

Task-structured Modularity Emerges in Artificial Networks and Aligns with Brain Architecture

Shi Gu

gus@zju.edu.cn

Zhejiang University

Yuhang Wu

University of Electronic Science and Technology of China

Shikuang Deng

University of Electronic Science and Technology of China

Kangrui Du

University of Electronic Science and Technology of China

Marcelo Mattar

New York University

Yifan Wu

University of Pennsylvania

Dani Bassett

University of Pennsylvania <https://orcid.org/0000-0002-6183-4493>

Huajin Tang

Gang Pan

Zhejiang University <https://orcid.org/0000-0002-4049-6181>

Article

Keywords:

Posted Date: July 31st, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-6987147/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Machine Intelligence on September 28th, 2026. See the published version at <https://doi.org/10.1038/s42256-026-01306-9>.

Task-structured Modularity Emerges in Artificial Networks and Aligns with Brain Architecture

Yuhang Wu¹, Shikuang Deng¹, Kangrui Du¹, Marcelo G. Mattar², Yifan Wu³,
Dani S. Bassett^{3,4,5}, Huajin Tang^{6,7}, Gang Pan^{6,7*}, Shi Gu^{6,7*}

¹Department of Computer Science and Engineering, University of Electronic Science and
Technology of China, Chengdu, China.

²Department of Psychology, New York University, New York, United States.

³Department of Bioengineering, University of Pennsylvania, Philadelphia, United States.

⁴Department of Electrical and Systems Engineering, University of Pennsylvania,
Philadelphia, United States.

⁵Santa Fe Institute, Santa Fe, United States.

⁶College of Computer Science and Technology, Zhejiang University, Hangzhou, China.

⁷State Key Laboratory of Brain Machine Intelligence, Zhejiang University, Hangzhou, China.

*Corresponding author(s). E-mail(s): gpan@zju.edu.cn; gus@zju.edu.cn;

Abstract

Understanding how neural systems develop modular organization is fundamental to both neuroscience and artificial intelligence. Although modular architectures can improve adaptability and cognitive performance, the processes leading to their emergence are poorly understood. Here we demonstrate that multitask and incremental learning enhance modularity in recurrent neural networks (RNNs) compared to single-task learning, revealing how functional demands influence the structural organization of neural networks. We trained RNNs on cognitive tasks under three distinct learning paradigms: single-task, simultaneous multitask, and incremental multitask learning. Our results suggest that networks adopting multitask learning show an enhanced degree of modularity compared to single-task training, especially when the task load exceeds the network’s representation capacity. Those trained with incremental multitask learning, in particular, develop the highest degree of modularity, maintaining superior performance even when connections are selectively pruned. Furthermore, these task-induced networks exhibit structural properties more closely resembling those of biological brain networks than those based solely on spatial constraints, particularly in clustering coefficients and edge length distributions. Collectively, these findings suggest that modular brain architecture emerges not only from physical constraints but as an adaptive response to the sequential introduction of complex cognitive tasks, providing a causal explanation for how functional demands shape network topology. Our findings also illustrate how the developmental and evolutionary context of the brain, wherein multiple tasks are learned, can inform the design of artificial systems well-suited for similar environments.

Introduction

Brain networks display strikingly modular patterns in both structural and functional connectomes, with groups of brain regions densely connected within themselves but sparsely connected to other modules [1]. While a single node in a network has limited representational capacity, modules have been hypothesized to enrich the brain’s ability to integrate and process information, thereby contributing to learning and adaptability [2]. The modular structure of the brain is also related to the maturity of the central nervous system [3] and cognitive abilities in different species and different ages [4, 5]. Yet, despite considerable research exploring the role of brain modules in cognition, the causal processes leading to the emergence of modular structure remain insufficiently understood.

Prior work has posited that modularity is an evolutionary response to physical constraints, such as the need to minimize wiring costs [6–8]. However, the clustering of representations in neurons and brain regions, particularly their functional and temporal variations, extends beyond what can be explained by network geometry alone. In addition to spatial constraints, biological organisms must constantly adapt to multiple complex tasks composed of elementary sensory, motor, and cognitive processes [9]. Yet, a critical question remains largely unexplored: how does the compositional nature of task representation influence the modularization of brain networks? From an evolutionary perspective, we can observe that gradual adaptation to an increasing variety of tasks has shaped neural systems into their current state [10–12]. However, whether and how multitask learning and incremental execution causally influence network topology throughout evolution is presently unknown.

