Refining RDoC Using Individual-Level Task fMRI Factor Models Reveals Reproducible Brain-wide Motifs

S.K.L. Quah[1], S. Madsen[1], S. Pirzada[1], B. Jo[1], L.Q. Uddin[2], J.A. Mumford[3], D.M. Barch[4], I.H. Gotlib[3], D.A. Fair[5], R.A. Poldrack[3,6], M. Saggar[1,6]

[1]Department of Psychiatry & Behavioral Sciences, Stanford University, Stanford, CA, USA
[2]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA, USA
[3]Department of Psychology, Stanford University, Stanford, CA, USA
[4]Departments of Psychological & Brain Sciences, Psychiatry, and Radiology, Washington University in St. Louis, St Louis, MO, USA
[5]Department of Pediatrics, University of Minnesota Medical School, Minneapolis, MN, USA
[6]Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA

Corresponding author: Manish Saggar (saggar@stanford.edu).

Abstract

The Research Domain Criteria (RDoC) framework was introduced to guide psychiatric research using biologically grounded, dimensional constructs of mental function. However, its current hierarchical domain structure remains largely unvalidated against individual-level brain imaging data. Building on our prior group-level work showing that data-driven bifactor models outperform RDoC-based models, we applied a multi-stage validation framework to Human Connectome Project (HCP) task fMRI data to test whether individual-level, empirically derived models more accurately reflect the intrinsic organization and behavioral relevance of brain activity. Using confirmatory factor analysis in two independent cohorts, we found that individual-level, data-driven bifactor models consistently outperformed RDoC-based models across multiple fit indices in both training and validation sets. The general factor derived from these models revealed a reproducible macroscale gradient spanning visual–attentional to auditory–default mode networks, aligning with canonical resting-state gradients and supporting its interpretation as a domain-general axis of large-scale brain organization. Applying community detection to subject-specific factor representations revealed four spatial motifs whose centroids corresponded to interpretable brain systems and were robustly reproduced across cohorts. Similarity to these centroids predicted individual task performance in working memory and relational reasoning, as measured by both raw accuracy and latent performance factors. To further assess organizational validity, we applied Mapper—a topological data analysis method—to contrast maps, generating unsupervised graph representations of task-evoked brain activity. Mapper graphs annotated with data-driven centroids showed greater modularity than those annotated with RDoC domains, suggesting that the data-driven framework better captures the topological structure of individual functional brain states. Together, these findings demonstrate that individual-level, data-driven factor models provide a more accurate, interpretable, and behaviorally relevant account of brain organization than the current RDoC framework. By modeling inter-individual variability directly from neuroimaging data, this approach advances precision neuroscience and supports the empirical refinement of dimensional psychiatric frameworks such as RDoC.

1. Introduction

The Research Domain Criteria (RDoC) framework[1,2] was developed by the National Institute of Mental Health (NIMH) to advance precision psychiatry by grounding mental health constructs in neurobiological systems. RDoC seeks to reconceptualize mental disorders based on dimensional psychological constructs and their associated neurobiological systems, moving beyond traditional categorical diagnoses[3,4]. Over the past decade, RDoC has provided a guiding ontology to organize psychiatric research across multiple levels of analysis—from genes to behavior—structured into domains and constructs reflecting core functional systems, including negative valence, positive valence, cognitive, social, arousal/regulatory, and sensorimotor processes.

Despite its utility, recent critiques argue that the RDoC framework lacks empirical specificity, particularly in its mapping of domains to underlying neural circuits (Beam et al., 2021). For example, a large meta-analysis of emotion regulation tasks found that the associated neural circuitry overlapped substantially across all RDoC domains, suggesting limited domain-level specificity in patterns of brain activation[5]. Beyond the emotion (or negative valence systems) domain, other domains—such as cognitive systems—may also be overly broad, encompassing heterogeneous processes that activate diverse functional brain systems[6,7]. Moreover, RDoC's top-down hierarchical structure, based on expert consensus, has yet to be comprehensively validated against large-scale meta-analytic or empirical neuroimaging data. As Cuthbert[8] notes, the RDoC working group at NIMH views the constructs as heuristic starting points—not fixed categories—and emphasizes that empirical refinement is a necessary step in the framework's evolution.

To address these limitations, researchers have increasingly turned to data-driven approaches that extract latent dimensions directly from neuroimaging and behavioral data—without relying on predefined domain boundaries[6,7,9]. These bottom-up methods offer more flexible modeling of functional brain organization and may reveal neurobiologically grounded factors that explain individual variability better than the RDoC framework. However, most existing approaches operate at the group level, leaving open questions about whether such models capture subject-specific patterns of functional brain activation and generalize across datasets or behavioral phenotypes. In a prior study[7], we used group-level exploratory and confirmatory factor analysis to compare RDoC-based models with data-driven models of whole-brain task-based fMRI activation. The data-driven models consistently demonstrated superior fit compared to models constrained by RDoC-defined assignments, and this performance advantage was replicated in an independent validation sample. Yet, because those analyses were limited to group-averaged data, it remains unclear whether the same structure holds at the level of individual participants, where idiosyncratic variation may be more pronounced.

Validating RDoC dimensions at the level of individual brain function is critical for advancing the framework's translational utility. Precision neuroscience seeks to tailor cognitive and clinical insights to the person by modeling inter-individual variability in brain function. Although RDoC was designed to support such personalized approaches, its domain definitions have not been validated against the fine-grained neural patterns evident at the single-subject level. Leveraging a comprehensive RDoC-aligned task-fMRI dataset, we evaluate whether individual-level, data-driven representations can offer biologically grounded alternatives to fixed cognitive taxonomies.

Here, we test whether data-driven factor models derived from individual-level task-fMRI data outperform RDoC-based models in explaining functional brain organization (**Fig. 1**). We apply this approach to the Human Connectome Project (HCP), which includes high-resolution task-fMRI data from seven tasks spanning five core RDoC domains. Using confirmatory factor analysis (CFA), we compared the fit of four model types: data-driven bifactor, data-driven specific factor, RDoC bifactor, and RDoC specific factor models. Data-driven models were constructed by first identifying factors through individual-level exploratory factor analysis (EFA), followed by CFA using either bifactor or specific-factor structures. RDoC-based models were defined based on task-to-domain assignments, with both bifactor and specific-factor variants estimated to enable direct comparison of models with and without a general factor. To evaluate generalizability, we divided the dataset into two independent cohorts: Cohort 1 (N = 412) and Cohort 2 (N = 329), with no related participants within cohorts. Factor models were derived in Cohort 1 and validated in Cohort 2—and vice versa—using a principal component–based projection approach.

To better understand the structure underlying the data-driven models, we first examined the spatial pattern of the general factor recovered from the data-driven bifactor model. Building on this, we applied two independent unsupervised approaches to the specific factors: community detection and a topological data analysis (TDA)-based Mapper method. In the community detection approach, we used modularity maximization on individual-level factor representations to identify shared population-wide centroids—putative latent factors or recurrent configurations of functional brain activity. Based on prior evidence that similarity between individual activation patterns and canonical network configurations is associated with task performance[10], we hypothesized that individuals whose task-evoked activation patterns more closely aligned with these centroids would show better cognitive task performance. To further validate the centroids, we compared them with previously published group-level latent factors derived from group-averaged task-fMRI maps in independent datasets[7].

We next applied a TDA-based Mapper approach[11,12] to visualize the global organization of task-related brain activation patterns across individuals. Mapper is a powerful topological method that produces simplified graph representations of high-dimensional data while preserving its intrinsic shape and connectivity. This allows us to assess whether latent structures identified by data-driven models or RDoC domains align with the overall topology of functional brain states. To compare these frameworks, we annotated Mapper graphs using either RDoC task domains or the data-driven community centroids described above. We hypothesized that data-driven annotations would exhibit stronger correspondence with the topological structure of individual-level brain activity, providing further evidence that individual-level, data-driven models offer a more biologically grounded account of functional brain organization.

This multi-stage validation framework enables us to assess both the internal coherence and external relevance of data-driven factor models compared to RDoC-based models. Our findings provide new insights into the dimensional organization of task-evoked brain activity and suggest that individual-level, data-driven models can inform empirically grounded refinements to the theoretical structure of the RDoC framework. More broadly, our study builds on prior work showing that activation patterns from diverse tasks reveal low-dimensional, continuous axes that

transcend traditional psychological constructs—supporting a shift toward biologically constrained cognitive taxonomies[13,14].

