_{\text{prop}} > 0$ instead of the population prior; Eq. (4) then gives $\widetilde{\boldsymbol{\theta}}^{(a)}_{s,g}$ and $\widetilde{\boldsymbol{\theta}}^{(b)}_{s,g}$. Their within-subject agreement defines

$$S_{\text{prop}}(g) = \frac{1}{|\mathcal{D}_{\text{prop}}|\, d_m} \sum_{s \in \mathcal{D}_{\text{prop}}} \left\langle \widetilde{\boldsymbol{\theta}}^{(a)}_{s,g}, \widetilde{\boldsymbol{\theta}}^{(b)}_{s,g} \right\rangle, \tag{6}$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product; the score rewards pronounced, directionally consistent within-subject configurations without requiring directions to agree across subjects. We nominate the K_m highest-scoring groups per size:

$$\mathcal{G}_{\text{cand}} = \bigcup_{m \in \mathcal{M}} \text{TopK}_{K_m}(\mathcal{U}_m; S_{\text{prop}}). \tag{7}$$

Quadruplets are selected directly from $\mathcal{U}_4$, not by extending triplets. Maximum-entropy fitting and surrogate calibration are applied only to $\mathcal{G}_{\text{cand}}$ (Appendix C.2).

Confirmation. On $\mathcal{D}_{\text{conf}}$, *group evidence* compares the mean observed $J^{\text{HO}}_{s,g}$ with the corresponding mean for each of the M_{conf} surrogate replicates to obtain a one-sided Monte Carlo p-value p_g. Benjamini–Hochberg adjustment (Benjamini & Hochberg, 1995) of these p-values over all of $\mathcal{G}_{\text{cand}}$ gives the false-discovery-rate-adjusted value q_g. Group evidence does not ensure stable individual differences, on which subject-level prediction relies. For *split-half reliability*, compute $\mathbf{r}^{(\text{HO},a)}_{s,g}$ and $\mathbf{r}^{(\text{HO},b)}_{s,g}$ from two fixed, non-overlapping halves a, b of each confirmation scan, using the fixed priors and shrinkage weights and surrogate calibration at each half's length. Rank each coordinate across confirmation subjects and center each rank column to form $\mathbf{K}^{(a)}_g, \mathbf{K}^{(b)}_g \in \mathbb{R}^{|\mathcal{D}_{\text{conf}}| \times d_{m_g}}$. We define

$$\text{Rel}_g = \frac{\langle \mathbf{K}^{(a)}_g, \mathbf{K}^{(b)}_g \rangle_{\text{F}}}{\|\mathbf{K}^{(a)}_g\|_{\text{F}} \|\mathbf{K}^{(b)}_g\|_{\text{F}}}, \qquad \mathcal{E}_{\text{rel}} = \{g \in \mathcal{G}_{\text{cand}} : q_g \le q_0,\ \text{Rel}_g \ge \rho_0\}, \tag{8}$$

where $\langle \cdot, \cdot \rangle_{\text{F}}$ and $\| \cdot \|_{\text{F}}$ are the Frobenius inner product and norm; set $\text{Rel}_g = 0$ if either norm vanishes. Here q_0 and ρ_0 are the evidence and reliability thresholds, and Rel_g is a multivariate Spearman correlation of between-subject rankings across scan halves. The possibly empty retained set defines $\mathcal{H}^{\text{HO}}_s = (\mathcal{V}, \mathcal{E}_{\text{rel}}, \{(\mathbf{r}^{\text{HO}}_{s,g}, \text{Rel}_g)\}_{g \in \mathcal{E}_{\text{rel}}})$ with shared topology and subject-specific attributes. Topology, priors, and reliabilities are fixed before supervised training (Appendix C.3).

3.4 Residual Hypergraph Learning on a Frozen Pairwise Predictor

To separate the interaction branch's contribution from backbone re-optimization, we learn residual corrections to a fixed pairwise predictor.

Pairwise backbone. A four-channel Brain Network Transformer (Kan et al., 2022) maps $\mathbf{S}^{(2)}_s$ to ROI representations and base logits:

$$(\mathbf{H}^{(2)}_s, \boldsymbol{\ell}^{(2)}_s) = f_{\text{pair}}(\mathbf{S}^{(2)}_s), \tag{9}$$

where $\mathbf{H}^{(2)}_s = (\mathbf{h}^{(2)}_{s,v})_{v \in \mathcal{V}} \in \mathbb{R}^{N \times d}$, $\boldsymbol{\ell}^{(2)}_s \in \mathbb{R}^C$, and d is the representation width. The backbone is trained with cross-entropy on $\mathcal{D}_{\text{tr}}$, selected on $\mathcal{D}_{\text{val}}$, and then frozen. Self-attention combines pairwise inputs without adding information absent from them; interaction vectors enter only through the residual branch (Appendix D.1).

Group context. Let ϕ and ρ be two-layer GELU networks shared across groups and subjects. For a hyperedge g and member $v \in g$, define

$$\begin{aligned} \mathbf{c}_{s,g} &= \frac{1}{m_g} \sum_{v \in g} \mathbf{h}^{(2)}_{s,v}, & \mathbf{t}_{s,v,g} &= \mathbf{h}^{(2)}_{s,v} - \mathbf{c}_{s,g}, \\ \mathbf{a}_{s,g} &= \frac{1}{m_g} \sum_{v \in g} \phi(\mathbf{t}_{s,v,g}), & \mathbf{q}_{s,g} &= \rho([\mathbf{c}_{s,g} \| \mathbf{a}_{s,g} \| \log m_g]), \end{aligned} \tag{10}$$

where $\|$ denotes concatenation. The center $\mathbf{c}_{s,g}$ captures pairwise context, while deviations $\mathbf{t}_{s,v,g}$ encode internal variation and are pooled after the nonlinearity because their linear mean is zero; $\log m_g$ makes summaries comparable across sizes.

Interaction-gated messages. Let ψ be a two-layer GELU network shared across groups, subjects, and receiving ROIs. Project the interaction vector and modulate it for each receiver:

$$\begin{aligned} \mathbf{u}_{s,g} &= \mathbf{P}_{m_g}\mathbf{r}^{\mathrm{HO}}_{s,g}, & \mathbf{b}_{s,g,v} &= \psi([\mathbf{q}_{s,g}\|\mathbf{t}_{s,v,g}]), \\ \mathbf{m}^{\mathrm{HO}}_{s,g\to v} &= \mathrm{Rel}_g\ \mathbf{W}_o\,(\mathbf{u}_{s,g}\odot\mathbf{b}_{s,g,v})\,. \end{aligned} \tag{11}$$

Here $\mathbf{P}_m \in \mathbb{R}^{d\times d_m}$ is shared by groups of size m, and both $\mathbf{P}_m$ and $\mathbf{W}_o \in \mathbb{R}^{d\times d}$ are bias-free. With $\odot$ denoting elementwise multiplication, the interaction vector supplies the signal, $\mathbf{b}_{s,g,v}$ controls receiver-specific expression, and Rel_g weights messages by reproducibility.

Zero-preserving readout. Aggregate messages and read out residual logits:

$$\Delta\mathbf{h}_{s,v} = \sum_{g\ni v}\frac{\mathbf{m}^{\mathrm{HO}}_{s,g\to v}}{\sqrt{d_v m_g}}, \qquad \mathbf{z}^{\mathrm{HO}}_s = \frac{1}{N}\sum_{v\in\mathcal{V}}\chi_0(\Delta\mathbf{h}_{s,v}), \qquad \Delta\boldsymbol{\ell}^{\mathrm{HO}}_s = \mathbf{W}_{\mathrm{cls}}\,\mathbf{z}^{\mathrm{HO}}_s, \tag{12}$$

where the sum uses retained groups and $d_v = |\{g \in \mathcal{E}_{\mathrm{rel}} : v \in g\}|$, with $\Delta\mathbf{h}_{s,v} = \mathbf{0}$ for $d_v = 0$. Zero interaction vectors remain attributes on the shared hyperedges and do not alter d_v. With bias-free $\mathbf{W}_{\chi,1}$, $\mathbf{W}_{\chi,2}$, and $\mathbf{W}_{\mathrm{cls}} \in \mathbb{R}^{C\times d}$, the map $\chi_0(\mathbf{x}) = \mathbf{W}_{\chi,2}\,\mathrm{GELU}(\mathbf{W}_{\chi,1}\mathbf{x})$ satisfies $\chi_0(\mathbf{0}) = \mathbf{0}$. Pairwise features affect messages only through $\mathbf{b}_{s,g,v}$; the readout receives messages alone.

Residual prediction and training. The final logits and training objective are

$$\boldsymbol{\ell}_s = \mathrm{sg}(\boldsymbol{\ell}^{(2)}_s) + \alpha\,\Delta\boldsymbol{\ell}^{\mathrm{HO}}_s, \qquad \mathcal{L}_{\mathrm{HO}} = \frac{1}{|\mathcal{D}_{\mathrm{tr}}|}\sum_{s\in\mathcal{D}_{\mathrm{tr}}}\mathrm{CE}(y_s,\boldsymbol{\ell}_s), \tag{13}$$

where sg stops gradients, CE denotes cross-entropy, and $\alpha \in [0,1]$ controls residual strength. At $\alpha = 1$, we minimize $\mathcal{L}_{\mathrm{HO}}$ over only the parameters Ω of the branch maps in Eqs. (10)–(12), keeping the backbone, topology, reliabilities, and interaction vectors fixed. Only $\mathbf{W}_{\mathrm{cls}}$ is zero-initialized, so training starts exactly at the pairwise predictor; α is then selected on $\mathcal{D}_{\mathrm{cal}}$ (Appendix D.4).

Proposition 1 (Zero-preserving residual path)**.** *Fix $\mathbf{H}^{(2)}_s$, $\boldsymbol{\ell}^{(2)}_s$, $\mathcal{E}_{\mathrm{rel}}$, the reliabilities, the normalization factors, and Ω, all finite. If $\mathbf{r}^{\mathrm{HO}}_{s,g} = \mathbf{0}$ for every $g \in \mathcal{E}_{\mathrm{rel}}$, or $\mathcal{E}_{\mathrm{rel}} = \emptyset$, then $\Delta\boldsymbol{\ell}^{\mathrm{HO}}_s = \mathbf{0}$ and $\boldsymbol{\ell}_s = \boldsymbol{\ell}^{(2)}_s$.*

Proof. Bias-free message maps give $\mathbf{m}^{\mathrm{HO}}_{s,g\to v} = \mathbf{0}$ and hence $\Delta\mathbf{h}_{s,v} = \mathbf{0}$; since $\chi_0(\mathbf{0}) = \mathbf{0}$ and $\mathbf{W}_{\mathrm{cls}}$ is bias-free, Eqs. (12)–(13) yield $\Delta\boldsymbol{\ell}^{\mathrm{HO}}_s = \mathbf{0}$ and $\boldsymbol{\ell}_s = \boldsymbol{\ell}^{(2)}_s$. □

By Eq. (11), a group with a zero interaction vector sends no message to any member ROI. With pairwise inputs, topology, reliabilities, normalization, and all model parameters fixed, profile replacement leaves the backbone outputs and context gates unchanged (Eqs. (9)–(11)); only the interaction-input path changes.

3.5 Certifying Higher-Order Use

The controls address two alternative explanations for predictive gains: added capacity (P2), and a branch responding to an auxiliary input's presence regardless of its correct subject-profile association (P3). Every risk $\mathcal{R}$ is the test-subject log-loss after calibration fitted on $\mathcal{D}_{\mathrm{cal}}$.

Matched profile replacement. Let π be a label-free, one-to-one subject reassignment with $\pi(s) \neq s$, matched on $\mathbf{S}^{(2)}$ (Appendix E.1). Replacing $\mathbf{F}^{\mathrm{HO}}_s$ by $\mathbf{F}^{\mathrm{HO}}_{\pi(s)}$ preserves hyperedge ROI identities, sizes, coordinate order, and the profile marginal distribution, but breaks the original subject-profile association while limiting pairwise-context shift.

Four conditions. We evaluate **Reduced** ($\mathcal{R}_{\mathrm{R}}$), the frozen pairwise predictor $\boldsymbol{\ell}^{(2)}_s$ alone; **Factual** ($\mathcal{R}_{\mathrm{F}}$), Eq. (13) with the correctly associated $\mathbf{F}^{\mathrm{HO}}_s$; **Capacity-Matched Null** ($\mathcal{R}_{\mathrm{N}}$), independently trained and tested with within-split reassigned profiles under the same architecture, frozen backbone, and optimization/calibration protocol as Factual; and **Counterfactual** ($\mathcal{R}_{\mathrm{CF}}$), the Factual checkpoint evaluated with matched test-set profiles $\mathbf{F}^{\mathrm{HO}}_{\pi'(s)}$, with parameters and calibration fixed.

Table 1: Classification accuracy (%) on ABIDE, REST-meta-MDD, ADNI (MCI vs NC), and ADHD-200 across different parcellations. **Bold** indicates the best result. Results are reported as mean ± standard deviation across the five outer folds after averaging the three model seeds within each fold.

Method	ABIDE		REST-meta-MDD		ADNI (MCI vs NC)		ADHD-200	
	AAL116	CC200	AAL116	CC200	AAL116	CC200	AAL116	CC200
BrainNetCNN	$63.12_{\pm 6.34}$	$62.45_{\pm 6.81}$	$61.87_{\pm 5.92}$	$63.05_{\pm 6.14}$	$59.87_{\pm 6.12}$	$59.21_{\pm 6.27}$	$60.15_{\pm 6.42}$	$60.52_{\pm 6.65}$
BrainGNN	$64.85_{\pm 5.72}$	$64.12_{\pm 6.15}$	$63.58_{\pm 5.48}$	$64.71_{\pm 5.63}$	$61.23_{\pm 5.94}$	$60.84_{\pm 6.13}$	$61.34_{\pm 6.15}$	$61.85_{\pm 6.35}$
BNT	$67.68_{\pm 5.12}$	$68.32_{\pm 5.25}$	$67.15_{\pm 4.88}$	$67.88_{\pm 4.95}$	$65.12_{\pm 5.14}$	$65.78_{\pm 5.24}$	$65.45_{\pm 5.82}$	$65.15_{\pm 5.95}$
STAGIN	$67.45_{\pm 5.16}$	$68.05_{\pm 5.44}$	$66.72_{\pm 4.95}$	$67.54_{\pm 5.08}$	$64.82_{\pm 5.38}$	$65.37_{\pm 5.41}$	$64.12_{\pm 5.95}$	$64.85_{\pm 6.12}$
HyBRiD	$68.12_{\pm 5.04}$	$68.94_{\pm 5.28}$	$67.85_{\pm 4.76}$	$68.41_{\pm 4.92}$	$65.71_{\pm 5.16}$	$66.42_{\pm 5.29}$	$66.24_{\pm 5.45}$	$65.92_{\pm 5.58}$
HOC	$68.85_{\pm 4.92}$	$69.42_{\pm 5.15}$	$68.15_{\pm 4.65}$	$69.02_{\pm 4.81}$	$66.38_{\pm 4.95}$	$67.08_{\pm 5.11}$	$66.85_{\pm 5.32}$	$66.41_{\pm 5.45}$
MvHo-IB	$71.45_{\pm 4.63}$	$72.18_{\pm 4.84}$	$70.92_{\pm 4.41}$	$71.65_{\pm 4.58}$	$69.15_{\pm 4.42}$	$69.87_{\pm 4.62}$	$69.45_{\pm 4.85}$	$69.12_{\pm 4.92}$
HOI-Brain	$70.21_{\pm 4.78}$	$71.05_{\pm 4.96}$	$69.58_{\pm 4.54}$	$70.34_{\pm 4.72}$	$67.92_{\pm 4.68}$	$68.71_{\pm 4.84}$	$67.32_{\pm 5.12}$	$67.95_{\pm 5.25}$
HGFD	$69.54_{\pm 4.85}$	$70.12_{\pm 5.07}$	$69.15_{\pm 4.61}$	$69.87_{\pm 4.79}$	$67.14_{\pm 4.83}$	$67.86_{\pm 4.97}$	$67.54_{\pm 5.24}$	$67.12_{\pm 5.35}$
HGST	$70.81_{\pm 4.71}$	$71.35_{\pm 4.92}$	$70.24_{\pm 4.48}$	$71.18_{\pm 4.65}$	$68.45_{\pm 4.56}$	$69.14_{\pm 4.73}$	$68.85_{\pm 4.95}$	$68.42_{\pm 5.15}$
HyperDiag	$71.23_{\pm 4.68}$	$71.86_{\pm 4.88}$	$71.05_{\pm 4.43}$	$71.52_{\pm 4.61}$	$68.96_{\pm 4.48}$	$69.63_{\pm 4.58}$	$69.25_{\pm 4.88}$	$68.85_{\pm 4.98}$
HA-STA	$71.95_{\pm 4.56}$	$72.41_{\pm 4.75}$	$71.68_{\pm 4.35}$	$72.15_{\pm 4.52}$	$69.81_{\pm 4.35}$	$70.39_{\pm 4.44}$	$69.15_{\pm 4.75}$	$69.84_{\pm 4.85}$
HyPI	$\mathbf{72.84}_{\pm 4.93}$	$\mathbf{73.31}_{\pm 5.17}$	$\mathbf{72.47}_{\pm 4.82}$	$\mathbf{72.95}_{\pm 4.61}$	$\mathbf{70.59}_{\pm 4.13}$	$\mathbf{71.16}_{\pm 4.29}$	$\mathbf{71.47}_{\pm 4.52}$	$\mathbf{71.28}_{\pm 4.65}$

The two criteria are

$$A_{\mathrm{HO}} = \min\{\mathcal{R}_{\mathrm{R}}, \mathcal{R}_{\mathrm{N}}\} - \mathcal{R}_{\mathrm{F}}, \tag{14}$$

$$D_{\mathrm{HO}} = \mathcal{R}_{\mathrm{CF}} - \mathcal{R}_{\mathrm{F}}. \tag{15}$$

Positive A_{HO} means the Factual model beats both the pairwise predictor and the capacity-matched null; D_{HO} measures the risk change under the fixed-checkpoint Counterfactual condition. We certify higher-order use only when the lower bounds of both quantities, obtained by paired subject-level bootstrap (Appendix E.2), exceed zero.