Unfortunately, studying the effect of multitasking and incremental learning on network topology presents significant challenges. In biological organisms, such studies would require imposing different functional demands (e.g., different task sets) on an organism and observing the resulting effects on the structural connectome across many generations. In fact, it has also proven difficult to characterize the precise way in which modularity influences task performance, as this would require causally disrupting such modular structure while maintaining the general integrity of the network. Our study circumvents these methodological issues by modeling the brain through artificial neural networks [9, 13]. By systematically varying model conditions and learning paradigms for multiple tasks, we can track learning curves and identify the impact of tasks on network topology throughout the learning process and across different model instances.

In this work, we examine three fundamental questions related to network modularization: (1) whether and how multi-task learning leads to the emergence of modular patterns in underlying networks; (2) whether and how different task-learning approaches affect modular patterns; and (3) whether task-learning-induced networks more closely resemble biological brain networks than models based solely on spatial constraints. To

66 answer these questions, we trained recurrent neural networks (RNNs) with varying numbers of hidden-layer
 67 units to execute 20 cognitive tasks. We explored how the RNNs learn the tasks in three different modes. In
 68 Training Mode I, we train each RNN to learn on a single task. In Training Mode II, we train each RNN on
 69 all 20 tasks simultaneously. In Training Mode III, we train each RNN on all 20 tasks while we incrementally
 70 update the network structure from partial to complete architecture, mimicking the evolution of neural systems
 71 with an increasing number of synapses. Through a systematic examination of these three training modes, we
 72 find that the incremental learning of multiple tasks leads to network topologies with higher modularity than
 73 those trained on single tasks. We also find that the network with edge locations induced from the learned
 74 module maintains a higher performance than the randomized masked network with the same amount of
 75 edges. Finally, we demonstrate that the multi-task-learning-induced weights improve the generative model of
 76 network formation with a higher similarity to the brain network connectome as measured by the distribution
 77 of graph measurements. Together, our results provide a causal structural explanation for the emergence and
 78 function of the brain's modular topology.

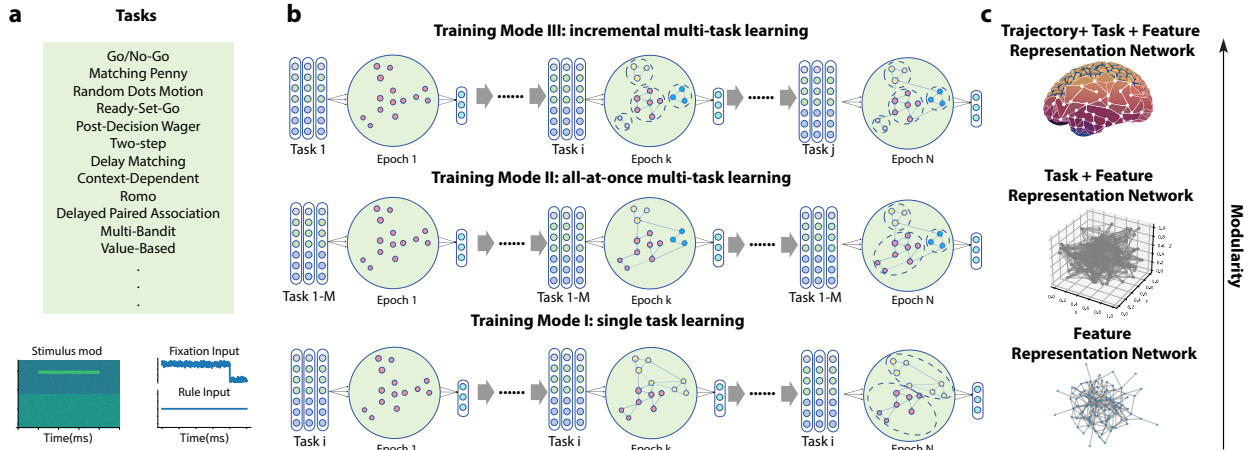


Fig. 1: Recurrent Neural Networks in Different Training Modes. Here we demonstrate the architecture and training progression of the recurrent neural network. **a**, Task configurations range from simple cognitive tasks like Go/No-Go to complex ones like value-based decision-making. Inputs include fixation and rule-based stimuli. **b**, The models are trained in three modes. Mode I: Single-Task Learning, where the network learns tasks individually; Mode II: All-at-Once Multi-Task Learning, which involves simultaneous training on multiple tasks; Mode III: Connection-incremental Learning, where tasks are introduced progressively over time. The network undergoes training through multiple epochs (Epoch 1 to Epoch N) to capture the development of task-specific and generalized neural modules. **c**, Correspondingly, the trained networks are categorized into feature representation network, task and feature representation network, and trajectory, task, and feature representation network, which denote the factors that affect the emergent modular structure.