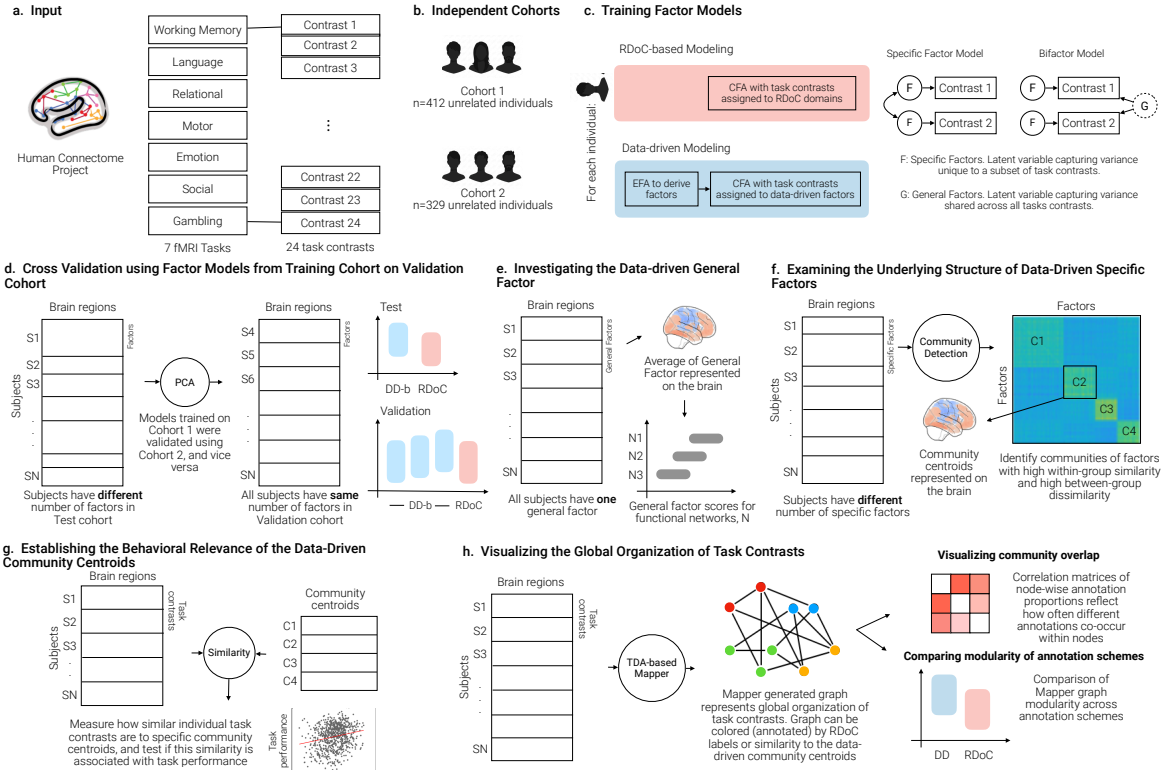

Figure 1. Overview of approach. Using the Human Connectome Project (HCP) task-based fMRI data, we examined whether data-driven factor models derived from individual-level task-fMRI data outperform RDoC-based models in explaining functional brain organization. (a) Preprocessed and parcellated contrast maps from seven HCP tasks were used as input. (b) Data from 741 unrelated subjects were split into two cohorts (Cohort 1: N = 412, and Cohort 2: N = 329). (c) Four CFA model types were trained: RDoC-specific, RDoC-bifactor, data-driven specific, and data-driven bifactor. (d) Cross-validation across the two cohorts was conducted such that the models trained on one cohort were validated on held-out data from another cohort. Principal components analysis (PCA) was applied to data-driven specific factors from the data-driven bifactor model to derive the same number of factors across all individuals (detailed in Methods and Supplementary Figure 4). (e) Mean general factor scores from the data-driven bifactor models were mapped onto the brain. Reproducibility across cohorts, correspondence with resting-state gradients[15], and network-wise distributions were evaluated. (f) To examine the underlying structure of the data-driven specific factors, an unsupervised community detection algorithm was used to derive communities. (g) Associations between similarity of individual task activation maps to specific community centroids and behavioral performance were estimated with linear mixed-effects models. Similarity was entered as a fixed effect, with Parental ID included as a random intercept to control for genetic relatedness. (h) Finally, we used the TDA-based Mapper approach to visualize the global organization of task contrasts and examined whether RDoC or data-driven groupings (annotation) better capture this structure.

2. Results

We analyzed task-based fMRI data from the HCP dataset of 741 unrelated healthy adults, divided into Cohort 1 (N=412) and Cohort 2 (N=329). Each subject completed seven tasks spanning five core RDoC domains. The whole-brain contrast maps were parcellated into 347 brain regions. For each subject, we treated the 24 task-contrast maps as variables and the 347 parcellated brain regions as observations, yielding a 347 × 24 data matrix on which the factor analyses were performed. We first constructed individual-level RDoC-based factor models, defined by expert task–domain assignments, and data-driven factor models, derived from exploratory and confirmatory factor analysis. To examine shared variance across all task contrasts, we analyzed the spatial organization of the general factor recovered from the data-driven bifactor model. We then applied community detection to the individual-specific factor score maps to identify reproducible spatial motifs and evaluated their functional relevance by testing whether similarity between these motifs and individual activation patterns predicted task performance. We further characterized the global structure of inter-individual variation in task contrasts using topological data analysis.

2.1 Data-driven bifactor models outperform RDoC models in both training and validation cohorts

To evaluate whether individual-level data-driven models better capture the structure of task-evoked brain activity than RDoC-based models, we applied CFA to each subject's activation maps, treating contrast maps as variables and brain parcels as observations. Data-driven models were constructed using EFA followed by bifactor CFA, while RDoC models used fixed domain assignments. Model fit was assessed using four standard indices: AIC, RMSEA, CFI, and TLI.

In Cohort 1, data-driven bifactor models demonstrated significantly superior model fit compared with RDoC-based models across all evaluated indices. For RMSEA and AIC, data-driven models showed markedly better fit (RMSEA: $M = .159$, $SD = .014$ vs. $.207$, $SD = .012$; $t(411) = –75.10$, $p < .001$, $d = –3.70$; AIC: $M = 16{,}275.1$, $SD = 1040.3$ vs. $17{,}898.8$, $SD = 1141.6$; $t(411) = –79.05$, $p < .001$, $d = –3.89$) (Figure 2). Similar advantages were observed for the CFI (Data-driven: $M = .789$, $SD = .032$; RDoC: $M = .613$, $SD = .060$; $t(411) = 66.66$, $p < .001$; $d = 3.28$) and TLI (Data-driven: $M = .743$, $SD = .039$; RDoC: $M = .562$, $SD = .068$; $t(411) = 60.54$, $p < .001$; $d = 2.98$) (Supplementary Figure 1).

Swapping the training and validation cohorts yielded the same pattern: data-driven bifactor models consistently outperformed RDoC-based models across all indices. This replication supports the robustness of the data-driven bifactor solution across independent datasets.

To test generalizability more directly, we projected principal components from the data-driven bifactor model in Cohort 1 onto Cohort 2 and repeated the procedure in reverse. These cross-cohort validations focused on comparing the data-driven bifactor model with the canonical RDoC-specific model, allowing us to isolate the contribution of the general factor. In both directions and across all tested component numbers, the data-driven bifactor model outperformed the RDoC model on all fit indices ($p < .001$; Supplementary Tables 1 & 2).

To formally compare all four model types—(i) RDoC-specific, (ii) RDoC bifactor, (iii) data-driven specific, and (iv) data-driven bifactor—we conducted repeated-measures ANOVAs. Mauchly's test revealed violations of sphericity for all fit indices ($p < .001$), so Greenhouse–Geisser corrections were applied. Post hoc Tukey-adjusted comparisons confirmed that the data-driven bifactor model consistently yielded the best fit (all $^{**}p < .001$). Full results, including detailed statistics and pairwise comparisons, are reported in Supplementary Figure 1 and Supplementary Table 3.

Together, these findings demonstrate that data-driven bifactor models provide a significantly better and more generalizable account of task-evoked brain activity than expert-defined RDoC models.

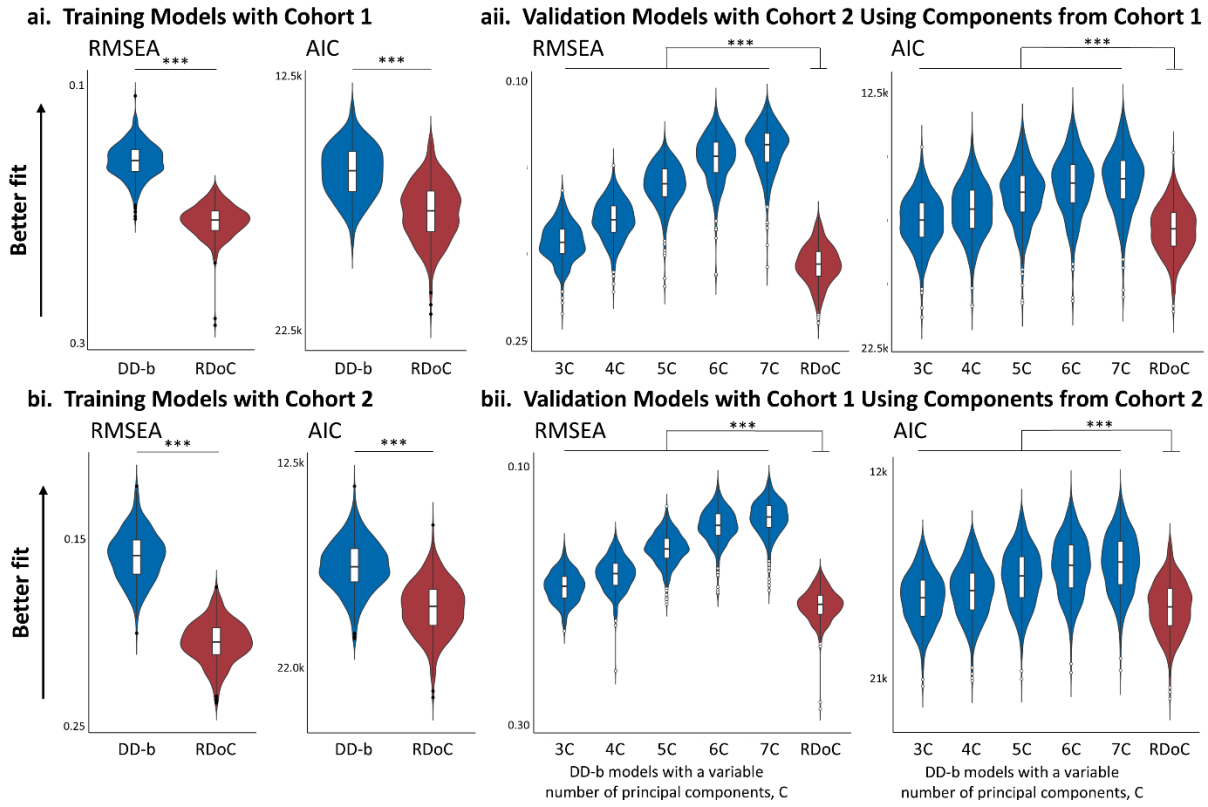

Figure 2. Data-driven bifactor models consistently showed superior model fit compared to RDoC-based models across training and validation comparisons. Root Mean Square Error of Approximation (RMSEA) and Akaike Information Criterion (AIC) were used to assess model performance. (ai, bi) Within-cohort model fit comparison for data-driven bifactor (DD-b) and RDoC models trained using Cohort 1 and Cohort 2, respectively. (aii) Models derived from Cohort 1 and validated in Cohort 2 using principal component-based projection. (bii) Models derived from Cohort 2 and validated in Cohort 1. For validation, data-driven bifactor models were tested with 3 to 7 principal components (3C–7C). Violin plots show distributions of fit indices across individuals. Data-driven bifactor models consistently outperformed RDoC-based models across all comparisons. $^{***}p < .001$.