4 Experiments

Datasets. We evaluate four rs-fMRI cohorts: ABIDE (Di Martino et al., 2014) for autism spectrum disorder, REST-meta-MDD (Yan et al., 2019) for major depressive disorder, ADNI (Jack Jr et al., 2015) for mild cognitive impairment (MCI), and ADHD-200 (ADHD-200 Consortium, 2012) for attention-deficit/hyperactivity disorder (ADHD). Each cohort is evaluated separately under AAL116 (Tzourio-Mazoyer et al., 2002) and CC200 (Craddock et al., 2012), yielding eight cohort–parcellation–task settings. Within each dataset and parcellation, all methods share the same preprocessed ROI time series and subject-wise five-fold partitions. Interaction discovery is restricted to the fitting subjects; evaluated subjects receive their own interaction vectors under the fixed discovery rules. Cohort statistics and dataset-specific preprocessing are described in Appendix G.

Baselines. Pairwise baselines are BrainNetCNN, BrainGNN, BNT, and STAGIN (Kawahara et al., 2017; Li et al., 2021; Kan et al., 2022; Kim et al., 2021). Explicit higher-order baselines are HyBRiD, HOC, MvHo-IB, and HOI-Brain (Qiu et al., 2024; Santoro et al., 2024; Zhang et al., 2025; Zhao et al., 2026). Brain hypergraph baselines are HGFD, HGST, HyperDiag, and HA-STA (Jin et al., 2025; Han et al., 2026; Ma et al., 2026; Hu et al., 2026). Implementation and hyperparameters are given in Appendix H.

Evaluation and Certification Protocol. We report accuracy and assess higher-order use with the four conditions of Section 3.5: Reduced, Factual, Capacity-Matched Null, and Counterfactual. A_{HO} and D_{HO} are computed from held-out log-loss via Eqs. (14) and (15), and certification requires both lower confidence bounds to exceed zero. Certification applies only to models with a subject-specific higher-order input that can be removed or replaced, namely HyPI and five baselines (MvHo-IB, HOI-Brain, HGST, HyperDiag, and HA-STA); for these baselines, Reduced and Counterfactual remove or replace that input (Appendix E.3).

4.1 Predictive Performance and Higher-Order Use

Table 1 shows that HyPI achieves the highest mean accuracy across four datasets under two parcellations. Figure 3(a) compares controlled advantage A_{HO} and functional dependence D_{HO} for the six certifiable methods on ABIDE (AAL116). HyPI occupies the upper-right region at

$(A_{\mathrm{HO}}, D_{\mathrm{HO}}) = (5.48, 6.47) \times 10^{-3}$ nats, giving the largest point estimates on both axes. Among these six methods, only HyPI meets the joint certification criterion: its 95% confidence-interval lower endpoints are 1.25×10^{-3} and 2.04×10^{-3} nats for A_{HO} and D_{HO}, respectively. MvHo-IB lies high but farther left: its positive D_{HO} interval supports functional dependence, while its A_{HO} interval spans zero. HA-STA attains similar accuracy but lies close to $D_{\mathrm{HO}} = 0$, with an interval spanning zero. HGST, HyperDiag, and HOI-Brain cluster near $A_{\mathrm{HO}} = 0$ with nonpositive D_{HO}; both intervals include zero for all three. Thus, comparable accuracy can mask distinct patterns of controlled advantage and profile dependence. Failure to meet the criterion does not establish an absence of dependence. Full intervals are reported in Appendix E.4.

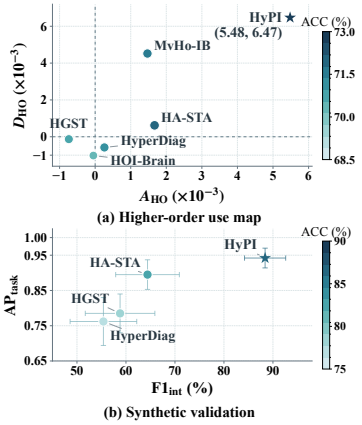

Figure 3: (a) Controlled advantage A_{HO} against functional dependence D_{HO} (10^{-3} nats) on ABIDE (AAL116); dashed lines mark zero, and 95% intervals are listed in Appendix E.4. (b) Synthetic validation: interaction-recovery $\mathrm{F1}_{\mathrm{int}}$ against task-relevance $\mathrm{AP}_{\mathrm{task}}$; error bars show one standard deviation across ten datasets.

4.2 Synthetic Validation

Because true interactions are unknown in rs-fMRI, we evaluate discovery and task relevance on ten synthetic datasets. Each contains 50 subjects, 30 ROIs, and 50 time points per subject, with six planted pairwise-irreducible triplets, three of which carry the label. Interaction recovery $\mathrm{F1}_{\mathrm{int}}$ measures agreement between retained and planted hyperedges (P1), while $\mathrm{AP}_{\mathrm{task}}$ evaluates the ranking of label-carrying interactions. In Figure 3(b), HyPI scores highest on both metrics and on accuracy. HA-STA approaches its task-relevance ranking but has substantially lower recovery F1; HGST and HyperDiag perform poorly on both metrics. Only HyPI performs strongly on both metrics among the compared methods, distinguishing interaction recovery from task-relevance ranking.

4.3 Ablation Study

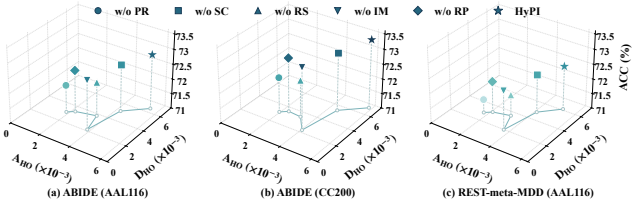

Figure 4: Ablation on ABIDE (AAL116), ABIDE (CC200), and REST-meta-MDD (AAL116). Axes are controlled advantage A_{HO}, functional dependence D_{HO} (both 10^{-3} nats), and accuracy; each marker removes one component, and the star is the full model.

We individually remove the pairwise reference (PR), surrogate-based temporal-null calibration (SC), reliability screening (RS), interaction-gated messages (IM), and residual prediction (RP). Figure 4 shows consistent trends across the three settings. Without PR, groups are no longer defined against a pairwise reference, and A_{HO} largely disappears, indicating little improvement over the pairwise and capacity-matched controls. Removing RS lowers all three measures, especially under CC200, whose candidate space is larger. SC provides smaller but consistent gains, in line with its role in correcting magnitude inflation rather than directly selecting groups. Without IM, corrections can arise from regional context alone: accuracy stays close to the full model, but D_{HO} is lowest in all three settings. Removing RP reduces A_{HO} as the branch competes with, rather than corrects, the pairwise predictor. Discovery and learning play complementary roles, with IM most clearly separating accuracy from profile dependence.

4.4 Sensitivity Study

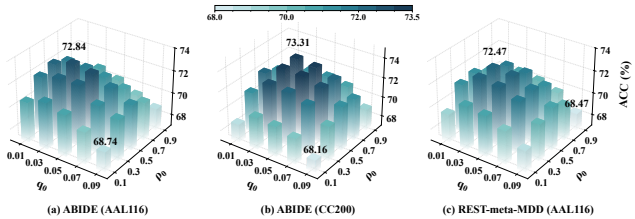

Figure 5: Accuracy of HyPI as a function of the group-evidence level q_0 and reliability threshold ρ_0 on ABIDE (AAL116), ABIDE (CC200), and REST-meta-MDD (AAL116).

We jointly vary the group-evidence level q_0 and reliability threshold ρ_0. Figure 5 shows a broad interior plateau across the three displayed settings: accuracy remains nearly unchanged at moderate thresholds and degrades at the

Table 2: Mean log-loss changes relative to Factual (10^{-3} nats; pointwise 95% paired-bootstrap CIs). All three interventions preserve the receiver's zero/nonzero pattern.

Intervention	ABIDE		REST-meta-MDD	
	AAL116	CC200	AAL116	CC200
Magnitude	3.82 [0.94, 6.71]	4.35 [1.32, 7.18]	2.95 [0.81, 5.12]	3.68 [1.05, 6.22]
Configuration	2.41 [0.45, 4.38]	2.18 [0.52, 4.05]	3.15 [0.92, 5.28]	1.92 [0.38, 3.55]
Joint	5.45 [1.56, 9.22]	6.02 [1.77, 10.15]	6.21 [1.85, 10.45]	4.85 [1.22, 8.41]

extremes. Looser q_0 and ρ_0 admit groups with weaker temporal-null evidence and less reproducible individual differences, respectively. Very strict reliability thresholds retain few groups, consistent with the model moving toward its pairwise predictor. CC200 favors a stricter reliability threshold than AAL116, consistent with its larger candidate space. The thresholds thus balance coverage against reliability while maintaining stable performance over a moderate range.

4.5 Dependence on Configuration Beyond Magnitude

We probe (P3)-type dependence by reassigning interaction norms, normalized directions, or both using a fixed pairwise-matched test-set mapping. Only groups with nonzero receiver and donor vectors are modified; all other inputs, topology, model parameters, normalization, and calibration remain fixed. For triplets, direction is the sign of the single coefficient (Section 3.2). Configuration reassignment preserves every norm and zero/nonzero pattern, yet raises log-loss by $(1.92\text{–}3.15) \times 10^{-3}$ nats across ABIDE and REST-meta-MDD under both atlases, with every reported 95% interval excluding zero (Table 2). Magnitude reassignment also increases risk; joint reassignment has the largest point estimate in each setting. Unlike the full Counterfactual, joint reassignment leaves zero-status mismatches unchanged. These results support configuration-specific dependence of the fixed model, not controlled advantage over a retrained magnitude-only model (Appendix E.5).

4.6 Interpretability Analysis

Figure 6 compares representative regional groups on ABIDE (AAL116), with complete ROI memberships and anatomical context provided in Appendix I. HyPI retains two distributed patterns: medial superior frontal–posterior cingulate–precuneus regions, consistent with reported default-mode dysconnectivity in ASD (Monk et al., 2009; Assaf et al., 2010), and a calcarine–Heschl's–precentral–middle frontal group spanning visual, auditory, motor, and frontal systems (Nebel et al., 2016; Holiga et al., 2019). The displayed baseline groups more often cluster anatomically related regions, including cerebellar/basal-ganglia, bilateral precuneus–posterior cingulate, and thalamus–caudate groups. These anatomical patterns are descriptive: HyPI admits interactions through pairwise-reference and reliability criteria, not spatial separation.

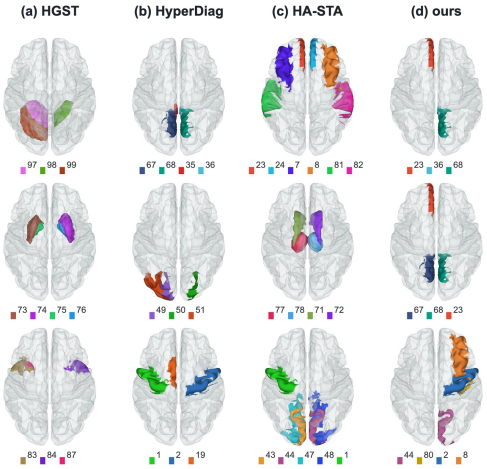

Figure 6: Representative regional groups on ABIDE (AAL116). Highlighted ROIs form one group; numbers are AAL116 indices.

5 Conclusion

We asked what brain hypergraphs learn beyond pairwise connectivity and argued that the question cannot be answered by construction. Admitting interactions only relative to a pairwise maximum-entropy reference, learning them as a zero-preserving residual over a frozen pairwise predictor, and certifying their use against a capacity-matched null and a same-checkpoint counterfactual turn higher-order modeling into a verifiable claim. Across four rs-fMRI cohorts and two parcellations, HyPI attains the highest mean accuracy, and the controlled evaluations separate predictive performance from evidence of higher-order use, motivating their joint assessment in higher-order brain models.

AI USE STATEMENT

Large language models were only used during manuscript revision for language polishing. The authors reviewed and revised the AI-assisted content and take full responsibility for the research ideas, experimental design, data analysis, and scientific conclusions.

ETHICS STATEMENT

This study performs secondary analysis of existing de-identified neuroimaging datasets. No new human participants were recruited and no new data were collected. We use each dataset in accordance with its applicable data-access and usage requirements. Dataset sources, cohort definitions, and preprocessing procedures are documented in Appendix G.

REPRODUCIBILITY STATEMENT

We provide detailed specifications for reproducing the reported experiments. Appendices B.1–C.3 describe interaction estimation and discovery; Appendix D.1 specifies the model architecture; Appendix E.1 details the certification protocol; and Appendices F and G report experimental configurations, preprocessing, and data splits. Baseline implementations are documented in Appendix H.

REFERENCES

ADHD-200 Consortium. The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in systems neuroscience*, 6:62, 2012.

S-I Amari. Information geometry on hierarchy of probability distributions. *IEEE transactions on information theory*, 47(5):1701–1711, 2001.

Michal Assaf, Kanchana Jagannathan, Vince D Calhoun, Laura Miller, Michael C Stevens, Robert Sahl, Jacqueline G O'Boyle, Robert T Schultz, and Godfrey D Pearlson. Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. *Neuroimage*, 53 (1):247–256, 2010.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

Rasmus M Birn, Erin K Molloy, Rémi Patriat, Taurean Parker, Timothy B Meier, Gregory R Kirk, Veena A Nair, M Elizabeth Meyerand, and Vivek Prabhakaran. The effect of scan length on the reliability of resting-state fmri connectivity estimates. *Neuroimage*, 83:550–558, 2013.

Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541, 1995.

Eli Chien, Chao Pan, Jianhao Peng, and Olgica Milenkovic. You are allset: A multiset function framework for hypergraph neural networks. *arXiv preprint arXiv:2106.13264*, 2021.