2.2 Data-driven General Factor Reflects a Gradient from Visual–Attentional to Auditory–Default Mode Networks

We next examined the spatial topography of the general factor recovered from the data-driven bifactor models in both cohorts. Whole-brain general factor maps showed reproducible activation patterns, with prominent positive loadings in the visual and parietal cortices (Figure 3a). This spatial organization resembled the visual-to-somatomotor-to-auditory axis—also known as Gradient 2—previously described in resting-state connectivity gradients by Margulies et al.[15] (Figure 3b). These gradients describe dominant modes of functional connectivity variation in the cortex.

To quantitatively assess convergence with canonical cortical hierarchies, we computed parcel-wise correlations between the general factor maps and the first five resting-state gradients reported by Margulies et al.[15]. In both cohorts, the general factor showed its strongest positive association with Gradient 2 and a moderate positive association with Gradient 3. Weaker negative associations were observed with Gradient 4, while correlations with Gradients 1 and 5 were not significant (Figure 3c; adjusted p-values in Supplementary Table 4).

To further contextualize the general factor within canonical functional networks, we examined its distribution across the Gordon[16] network parcellation. The factor exhibited a graded pattern: highest positive scores were observed in visual and dorsal attention networks, while lower or negative scores were found in the default mode, auditory, and somatomotor networks (Figure 3d–e).

Together, these results demonstrate that the general factor—capturing variance shared across all task contrasts—exhibits stable spatial organization across cohorts and aligns with established macroscale connectivity gradients. These findings confirm that task-derived general factors align with established macroscale cortical hierarchies, providing a stable and interpretable axis of organization that links task-evoked activity to intrinsic connectivity structure.

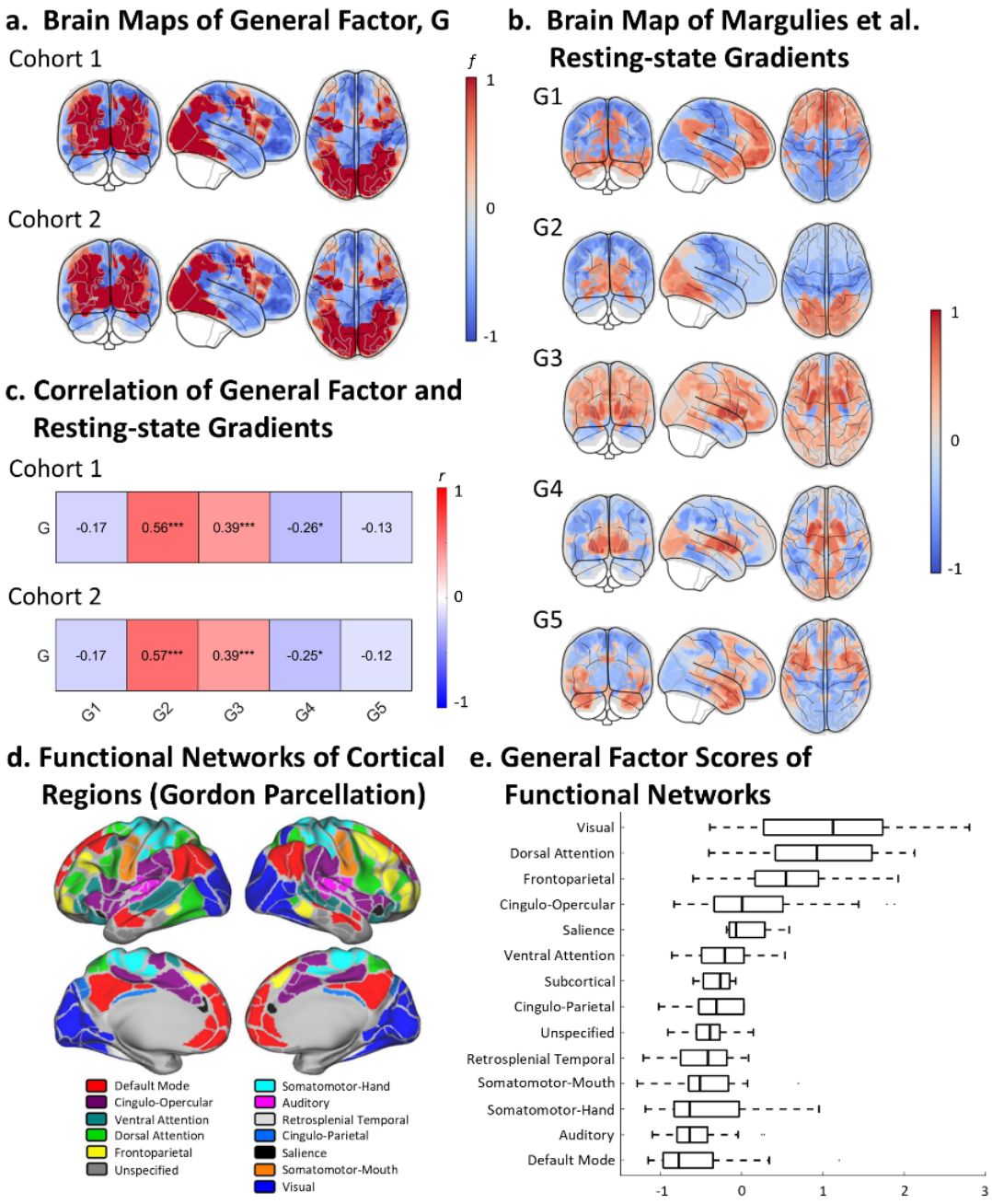

Figure 3. The general factor shows reproducible topography, aligns with resting-state gradients, and maps onto a visual–dorsal attention to auditory–default mode axis. (a) Whole-brain spatial maps of the general factor from the data-driven bifactor models in both cohorts, with reproducible organization across cohorts. (b) Resting-state gradients 1–5 from Margulies et al.[15] are shown for comparison. (c) Correlations between general factor maps and the resting-state gradients. The general factor was most strongly associated with gradient 2 in both cohorts, followed by a moderate association with gradient 3. Weaker negative correlations were observed with gradient 4. P-values were corrected for spatial autocorrelation. (d) Functional networks of cortical regions based on the Gordon parcellation. Figure adapted from Vazquez-Trejo et al.[17]. (e) Distribution of general factor scores across regions of different functional networks. Networks have been ordered by their median factor scores. Activation follows a gradient from the highest positive values in visual and dorsal attention networks to the lowest/most negative values in somatomotor, auditory, and default mode networks. *$p < .05$, ***$p < .001$.

2.3 Reproducible Brain-Wide Communities in Individual Factor Representations

Community detection applied to individual-specific factor scores from both cohorts replicated topographically distinct communities (Fig. 4a, C1–C4). The modal partition across 1,000 Louvain iterations revealed four communities in Cohort 1 and five in Cohort 2. However, the final consensus solution contained five communities in both cohorts, indicating that a frequently occurring community (>50% of runs) in Cohort 1 was further subdivided in the consensus.

Stability of the recovered community structure was high in both cohorts, with adjusted Rand index (ARI) values indicating strong agreement between individual runs and the consensus (Cohort 1: ARI = 0.893, SD = 0.127; Cohort 2: ARI = 0.956, SD = 0.095). These spatial patterns were consistent across both training and validation cohorts, with high reproducibility (Pearson correlation coefficient, r = .5-.99) between matched centroids of Cohort 1 and Cohort 2 (Fig. 4b). Each community centroid exhibited a characteristic spatial topography (Fig. 4c): C1 was predominantly prefrontal and occipital activation, C2 engaged somatomotor and subcortical regions, C3 included the parietal cortex and fusiform gyri, and C4 was centered in temporal and subcortical structures.

To evaluate external validity, we compared these data-driven centroids with group-level factors from an independent dataset based on group-averaged task-fMRI maps from Neurovault[7]. The centroids showed significant and differential alignment with both data-driven and RDoC-specific factor maps from that study (Fig. 5), supporting the interpretability and robustness of the identified community structures. Notably, centroids in both cohorts aligned more strongly (both positively and negatively) with RDoC-defined factors related to the sensorimotor systems and positive valence systems, and less so with those for social processes, cognitive systems, or negative valence systems (Figure 5aii, bii). This suggests that the former domains are associated with more stereotyped and robust functional architectures in the HCP tasks, making them more detectable in both individual- and group-level analyses.

To further evaluate convergence with other established functional brain topographies, we correlated task-derived community centroids with the gradients of resting-state fMRI reported by Margulies et al.[15]. In both cohorts, the centroids showed significant associations with the first four gradients. Specifically, centroid 1 in Cohort 1 and 1a in Cohort 2 are positively associated with the principal gradient, gradient 1. Additionally, centroid 1 in Cohort 1 is positively associated with gradient 2. Centroid 1b in Cohort 2 is positively associated with gradients 2 and 4 and negatively associated with gradient 3. Centroid 2 in both cohorts is positively associated with gradient 3 and negatively associated with gradients 1 and 2. Finally, centroids 3 and 4 in both cohorts are negatively and positively associated with gradient 4, respectively. All p-values were corrected for multiple comparisons using the Benjamini–Hochberg false discovery rate (FDR) procedure (adjusted values in Supplementary Table 5). These results support the interpretation of task-derived communities as canonical axes of functional brain organization, grounded in both task and resting-state systems.

Finally, we assessed within-subject consistency of community assignments by computing Shannon entropy of factor map distributions across communities. In both the four-community solution from Cohort 1 (mean = 1.83, median = 1.92; maximum possible = 2.00) and the five-

community solution from Cohort 2 (mean = 2.13, median = 2.24; maximum possible = 2.32), entropy values were close to the theoretical maximum. This indicates that most participants exhibited diverse factor–community mappings, with maps distributed across multiple modules rather than concentrated in a single one.

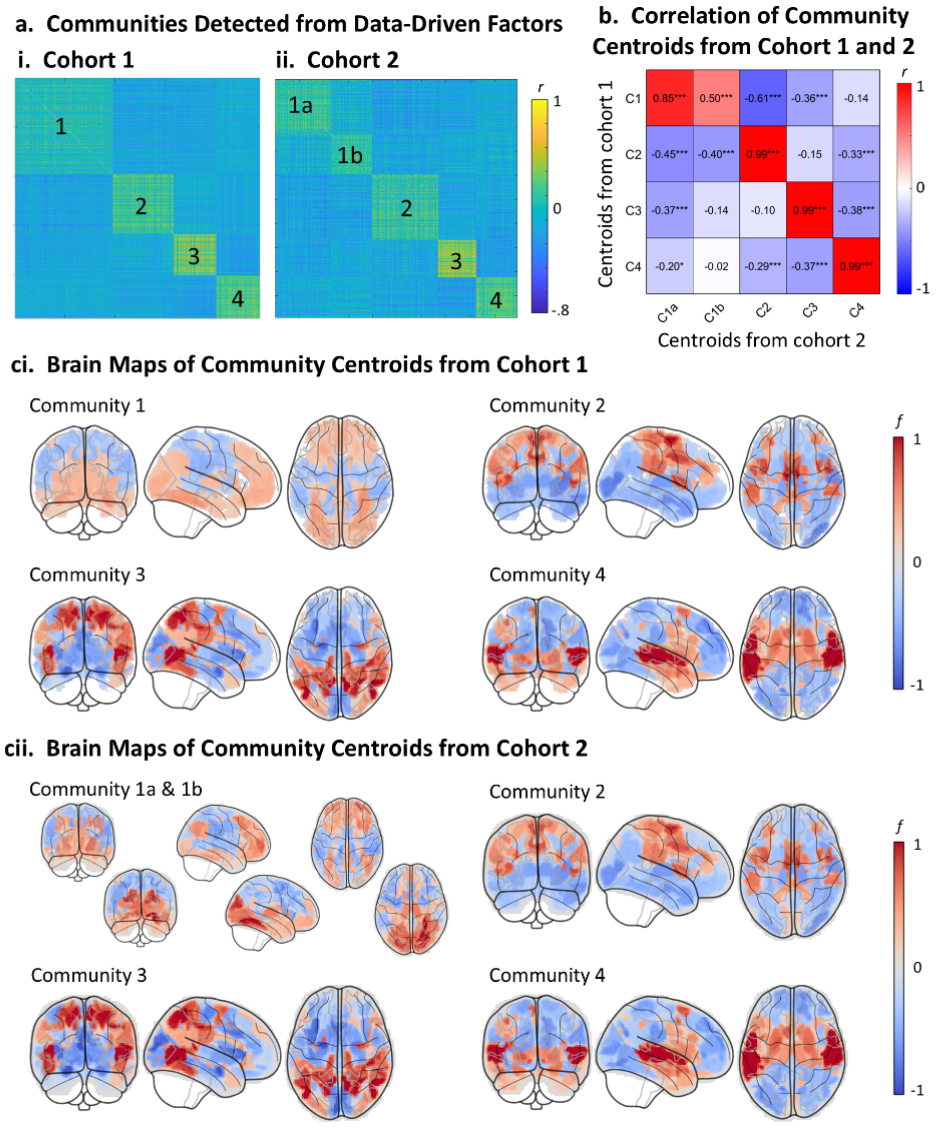

Figure 4. Community detection using data-driven factors reveals reproducible and functionally distinct factor communities across cohorts. (ai) Consensus clustering of correlation matrices from data-driven factor scores in Cohort 1 identified four stable communities. (aii) Community detection in Cohort 2 yielded a similar five-community structure, with Community 1 from Cohort 1 corresponding to two communities (1a, 1b). (b) Correlations between community centroid maps derived from Cohort 1 and Cohort 2. Centroids from both cohorts exhibit strong cross-cohort correspondence, particularly for C2, C3 and C4 ($r > 0.99$). C1-4: Community 1-4. (ci, cii) Brain maps of the resulting community centroids from Cohort 1 and Cohort 2, respectively. Each centroid reflects the mean spatial topography of factor maps assigned to that community.

Consistent patterns of activation were observed across cohorts, indicating stable and reproducible community structure. Warm colors denote positive loadings; cool colors denote negative loadings.

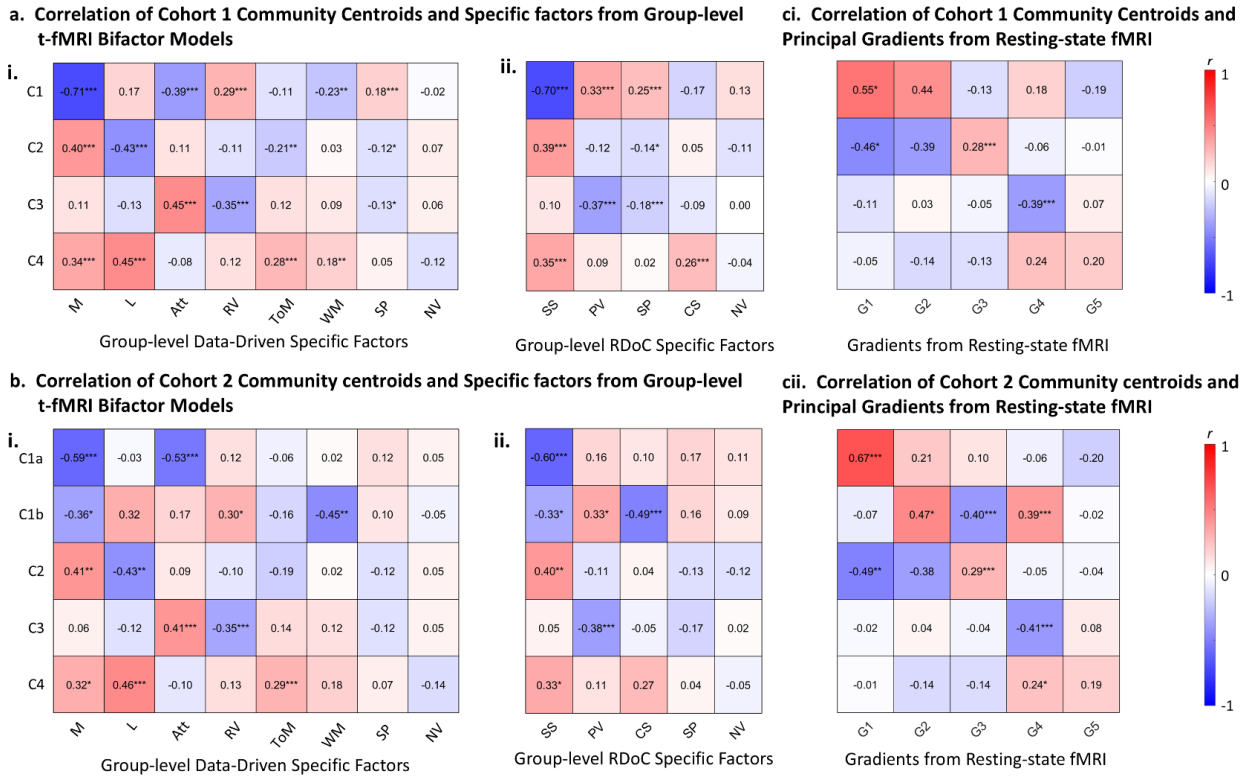

Figure 5. Community centroids show high cross-cohort correspondence and align with specific factors from group-level bifactor models. (a) Correlation of community centroids from Cohort 1 with group-level bifactor model factors from Quah et al.[7]. (ai) Correlations with data-driven specific factors. (aii) Correlations with RDoC-based specific factors. (b) Correlation of community centroids from Cohort 2 with group-level bifactor model factors from Quah et al.[7]. (bi) Correlations with data-driven specific factors. (bii) Correlations with RDoC-based specific factors. Columns for (a) and (b) have been ordered from highest total correlation coefficient to lowest. (ci–cii) Correlations between task-derived community centroids and the resting-state connectivity gradients from Margulies et al.[15]. P-values have been adjusted for spatial autocorrelation[18]. Additionally, p-values here were corrected for multiple comparisons using the Benjamini–Hochberg false discovery rate (FDR) procedure. C1-4: Community 1-4; M: Motor; L: Language; Att: Attention; ToM: Theory-of-Mind; RV: Reward Valuation; WM: Working Memory; SP: Social Processing; NV: Negative Valence. G1-5: Gradient 1-5. *$p_{adj} < .05$, **$p_{adj} < .01$; ***$p_{adj} < .001$.