R Cameron Craddock, G Andrew James, Paul E Holtzheimer III, Xiaoping P Hu, and Helen S Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.

Adriana Di Martino, Clare Kelly, Rebecca Grzadzinski, Xi-Nian Zuo, Maarten Mennes, Maria Angeles Mairena, Catherine Lord, F Xavier Castellanos, and Michael P Milham. Aberrant striatal functional connectivity in children with autism. *Biological psychiatry*, 69(9):847–856, 2011.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

Federico Errica, Marco Podda, Davide Bacciu, and Alessio Micheli. A fair comparison of graph neural networks for graph classification. *arXiv preprint arXiv:1912.09893*, 2019.

Joshua Faskowitz, Farnaz Zamani Esfahlani, Youngheun Jo, Olaf Sporns, and Richard F Betzel. Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture. *Nature neuroscience*, 23(12):1644–1654, 2020.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 3558–3565, 2019.

Aaron Fisher, Cynthia Rudin, and Francesca Dominici. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of machine learning research*, 20(177):1–81, 2019.

Marilyn Gatica, Rodrigo Cofré, Pedro AM Mediano, Fernando E Rosas, Patricio Orio, Ibai Diez, Stephan P Swinnen, and Jesus M Cortes. High-order interdependencies in the aging brain. *Brain connectivity*, 11(9):734–744, 2021.

Chad Giusti, Robert Ghrist, and Danielle S Bassett. Two's company, three (or more) is a simplex: Algebraic-topological tools for understanding higher-order structure in neural data. *Journal of computational neuroscience*, 41(1):1–14, 2016.

Stephen J Gotts, W Kyle Simmons, Lydia A Milbury, Gregory L Wallace, Robert W Cox, and Alex Martin. Fractionation of social brain circuits in autism spectrum disorders. *Brain*, 135(9): 2711–2725, 2012.

Xiangmin Han, Mengqi Lei, and Junchang Li. Hypergraph-based semantic and topological self-supervised learning for brain disease diagnosis. *Pattern Recognition*, 169:111921, 2026.

Jesse Hemerik and Jelle Goeman. Exact testing with random permutations. *Test*, 27(4):811–825, 2018.

Štefan Holiga, Joerg F Hipp, Christopher H Chatham, Pilar Garces, Will Spooren, Xavier Liogier D'Ardhuy, Alessandro Bertolino, Céline Bouquet, Jan K Buitelaar, Carsten Bours, et al. Patients with autism spectrum disorders display reproducible functional connectivity alterations. *Science Translational Medicine*, 11(481):eaat9223, 2019.

Giles Hooker, Lucas Mentch, and Siyu Zhou. Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. *Statistics and Computing*, 31(6):82, 2021.

Wenqi Hu, Xuerui Su, Guanliang Li, Yidi Pan, and Aijing Lin. Learning high-order relationships with hypergraph attention-based spatio-temporal aggregation for brain disease analysis. *IEEE Journal of Biomedical and Health Informatics*, 2026.

Clifford R Jack Jr, Josephine Barnes, Matt A Bernstein, Bret J Borowski, James Brewer, Shona Clegg, Anders M Dale, Owen Carmichael, Christopher Ching, Charles DeCarli, et al. Magnetic resonance imaging in alzheimer's disease neuroimaging initiative 2. *Alzheimer's & Dementia*, 11(7):740–756, 2015.

Sarthak Jain and Byron C Wallace. Attention is not explanation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 3543–3556, 2019.

Jiantao Jiao, Kartik Venkat, Yanjun Han, and Tsachy Weissman. Minimax estimation of functionals of discrete distributions. *IEEE Transactions on Information Theory*, 61(5):2835–2885, 2015.

Biao Jie, Chong-Yaw Wee, Dinggang Shen, and Daoqiang Zhang. Hyper-connectivity of functional networks for brain disease diagnosis. *Medical image analysis*, 32:84–100, 2016.

Tao Jin, Yidan Xu, Yuhan Gao, Xichun Sheng, Chenggang Yan, Yaoqi Sun, Xiangmin Han, and Yue Gao. Hypergraph-guided federated distillation learning for efficient and robust multi-center fmri data analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 294–303. Springer, 2025.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer. *Advances in Neural Information Processing Systems*, 35:25586–25599, 2022.

Jeremy Kawahara, Colin J Brown, Steven P Miller, Brian G Booth, Vann Chau, Ruth E Grunau, Jill G Zwicker, and Ghassan Hamarneh. Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.

Amanda J Khan, Aarti Nair, Christopher L Keown, Michael C Datko, Alan J Lincoln, and Ralph-Axel Müller. Cerebro-cerebellar resting-state functional connectivity in children and adolescents with autism spectrum disorder. *Biological psychiatry*, 78(9):625–634, 2015.

Byung-Hoon Kim, Jong Chul Ye, and Jae-Jin Kim. Learning dynamic graph representation of brain connectome with spatio-temporal attention. *Advances in Neural Information Processing Systems*, 34:4314–4327, 2021.

Nikolaus Kriegeskorte, W Kyle Simmons, Patrick SF Bellgowan, and Chris I Baker. Circular analysis in systems neuroscience: the dangers of double dipping. *Nature neuroscience*, 12(5):535–540, 2009.

Zach Ladwig, Benjamin A Seitzman, Ally Dworetsky, Yuhua Yu, Babatunde Adeyemo, Derek M Smith, Steven E Petersen, and Caterina Gratton. Bold cofluctuation 'events' are predicted from static functional connectivity. *NeuroImage*, 260:119476, 2022.

Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H Staib, Pamela Ventola, and James S Duncan. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical image analysis*, 74:102233, 2021.

Andrea I Luppi, Pedro AM Mediano, Fernando E Rosas, Negin Holland, Tim D Fryer, John T O'Brien, James B Rowe, David K Menon, Daniel Bor, and Emmanuel A Stamatakis. A synergistic core for human brain evolution and cognition. *Nature neuroscience*, 25(6):771–782, 2022.

Yulan Ma, Fangkun Li, Wenchao Yang, Qian Si, Chenglong Yu, and Yang Li. Hyperdiag: Temporal–regional hypergraph learning via topology-enhanced state propagation for brain disease diagnosis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 7874–7882, 2026.

Christopher S Monk, Scott J Peltier, Jillian Lee Wiggins, Shih-Jen Weng, Melisa Carrasco, Susan Risi, and Catherine Lord. Abnormalities of intrinsic functional connectivity in autism spectrum disorders. *Neuroimage*, 47(2):764–772, 2009.

Stewart H Mostofsky, Stephanie K Powell, Daniel J Simmonds, Melissa C Goldberg, Brian Caffo, and James J Pekar. Decreased connectivity and cerebellar activity in autism during motor task performance. *Brain*, 132(9):2413–2425, 2009.

Aarti Nair, Jeffrey M Treiber, Dinesh K Shukla, Patricia Shih, and Ralph-Axel Müller. Impaired thalamocortical connectivity in autism spectrum disorder: a study of functional and anatomical connectivity. *Brain*, 136(6):1942–1955, 2013.

Mary Beth Nebel, Ani Eloyan, Carrie A Nettles, Kristie L Sweeney, Katarina Ament, Rebecca E Ward, Ann S Choe, Anita D Barber, James J Pekar, and Stewart H Mostofsky. Intrinsic visual-motor synchrony correlates with social deficits in autism. *Biological psychiatry*, 79(8):633–641, 2016.

Leonardo Novelli and Adeel Razi. A mathematical perspective on edge-centric brain functional connectivity. *Nature Communications*, 13(1):2693, 2022.

Giovanni Petri, Paul Expert, Federico Turkheimer, Robin Carhart-Harris, David Nutt, Peter J Hellyer, and Francesco Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101):20140873, 2014.

Dean Prichard and James Theiler. Generating surrogate data for time series with several simultaneously measured variables. *Physical review letters*, 73(7):951, 1994.

Weikang Qiu, Huangrui Chu, Selena Wang, Haolan Zuo, Xiaoxiao Li, Yize Zhao, and Rex Ying. Learning high-order relationships of brain regions. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=WC14xZIaC2.

Fernando E Rosas, Pedro AM Mediano, Michael Gastpar, and Henrik J Jensen. Quantifying high-order interdependencies via multivariate extensions of the mutual information. *Physical Review E*, 100(3):032305, 2019.

Fernando E Rosas, Pedro AM Mediano, Andrea I Luppi, Thomas F Varley, Joseph T Lizier, Sebastiano Stramaglia, Henrik J Jensen, and Daniele Marinazzo. Disentangling high-order mechanisms and high-order behaviours in complex systems. *Nature Physics*, 18(5):476–477, 2022.

Andrea Santoro, Federico Battiston, Giovanni Petri, and Enrico Amico. Higher-order organization of multivariate time series. *Nature physics*, 19(2):221–229, 2023.

Andrea Santoro, Federico Battiston, Maxime Lucas, Giovanni Petri, and Enrico Amico. Higher-order connectomics of human brain function reveals local topological signatures of task decoding, individual identification, and behavior. *Nature Communications*, 15(1):10244, 2024.

Andrea Santoro, Matteo Neri, Simone Poetto, Davide Orsenigo, Matteo Diano, Marilyn Gatica, and Giovanni Petri. Charting higher-order models of brain function beyond pairwise interactions. *Nature Communications*, 2026.

Elad Schneidman, Susanne Still, Michael J Berry, and William Bialek. Network information and connected correlations. *Physical review letters*, 91(23):238701, 2003.

Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

Stephen M Smith, Peter T Fox, Karla L Miller, David C Glahn, P Mickle Fox, Clare E Mackay, Nicola Filippini, Kate E Watkins, Roberto Toro, Angela R Laird, et al. Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the national academy of sciences*, 106(31):13040–13045, 2009.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Octave Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

Thomas F Varley, Maria Pope, Joshua Faskowitz, and Olaf Sporns. Multivariate information theory uncovers synergistic subsystems of the human cerebral cortex. *Communications biology*, 6(1):451, 2023.

Takamitsu Watanabe, Satoshi Hirose, Hiroyuki Wada, Yoshio Imai, Toru Machida, Ichiro Shirouzu, Seiki Konishi, Yasushi Miyashita, and Naoki Masuda. A pairwise maximum entropy model accurately describes resting-state human brain networks. *Nature communications*, 4(1):1370, 2013.

Sarah Wiegreffe and Yuval Pinter. Attention is not not explanation. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pp. 11–20, 2019.

Chao-Gan Yan, Xiao Chen, Le Li, Francisco Xavier Castellanos, Tong-Jian Bai, Qi-Jing Bo, Jun Cao, Guan-Mao Chen, Ning-Xuan Chen, Wei Chen, et al. Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. *Proceedings of the National Academy of Sciences*, 116(18):9078–9083, 2019.

Kunyu Zhang and Tianxiang Xu. Brainriem: Riemannian prototype learning for source-free cross-site brain network diagnosis. In *European Conference on Computer Vision*, pp. 308–325. Springer, 2026.

Kunyu Zhang, Qiang Li, and Shujian Yu. Mvho-ib: Multi-view higher-order information bottleneck for brain disorder diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 407–417. Springer, 2025.

Dengyi Zhao, Zhiheng Zhou, Guiying Yan, Dongxiao Yu, and Xingqin Qi. Hoi-brain: A novel multi-channel transformers framework for brain disorder diagnosis by accurately extracting signed higher-order interactions from fmri data. *Medical Image Analysis*, pp. 104009, 2026.

Chen Zu, Yue Gao, Brent Munsell, Minjeong Kim, Ziwen Peng, Jessica R Cohen, Daoqiang Zhang, and Guorong Wu. Identifying disease-related subnetwork connectome biomarkers by sparse hypergraph learning. *Brain imaging and behavior*, 13(4):879–892, 2019.

A NOTATION

Table 3 summarizes the key notation used in the main text and implementation appendices. Shared discovery quantities are fixed within each outer fold; interaction attributes are subject-specific.

Table 3: Summary of key notation.

Symbol	Description
Data, regional signals, and pairwise inputs	
$\mathcal{D}$, n, s	Subject dataset, number of subjects, and subject index.
$\mathbf{X}_s$, y_s	Retained BOLD sequence $\mathbf{X}_s \in \mathbb{R}^{T_s \times N}$ and diagnostic label of subject s.
T_s, N, C, $\mathcal{V}$	Number of retained time points, number of ROIs, number of classes, and the shared ROI set $\mathcal{V} = \{1, \dots, N\}$.
g, m_g, $\mathcal{M}$	Regional group, its size $m_g = \lvert g \rvert$, and the allowed size set $\mathcal{M} = \{3, 4\}$.
$\mathcal{D}_{\mathrm{tr}}, \mathcal{D}_{\mathrm{val}}, \mathcal{D}_{\mathrm{cal}}$	Subject sets for supervised fitting, checkpoint selection, and post-training calibration, respectively.
$\mathcal{D}_{\mathrm{prop}}, \mathcal{D}_{\mathrm{conf}}$	Label-free, subject-disjoint proposal and confirmation subsets, with $\mathcal{D}_{\mathrm{tr}} = \mathcal{D}_{\mathrm{prop}} \dot{\cup} \mathcal{D}_{\mathrm{conf}}$.
$Z_{s,i}(t)$, $B_{s,i}(t)$	Standardized signal of ROI i and its binary activity state $B_{s,i}(t) = \mathbb{I}[Z_{s,i}(t) > 0]$.
$\mathbf{S}_s^{(2)}$, Ψ_2	Pairwise input representation and the mapping from a subject's ROI sequence to that representation.
$\mathbf{R}_s$, $E_{s,ij}(t)$	Static Fisher-z FC matrix and edge co-fluctuation $E_{s,ij}(t) = Z_{s,i}(t) Z_{s,j}(t)$.
$\mathbf{V}_s$, $\mathbf{U}_s$, $\mathbf{L}_s$	Pairwise temporal summaries: co-fluctuation standard deviation, 0.9 quantile of absolute co-fluctuation, and lag-one co-fluctuation correlation.
Joint distributions and interaction representation	
$\mathbf{b}$, $n_{s,g}(\mathbf{b})$	Binary state of group g, with $\mathbf{b} \in \{0,1\}^{m_g}$, and its observed occurrence count.
$P_{s,g}$, $\widehat{P}_{s,g}$	Joint binary-state distribution and its population-informed, smoothed estimate for subject s and group g.
$\overline{P}_g$, λ_0, τ_g	Positive population prior estimated on $\mathcal{D}_{\mathrm{prop}}$, its per-state pseudocount, and the total subject-level shrinkage weight.
$\widehat{Q}_{s,g}^{(2)}$	Pairwise maximum-entropy reference matching all univariate and pairwise binary-state moments of $\widehat{P}_{s,g}$.
$\mathbf{f}_g$, $\boldsymbol{\mu}_{s,g}$	Vector of univariate and pairwise state statistics, and its expectation under $\widehat{P}_{s,g}$ used as the moment-matching target.
$\boldsymbol{\eta}$, h_i, w_{ij}, Z_g	Pairwise-reference parameter vector, fields, pairwise couplings, and partition function $Z_g(\boldsymbol{\eta})$.
$H(Q)$, $D_{\mathrm{KL}}(P \Vert Q)$	Shannon entropy and Kullback–Leibler divergence, evaluated with natural logarithms.
$J_{s,g}^{\mathrm{HO}}$	Raw pairwise-irreducible magnitude $D_{\mathrm{KL}}(\widehat{P}_{s,g} \Vert \widehat{Q}_{s,g}^{(2)})$, measured in nats.
$\theta_{s,g,A}$, $\boldsymbol{\theta}_{s,g}^{\mathrm{HO}}$	Log-linear coefficient associated with $A \subseteq g$, and the vector retaining coefficients with $\lvert A \rvert \geq 3$.
d_m	Interaction-vector dimension $2^m - 1 - m - \binom{m}{2}$; $d_3 = 1$ and $d_4 = 5$.
M_{cal}, M_{conf}	Numbers of surrogate realizations per full scan or half for feature calibration, and per confirmation subject for group-evidence testing.
$\mu_{s,g}^{\mathrm{null}}$, $\sigma_{s,g}^{\mathrm{null}}$	Mean and standard deviation of surrogate divergences used to calibrate $J_{s,g}^{\mathrm{HO}}$.
$\mathbf{r}_{s,g}^{\mathrm{HO}}$	Interaction vector combining nonnegative null-calibrated magnitude with normalized higher-order configuration.
$[\cdot]_+$, $\mathcal{N}_0$, ϵ	Positive-part operator, zero-preserving unit-norm normalization, and positive numerical stabilizer.
Candidate proposal, confirmation, and shared hypergraph	
a, b; $T_s^{(a)}, T_s^{(b)}$	Two fixed non-overlapping scan segments and their respective observation lengths.