2.4 Community Alignment Predicts Individual Cognitive Performance

We next tested whether individual alignment with community-level factor representations was associated with behavioral performance on working memory and relational reasoning tasks. These tasks were selected because they were the only ones in the HCP dataset with well-quantified and sufficiently variable performance measures (see Supplementary Methods 1). Other tasks lacked usable metrics, showed ceiling effects, or exhibited low variability, limiting their suitability for individual-level correlation analyses.

To maximize statistical power, this analysis included all participants from the HCP dataset (N = 962). For each subject, we computed two similarity indices between their average task activation map and a community centroid: (i) Pearson correlation and (ii) continuous Dice similarity. We selected centroids from Community 4 (C4) and Community 3 (C3) in Cohort 1, as these showed the strongest correlations with group-level factors related to working memory and attention, respectively (Figure 5bi). Associations between similarity and behavior were estimated using linear mixed-effects models, with Parental ID included as a random intercept to control for genetic relatedness.

Higher alignment to community centroids was significantly associated with better behavioral performance. Specifically, similarity to the C4 centroid was positively associated with accuracy on the working memory (β = 0.115, SE = 0.030, t(941) = 3.81, p*** <.001; Dice: β = 0.174, SE = 0.032, t(957) = 5.68, p***< .001). Similarly, similarity to the C3 centroid was positively associated with relational reasoning accuracy (β = 0.088, SE = 0.032, t(957) = 2.72, p** = .007). The corresponding Dice-based association was positive but marginal (β = 0.065, SE = 0.033, t(953) = 1.95, p = .051) (Figure 6ai, bi; full Dice results in Supplementary Figure 3).

We next examined whether these similarity measures also predicted latent cognitive ability, derived using EFA from both accuracy and reaction time metrics. For working memory, the extracted factor loaded negatively on accuracy and positively on reaction time, indicating that higher scores reflected slower and less accurate performance. To aid interpretation, we inverted these scores so that higher values corresponded to better performance (i.e., faster and more accurate responses). For relational processing, both accuracy and reaction time loaded positively, meaning higher scores reflected greater accuracy with longer response times—a speed–accuracy tradeoff factor.

Similarity to the C4 and C3 centroids significantly predicted these latent performance factor scores (Figure 6aii, bii). Specifically, similarity to the C4 centroid was positively associated with working memory factor scores (Pearson's: β = 0.083, SE = 0.032, t(946) = 2.63, p** = .009; Dice: β = 0.131, SE = 0.032, t(942) = 4.06, p*** < .001), and similarity to the C3 centroid was positively associated with relational processing factor scores (Pearson's: β = 0.198, SE = 0.032, t(956) = 6.14, p*** < .001; Dice: β = 0.190, SE = 0.033, t(952) = 5.73, p*** < .001).

These findings confirm that individual similarity to shared spatial motifs derived from data-driven communities is not only associated with raw task accuracy but also reflects generalizable latent cognitive abilities across tasks. Dice-based results are shown in Supplementary Figure 3.

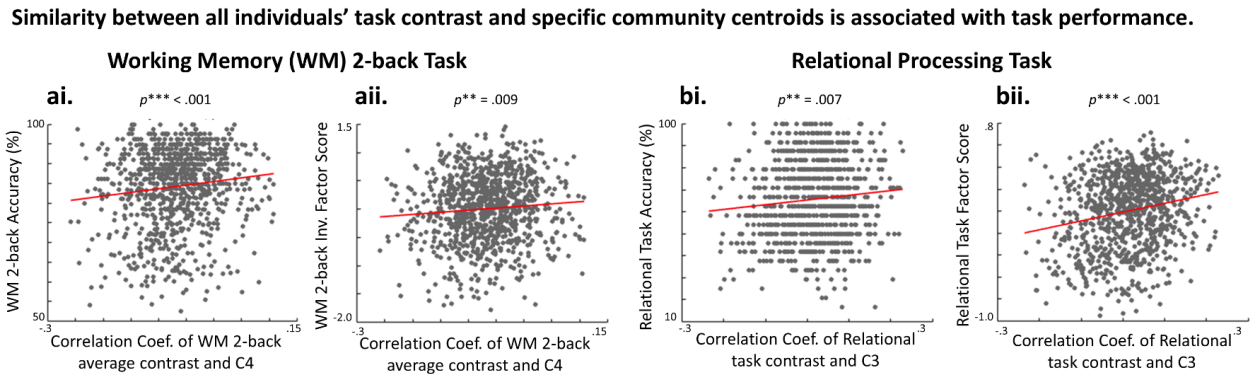

Figure 6. Similarity between individuals' task contrast and specific community centroids predicts behavioral performance in mixed-effects models. Similarity (quantified by Pearson correlation) was entered as a fixed effect, with Parental ID included as a random intercept to control for genetic relatedness. Higher similarity between an individual's average task map and community centroid was associated with better performance on the working memory and relational processing tasks. This association was observed both in raw task accuracy (panels ai–ii) and in task performance factor scores derived using EFA (panels bi–ii). Subpanels (i) show predictions for raw accuracy; subpanels (ii) show predictions for latent performance factor scores derived via EFA of accuracy and reaction time (working-memory factor inverted so higher = more accurate but slower responses).

2.5 Data-driven Community Centroids Capture Topological Structure Better than RDoC Domains

To evaluate how well different annotation schemes capture the global organization of individual brain activity, we applied the Mapper algorithm from topological data analysis to the contrast maps in Cohort 1 and Cohort 2, separately. Mapper constructs a simplified graph representation of high-dimensional data by applying binning and local clustering to a low-dimensional embedding. In the resulting graphs, nodes represent clusters of similar contrast maps, and edges connect nodes with overlapping contrasts. We annotated each node using two alternative schemes: (i) RDoC domain labels, based on the task-condition mappings; and (ii) data-driven community centroids, based on the highest spatial correlation between contrast maps and centroid patterns (Figure 7a).

To assess the distinctiveness of each annotation scheme, we computed node-wise assignment proportions and their correlation matrices across nodes (Fig. 7b). For each Mapper node, we calculated the proportion of contrasts assigned to each RDoC domain or data-driven community. The resulting correlation matrices reflect how often different categories co-occur within the same topological neighborhood. Compared to RDoC annotations, data-driven communities showed more moderate off-diagonal correlations, indicating more balanced and less redundant structure. In contrast, RDoC domains showed extreme values (high highs and low lows), reflecting greater inconsistency—some domains were overrepresented across nodes, while others were nearly absent. This pattern suggests that RDoC labels are more unevenly distributed in the intrinsic topological space, whereas data-driven annotations yield more consistent and balanced partitions of brain states.

Crucially, across both cohorts, Mapper graphs annotated with community centroids exhibited higher modularity than those annotated with RDoC domains (Fig. 7c). This indicates that the community-based annotation more closely aligns with the underlying topological organization of individual contrast maps. Together, these findings support the conclusion that data-driven community centroids offer a more accurate and functionally coherent framework for summarizing task-evoked brain activity than expert-defined RDoC categories.

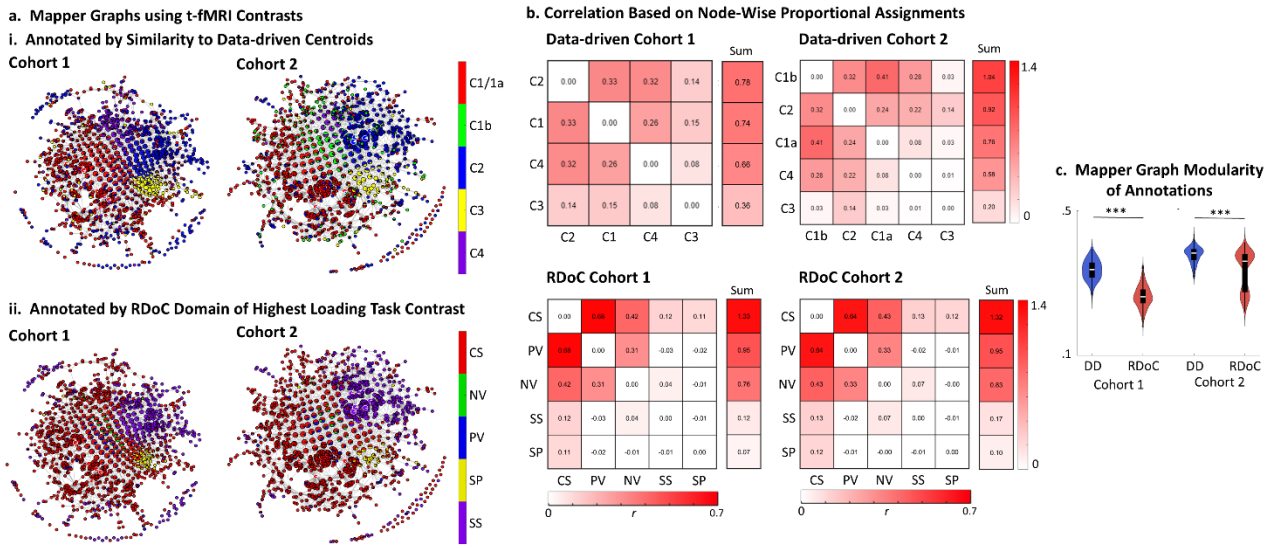

Figure 7. Topological analysis of individual contrast maps using Mapper. (a) Mapper graphs constructed from individual contrast maps from Cohort 1 and Cohort 2, annotated using either data-driven community centroids (top) or RDoC domain assignments (bottom). Each node represents a cluster of similar activation patterns, with edges connecting nodes sharing overlapping contrast maps. (b) Correlation matrices showing the similarity between annotation groups, based on node-wise proportions. For each node in the Mapper graph, we calculated the proportion of contrasts belonging to each community or RDoC domain. The resulting matrices reflect how often different annotations co-occur within nodes. Lower off-diagonal correlations in the data-driven models indicate that the communities are more distinct and less overlapping. The final column ("Sum") shows how much each group overlaps with all others, with lower values indicating greater separation. (c) Data-driven community annotations exhibit higher Mapper graph modularity than RDoC domain labels. Violin plots show the distribution of modularity values for data-driven (DD, blue) and RDoC-based (red) annotations applied to Mapper graphs in Cohort 1 and Cohort 2. Modularity quantifies the degree to which annotations align with topological community structure. In both cohorts, data-driven annotations yielded significantly higher modularity, indicating stronger alignment with intrinsic graph structure. ***$p < .001$.

3. Discussion

3.1 Validating the RDoC at the single-participant level

The RDoC framework has been widely applied in psychiatric neuroimaging research, yet its empirical grounding—particularly the mapping between functional domains and neural circuits—remains underdeveloped. Here, we present a latent factor analysis of task-based fMRI data aimed at refining RDoC to more closely align with the functional organization of brain activity and to identify the large-scale organization of task-evoked brain circuits. While prior efforts have derived latent dimensions from neuroimaging and behavioral data at the group level, it is unclear how well such models capture subject-specific activation patterns or generalize across datasets and behavioral phenotypes.

Our study systematically evaluated the structure, generalizability, and behavioral relevance of data-driven factor models estimated at the individual level, thereby offering a precision neuroscience framework for modeling brain-behavior associations. These models consistently outperformed RDoC-based alternatives in statistical fit, replicated across cohorts, and predicted task performance. The general factor derived from the bifactor models revealed a reproducible macroscale gradient from visual–attentional to auditory–default mode networks, aligning with canonical resting-state gradients and providing evidence for a domain-general activation dimension. Beyond the general factor, the data-driven models also recover stable, topographically distinct functional motifs that are associated with performance on high-demand cognitive tasks. Together, these findings provide a scalable framework for advancing precision neuroscience and refining psychiatric ontologies such as RDoC.

We first showed that bifactor models derived from EFA of individual-level tfMRI data offered significantly better model fit than RDoC-based models across multiple indices (AIC, RMSEA, CFI, TLI). This advantage was consistent across two independent cohorts, and held even when data-driven solutions from one cohort were projected onto the other. These results replicate prior group-level findings[7] and extend them by showing that subject-specific activation patterns retain a coherent, generalizable structure. This supports recent meta-analytic work suggesting that data-driven ontologies yield more reproducible brain–behavior mappings than theory-defined constructs[6,14].

3.2 Evidence for the general factor

The general factor itself reflected a core axis of task activation, with positive scores in visual, dorsal attention, and frontoparietal networks, and negative scores in default mode, auditory, and somatomotor networks. This configuration aligns closely with Gradient 2 from Margulies et al.[15], and supports a unifying framework that shapes both spontaneous and evoked activity. Functionally, the emergence of a single, brain-wide factor spanning multiple task domains supports the presence of a domain-general activation dimension[7]. This interpretation is supported by prior work showing that the dominant component of variance across diverse tasks reflects a shared visual-dorsal attention activation coupled with concomitant auditory–default mode deactivation[19]. More broadly, these findings reinforce emerging views that the brain's functional architecture is best described along continuous gradients and hierarchies, rather than rigid domain partitions[20]. Our general factor integrates systems often cast as antagonistic (e.g., visual–attention vs. default mode, task-positive vs. task-negative) into a unified continuum.

3.3 Evidence for specific motifs and communities

We next applied community detection to uncover shared structure among specific factors. This analysis identified four stable community motifs in Cohort 1, each with distinct spatial signatures. In Cohort 2, these patterns were replicated, with one community (C1) subdivided into two closely related subcommunities (C1a and C1b). The most prominent community (C1) involved co-activation of primary visual cortex and DMN regions—notably posterior cingulate and medial prefrontal cortex. This topography aligns with the principal gradient of cortical organization[15] and with the first complex principal component in resting-state data[21], supporting its interpretation as a core organizational axis of human brain function. The presence of visual–DMN co-activation likely reflects the strong visual demands embedded in many HCP tasks.

Importantly, these spatial motifs (or communities) were reproducible across cohorts and aligned with both group-level data-driven factors and canonical task contrasts. This suggests that they reflect recurrent, interpretable modes of co-activation, transcending individual tasks or datasets. The presence of these discrete motifs supports the formulation that large-scale brain activity can be decomposed into a relatively small number of interpretable functional modes. Notably, these communities aligned most strongly with RDoC domains associated with sensorimotor and positive valence systems, and less so with cognitive, social, or negative valence domains—replicating a pattern seen in our prior group-level work[7]. The stronger alignment for sensorimotor and reward domains may reflect their greater spatial stereotypy, consistent with prior meta-analyses[22,23], while more distributed and context-sensitive processes (e.g., negative valence, cognitive control) may resist simple spatial encapsulation[6,24].

Our finding that task-derived community centroids align with the principal gradients of resting-state functional connectivity described by Margulies et al.[15] provides strong evidence for their biological validity. The positive alignment of fronto-occipital centroids (C1/1a) with the principal gradient indicates that the task-derived factors here recapitulate the unimodal–to–transmodal axis of brain organization observed at rest, whereas the associations of centroids 2–4 with higher-order gradients indicate sensitivity to additional axes of functional specialization beyond the dominant unimodal–transmodal spectrum. The reproducibility of these relationships across independent cohorts further supports that the task-derived centroids here reflect stable, canonical dimensions of brain functional organization. Together, these findings support the view that individual-level task activation patterns are constrained by the same large-scale topographic gradients that govern intrinsic functional connectivity.

3.4 Behavioral associations

Behaviorally, we found that individuals whose activation patterns aligned more closely with certain centroids—especially C4 and C3—showed better performance on working memory and relational reasoning tasks. These associations held for both raw accuracy and EFA-derived latent performance scores. This supports the idea that spatial similarity to population-level motifs, rather than overall activation magnitude or connectivity, offers a meaningful index of task-relevant network engagement. These results extend prior findings[10,25] by demonstrating that centroid similarity predicts trait-like cognitive variation and may offer a scalable approach to individualized phenotyping.

3.5 Mapper and topological validation

Finally, we used TDA to show that data-driven community centroids better capture the functional organization of task-evoked brain states than predefined RDoC domains. Mapper graphs annotated with community centroids exhibited higher modularity, suggesting better alignment with the data's natural topological structure. In contrast, RDoC annotations—especially those for cognitive and valence domains—showed greater inter-domain overlap and more uneven distribution. These findings echo prior work showing that Mapper reveals modular and behaviorally relevant brain states[11], and they reinforce the view that empirically derived models offer more topologically coherent functional partitions than expert-defined taxonomies.

3.6 Implications for refining RDoC

These findings offer actionable insights for the refinement of cognitive ontologies such as RDoC. First, our results suggest that RDoC constructs could benefit from greater empirical grounding through systematic bottom-up modeling of neuroimaging data at the individual level. Instead of defining domains solely through top-down consensus or behavioral theory, a hybrid approach that integrates expert annotation with data-driven spatial motifs may yield more biologically valid and reproducible constructs. Second, the consistent emergence of a domain-general factor spanning multiple tasks points to the utility of incorporating continuous dimensions—such as macroscale activation gradients—alongside discrete domain categories. Such gradients may serve as organizing axes that unify multiple constructs or identify shared variance across domains. Third, adopting a graph-based or topological lens, as shown through Mapper analysis, provides a promising route for assessing whether cognitive taxonomies align with the intrinsic geometry of brain function. We propose that future iterations of RDoC include empirical benchmarking tools such as factor coherence, spatial reproducibility, behavioral relevance, and topological modularity, enabling systematic validation and revision of the proposed domains. Together, these principles can help re-anchor psychiatric constructs in neurobiologically informed, scalable frameworks—advancing RDoC's original mission of building a precision mental health science.

3.7 Limitations & Future directions

Several limitations should be acknowledged. First, the HCP dataset includes only healthy young adults and a constrained set of tasks. Generalization to clinical populations and ecologically valid paradigms remains a critical next step. Second, future work could explore nonlinear embeddings or alternative dimensionality reduction techniques. Third, while Mapper offers rich visualizations of data topology, its parameters (e.g., resolution, gain) are not biologically grounded and may influence results. Future work should evaluate parameter sensitivity, cross-dataset replication, and developmental or genetic correlates of the identified motifs. Fourth, while our models predict behavior, variability in reaction time can itself modulate BOLD responses[26], which may confound associations between factor scores and latent performance. Finally, this study did not include tasks targeting underrepresented RDoC domains (e.g., arousal/regulatory systems), nor did it integrate multimodal data (e.g., genetics, physiology, behavioral traits), which would further strengthen and contextualize our conclusions.