Continued on next page.

Table 3: Summary of key notation (continued).

Symbol	Description
$\widetilde{P}_{s,g}^{(h)}$, $\widetilde{\boldsymbol{\theta}}_{s,g}^{(h)}$, λ_{prop}	Proposal-segment state estimate, its higher-order coefficient vector, and the symmetric per-state pseudocount; $h \in \{a, b\}$.
$\mathcal{U}_m$, K_m, $\mathcal{G}_{\mathrm{cand}}$	All size-m ROI groups, nomination budget for that size, and the combined shortlisted family.
$S_{\mathrm{prop}}(g)$	Proposal score: dimension-normalized, mean within-subject inner product of the two segment coefficient vectors.
n_{conf}, T_g^{obs}, $T_g^{(u)}$	Number of confirmation subjects and their mean observed or replicate-u surrogate divergence for group g.
p_g, q_g, K_{test}	Monte Carlo p-value, Benjamini–Hochberg-adjusted value, and size $\lvert\mathcal{G}_{\mathrm{cand}}\rvert$ of the tested family.
$\mathbf{K}_g^{(a)}$, $\mathbf{K}_g^{(b)}$	Column-centered, coordinatewise subject-rank matrices for the calibrated interaction vectors from the two scan halves.
Rel_g, q_0, ρ_0	Split-half rank agreement, group-evidence threshold, and minimum reliability threshold.
$\mathcal{E}$, $\mathcal{E}_{\mathrm{rel}}$	Generic hyperedge set and the retained set satisfying $q_g \le q_0$ and $\mathrm{Rel}_g \ge \rho_0$, shared within an outer fold.
$\mathcal{H}_s^{\mathrm{HO}}$, $\mathbf{F}_s^{\mathrm{HO}}$	Subject-attributed hypergraph and the subject's interaction profile $\{\mathbf{r}_{s,g}^{\mathrm{HO}}\}_{g \in \mathcal{E}_{\mathrm{rel}}}$.
Frozen backbone and residual hypergraph learning	
f_{pair}, $\mathbf{H}_s^{(2)}$	Pairwise predictor and its ROI representations, with $\mathbf{H}_s^{(2)} \in \mathbb{R}^{N \times d}$ and row $\mathbf{h}_{s,v}^{(2)}$.
$\boldsymbol{\ell}_s^{(2)}$, $\Delta\boldsymbol{\ell}_s^{\mathrm{HO}}$, $\boldsymbol{\ell}_s$	Frozen pairwise logits, higher-order residual logits, and their final combined logits, each in $\mathbb{R}^C$.
d, d_{h}, Ω	ROI/message width, hidden-network width, and trainable parameters of the residual interaction branch.
$\mathbf{c}_{s,g}$, $\mathbf{t}_{s,v,g}$	Mean pairwise ROI context of a group and the deviation of member v from that mean.
$\mathbf{a}_{s,g}$, $\mathbf{q}_{s,g}$	Nonlinearly pooled member deviations and the learned group summary incorporating center, variation, and size.
ϕ, ρ, ψ	Shared two-layer GELU networks for member deviations, group context, and receiver-specific modulation, respectively.
$\mathbf{P}_m$, $\mathbf{u}_{s,g}$, $\mathbf{W}_o$	Bias-free size-specific interaction projection, projected interaction vector, and bias-free message output map.
$\mathbf{b}_{s,g,v}$, $\mathbf{m}_{s,g\to v}^{\mathrm{HO}}$	Receiver-specific context modulation vector and the interaction-gated, reliability-weighted message to ROI v.
d_v, $\Delta\mathbf{h}_{s,v}$	Number of retained hyperedges incident to ROI v and its normalized aggregate message; zero interaction attributes do not change d_v.
χ_0, $\mathbf{z}_s^{\mathrm{HO}}$, $\mathbf{W}_{\mathrm{cls}}$	Zero-preserving node map, mean-pooled residual representation, and bias-free residual classifier.
α, $\mathcal{L}_{\mathrm{HO}}$	Residual strength selected after training and cross-entropy training objective for the interaction branch.
$\odot$, $\lVert$, sg	Elementwise multiplication, concatenation, and stop-gradient operator.
Matched interventions and higher-order-use evaluation	
$\mathcal{P}$, $\mathbf{s}_s$, $d(s, s')$	Matching pool, vectorized pairwise channels, and Euclidean pairwise-context distance between subjects.
π, π'	Label-free matched subject derangement for the capacity-matched Null and an independent test-pool derangement for Counterfactual evaluation.
$\ell_c(s)$, $\mathcal{R}_c$	Seed-averaged calibrated log-loss of subject s and its test-subject mean under condition $c \in \{\mathrm{R}, \mathrm{F}, \mathrm{N}, \mathrm{CF}\}$.
$\mathcal{R}_{\mathrm{R}}$, $\mathcal{R}_{\mathrm{F}}$, $\mathcal{R}_{\mathrm{N}}$, $\mathcal{R}_{\mathrm{CF}}$	Risks of Reduced, Factual, independently trained capacity-matched Null, and same-checkpoint Counterfactual conditions.
A_{HO}, D_{HO}	Controlled advantage $\min\{\mathcal{R}_{\mathrm{R}}, \mathcal{R}_{\mathrm{N}}\} - \mathcal{R}_{\mathrm{F}}$ and fixed-checkpoint risk change $\mathcal{R}_{\mathrm{CF}} - \mathcal{R}_{\mathrm{F}}$.
B, n_{test}	Number of paired bootstrap resamples and number of evaluated test subjects.

Continued on next page.

Table 3: Summary of key notation (continued).

Symbol	Description
I_{HO}	Conditional task information $I(Y;\mathbf{F}^{\mathrm{HO}} \mid \mathbf{S}^{(2)})$; distinct from the finite-model risk criteria A_{HO} and D_{HO}.

B Interaction Estimation and Configuration Encoding

This section specifies the quantities of Section 3.2: activity states and their shrinkage estimate, the pairwise maximum-entropy reference and its fitting, the higher-order coefficient coordinates, and the temporal surrogates. A group $g = \{i_1, \dots, i_m\}$ is stored with $i_1 < \cdots < i_m$, and every state vector, coefficient vector, and feature row of g follows this order. All logarithms are natural.

B.1 Activity States and Population-Informed Estimation

For subject s and ROI i, the standardized signal and binary state are

$$Z_{s,i}(t) = \frac{X_{s,i}(t) - \mu_{s,i}}{\sigma_{s,i} + \epsilon}, \qquad B_{s,i}(t) = \mathbb{I}[Z_{s,i}(t) > 0], \tag{16}$$

where $\mu_{s,i}$ and $\sigma_{s,i}$ are the temporal mean and standard deviation of that ROI signal within the scan being represented, and $\epsilon > 0$ is a small constant. Let $\mathcal{I}_s$ be the set of time indices used for an estimate and $T_s^{\mathcal{I}} = |\mathcal{I}_s|$; full-scan estimates use all indices, and split-half estimates use two non-overlapping index sets fixed once per scan. The state count and the population-informed estimate are

$$n_{s,g}^{\mathcal{I}}(\mathbf{b}) = \sum_{t \in \mathcal{I}_s} \mathbb{I}\big[(B_{s,i_1}(t), \dots, B_{s,i_m}(t)) = \mathbf{b}\big], \qquad \widehat{P}_{s,g}^{\mathcal{I}}(\mathbf{b}) = \frac{n_{s,g}^{\mathcal{I}}(\mathbf{b}) + \tau_g \overline{P}_g(\mathbf{b})}{T_s^{\mathcal{I}} + \tau_g}. \tag{17}$$

The population prior is the pooled, smoothed state frequency over proposal subjects,

$$\overline{P}_g(\mathbf{b}) = \frac{\sum_{s \in \mathcal{D}_{\mathrm{prop}}} n_{s,g}(\mathbf{b}) + \lambda_0}{\sum_{s \in \mathcal{D}_{\mathrm{prop}}} T_s + 2^m \lambda_0}, \qquad \lambda_0 > 0, \tag{18}$$

which is strictly positive and sums to one. The prior and the shrinkage weight τ_g are fixed before confirmation and reused for the two scan segments, the temporal surrogates, and all subsequently evaluated subjects. The superscript $\mathcal{I}$ is omitted for full-scan estimates.

B.2 Pairwise Maximum-Entropy Reference

Collect the first- and second-order statistics of g as

$$\mathbf{f}_g(\mathbf{b}) = \big[(b_i)_{i \in g} \,\|\, (b_i b_j)_{i<j \in g}\big]^\top \in \{0,1\}^{m + \binom{m}{2}}, \qquad \boldsymbol{\mu}_{s,g} = \sum_{\mathbf{b}} \widehat{P}_{s,g}(\mathbf{b})\, \mathbf{f}_g(\mathbf{b}). \tag{19}$$

The maximum-entropy problem of Eq. (2) has the exponential-family solution

$$Q_{\boldsymbol{\eta},g}(\mathbf{b}) = \exp\big(\boldsymbol{\eta}^\top \mathbf{f}_g(\mathbf{b}) - \log \mathcal{Z}_g(\boldsymbol{\eta})\big), \qquad \mathcal{Z}_g(\boldsymbol{\eta}) = \sum_{\mathbf{b} \in \{0,1\}^m} \exp\big(\boldsymbol{\eta}^\top \mathbf{f}_g(\mathbf{b})\big), \tag{20}$$

where $\boldsymbol{\eta} = [(h_i)_{i \in g} \| (w_{ij})_{i<j,\ i,j \in g}]^\top$ collects the fields and pairwise couplings of the pairwise reference, and Z_g is its normalizer. Equivalently,

$$Q_{\boldsymbol{\eta},g}(\mathbf{b}) = \frac{\exp\left(\sum_{i \in g} h_i b_i + \sum_{\substack{i,j \in g \\ i<j}} w_{ij} b_i b_j\right)}{Z_g(\boldsymbol{\eta})}.$$

The notation w_{ij} distinguishes pairwise couplings from the divergence $J_{s,g}^{\mathrm{HO}}$. The parameters are the unique minimizer of the convex objective

$$\begin{aligned} F_{s,g}(\boldsymbol{\eta}) &= \log \mathcal{Z}_g(\boldsymbol{\eta}) - \boldsymbol{\eta}^\top \boldsymbol{\mu}_{s,g}, \\ \nabla F_{s,g}(\boldsymbol{\eta}) &= \mathbb{E}_{Q_{\boldsymbol{\eta},g}}[\mathbf{f}_g] - \boldsymbol{\mu}_{s,g}, \qquad \nabla^2 F_{s,g}(\boldsymbol{\eta}) = \mathrm{Cov}_{Q_{\boldsymbol{\eta},g}}(\mathbf{f}_g), \end{aligned} \tag{21}$$

and $\widehat{Q}^{(2)}_{s,g} = Q_{\boldsymbol{\eta}^*_{s,g},g}$. The gradient vanishes exactly at moment matching. Because the state space has $2^m \le 16$ elements, $\mathcal{Z}_g$, the gradient, and the Hessian are evaluated by explicit enumeration with log-sum-exp, and Newton iterations from $\boldsymbol{\eta} = \mathbf{0}$ are run until the moment residual $\delta^{\mathrm{fit}}_{s,g} = \|\mathbb{E}_{\widehat{Q}^{(2)}_{s,g}}[\mathbf{f}_g] - \boldsymbol{\mu}_{s,g}\|_\infty$ falls below a fixed tolerance. The same routine and tolerance are used for observed and surrogate distributions.

The divergence of Eq. (3) is evaluated over the same 2^m states. Because $\log \widehat{Q}^{(2)}_{s,g}$ is affine in $\mathbf{f}_g$, its expectation is the same under $\widehat{P}_{s,g}$ and under $\widehat{Q}^{(2)}_{s,g}$ once the moments match, so

$$J^{\mathrm{HO}}_{s,g} = H(\widehat{Q}^{(2)}_{s,g}) - H(\widehat{P}_{s,g}), \tag{22}$$

which provides an algebraic check of the fit and shows that $J^{\mathrm{HO}}_{s,g}$ is the entropy that pairwise statistics leave unexplained.

B.3 Higher-Order Coefficient Coordinates

For $U \subseteq g$, let $\mathbf{b}^U_g$ denote the state with ones on U and zeros on $g \setminus U$. Inverting the expansion of Eq. (4) gives the Möbius formula

$$\theta_{s,g,A} = \sum_{U \subseteq A} (-1)^{|A|-|U|} \log \widehat{P}_{s,g}(\mathbf{b}^U_g), \qquad A \subseteq g, \tag{23}$$

so $\theta_{s,g,\emptyset} = \log \widehat{P}_{s,g}(\mathbf{0})$ and, for a triplet,

$$\begin{aligned}\theta_{s,g,\{i,j,k\}} = {} & \log \widehat{P}(1,1,1) - \log \widehat{P}(1,1,0) - \log \widehat{P}(1,0,1) - \log \widehat{P}(0,1,1) \\ & + \log \widehat{P}(1,0,0) + \log \widehat{P}(0,1,0) + \log \widehat{P}(0,0,1) - \log \widehat{P}(0,0,0),\end{aligned} \tag{24}$$

where $\widehat{P}$ abbreviates $\widehat{P}_{s,g}$. The higher-order vector $\boldsymbol{\theta}^{\mathrm{HO}}_{s,g}$ collects the coordinates with $|A| \ge 3$, ordered first by $|A|$ and then lexicographically by the ordered ROI tuple; its dimension is $d_m = \sum_{j=3}^{m} \binom{m}{j} = 2^m - 1 - m - \binom{m}{2}$. For $g = \{i,j,k,l\}$,

$$\boldsymbol{\theta}^{\mathrm{HO}}_{s,g} = \big(\theta_{s,g,\{i,j,k\}}, \theta_{s,g,\{i,j,l\}}, \theta_{s,g,\{i,k,l\}}, \theta_{s,g,\{j,k,l\}}, \theta_{s,g,\{i,j,k,l\}}\big)^\top, \tag{25}$$

all computed from the same four-variable distribution rather than from separate three-variable marginals. Computationally, the vector equals $\mathbf{L}_m \log \widehat{\mathbf{p}}_{s,g}$, where $\widehat{\mathbf{p}}_{s,g}$ lists the 2^m state probabilities in the fixed order and the entry of $\mathbf{L}_m$ in row A and column U is $(-1)^{|A|-|U|}$ if $U \subseteq A$ and zero otherwise. The matrix $\mathbf{L}_m$ is shared by all groups of size m and used for proposal, confirmation, segment, and test-subject estimates alike.

Lemma 1 (Sign of the triplet coefficient)**.** *Let $g = \{i,j,k\}$ and let $\widehat{Q}^{(2)}_{s,g}$ match all univariate and pairwise moments of $\widehat{P}_{s,g}$. Then* $\operatorname{sign} \theta_{s,g,\{i,j,k\}} = \operatorname{sign}\big(\widehat{P}_{s,g}(1,1,1) - \widehat{Q}^{(2)}_{s,g}(1,1,1)\big)$.