3.8 Conclusions

Despite these limitations, our findings provide strong evidence that individual-level, data-driven factor models outperform predefined RDoC models in explaining the structure of task-evoked

brain activity. These models reveal two complementary levels of organization: (i) a reproducible general factor that reflects a macroscale gradient spanning visual–attentional to auditory–default mode networks, and (ii) topographically distinct, behaviorally relevant communities that generalize across individuals and align with large-scale functional gradients. By embedding subject-specific brain activation patterns in scalable, interpretable, and empirically derived frameworks, this work offers a concrete step toward developing neurobiologically grounded cognitive ontologies and supports ongoing efforts to refine and operationalize RDoC in the service of precision psychiatry[27].

4. Methods

4.1 Study Cohort. We used task-fMRI data from the publicly available Human Connectome Project (HCP) Young Adult dataset[28]. A total of 962 healthy adults aged 22–35 years with complete task-fMRI data across seven tasks were included; participants with missing data for any task were excluded. Two age-, sex-, race-, and income-matched cohorts were defined for cross-validation: Cohort 1 (N = 412; 52.4% female; mean age = 28.5 years; racial composition: 65.7% White, 14.1% Black, 10.7% Hispanic/Latino, 7.5% Asian, 1.9% Other; income distribution: 46.5% high / 53.5% low) and Cohort 2 (N = 329; 55.0% female; mean age = 28.8 years; racial composition: 69.3% White, 13.5% Black, 8.9% Hispanic/Latino, 7.1% Asian, 1.2% Other; income distribution: 54.3% high / 45.7% low). All participants were unrelated within each cohort.

The HCP task battery included seven paradigms aligned with core RDoC domains: (1) a 2-back task assessing working memory; (2) a motor response task targeting motor function; (3) an auditory story task probing language processing; (4) a stimulus matching task assessing emotional processing; (5) a social cognition task probing theory of mind; (6) a relational matching task assessing higher-order reasoning; and (7) a gambling task assessing reward sensitivity. These tasks are described in detail by Barch et al.[29].

4.2 Contrast Map Processing. Individual-level task contrast maps were downloaded from the HCP database in CIFTI format. For each participant, contrast maps were computed by combining parameter estimates from both the left–right (LR) and right–left (RL) phase-encoding runs. This approach integrates data from independent runs into a single subject-level contrast, helping to mitigate run-specific dependencies. The resulting maps were processed and parcellated into a total of 347 regions: 333 cortical parcels from the Gordon atlas[16] and 14 subcortical regions from the Harvard–Oxford atlas[30]. These atlases were selected for their fine-grained, functionally informed parcellations, which have been widely validated in both task-based and resting-state fMRI studies.

4.3 Factor Models. We compared four model types: (i) an RDoC-specific factor model (canonical RDoC model), (ii) an RDoC bifactor model, (iii) a data-driven specific factor model, and (iv) a data-driven bifactor model. In the bifactor variants, a general factor (onto which all task contrasts loaded) was included alongside domain- or EFA-derived specific factors. In the specific factor variants, only domain- or data-driven specific factors were modeled, without a

general factor. The general factor captures variance shared across all task contrasts, while specific factors reflect variance unique to subsets of contrasts. For both bifactor models, orthogonality constraints were imposed to ensure separation between shared and specific variance components[31].

Data-driven models were estimated using a two-step process involving exploratory factor analysis (EFA) followed by confirmatory factor analysis (CFA). In contrast, RDoC-based models were estimated directly via CFA, using predefined domain-to-task assignments. All factor analyses were implemented in R (version 4.3.0) using the psych (version 2.3.3) and lavaan (version 0.6.15) packages.

4.3.1 Data-driven Factor Analysis for each subject. Subject-specific factor models were derived using the following steps: (1) Horn's parallel analysis to determine the optimal number of specific factors (see below); (2) EFA using principal axis factoring and oblimin rotation; and (3) CFA specifying either a bifactor or specific factor structure.

Each participant's data consisted of 24 contrast maps (i.e., task conditions), treated as variables, and 347 brain parcels (333 cortical + 14 subcortical) as observations—resulting in a 347 × 24 matrix per subject. EFA was applied to model the covariance structure among the contrast maps based on spatial similarity across parcels. A high factor loading indicated that a task contrast's spatial activation pattern closely aligned with the topography represented by that factor.

The number of factors to extract was determined using Horn's parallel analysis[32], which identifies the point where eigenvalues from the real data intersect those from randomly generated data[33]. Principal axis factoring with oblimin rotation was used to allow for weak correlations among factors. In the subsequent CFA, we used robust maximum likelihood estimation to account for non-normality. Two data-driven models were specified: (i) a bifactor model, where all contrasts loaded on a general factor and one or more EFA-derived specific factors; and (ii) a specific factor model, where only EFA-derived specific factors were included. Consistent with prior literature[34], specific factors were defined using a threshold of absolute loadings ≥ 0.4.

4.3.2 RDoC Domain Factor Analysis for each subject. RDoC-based models were constructed by grouping whole-brain activation maps into domain-specific factors, based on expert mappings of task descriptions to RDoC domains (Supplementary Table 6). As with data-driven models, we tested both (i) a specific factor model (contrasts loaded only onto their assigned RDoC domains) and (ii) a bifactor model (all contrasts additionally loaded onto a general factor). All RDoC-based models were estimated using CFA with robust maximum likelihood estimation (Fig. 2ai and 2aii).

4.4 Training and Validation Strategy. To ensure a fair comparison between data-driven and RDoC-based models, we derived the data-driven factor structure from one cohort and tested its generalizability in the other (Supplementary Fig. 4). This cross-cohort approach avoids circularity, as data-driven models are inherently optimized to fit the data they are trained on. One cohort was designated as the training set, from which bifactor models were estimated via EFA and CFA. Specific factors were then reduced via principal component analysis (PCA) to extract the top 3–7 components, yielding a low-dimensional representation of the training data.

These PCA components were projected onto the validation cohort by computing a correlation matrix between individual activation maps and the training-derived component scores. A dynamic threshold, set to the p-th quantile (p = 1/k, where k is the number of components), was applied to assign activation maps to components, ensuring balanced factor assignment. This generated individual-level factor scores in the validation set, grounded entirely in the training cohort's structure.

In contrast, RDoC-based models—being predefined—were directly estimated in each cohort using CFA, with identical factor assignments across training and validation analyses.

4.5 Model Performance. Model fit was evaluated using robust versions of the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Tucker–Lewis Index (TLI), which account for potential non-normality in the data. Additionally, we used the Akaike Information Criterion (AIC), an information-theoretic measure that balances model fit with complexity. Lower AIC values indicate a more favorable trade-off between goodness of fit and parsimony[35].

4.6 Community Detection. To identify shared structure across individuals' factor representations, we applied Louvain modularity maximization[36] to the specific factor scores derived from the data-driven bifactor models. Subject-level factor maps were concatenated into a matrix of size (subjects × factors) × regions, and community structure was estimated using an asymmetric treatment of negative edges. This variant preserves meaningful modularity by asymmetrically penalizing anti-correlated nodes. To address the degeneracy of modularity (Q) solutions, we ran the algorithm 1,000 times with a fixed resolution parameter ($\gamma = 1$). Each run produced a community assignment and associated modularity value. We then computed an agreement matrix based on the frequency of co-assignment across runs, weighted by the modularity of each partition. A consensus clustering procedure was applied to this matrix to derive a stable community solution. Stability was quantified as the adjusted Rand index (ARI) between each run's partition and the consensus solution, summarized by the mean and standard deviation across all runs.

The resulting communities grouped factor maps with similar spatial profiles. We computed centroid maps—the average activation pattern within each community—for use in downstream analyses, including cross-cohort comparisons and behavioral prediction.

4.6.1 Within-Subject Consistency of Community Assignments. To evaluate the consistency of community assignments across factor maps from the same individual, we computed subject-wise Shannon entropy. For each participant, we tallied how many of their factor maps were assigned to each community and converted these counts into a probability distribution. Entropy was then computed as:

$$H = -\sum_{\{i=1\}}^{k} p_i log_2(p_i)$$

where p_i is the proportion of maps assigned to the *i-th* community, and k is the total number of detected communities. Entropy approaches zero when all maps from an individual fall within a single community and reaches a maximum $\log_2(k)$ when maps are evenly distributed across communities.

4.7 Topological Data Analysis. We applied the Mapper algorithm[11,37], a TDA method, to visualize the intrinsic structure of individual task-evoked brain activation. Mapper constructs a graph-based summary of high-dimensional data, revealing topological features that are often missed by traditional dimensionality reduction techniques.

We began by computing a pairwise Euclidean distance matrix across all contrast maps. A k-nearest neighbor (kNN) graph was constructed to approximate the underlying data manifold, with k set to the square root of the number of data points, following the empirical rule-of-thumb[38]. From the kNN graph, we computed geodesic distances between all data points to better capture the intrinsic structure while preserving local neighborhood relationships. Non-Metric Multidimensional Scaling (NMDS) was then applied using the Sammon stress criterion to obtain a low-dimensional embedding for filtering, emphasizing preservation of local distances.

Mapper was run on this embedded space using overlapping bins and local clustering. The resulting graph consisted of nodes—representing local subgroups of contrast maps—and edges, which indicated shared data points across nodes.

Nodes were annotated using either RDoC domain labels (based on task-condition mappings; Supplementary Table 6) or data-driven community centroid labels (assigned via maximal spatial correlation). To assess the stability of modularity under both annotation schemes, we performed jackknife resampling across subjects. For each iteration, one subject was omitted, and Mapper was re-run on the remaining participants using identical parameters (number of bins, overlap, resolution, gain, k, and embedding dimensionality). Modularity was then computed for each resampled graph after annotating nodes with either RDoC or community labels. This produced empirical distributions of modularity values for both labeling schemes, separately for each cohort.