Proof. Write $\delta(\mathbf{b}) = \widehat{P}_{s,g}(\mathbf{b}) - \widehat{Q}^{(2)}_{s,g}(\mathbf{b})$ and $|\mathbf{b}|$ for the number of ones in $\mathbf{b}$. Moment matching means $\sum_{\mathbf{b}} \delta(\mathbf{b}) \prod_{i \in A} b_i = 0$ for every A with $|A| \le 2$, seven independent linear constraints on eight values, so δ lies in a one-dimensional space. The function $\chi(\mathbf{b}) = (-1)^{|\mathbf{b}|}$ satisfies these constraints, because $\sum_{\mathbf{b}} \chi(\mathbf{b}) \prod_{i \in A} b_i = \prod_{i \in A} \big(\sum_{b_i} (-1)^{b_i} b_i\big) \prod_{i \notin A} \big(\sum_{b_i} (-1)^{b_i}\big)$ contains the factor $\sum_{b_i} (-1)^{b_i} = 0$ whenever $|A| < 3$. Hence $\delta = c\,\chi$ for some $c \in \mathbb{R}$, and $\widehat{P}_{s,g}(1,1,1) - \widehat{Q}^{(2)}_{s,g}(1,1,1) = c\,\chi(1,1,1) = -c$. Since $\widehat{Q}^{(2)}_{s,g}$ has no third-order term, $\theta_{s,g,\{i,j,k\}}$ equals the Möbius sum of Eq. (24) applied to $\log \widehat{P}_{s,g} - \log \widehat{Q}^{(2)}_{s,g}$. Each term $\log \widehat{P}_{s,g}(\mathbf{b}^U_g) - \log \widehat{Q}^{(2)}_{s,g}(\mathbf{b}^U_g)$ has the sign of $c\,(-1)^{|U|}$ and carries the coefficient $(-1)^{3-|U|}$, so every term has the sign of $-c$. Therefore $\operatorname{sign} \theta_{s,g,\{i,j,k\}} = -\operatorname{sign} c = \operatorname{sign}\big(\widehat{P}_{s,g}(1,1,1) - \widehat{Q}^{(2)}_{s,g}(1,1,1)\big)$, with both sides zero when $c = 0$. □

B.4 Temporal Surrogates and Null Calibration

The same phase-randomization procedure is used with M_{cal} realizations for feature calibration and M_{conf} realizations per confirmation subject for group-level testing. Let $F_{s,i}(f)$ be the discrete

Table 4: Data allocation within an outer evaluation fold.

Subject set	Role
$\mathcal{D}_{\text{prop}}$	Candidate ranking; estimation of the population priors.
$\mathcal{D}_{\text{conf}}$	Group-level temporal-null evidence and split-half reliability of the nominated groups.
$\mathcal{D}_{\text{tr}} = \mathcal{D}_{\text{prop}} \dot{\cup} \mathcal{D}_{\text{conf}}$	Supervised training of the backbone and of the interaction branch after discovery is fixed.
$\mathcal{D}_{\text{val}}$	Checkpoint selection for the backbone and the interaction branch.
$\mathcal{D}_{\text{cal}}$	Post-training probability calibration and selection of α.
Test set	Final predictions, risks, and certification statistics.

Fourier transform of $Z_{s,i}$ at frequency index f. For each surrogate realization u and each positive non-Nyquist frequency, draw one phase $\varphi_{s,u,f} \sim \text{Unif}[0, 2\pi)$ and apply it to every ROI:

$$\widetilde{F}^{(u)}_{s,i}(f) = F_{s,i}(f) \exp(\mathrm{i}\varphi_{s,u,f}), \qquad i = 1, \dots, N, \tag{26}$$

assign the complex conjugate to the negative frequency, keep the DC and Nyquist components, and invert the transform to obtain a real-valued surrogate scan of the same length. Sharing the phase across ROIs preserves every auto- and cross-spectral product, $\widetilde{F}^{(u)}_{s,i}(f)\overline{\widetilde{F}^{(u)}_{s,j}(f)} = F_{s,i}(f)\overline{F_{s,j}(f)}$, and hence the corresponding circular auto- and cross-correlations of the represented discrete sequences, while randomizing common Fourier phases across frequencies. Phases are drawn independently across surrogates and subjects; surrogates are generated for the whole multivariate scan and groups are extracted afterward.

Each surrogate is processed exactly as the observed scan: thresholding, the fixed population prior and τ_g, a pairwise maximum-entropy reference refitted to the surrogate's own binary moments, and the divergence from that reference. Every realization is thus compared with its own reference, so $J^{(\text{HO},u)}_{s,g}$ is the statistic $J^{\text{HO}}_{s,g}$ evaluated on a phase-rotated scan, never a divergence from the observed reference. For feature calibration, let $J^{(\text{HO},u)}_{s,g}$ denote the divergence of the uth calibration surrogate, with $u = 1, \dots, M_{\text{cal}}$ and $M_{\text{cal}} \geq 2$:

$$\mu^{\text{null}}_{s,g} = \frac{1}{M_{\text{cal}}} \sum_{u=1}^{M_{\text{cal}}} J^{(\text{HO},u)}_{s,g}, \qquad \left(\sigma^{\text{null}}_{s,g}\right)^2 = \frac{1}{M_{\text{cal}} - 1} \sum_{u=1}^{M_{\text{cal}}} \left(J^{(\text{HO},u)}_{s,g} - \mu^{\text{null}}_{s,g}\right)^2, \tag{27}$$

and $\mathbf{r}^{\text{HO}}_{s,g}$ follows from Eq. (5). Each full scan and each scan half uses $M_{\text{cal}} = 500$ surrogate realizations at its own observation length. The calibrated magnitude is dimensionless, whereas $J^{\text{HO}}_{s,g}$ is in nats.

C Candidate Proposal and Reliable Interaction Discovery

C.1 Subject-Level Data Allocation

Within each outer fold, the held-out test subjects are separated before discovery, and the remaining subjects are divided into $\mathcal{D}_{\text{tr}}$, $\mathcal{D}_{\text{val}}$, and $\mathcal{D}_{\text{cal}}$, with $\mathcal{D}_{\text{tr}} = \mathcal{D}_{\text{prop}} \dot{\cup} \mathcal{D}_{\text{conf}}$. Table 4 lists the role of each set. Proposal and confirmation never access labels; after discovery, both fitting subsets are used for supervised training. Split identifiers and seeds are retained. Confirmation does not update the proposal ranking or refit the prior, and after confirmation the selected groups, coordinate order, priors, and reliabilities are reused unchanged for every subject in every subset.

C.2 Independent Proposal for Each Group Size

For $m \in \mathcal{M}$ and a proposal subject s with segments $h \in \{a, b\}$ of length $T^{(h)}_s$, the proposal distribution uses a symmetric pseudocount,

$$\widetilde{P}^{(h)}_{s,g}(\mathbf{b}) = \frac{n^{(h)}_{s,g}(\mathbf{b}) + \lambda_{\text{prop}}}{T^{(h)}_s + 2^m \lambda_{\text{prop}}}, \qquad \lambda_{\text{prop}} > 0, \tag{28}$$

and $\widetilde{\boldsymbol{\theta}}_{s,g}^{(h)} = \mathbf{L}_m \log \widetilde{\mathbf{p}}_{s,g}^{(h)}$. The score $S_{\mathrm{prop}}(g)$ of Eq. (6) is a ranking statistic, not a calibrated magnitude. For each size, the $\min(K_m, |\mathcal{U}_m|)$ groups with the largest scores are retained, ties broken by the canonical ROI tuple. The three- and four-region searches are independent: a nominated quadruplet need not contain a nominated triplet, a strong FC edge, or adjacent ROIs. The distribution $P(x_0, x_1, x_2, x_3) = \frac{1}{16}(1 + \gamma x_0 x_1 x_2 x_3)$ on $x_i \in \{-1, +1\}$ with $0 < |\gamma| < 1$ illustrates why: every three-variable and lower-order marginal is uniform, so no triplet shows any interaction, yet the four-variable distribution is nonuniform.

Streaming evaluation. The search space contains $\binom{N}{3}$ triplets and $\binom{N}{4}$ quadruplets: 253,460 and 7,160,245 for $N = 116$, and 1,313,400 and 64,684,950 for $N = 200$. For each size, the implementation caches the proposal subjects' binary sequences and $\mathbf{L}_m$, enumerates canonical tuples in blocks, forms state counts and segment coefficients for the block, accumulates Eq. (6), and keeps the current top-K_m tuples in a heap before releasing the block. Exact maximum-entropy fits and surrogates are computed only for the $|\mathcal{G}_{\mathrm{cand}}| = \sum_m K_m$ nominated groups.

C.3 Confirmation Against the Temporal Reference

Let $n_{\mathrm{conf}} = |\mathcal{D}_{\mathrm{conf}}|$. Group-level confirmation uses $M_{\mathrm{conf}} = 20{,}000$ full-scan surrogate realizations per confirmation subject, generated by the procedure in Appendix B.4. Here $J_{s,g}^{(\mathrm{HO},u)}$ denotes the divergence from the uth confirmation surrogate, and the observed statistic, surrogate statistics, and one-sided Monte Carlo p-value are

$$T_g^{\mathrm{obs}} = \frac{1}{n_{\mathrm{conf}}} \sum_{s \in \mathcal{D}_{\mathrm{conf}}} J_{s,g}^{\mathrm{HO}}, \qquad T_g^{(u)} = \frac{1}{n_{\mathrm{conf}}} \sum_{s \in \mathcal{D}_{\mathrm{conf}}} J_{s,g}^{(\mathrm{HO},u)}, \quad u = 1, \ldots, M_{\mathrm{conf}},$$
$$p_g = \frac{1 + \sum_{u=1}^{M_{\mathrm{conf}}} \mathbb{I}[T_g^{(u)} \geq T_g^{\mathrm{obs}}]}{M_{\mathrm{conf}} + 1}. \tag{29}$$

Surrogate replicate u uses one complete surrogate scan per confirmation subject, shared across that subject's candidate groups, so observed and surrogate statistics use the same subjects, scan lengths, priors, and fitting conventions.

Null hypothesis and validity. Assume independent subjects. For each $g \in \mathcal{G}_{\mathrm{cand}}$, p_g tests the temporal null $H_{0,g}$: for every confirmation subject, the joint law of the standardized signals of the ROIs in g is invariant under the shared-phase rotations of Eq. (26). Conditional on the quantities fixed on $\mathcal{D}_{\mathrm{prop}}$ (the candidate family, P_g, and τ_g), $T_g^{\mathrm{obs}}, T_g^{(1)}, \ldots, T_g^{(M_{\mathrm{conf}})}$ are exchangeable under $H_{0,g}$, so $\Pr(p_g \leq \alpha) \leq \alpha$ for all $\alpha \in [0, 1]$ (Hemerik & Goeman, 2018). The invariance holds exactly for circularly stationary Gaussian processes and approximately for finite stationary Gaussian scans. $H_{0,g}$ is therefore not the hypothesis that the binarized joint states are pairwise-explained: a rejection indicates a departure from this invariance, which beyond-pairwise structure can produce, as can non-Gaussian marginals, non-stationarity, or transient artifacts.

With $K_{\mathrm{test}} = |\mathcal{G}_{\mathrm{cand}}|$ and sorted values $p_{(1)} \leq \cdots \leq p_{(K_{\mathrm{test}})}$, the Benjamini–Hochberg adjusted values are

$$q_{(r)} = \min\left\{1, \min_{j \geq r} \frac{K_{\mathrm{test}}}{j} p_{(j)}\right\}, \tag{30}$$

mapped back to their groups; adjustment is performed over the combined nominated family before reliability filtering.

Monte Carlo resolution. The smallest attainable p_g is $1/(M_{\mathrm{conf}} + 1)$. For this minimum to reach the first BH comparison level q_0/K_{test}, the budget must satisfy $M_{\mathrm{conf}} \geq \lceil K_{\mathrm{test}}/q_0 \rceil - 1$. With $K_{\mathrm{test}} = 512$ and $q_0 = 0.03$, this requires $M_{\mathrm{conf}} \geq 17{,}066$. Our setting $M_{\mathrm{conf}} = 20{,}000$ gives $p_{\min} = 1/20{,}001 \approx 5.00 \times 10^{-5}$, below $q_0/K_{\mathrm{test}} \approx 5.86 \times 10^{-5}$. This establishes numerical resolution of the first comparison level, not that any particular group is rejected; several small p-values may also pass later step-up levels with smaller budgets.

C.4 Split-Half Configuration Reliability

For each confirmation subject, $\mathbf{r}_{s,g}^{(\mathrm{HO},a)}$ and $\mathbf{r}_{s,g}^{(\mathrm{HO},b)}$ are computed from the two fixed segments by repeating the shrinkage estimate, the maximum-entropy fit, surrogate calibration with $M_{\mathrm{cal}} = 500$

realizations at each segment's length, and the normalization, with the same coordinate order and prior. For coordinate $j = 1, \ldots, d_{m_g}$ and segment h, let $R^{(h)}_{g,sj}$ be the rank of $r^{(\mathrm{HO},h)}_{s,g,j}$ among confirmation subjects, ties receiving their average rank. The centered rank matrices have entries

$$K^{(h)}_{g,sj} = R^{(h)}_{g,sj} - \frac{n_{\mathrm{conf}} + 1}{2}, \tag{31}$$

so a coordinate that is constant across subjects becomes a zero column. Rel_g of Eq. (8) compares corresponding coordinatewise subject rankings across segments and reduces to the split-half Spearman correlation for a scalar interaction. Segments are non-overlapping halves of one scan, so Rel_g measures within-scan consistency rather than test–retest reliability across sessions. No group is added to force a nonempty $\mathcal{E}_{\mathrm{rel}}$. After selection, full-scan attributes are computed under the fixed rules for every fitting, validation, calibration, and test subject; a zero interaction vector remains a zero attribute on the shared hyperedge and does not alter the topology.

D Backbone and Interaction-Branch Implementation

D.1 Reference Channels and Pairwise Backbone

Let $\mathbf{R}_s \in \mathbb{R}^{N\times N}$ be the static FC matrix from the preprocessing pipeline. From the standardized signals define the edge co-fluctuation $E_{s,ij}(t) = Z_{s,i}(t)Z_{s,j}(t)$ and the dynamic pairwise summaries

$$V_{s,ij} = \mathrm{Std}_t\big(E_{s,ij}(t)\big), \quad U_{s,ij} = Q_{0.9}\big(|E_{s,ij}(t)|\big), \quad L_{s,ij} = \mathrm{Corr}_{t\geq 2}\big(E_{s,ij}(t), E_{s,ij}(t-1)\big), \tag{32}$$

where $Q_{0.9}$ is the 0.9 quantile over time. The pairwise inputs are $\mathbf{S}^{(2,0)}_s = \{\mathbf{R}_s\}$ (static) and $\mathbf{S}^{(2,1)}_s = \{\mathbf{R}_s, \mathbf{V}_s, \mathbf{U}_s, \mathbf{L}_s\}$ (static and dynamic); every channel is a function of one pair of ROIs at a time.

Default configuration. Unless otherwise stated, the reported HyPI experiments use $\mathbf{S}^{(2)}_s = \mathbf{S}^{(2,1)}_s = \{\mathbf{R}_s, \mathbf{V}_s, \mathbf{U}_s, \mathbf{L}_s\}$. The four channels comprise static Fisher-z functional connectivity and the three pairwise temporal summaries defined above. For each experimental run, the Reduced, Factual, capacity-matched Null, and Counterfactual conditions share the same frozen pairwise backbone checkpoint. The interaction-projection channels are not included in this default input.

As a representation control, an interaction-projection input $\mathbf{S}^{(2,2)}_s$ appends

$$\Pi^{\pm}_{s,ij} = \frac{1}{\max(c_{ij}, 1)} \sum_{g\in\mathcal{E}_{\mathrm{rel}},\, \{i,j\}\subset g} \mathrm{Rel}_g \left\|[\pm \mathbf{r}^{\mathrm{HO}}_{s,g}]_+\right\|_2, \qquad c_{ij} = \big|\{g \in \mathcal{E}_{\mathrm{rel}} : \{i,j\} \subset g\}\big|, \tag{33}$$

which folds the discovered interactions into matrices and evaluates the form in which higher-order information is presented rather than adding pairwise information.

For an ordered channel collection $\mathcal{C}$, the token of ROI i is $\mathbf{x}_{s,i} = \mathbf{W}_{\mathrm{in}}\big[\|_{\mathbf{C}\in\mathcal{C}}\mathbf{C}_{s,i,:}\big] + \mathbf{e}_i$, where $\mathbf{e}_i$ is an atlas-specific ROI embedding and $\mathbf{W}_{\mathrm{in}}$ accommodates the channel count. A Brain Network Transformer Kan et al. (2022) with its subject readout produces $\mathbf{H}^{(2)}_s \in \mathbb{R}^{N\times d}$ and $\boldsymbol{\ell}^{(2)}_s \in \mathbb{R}^C$. Backbones are fitted separately for each outer fold and input, trained with cross-entropy on $\mathcal{D}_{\mathrm{tr}}$, selected on $\mathcal{D}_{\mathrm{val}}$, and thereafter kept in evaluation mode with fixed parameters; their ROI representations and logits are cached per fold and input.

D.2 Network Definitions and Parameter Sharing

Each of ϕ, ρ, and ψ is a two-layer network $f(\mathbf{x}) = \mathbf{W}_2\,\mathrm{GELU}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2$ with its own parameters, shared across subjects, groups, and receiving ROIs. These networks may carry biases because their outputs enter only the multiplicative gate of Eq. (11). The projections $\mathbf{P}_3 \in \mathbb{R}^{d\times 1}$ and $\mathbf{P}_4 \in \mathbb{R}^{d\times 5}$, the output map $\mathbf{W}_o$, the readout maps $\mathbf{W}_{\chi,1}, \mathbf{W}_{\chi,2}$, and $\mathbf{W}_{\mathrm{cls}}$ have no biases. The branch contains no ROI-specific or hyperedge-specific parameters; ROI identity enters only through the frozen backbone representations. Table 5 lists the shapes at width d with hidden width d_{h}.

D.3 Aggregation and Zero-Preserving Readout

For a batch of B_{batch} subjects and E_m retained groups of size m, the member-index array has shape $[E_m, m]$, gathered ROI features $[B_{\mathrm{batch}}, E_m, m, d]$, and interaction attributes $[B_{\mathrm{batch}}, E_m, d_m]$.

Table 5: Interaction-branch operators. Context networks have one hidden GELU layer of width d_{h}.

Operator	Dimensions	Bias	Shared across
ϕ	$d \to d_{\mathrm{h}} \to d$	yes	members, groups, subjects
ρ	$(2d+1) \to d_{\mathrm{h}} \to d$	yes	groups, subjects
ψ	$2d \to d_{\mathrm{h}} \to d$	yes	receivers, groups, subjects
$\mathbf{P}_3$	$1 \to d$	no	three-region groups
$\mathbf{P}_4$	$5 \to d$	no	four-region groups
$\mathbf{W}_o$	$d \to d$	no	all messages
χ_0	$d \to d_{\mathrm{h}} \to d$	no	all ROIs
$\mathbf{W}_{\mathrm{cls}}$	$d \to C$	no	all subjects

Centers, contexts, and projected interactions have shape $[B_{\mathrm{batch}}, E_m, d]$ and are broadcast along the member dimension to form messages of shape $[B_{\mathrm{batch}}, E_m, m, d]$, which are scattered by incidence into a node-correction tensor $[B_{\mathrm{batch}}, N, d]$ with the normalization $1/\sqrt{d_v m_g}$ applied on actual incidence pairs only. Both size buckets write into the same tensor; empty buckets are skipped. A single propagation block is used, after which χ_0, the ROI average, and $\mathbf{W}_{\mathrm{cls}}$ are applied once. Each group is stored once as a canonical tuple, so reordering the group list changes the result only by floating-point accumulation. Under zero interaction inputs the computation follows $\mathbf{r}^{\mathrm{HO}}_{s,g} = \mathbf{0}\ \forall g \Rightarrow \mathbf{m}^{\mathrm{HO}}_{s,g\to v} = \mathbf{0}\ \forall (g,v) \Rightarrow \Delta\mathbf{h}_{s,v} = \mathbf{0}\ \forall v \Rightarrow \Delta\boldsymbol{\ell}^{\mathrm{HO}}_s = \mathbf{0}$, using $\mathrm{GELU}(0) = 0$ and the bias-free readout; the implementation checks this identity, and the empty-hypergraph case, with the same frozen backbone outputs and normalization factors before and after the intervention.

D.4 Initialization, Freezing, Optimization, and Calibration

During interaction training, backbone parameters are excluded from optimization and the backbone stays in evaluation mode, fixing dropout and normalization statistics. The trainable parameters are Ω of Eq. (13); topology, reliabilities, and interaction vectors are fixed data. Only $\mathbf{W}_{\mathrm{cls}}$ is initialized to zero, so the initial predictor equals the pairwise predictor; the remaining weights use a standard nonzero initialization, so gradients reach the upstream modules as soon as $\mathbf{W}_{\mathrm{cls}}$ becomes nonzero. The objective of Eq. (13) is minimized at $\alpha = 1$ on $\mathcal{D}_{\mathrm{tr}}$ and the checkpoint is selected on $\mathcal{D}_{\mathrm{val}}$.

On $\mathcal{D}_{\mathrm{cal}}$, a scalar temperature is fitted to the combined logits and $\alpha \in [0,1]$ is selected on a fixed grid by minimizing the calibrated log-loss; both are then frozen. Test predictions use $\boldsymbol{\ell}_s = \boldsymbol{\ell}^{(2)}_s + \alpha\Delta\boldsymbol{\ell}^{\mathrm{HO}}_s$ with the fitted temperature. Candidate rankings, priors, and reliabilities are never refitted on validation, calibration, or test subjects.

E Certification Protocol Details

Task information and predictive risk. The task information carried by the profile beyond the pairwise representation is the conditional mutual information $I_{\mathrm{HO}} = I(Y; \mathbf{F}^{\mathrm{HO}} \mid \mathbf{S}^{(2)})$, which equals the reduction in Bayes log-loss from adding $\mathbf{F}^{\mathrm{HO}}$ to $\mathbf{S}^{(2)}$. Trained predictors need not attain the Bayes risk, so Section 3.5 measures held-out log-loss differences under the controls that (P2) and (P3) require.

E.1 Matched Reassignment

Within a pool of subjects $\mathcal{P}$, let $\mathbf{s}_s$ be the vector of upper-triangular entries of the channels in $\mathbf{S}^{(2)}_s$, and $d(s,s') = \|\mathbf{s}_s - \mathbf{s}_{s'}\|_2$. The matched reassignment is the minimum-cost derangement

$$\pi = \arg \min_{\pi \in \mathrm{Sym}(\mathcal{P}),\ \pi(s)\neq s\ \forall s} \sum_{s\in\mathcal{P}} d(s, \pi(s)), \tag{34}$$

solved as a linear assignment problem with the diagonal cost set to infinity. It uses only pairwise inputs and never labels.

Partition-specific matching. Within each outer fold, let $\mathcal{D}_{\mathrm{te}}$ denote the held-out test set. The matching problem above is solved separately for each pool $\mathcal{P} \in \{\mathcal{D}_{\mathrm{tr}}, \mathcal{D}_{\mathrm{val}}, \mathcal{D}_{\mathrm{cal}}, \mathcal{D}_{\mathrm{te}}\}$. The resulting

Table 6: Higher-order input $\mathbf{\Phi}_s$ replaced under the matched reassignment for each baseline.

Method	$\mathbf{\Phi}_s$	Reduced condition
MvHo-IB	Subject-specific higher-order views	Pairwise view only
HOI-Brain	Signed higher-order interaction channels	FC channel only
HGST	Subject-specific hyperedge features	Pairwise graph branch only
HyperDiag	Temporal–regional hyperedge inputs	Regional graph only
HA-STA	Hyperedge attention inputs	Pairwise spatio-temporal attention only

map $\pi_{\mathcal{P}}$ assigns each receiver to a different subject within the same pool; donor profiles never cross partition boundaries. For receiver $s \in \mathcal{P}$, only $\mathbf{F}_s^{\mathrm{HO}}$ is replaced by $\mathbf{F}_{\pi_{\mathcal{P}}(s)}^{\mathrm{HO}}$; its own pairwise input $\mathbf{S}_s^{(2)}$ and label y_s remain unchanged. The notation $\pi(s)$ in the main text refers to the map for the set containing s.

Capacity-matched Null. The Null interaction branch is trained independently while sharing the frozen pairwise backbone with Factual. Reassigned profiles are used for fitting on $\mathcal{D}_{\mathrm{tr}}$, checkpoint selection on $\mathcal{D}_{\mathrm{val}}$, calibration on $\mathcal{D}_{\mathrm{cal}}$, and evaluation on $\mathcal{D}_{\mathrm{te}}$. Its architecture, training budget, checkpoint-selection rule, and calibration procedure match those of Factual. The residual coefficient and temperature are selected or fitted using the Null model's reassigned calibration inputs and the receivers' labels. The reported $\mathcal{R}_{\mathrm{N}}$ is therefore evaluated using reassigned test profiles, not the original test subject–profile associations.

Same-checkpoint Counterfactual. A matched reassignment π' within $\mathcal{D}_{\mathrm{te}}$ replaces each test subject's interaction profile at the Factual checkpoint. No retraining, checkpoint selection, or recalibration is performed; the Factual model's fitted residual coefficient and temperature are retained.

Replacing $\mathbf{F}_s^{\mathrm{HO}}$ by $\mathbf{F}_{\pi(s)}^{\mathrm{HO}}$ substitutes, for every hyperedge, the interaction vector of one subject by that of another subject on the same hyperedge, so ROI identity, size, coordinate order, and the marginal distribution of profiles are preserved.

During a same-checkpoint intervention, the hyperedge set, d_v, reliabilities, pairwise inputs, backbone outputs, normalization factors, branch parameters, temperature, and α are all fixed; only the interaction vectors change. Under the interaction-projection input of Appendix D.1, the cached $\mathbf{\Pi}_s^{\pm}$ channels are part of the fixed backbone input, so the intervention still changes only the interaction-input path.

E.2 Bootstrap Confidence Intervals

Let $\ell_c(s)$ denote the calibrated log-loss of test subject s under condition $c \in \{\mathrm{R}, \mathrm{F}, \mathrm{N}, \mathrm{CF}\}$, averaged over the model seeds. The reported risks are $\mathcal{R}_c = \frac{1}{n_{\mathrm{test}}} \sum_s \ell_c(s)$. For $b = 1, \ldots, B$, a resample of n_{test} subjects is drawn with replacement, the same resample is used for all four conditions, and $A_{\mathrm{HO}}^{(b)}$ and $D_{\mathrm{HO}}^{(b)}$ are computed from the resampled means. The reported lower bound is the 2.5th percentile over b; pairing across conditions removes between-subject variance that is common to all conditions.

E.3 Interventions for Baseline Models

For a baseline whose predictor consumes a subject-specific higher-order input $\mathbf{\Phi}_s$ in addition to pairwise inputs, the four conditions are defined by the same rule as for HyPI: Reduced removes $\mathbf{\Phi}_s$ or replaces it by its pairwise-only counterpart as defined by the baseline's own architecture; Factual is the published model; Null retrains the model with $\mathbf{\Phi}_{\pi(s)}$ under the same budget; Counterfactual evaluates the Factual checkpoint with $\mathbf{\Phi}_{\pi'(s)}$. Pairwise inputs and, for the Counterfactual, all parameters are unchanged. Table 6 lists the higher-order input replaced for each baseline.

E.4 Numerical Certification Results on ABIDE (AAL116)

Table 7 provides numerical estimates and confidence intervals for the higher-order-use comparison in Figure 3(a). Following Appendix E.2, calibrated test-subject log-losses are averaged over model seeds

Table 7: Higher-order-use certification on ABIDE (AAL116). Entries are point estimates with 95% percentile bootstrap confidence intervals in brackets. All estimates and interval endpoints are expressed in 10^{-3} nats of mean log-loss. ✓ denotes that both interval lower endpoints are strictly positive; × denotes that the joint criterion is not met.

Method	A_{HO} [95% CI]	D_{HO} [95% CI]	**Certified**
HyPI	$\mathbf{5.48}$ $[\mathbf{1.25}, \mathbf{9.31}]$	$\mathbf{6.47}$ $[\mathbf{2.04}, \mathbf{10.62}]$	✓
MvHo-IB	1.48 $[-0.92, 4.25]$	4.52 $[1.45, 7.31]$	×
HA-STA	1.68 $[-0.62, 4.47]$	0.58 $[-2.35, 3.18]$	×
HyperDiag	0.25 $[-2.41, 3.32]$	-0.35 $[-3.81, 2.85]$	×
HGST	-0.82 $[-3.52, 2.37]$	-0.15 $[-3.42, 2.74]$	×
HOI-Brain	-0.12 $[-2.95, 2.68]$	-0.65 $[-3.82, 2.14]$	×

before paired subject-level resampling with $B = 10{,}000$ bootstrap replicates. Each replicate uses the same subject indices for all four conditions and recomputes A_{HO} and D_{HO} from the resampled risks, including the minimum over the two controls in A_{HO}. The reported 95% intervals are the 2.5th and 97.5th percentiles of the resulting bootstrap distributions. Certification follows Section 3.5 and requires both interval lower endpoints to exceed zero.

Results and scope. Among the six evaluated methods, only HyPI meets the joint certification criterion on ABIDE (AAL116). MvHo-IB has a strictly positive interval for D_{HO}, but its interval for A_{HO} includes zero, supporting functional dependence without establishing controlled predictive advantage. For HA-STA, HyperDiag, HGST, and HOI-Brain, both intervals include zero, so none meets the joint criterion under the evaluated protocol. Failure to meet the criterion is not evidence that a model has no higher-order information or no functional dependence. These method-wise intervals do not constitute pairwise tests of differences between methods, and the results are specific to ABIDE (AAL116).

E.5 Magnitude- and Configuration-Selective Interventions

This section specifies the fixed-model audit of Section 4.5. The reported analysis covers ABIDE and REST-meta-MDD under AAL116 and CC200; it does not establish configuration-specific dependence for ADNI or ADHD-200.

Fixed evaluation conditions. For each outer fold, fix a label-free, pairwise-matched test-set mapping π' constructed as in Appendix E.1. The three restricted interventions use the same mapping and evaluated subjects. The Factual checkpoint, pairwise inputs, retained hyperedges, coefficient ordering, reliability weights, degree normalization, residual coefficient, and fitted temperature remain fixed. No discovery, training, or calibration is repeated.

Magnitude, direction, and eligibility. For each subject–group entry, define

$$a_{s,g} = \|\mathbf{r}^{\mathrm{HO}}_{s,g}\|_2, \qquad \mathbf{v}_{s,g} = \mathcal{N}_0(\mathbf{r}^{\mathrm{HO}}_{s,g}), \qquad \mathcal{I}_s = \{g \in \mathcal{E}_{\mathrm{rel}} : a_{s,g} > 0,\ a_{\pi'(s),g} > 0\}.$$

Thus $\mathbf{r}^{\mathrm{HO}}_{s,g} = a_{s,g}\mathbf{v}_{s,g}$, with unit direction on nonzero entries. The same eligible set $\mathcal{I}_s$ is used for all three restricted interventions. For $g \in \mathcal{I}_s$, set

$$\begin{aligned}
\widetilde{\mathbf{r}}^{\mathrm{mag}}_{s,g} &= a_{\pi'(s),g}\mathbf{v}_{s,g},\\
\widetilde{\mathbf{r}}^{\mathrm{cfg}}_{s,g} &= a_{s,g}\mathbf{v}_{\pi'(s),g},\\
\widetilde{\mathbf{r}}^{\mathrm{joint}}_{s,g} &= a_{\pi'(s),g}\mathbf{v}_{\pi'(s),g}.
\end{aligned}$$

For $g \notin \mathcal{I}_s$, all three inputs remain $\mathbf{r}^{\mathrm{HO}}_{s,g}$. Every restricted intervention therefore preserves the receiver's zero/nonzero pattern, and configuration reassignment additionally preserves every vector norm. For triplets, $d_3 = 1$ and a nonzero direction is -1 or $+1$; configuration reassignment changes an input only when donor and receiver signs differ. By Lemma 1, these signs indicate an all-active probability below or above the subject's pairwise reference. Quadruplet directions lie in $\mathbb{R}^5$ and combine the four third-order and one fourth-order coefficients from the same four-region distribution.