4.8 Statistical Analysis of Model Fit. To compare model performance, paired t-tests were conducted to compare the data-driven bifactor and RDoC models within each cohort for both training and cross-validation analyses. For the training analyses to compare all four model types (data-driven specific, data-driven bifactor, RDoC specific, and RDoC bifactor), we conducted repeated-measures ANOVAs separately for each fit index: robust RMSEA, robust CFI, robust TLI, and AIC. Model type was treated as a four-level within-subject factor, with each participant contributing a set of fit indices for all model types. Mauchly's test indicated that the assumption of sphericity was violated for all indices; therefore, Greenhouse–Geisser corrections were applied to the degrees of freedom in all tests. ANOVAs were followed by Tukey-adjusted pairwise comparisons to evaluate differences between model families. All analyses were implemented in R (version 4.3.0) using the afex (version 1.5.0) and emmeans (version 1.11.2.8) packages.

4.9 Correlations Between Factor Scores. We assessed the topographical stability of community centroids by computing parcel-wise Pearson correlations between centroids derived from Cohort 1 and Cohort 2. To evaluate alignment with existing cognitive ontologies, we also correlated cohort-specific centroid maps with previously published group-level activation maps from an independent dataset[7], encompassing both data-driven and RDoC-based representations.

All correlations were performed across the 347 parcels using unthresholded factor score maps. Statistical significance was evaluated using BrainSMASH[18], which generates spatially autocorrelated null maps that preserve the intrinsic spatial structure of the data. P-values were corrected for spatial autocorrelation (see Supplementary Table 7). Inter-parcel distances for BrainSMASH were computed using Euclidean distances between the MNI-space centroids of each parcel. All statistical tests were two-tailed.

4.10 Task Performance Correlates. We tested whether the similarity between an individual's contrast map and the corresponding community centroid predicted cognitive task performance. Similarity was quantified using both Pearson correlation and continuous (soft) Dice coefficients. While correlation captures linear similarity, soft Dice scores account for spatial overlap weighted by magnitude. For each task contrast map, the continuous Dice similarity[39] was computed as:

$$Continuous\ Dice = \frac{2\ \sum_i x_i y_i}{\sum_i x_i^2 + \sum_i y_i^2}$$

where x is the centroid vector and y is the individual's contrast map. This formulation generalizes the traditional Dice coefficient to continuous, real-valued data, allowing sensitivity to both spatial alignment and intensity variation.

Associations between map–centroid similarity and behavior were estimated using linear mixed-effects models. The fixed effect was map–centroid similarity, and a random intercept for Parental ID (defined as the combination of maternal and paternal IDs) accounted for genetic relatedness. Behavioral outcomes included both raw task accuracy and an EFA-derived latent performance score (described below). We focused on the working memory and relational reasoning tasks, which are the only HCP tasks with well-defined individual-level behavioral metrics (see Supplementary Methods 1 for discussion of other tasks).

To derive a general measure of task performance, we conducted EFAs separately for the working memory and relational tasks, analyzing accuracy and median reaction time using principal axis factoring. A single latent factor was extracted, capturing shared variance among the behavioral measures. Each participant's score on this factor was used as an index of their overall working memory/relational reasoning performance.

4.11 Sex Effects Analysis. To assess potential sex differences in task-evoked brain activation, we performed parcel-wise regressions of activation level on sex across all 24 contrasts and 347 parcels in each cohort. After false discovery rate (FDR) correction, 9.1% of tests remained significant in Cohort 1 and 2.4% in Cohort 2. Given the limited reproducibility and small effect sizes, sex was not included as a covariate in the main analyses in order to preserve model

parsimony and statistical power. Future work may explore demographic influences such as sex in more depth within this framework.

References

1. Cuthbert, B. N. & Insel, T. R. Toward new approaches to psychotic disorders: the NIMH Research Domain Criteria project. *Schizophr. Bull.* **36**, 1061–1062 (2010).
2. Insel, T. *et al.* Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).
3. Cuthbert, B. N. & Insel, T. R. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* **11**, 126 (2013).
4. Kozak, M. J. & Cuthbert, B. N. The NIMH research domain criteria initiative: Background, issues, and pragmatics. *Psychophysiology* **53**, 286–297 (2016).
5. Morawetz, C., Hemetsberger, F. J., Laird, A. R. & Kohn, N. Emotion regulation: From neural circuits to a transdiagnostic perspective. *Neurosci. Biobehav. Rev.* **168**, 105960 (2025).
6. Beam, E., Potts, C., Poldrack, R. A. & Etkin, A. A data-driven framework for mapping domains of human neurobiology. *Nat. Neurosci.* **24**, 1733–1744 (2021).
7. Quah, S. K. L. *et al.* A data-driven latent variable approach to validating the research domain criteria framework. *Nat. Commun.* **16**, 830 (2025).
8. Cuthbert, B. N. Research Domain Criteria (RDoC): Progress and Potential. *Curr. Dir. Psychol. Sci.* **31**, 107–114 (2022).
9. Rubin, T. N. *et al.* Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. *PLoS Comput. Biol.* **13**, e1005649 (2017).
10. Egli, T. *et al.* Identification of two distinct working memory-related brain networks in healthy young adults. *eNeuro* **5**, (2018).
11. Saggar, M. *et al.* Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature Communications 2018 9:1* **9**, 1–14 (2018).

12. Singh, G., Mémoli, F. & Carlsson, G. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. *Eurographics Symposium on Point-Based Graphics* (2007) doi:10.2312/SPBG/SPBG07/091-100.
13. Bolt, T., Nomi, J. S., Yeo, B. T. T. & Uddin, L. Q. Data-Driven Extraction of a Nested Model of Human Brain Function. *J. Neurosci.* **37**, 7263 (2017).
14. Bolt, T. *et al.* Ontological Dimensions of Cognitive-Neural Mappings. *Neuroinformatics* **18**, 451–463 (2020).
15. Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. U. S. A.* (2016) doi:10.1073/pnas.1608282113.
16. Gordon, E. M. *et al.* Precision Functional Mapping of Individual Human Brains. *Neuron* **95**, 791-807.e7 (2017).
17. Vazquez-Trejo, V., Nardos, B., Schlaggar, B. L., Fair, D. A. & Miranda-Dominguez, O. Use of connectotyping on task functional MRI data reveals dynamic network level cross talking during task performance. *Front. Neurosci.* **16**, 951907 (2022).
18. Burt, J. B., Helmer, M., Shinn, M., Anticevic, A. & Murray, J. D. Generative modeling of brain maps with spatial autocorrelation. *Neuroimage* **220**, 117038 (2020).
19. Shine, J. M. *et al.* Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nat. Neurosci.* **22**, 289–296 (2019).
20. Huntenburg, J. M., Bazin, P.-L. & Margulies, D. S. Large-scale gradients in human cortical organization. *Trends Cogn. Sci.* **22**, 21–31 (2018).
21. Bolt, T. *et al.* A parsimonious description of global functional brain organization in three spatiotemporal patterns. *Nat. Neurosci.* **25**, 1093–1103 (2022).

22. Bartra, O., McGuire, J. T. & Kable, J. W. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **76**, 412–427 (2013).
23. Hardwick, R. M., Rottschy, C., Miall, R. C. & Eickhoff, S. B. A quantitative meta-analysis and review of motor learning in the human brain. *Neuroimage* **67**, 283–297 (2013).
24. de la Vega, A., Yarkoni, T., Wager, T. D. & Banich, M. T. Large-scale meta-analysis suggests low regional modularity in lateral frontal cortex. *Cereb. Cortex* **28**, 3414–3428 (2018).
25. Stern, Y., Gazes, Y., Razlighi, Q., Steffener, J. & Habeck, C. A task-invariant cognitive reserve network. *Neuroimage* **178**, 36–45 (2018).
26. Mumford, J. A. *et al.* The response time paradox in functional magnetic resonance imaging analyses. *Nat. Hum. Behav.* **8**, 349–360 (2024).
27. Poldrack, R. A. & Yarkoni, T. From brain maps to cognitive ontologies: Informatics and the search for mental structure. *Annu. Rev. Psychol.* **67**, 587–612 (2016).
28. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79 (2013).
29. Barch, D. M. *et al.* Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).
30. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
31. Carroll, J. B. Human cognitive abilities: A survey of factor-analytic studies. **819**, (1993).
32. Horn, J. L. A RATIONALE AND TEST FOR THE NUMBER OF FACTORS IN FACTOR ANALYSIS. *Psychometrika* **30**, 179–185 (1965).

33. Hayton, J. C., Allen, D. G. & Scarpello, V. Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. *Organizational Research Methods* **7**, 191–205 (2004).
34. Stevens, J. (james P. *Applied Multivariate Statistics for the Social Sciences*. 629 (L. Erlbaum Associates, 1992).
35. Burnham, K. P. & Anderson, D. R. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociol. Methods Res.* **33**, 261–304 (2004).
36. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
37. Saggar, M., Shine, J. M., Liégeois, R., Dosenbach, N. U. F. & Fair, D. Precision dynamical mapping using topological data analysis reveals a hub-like transition state at rest. *Nature Communications 2022 13:1* **13**, 1–19 (2022).
38. Duda, R. O., Hart, P. E. & Stork, D. G. *Pattern Classification*. (John Wiley & Sons, Nashville, TN, 2000).
39. *Deep Learning and Data Labeling for Medical Applications*. (Springer International Publishing, Cham, Switzerland, 2016).