Risk estimation. For $k \in \{\mathrm{mag}, \mathrm{cfg}, \mathrm{joint}\}$, evaluate the calibrated mean log-loss $\mathcal{R}_k$ and report

$$\Delta_k = \mathcal{R}_k - \mathcal{R}_{\mathrm{F}}.$$

Following Appendix E.2, model-seed losses are averaged within each test subject before paired subject resampling with $B = 10{,}000$ replicates. Within each replicate, use the same subject indices for Factual and all three interventions and recompute the risk differences. The 2.5th and 97.5th percentiles define pointwise 95% intervals, conditional on the fitted models, partitions, and realized mapping; they do not include retraining or rematching variability. They are neither simultaneous intervals over all settings nor pairwise tests between interventions.

Relationship to the full Counterfactual. The full replacement uses $\widetilde{\mathbf{r}}^{\mathrm{CF}}_{s,g} = \mathbf{r}^{\mathrm{HO}}_{\pi'(s),g}$ for every retained group. Under an identical mapping, it differs from restricted joint reassignment only on

$$\mathcal{Z}_s = \left\{ g \in \mathcal{E}_{\mathrm{rel}} : \mathbb{I}[a_{s,g} > 0] \neq \mathbb{I}[a_{\pi'(s),g} > 0] \right\}.$$

For ABIDE (AAL116), Table 2 reports restricted joint reassignment as $5.45\ [1.56,\ 9.22] \times 10^{-3}$ nats, whereas Appendix E.4 reports the full Counterfactual as $6.47\ [2.04,\ 10.62] \times 10^{-3}$ nats. A direct paired contrast additionally requires the same checkpoints, subjects, calibration, and donor mapping in the two evaluations. Under these conditions,

$$D_{\mathrm{HO}} - \Delta_{\mathrm{joint}} = \mathcal{R}_{\mathrm{CF}} - \mathcal{R}_{\mathrm{joint}}$$

is the risk change from extending replacement to zero-status-mismatched entries. It is not a context-independent or additive contribution of the zero pattern: those entries also introduce or remove vector magnitudes and directions, and the predictor is nonlinear. Neither $D_{\mathrm{HO}} \geq \Delta_{\mathrm{joint}}$ nor $\Delta_{\mathrm{joint}} = \Delta_{\mathrm{mag}} + \Delta_{\mathrm{cfg}}$ is required.

Interpretation and limitations. The positive configuration intervals in Table 2 support (P3)-type dependence under a norm- and zero-pattern-preserving intervention. They do not establish (P2) advantage over a separately trained magnitude-only model or replace the joint certification criterion. Combining receiver magnitudes with donor directions can create sparsely supported inputs, a limitation of permutation-based model audits (Hooker et al., 2021). Pairwise matching does not guarantee preservation of the magnitude–direction distribution. Holding norms fixed also does not exclude nuisance information in directions or establish robustness to binarization, motion, scan length, or site effects. The pooled results do not establish separate effects for triplets and quadruplets.

F Configuration of the Reported Experiments

Table 8 lists the quantities that the method leaves free and their values in the reported experiments. The released configuration additionally records the selected ROI tuples, retained hyperedge counts by size, proposal block size, numerical precision, runtime, and software versions, together with the architectural checks for zero interaction inputs, empty hypergraphs, mixed-size aggregation, coordinate ordering, and preservation of the frozen backbone state.

We distinguish the feature-calibration budget M_{cal} from the group-confirmation budget M_{conf}. The former is used to estimate the temporal-null mean and standard deviation for each subject and group, with separate length-matched calibration for full scans and scan halves; the latter is used for group-level Monte Carlo testing on $\mathcal{D}_{\mathrm{conf}}$. The Factual and capacity-matched Null branches use the same branch optimization, checkpoint-selection, and calibration procedures; the Counterfactual condition involves no retraining.

G Datasets

G.1 Cohorts and Diagnostic Tasks

We evaluate HyPI on four resting-state fMRI cohorts: ABIDE (Di Martino et al., 2014), REST-meta-MDD (Yan et al., 2019), ADNI (Jack Jr et al., 2015), and ADHD-200 (ADHD-200 Consortium, 2012). Table 9 summarizes the final analysis cohorts used in the reported experiments. All four cohorts are evaluated separately under AAL116 and CC200. The binary tasks are ASD versus HC, MDD versus HC, MCI versus NC, and ADHD versus HC, respectively, yielding eight cohort–parcellation–task settings.

ABIDE. The Autism Brain Imaging Data Exchange (ABIDE) is a multisite resting-state fMRI resource for autism spectrum disorder (ASD) research (Di Martino et al., 2014). The final analysis

Table 8: Configuration of the reported experiments. Surrogate budgets are per subject at the indicated stage.

Quantity	Symbol	Value
Outer cross-validation folds	–	5
Model seeds	–	3
Group sizes	$\mathcal{M}$	$\{3, 4\}$
Proposal / confirmation split of $\mathcal{D}_{\mathrm{tr}}$	–	30% / 70%
Candidate budget per size	K_3, K_4	256, 256
Proposal pseudocount per state	λ_{prop}	0.5
Population-prior pseudocount per state	λ_0	0.3
Total subject-level shrinkage weight	τ_g	2^{m_g} (8 or 16)
Numerical constant	ϵ	10^{-6}
Feature-calibration surrogates	M_{cal}	500 per full scan or half
Group-confirmation surrogates	M_{conf}	20,000 per confirmation subject
MaxEnt moment-residual tolerance	δ^{fit}	10^{-8}
MaxEnt arithmetic	–	`float64`
MaxEnt optimizer	–	Newton with backtracking line search
Maximum Newton iterations	–	100
Evidence threshold	q_0	0.03
Reliability threshold	ρ_0	0.5 (AAL116), 0.7 (CC200)
Representation and hidden widths	d, d_{h}	64, 128
Pairwise-backbone optimizer	–	Adam
Pairwise-backbone learning-rate schedule	–	Cosine: $10^{-4} \to 10^{-5}$
Pairwise-backbone weight decay	–	10^{-4}
Pairwise-backbone epochs / batch size	–	200 / 16
Pairwise-backbone checkpoint selection	–	Selected on $\mathcal{D}_{\mathrm{val}}$
Interaction-branch optimizer	–	AdamW
Interaction-branch learning-rate schedule	–	Initial 10^{-3}; cosine decay
Interaction-branch weight decay	–	10^{-2}
Interaction-branch epoch cap / batch size	–	100 / 16
Interaction-branch checkpoint selection	–	Minimum log-loss on $\mathcal{D}_{\mathrm{val}}$
Residual-coefficient grid	α	$\{0, 0.1, \ldots, 1\}$ (11 values)
Bootstrap resamples	B	10,000

Table 9: Demographic statistics of the cohorts. Age is reported in years as mean ± standard deviation. HC and NC denote healthy control and normal control, respectively. MCI includes EMCI, LMCI, and MCI. ADHD includes all diagnostic subtypes. ADHD-200 subjects with missing diagnosis, sex, or age are excluded.

Dataset	Group	Subjects	Male	Female	Age (years)
ABIDE	ASD	539	474	65	17.01 ± 8.37
	HC	573	474	99	17.08 ± 7.72
	Total	**1112**	**948**	**164**	$\mathbf{17.05 \pm 8.04}$
REST-meta-MDD	MDD	1300	474	826	36.23 ± 14.62
	HC	1128	487	641	36.25 ± 15.58
	Total	**2428**	**961**	**1467**	$\mathbf{36.24 \pm 15.07}$
ADNI	MCI	412	220	192	72.75 ± 7.56
	NC	413	166	247	72.79 ± 7.88
	Total	**825**	**386**	**439**	$\mathbf{72.77 \pm 7.72}$
ADHD-200	ADHD	279	220	59	11.65 ± 2.98
	HC	487	257	230	12.25 ± 3.30
	Total	**766**	**477**	**289**	$\mathbf{12.04 \pm 3.20}$

cohort contains 1,112 subjects, including 539 individuals with ASD and 573 healthy controls (HC). We formulate ASD-versus-HC classification separately under the AAL116 and CC200 parcellations.

REST-meta-MDD. The REST-meta-MDD consortium aggregates multisite resting-state fMRI data for major depressive disorder (MDD) research (Yan et al., 2019). The final analysis cohort contains 2,428 subjects, including 1,300 individuals with MDD and 1,128 HC. We formulate MDD-versus-HC classification separately under the AAL116 and CC200 parcellations.

ADNI. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal, multicenter resource for studying Alzheimer's disease and related cognitive impairment (Jack Jr et al., 2015). For the present study, the analysis cohort is restricted to 825 subjects, comprising 412 with mild cognitive impairment (MCI) and 413 normal controls (NC). Subjects labeled as early MCI (EMCI), late MCI (LMCI), or MCI in the source metadata are pooled into the MCI class. We evaluate only MCI-versus-NC classification, separately under AAL116 and CC200; subjects labeled as AD are not included.

ADHD-200. For ADHD-200 (ADHD-200 Consortium, 2012), subjects with missing diagnosis, sex, or age are excluded before defining the analysis cohort. All retained ADHD diagnostic subtypes are pooled into a single ADHD class. The final analysis cohort contains 766 subjects, including 279 with ADHD and 487 healthy controls (HC). We evaluate ADHD-versus-HC classification separately under AAL116 and CC200.

G.2 Image Preprocessing and ROI Time Series

Image-level preprocessing follows the dataset-specific pipelines described below. Within each dataset and parcellation, all compared methods use the same preprocessed ROI time series and the same subject partitions.

ABIDE and ADNI. ABIDE and ADNI are processed using the same SPM-based resting-state fMRI pipeline. The pipeline consists of slice-timing correction, rigid-body realignment, spatial normalization to Montreal Neurological Institute (MNI) space, and spatial smoothing with a 6 mm full-width-at-half-maximum Gaussian kernel. Temporal processing applies band-pass filtering at 0.01–0.08 Hz and nuisance regression of head-motion parameters, white-matter signals, and cerebrospinal-fluid signals.

For both ABIDE and ADNI, regional mean BOLD time series are extracted separately under AAL116 and CC200.

REST-meta-MDD. For REST-meta-MDD, we use the consortium-provided preprocessed resting-state fMRI data (Yan et al., 2019). The consortium workflow discards the first ten volumes and includes slice-timing correction, motion correction, spatial normalization to MNI space, temporal filtering at 0.01–0.1 Hz, and nuisance regression of head-motion parameters, global signal, white-matter signal, cerebrospinal-fluid signal, and linear trends. Regional mean BOLD time series are obtained under both AAL116 and CC200. The different temporal-filtering bands used for ABIDE/ADNI and REST-meta-MDD follow their respective preprocessing workflows and are shared by all compared methods within each dataset.

ADHD-200 preprocessing and ROI time series. For ADHD-200, we use preprocessed resting-state fMRI derivatives and evaluate AAL116 and CC200 ROI time series separately. The atlas-specific ROI time series subsequently undergo the common ROI-level processing and functional-connectivity estimation procedures in Appendices G.3 and G.4. Within each parcellation, all compared methods use the same derivative inputs, retained ROI sequences, and subject partitions.

Parcellation. AAL116 follows the Automated Anatomical Labeling atlas (Tzourio-Mazoyer et al., 2002) and contains 116 regions, comprising 90 cerebral and 26 cerebellar ROIs. CC200 follows the functional parcellation of Craddock et al. (2012) and contains 200 ROIs. For each atlas, ROI ordering is fixed across subjects and is shared by interaction estimation, pairwise representations, model training, and all intervention conditions. For preprocessed 4D fMRI volumes, we average the voxelwise BOLD signal within each atlas ROI at every time point to obtain an ROI time-series matrix $\mathbf{X}_s \in \mathbb{R}^{T_s \times N}$; when atlas-specific ROI time-series derivatives are provided, we use them directly.

G.3 Common ROI-Level Processing

For each subject, any time point containing a non-finite value in at least one ROI is removed jointly from all ROI signals. Thus, all ROIs of a subject retain exactly the same time indices. Let the resulting sequence be

$$\mathbf{X}_s \in \mathbb{R}^{T_s \times N},$$

where T_s is the number of retained time points and $N = 116$ or 200 is determined by the parcellation. The retained sequence is used consistently for functional-connectivity estimation, binary-state interaction estimation, temporal summaries, and temporal-surrogate construction.

Each retained ROI signal is standardized within subject:

$$Z_{s,i}(t) = \frac{X_{s,i}(t) - \mu_{s,i}}{\sigma_{s,i} + \epsilon},$$

consistent with Appendix B.1. The standardized signals are retained for both the pairwise and higher-order representations.

G.4 Functional Connectivity and Pairwise Inputs

For static functional connectivity, Pearson correlation is computed between every pair of ROI time series over the jointly retained time points. The correlation coefficients are Fisher-z transformed:

$$C_{s,ij} = \mathrm{Corr}_t \left(X_{s,i}(t), X_{s,j}(t) \right),$$

$$(\mathbf{R}_s)_{ij} = \begin{cases} \dfrac{1}{2} \log \dfrac{1 + C_{s,ij}}{1 - C_{s,ij}}, & i \neq j, \\ 0, & i = j. \end{cases}$$

The resulting matrix

$$\mathbf{R}_s \in \mathbb{R}^{N \times N}$$

is the static pairwise FC channel used by the pairwise backbone.

No fixed-density graph sparsification is imposed by HyPI. In particular, HyPI does not construct a training-site-derived propagation support or split positive and negative FC weights into separate propagation branches. The pairwise backbone uses the dense pairwise representations defined in Appendix D.1.

The higher-order interaction estimator operates directly on the standardized ROI time series rather than on the FC matrix. Binary activity states are formed as

$$B_{s,i}(t) = \mathbb{I}[Z_{s,i}(t) > 0],$$

and are subsequently used to estimate the joint-state distributions of candidate regional groups. The same standardized ROI sequences are also used to construct the pairwise temporal summaries in Appendix D.1 and the Fourier phase-randomized temporal surrogates in Appendix B.4.

G.5 Subject-Wise Evaluation and Discovery Separation

All reported experiments use subject-wise five-fold cross-validation rather than leave-one-site-out evaluation. The same outer-fold subject assignments are used for all compared methods within each dataset–parcellation–task setting.

Within each outer fold, test subjects are held out before the remaining subjects are partitioned into the fitting, model-selection, and calibration sets $\mathcal{D}_{\mathrm{tr}}$, $\mathcal{D}_{\mathrm{val}}$, and $\mathcal{D}_{\mathrm{cal}}$. The fitting set is further divided into subject-disjoint proposal and confirmation subsets,

$$\mathcal{D}_{\mathrm{tr}} = \mathcal{D}_{\mathrm{prop}} \dot{\cup} \mathcal{D}_{\mathrm{conf}},$$

using the $30\%/70\%$ allocation reported in Appendix F.

Population priors and candidate rankings are estimated exclusively from $\mathcal{D}_{\mathrm{prop}}$. Group-level temporal-null evidence and split-half reliability are then evaluated on $\mathcal{D}_{\mathrm{conf}}$ without diagnostic labels. After confirmation, the retained hyperedge set, coefficient coordinate ordering, population priors, and reliability weights are fixed and are not re-estimated on $\mathcal{D}_{\mathrm{val}}$, $\mathcal{D}_{\mathrm{cal}}$, or test subjects.

Subject-specific interaction attributes are nevertheless estimated separately for every subject under these fixed discovery rules. For each fitting, validation, calibration, or test subject, the observed scan is used to compute its own joint-state distributions, maximum-entropy references, interaction coefficients, and calibrated interaction vectors. Temporal-null calibration is therefore subject-specific and uses surrogate scans generated from that subject's retained ROI sequence at the corresponding observation length. This subject-specific feature computation does not update the discovered hypergraph topology, population prior, or reliability statistics.

Each full scan and each scan half uses $M_{\mathrm{cal}} = 500$ length-matched surrogate realizations for feature-level temporal calibration. Group-level evidence on $\mathcal{D}_{\mathrm{conf}}$ uses $M_{\mathrm{conf}} = 20{,}000$ full-scan surrogate realizations per confirmation subject, with each realization shared across that subject's candidate groups, as specified in Appendix F.

H Baseline Details

Common Evaluation Protocol. We reproduce the official implementations whenever available and otherwise follow the architectures and training objectives described in the original papers. All methods use the same subject-wise data partitions and the same preprocessed BOLD time series or functional connectivity matrices. The input and output dimensions are adjusted to the AAL116 and CC200 parcellations without modifying the core model design. All representation construction, normalization, pretraining, and model selection procedures are performed within each training fold, and hyperparameters are selected using the validation subset only.

H.1 Pairwise Brain Network Models

BrainNetCNN Kawahara et al. (2017). BrainNetCNN operates directly on dense brain connectivity matrices using connectome-specific convolutional operators. Its edge-to-edge filters aggregate connections sharing a common ROI, followed by edge-to-node and node-to-graph transformations that progressively convert the connectivity matrix into a subject-level representation. In our experiments, each subject's functional connectivity matrix is provided as the network input, and the final prediction layer is configured for the corresponding binary classification task.

BrainGNN Li et al. (2021). BrainGNN represents ROIs as graph nodes and models their functional relationships using ROI-aware graph convolution. Its Ra-GConv layers account for both regional

identity and graph topology, while ROI-selection pooling retains task-relevant regions and supports region-level interpretation. Following the original model, we retain the unit, top-K pooling, and group-level consistency regularizers. The same subject-specific functional connectivity matrices and atlas definitions are used across all datasets.

BNT Kan et al. (2022). The Brain Network Transformer uses each ROI's connectivity profile as its node representation, providing both functional information and implicit positional identity. Transformer layers learn pairwise dependencies among ROIs through self-attention, after which the Orthonormal Clustering Readout groups ROIs into latent functional modules and produces a subject-level embedding. We retain the original Transformer and clustering-readout architecture and adjust only the number of input ROIs and output classes for each experimental setting.

STAGIN Kim et al. (2021). STAGIN models the temporal evolution of brain connectivity rather than a single static connectome. Sliding-window functional connectivity matrices are represented as a sequence of brain graphs, which are encoded using graph-based spatial aggregation and a Transformer-based temporal module. Spatial readout attention summarizes informative ROIs within each window, whereas temporal attention integrates dynamic patterns across windows. We construct all dynamic connectivity sequences from the same preprocessed BOLD signals and select the window configuration using the validation subset.

H.2 Higher-Order Interaction Methods

HyBRiD Qiu et al. (2024). HyBRiD learns high-order relationships under a maximally informative and minimally redundant objective. Its Constructor parameterizes a collection of hyperedges without enumerating the exponential space of regional subsets, while its Weighter estimates the contribution of each learned hyperedge. The resulting relationships are optimized using the multi-head drop-bottleneck information bottleneck objective. Following the standardized implementation used in the MvHo-IB comparison, we preserve the learnable Constructor and Weighter and encode the resulting hypergraph using the common HyperAtten backbone. The final prediction head is configured for binary disease classification.

HOC Santoro et al. (2024). We use HOC as shorthand for the Higher-Order Connectomics pipeline of Santoro et al. (2024). The method first standardizes regional BOLD signals and constructs instantaneous interaction signals by multiplying the signals of regional groups. These interaction weights define a time-varying weighted simplicial complex, from which local violating triangles and homological scaffolds are extracted through topological filtration and persistent homology. Since the original framework was designed for task decoding, fingerprinting, and brain–behavior analysis, we aggregate its local triangle and scaffold features into subject-level representations and train the common binary classification head under our data partitions.

MvHo-IB Zhang et al. (2025). MvHo-IB combines pairwise connectivity and explicit third-order interactions within a multi-view information bottleneck framework. The pairwise view is constructed using matrix-based Rényi mutual information and encoded by a Graph Isomorphism Network. The higher-order view is represented by a third-order $\mathcal{O}$-information tensor over regional triplets and processed using the purpose-built Brain3DCNN encoder. The two representations are jointly optimized with a multi-view information bottleneck objective that preserves task-relevant information while compressing view-specific redundancy. We use the original two-view formulation and retain its view-specific encoders and loss terms.

HOI-Brain Zhao et al. (2026). HOI-Brain extracts signed higher-order interactions from BOLD signals using the Multiplication of Temporal Derivatives measure. Positive and negative interaction patterns are represented as separate monotonic weighted simplicial complexes, whose multiscale topological structures are summarized through persistent homology. A multi-channel Transformer subsequently encodes and integrates the heterogeneous topological features for subject-level diagnosis. We follow the original signed-interaction construction and multi-channel fusion procedure, with atlas-dependent input dimensions adjusted to each dataset.

H.3 Brain Hypergraph Models

HGFD Jin et al. (2025). HGFD constructs a subject-specific brain hypergraph from functional connectivity profiles. Each ROI is represented by its row in the Pearson correlation matrix and is grouped with its most feature-similar ROIs to form a K-nearest-neighbor hyperedge. A two-layer BrainHGNN serves as the teacher model and transfers its high-order representation to a lightweight MLP student through knowledge distillation. In the full HGFD protocol, acquisition centers are treated as federated clients, and only the student parameters are aggregated using FedAvg. We reproduce the complete federated procedure when valid client partitions are available. For centralized comparisons, we evaluate only the BrainHGNN teacher and report it as *HGFD*.

HGST Han et al. (2026). HGST first constructs subject-level hypergraphs from ROI time series using the sparse-representation procedure provided by the official implementation. It then performs hypergraph self-supervised pretraining through two complementary tasks. The semantic-aware module restores masked node features through hypergraph link prediction, whereas the topology-aware module encourages representations of structurally related hyperedges to remain similar. A shared pretrained hypergraph neural network is subsequently fine-tuned for disease classification. All hypergraph construction and self-supervised pretraining are performed using the training subjects of each fold only.

HyperDiag Ma et al. (2026). HyperDiag jointly models temporal evolution and regional functional heterogeneity through a dual-level hypergraph architecture. Its temporally-evolving branch captures high-order dependencies within individual time windows and propagates information across consecutive windows. A complementary region-wise branch constructs functional hypergraphs within regional partitions to characterize spatially heterogeneous interactions. The temporal and regional representations are integrated through a topology-enhanced selective state-space propagation network before subject-level prediction. We follow the original temporal segmentation, regional grouping, and dual-branch fusion procedure.

HA-STA Hu et al. (2026). HA-STA learns both high-order brain structures and their temporal dynamics directly from fMRI signals. A multi-hyperedge binary mask module learns sparse regional memberships for multiple hyperedges, while a hypergraph self-attention aggregation module performs adaptive node-to-hyperedge and hyperedge-to-node information exchange. The resulting spatial features are combined with a spatio-temporal low-dimensional network for disease classification. Its information-bottleneck-inspired objective encourages the learned structures to retain predictive information while suppressing redundancy. We preserve the original mask-learning, hypergraph attention, temporal modeling, and regularization components.

I Additional Interpretability Analysis

This section provides the complete anatomical definitions and neuroscientific context of the regional groups visualized in Figure 6. The analysis concerns ABIDE under the AAL116 parcellation and covers three displayed groups from each of four methods: HGST, HyperDiag, HA-STA, and HyPI. We first document their exact ROI memberships and then relate their anatomical organization to previous ASD studies.

I.1 Regional Definitions and Visualization

Table 10 lists the ROI indices and atlas labels of all displayed groups. Anatomical names follow the AAL nomenclature, with `L` and `R` indicating the left and right hemispheres, respectively. The numerical indices reproduce the ROI ordering used in the main-text figure. Group identifiers g_1, g_2, and g_3 are method-specific, and $|g|$ denotes the number of constituent ROIs. Each brain rendering displays the members of one regional group; colors identify anatomical regions rather than interaction strength, direction, or predictive importance.

Table 10: Complete membership of the regional groups displayed in Figure 6. Atlas labels are listed in the same order as the ROI indices. Group identifiers are local to each method.

Method	Group	$\|g\|$	ROI indices	AAL labels
HGST	g_1	3	97, 98, 99	Cerebellum_4_5_L, Cerebellum_4_5_R, Cerebellum_6_L
	g_2	4	73, 74, 75, 76	Putamen_L, Putamen_R, Pallidum_L, Pallidum_R
	g_3	3	83, 84, 87	Temporal_Pole_Sup_L, Temporal_Pole_Sup_R, Temporal_Pole_Mid_L
HyperDiag	g_1	4	67, 68, 35, 36	Precuneus_L, Precuneus_R, Cingulum_Post_L, Cingulum_Post_R
	g_2	3	49, 50, 51	Occipital_Sup_L, Occipital_Sup_R, Occipital_Mid_L
	g_3	3	1, 2, 19	Precentral_L, Precentral_R, Supp_Motor_Area_L
HA-STA	g_1	6	23, 24, 7, 8, 81, 82	Frontal_Sup_Medial_L, Frontal_Sup_Medial_R, Frontal_Mid_L, Frontal_Mid_R, Temporal_Sup_L, Temporal_Sup_R
	g_2	4	77, 78, 71, 72	Thalamus_L, Thalamus_R, Caudate_L, Caudate_R
	g_3	5	43, 44, 47, 48, 1	Calcarine_L, Calcarine_R, Lingual_L, Lingual_R, Precentral_L
HyPI	g_1	3	23, 36, 68	Frontal_Sup_Medial_L, Cingulum_Post_R, Precuneus_R
	g_2	3	67, 68, 23	Precuneus_L, Precuneus_R, Frontal_Sup_Medial_L
	g_3	4	44, 80, 2, 8	Calcarine_R, Heschl_R, Precentral_R, Frontal_Mid_R

I.2 Neuroscientific Context

I.2.1 HyPI

Group g_1: medial frontal–posterior midline organization. The first group consists of the left medial superior frontal gyrus, right posterior cingulate cortex, and right precuneus. It combines a medial frontal region with posterior midline regions commonly examined in default-mode connectivity studies. Monk et al. (2009) reported weaker posterior cingulate–superior frontal connectivity in ASD, with weaker connectivity associated with poorer social functioning. Assaf et al. (2010) additionally identified reduced connectivity involving the precuneus and medial prefrontal/anterior cingulate regions within default-mode subnetworks. Our group overlaps with this anterior–posterior anatomical organization, providing a regional basis for relating the identified interaction to previously reported default-mode dysconnectivity.

Group g_2: bilateral precuneus with a medial frontal region. The second group contains the bilateral precuneus and left medial superior frontal gyrus. Assaf et al. (2010) found that connectivity involving precuneus and medial frontal default-mode regions was associated with the severity of social and communication difficulties in ASD. This provides context for the regional composition of g_2. Within our displayed groups, g_1 and g_2 share the left medial superior frontal gyrus and right precuneus but differ in their third member: the right posterior cingulate in g_1 and the left precuneus in g_2. They therefore represent two distinct combinations within an overlapping medial frontal–posterior midline anatomical setting.

Group g_3: right-hemisphere sensory, motor, and frontal organization. The third group consists of the right calcarine region, Heschl's gyrus, precentral gyrus, and middle frontal gyrus. Its membership spans visual- and auditory-associated anatomy, motor cortex, and a frontal association region. Nebel et al. (2016) reported increased intrinsic asynchrony between visual and motor systems in children with ASD, replicated the finding in an independent ABIDE sample, and related visual–motor

synchrony to social traits and imitation performance. Their component-based analysis particularly implicated lateral occipital visual and upper-limb motor components, providing systems-level context for the visual–motor aspect of our group. Holiga et al. (2019) further identified reproducible ASD-associated hypoconnectivity centered on sensory-motor regions and hyperconnectivity hubs in prefrontal and parietal cortices across four cohorts. Together, these studies motivate examining coordinated activity across the sensory, motor, and frontal regions represented in g_3.

I.2.2 HGST

Group g_1: cerebellar organization. This group contains bilateral cerebellar lobules IV–V and left lobule VI. Its membership is confined to cerebellar regions. Khan et al. (2015) reported atypical cerebro-cerebellar connectivity in children and adolescents with ASD, including increased connectivity for sensorimotor pathways and reduced connectivity for supramodal pathways in their domain-specific analyses. These findings establish the relevance of cerebellar organization to ASD and provide a systems-level context for the regions selected by HGST. The present group specifically represents relationships among cerebellar parcels rather than the cerebro-cerebellar connections evaluated in that study.

Group g_2: putamen–pallidum organization. The second group combines the bilateral putamen and pallidum, forming a subcortical basal-ganglia configuration. Di Martino et al. (2011) examined caudate- and putamen-seeded connectivity in children with ASD and reported increased connectivity in atypical striatal circuits involving associative and limbic regions. Their findings provide direct background for the striatal component of this group. The inclusion of the pallidum characterizes the broader basal-ganglia membership of the displayed HGST group.

Group g_3: anterior temporal organization. The third group comprises bilateral superior temporal-pole regions and the left middle temporal-pole region. It is therefore concentrated in anterior temporal anatomy. Gotts et al. (2012) identified altered organization of social brain circuits in ASD, including anterior temporal regions and reduced coupling between limbic-related and other social-processing circuits. This offers a relevant context for the temporal-pole membership of the HGST example.

I.2.3 HyperDiag

Group g_1: bilateral posterior midline organization. This group contains the bilateral precuneus and posterior cingulate cortex. These regions overlap with the posterior midline anatomy discussed in default-mode studies of ASD Monk et al. (2009); Assaf et al. (2010). Compared with the first two groups of HyPI, the displayed HyperDiag group concentrates on bilateral posterior regions, whereas both of our groups include the left medial superior frontal gyrus. The comparison therefore reveals different regional compositions within a partly shared anatomical context.

Group g_2: occipital organization. The second group combines bilateral superior occipital gyri with the left middle occipital gyrus. All three members are occipital parcels, giving this example a visual-system-oriented anatomical composition. Studies of visual–motor synchrony in ASD have implicated occipital visual components Nebel et al. (2016). The displayed HyperDiag group focuses on relationships among occipital regions, in contrast to the broader sensory–motor–frontal membership of our third group.

Group g_3: motor cortical organization. The third group comprises bilateral precentral gyri and the left supplementary motor area. Mostofsky et al. (2009) studied finger-tapping performance in children with autism and reported reduced connectivity across the motor execution network, together with greater supplementary motor area activation and reduced anterior cerebellar activation relative to controls. These task-fMRI findings provide complementary evidence for the relevance of the motor regions present in the HyperDiag group.

I.2.4 HA-STA

Group g_1: bilateral frontal–temporal organization. The first group contains bilateral medial superior frontal, middle frontal, and superior temporal regions. It spans frontal and temporal anatomy and is the largest displayed group, with six ROIs. Gotts et al. (2012) reported reduced integration

between affective and language/sensorimotor components of social brain circuitry in ASD. This provides a distributed-circuit context for considering frontal and temporal regions together. The HA-STA example differs from the more posterior-midline-focused first two groups of HyPI by including bilateral lateral frontal and superior temporal parcels.

Group g_2: thalamic–caudate organization. The second group combines the bilateral thalamus and caudate. Nair et al. (2013) reported abnormalities of anatomical and functional thalamocortical connectivity in ASD, while Di Martino et al. (2011) identified atypical caudate- and putamen-seeded functional circuits. These studies provide complementary background for the thalamic and striatal members of this group. Since the displayed group contains only subcortical regions, its anatomical description is thalamic–caudate rather than a complete cortico-striato-thalamic circuit.

Group g_3: occipital–motor organization. The third group contains bilateral calcarine and lingual regions together with the left precentral gyrus. It combines occipital visual-associated parcels with a motor cortical region. The visual–motor connectivity findings of Nebel et al. (2016) provide a systems-level context for this composition. Both this HA-STA example and the third group of HyPI include visual and motor anatomy, while our displayed group additionally contains right Heschl's and middle frontal regions.