79 Results

80 To examine how task requirements influence network structure, we trained RNN models (Fig. 1b) to execute
 81 a series of inter-related cognitive tasks. Most of these tasks are digital analogs of those frequently employed
 82 in neurophysiological research, encompassing variants of memory-guided responses[14], simple perceptual
 83 decision-making (DM)[15], context-dependent DM[16, 17], multi-sensory integration[18], parametric working
 84 memory[19], inhibitory control[20], delayed match-to-sample[21], and delayed match-to-category tasks[22].
 85 Following the approach described in a prior study [9], the network is subjected to three distinct categories
 86 of noisy inputs (Fig. 1a). It subsequently employs recurrent neurons to evaluate all pertinent information
 87 and delivers this data to the output layer, which comprises the fixation unit and motor units. The fixation
 88 unit determines if the network is ‘fixating’ or ‘responding’ to the inputs, whilst the motor units encode the
 89 response direction on a circular array of outputs. For additional information concerning the setup, please
 90 refer to the Materials and Methods section and the previous research undertaken by [9].

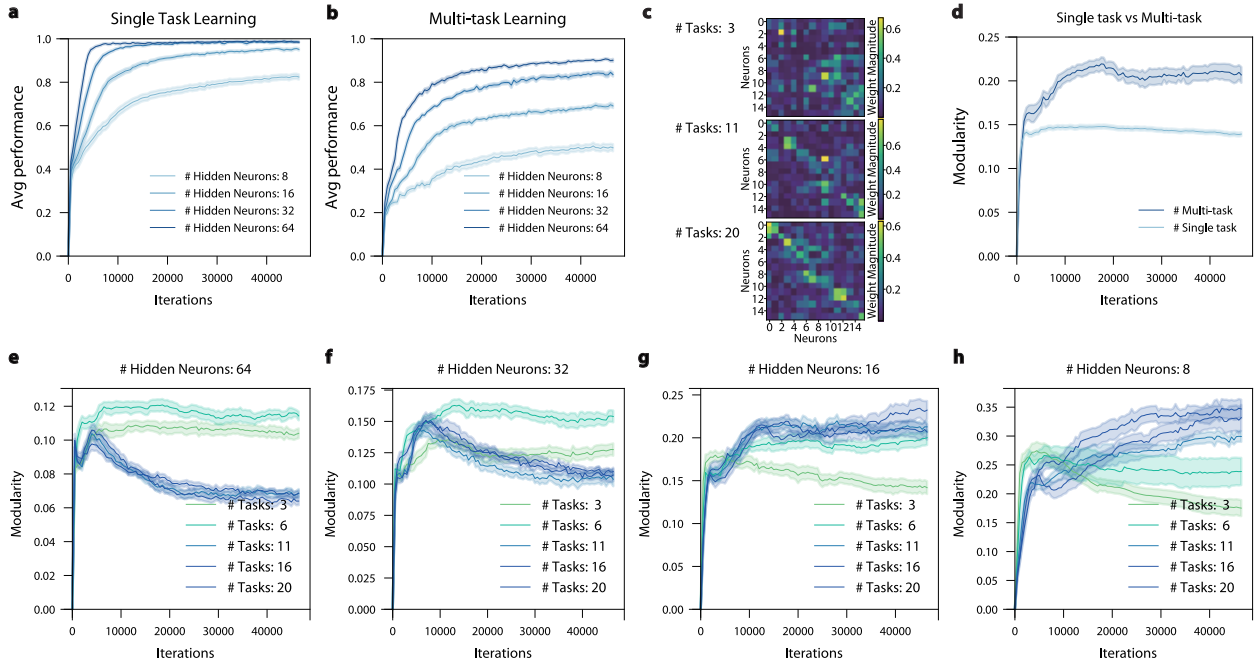


Fig. 2: Multitask Training Enhances Modularity. **a**, Performance of models of varying sizes (# Hidden Neurons) trained on a single task. **b**, Performance of models trained on multiple tasks all-at-once. **c**, Examples of correlation matrices of the hidden recurrent layer activation values for models trained with different numbers of tasks. **d**, Multitask learning leads to a more modularized network compared to that induced by single-task learning (the models in the example have # Hidden Neurons set to 16). **e-h**, the resulting modularity is co-affected by the number of hidden neurons and the number of tasks. (For details on the composition of these task sets, please refer to the Appendix.) Solid lines are the mean over 20 independent runs. Shaded areas indicate the standard error.

Multi-task Training Promotes Modularization via Co-embedding

Emergence of Task Co-embedding in Capacity-Limited RNNs

While the dimension of cognitive tasks could be infinite in theory, the brain maintains the tasks with limited dimensions of modularized functional representation. In order to characterize the formation of such modularized functional representation, we trained a series of RNN models of varying sizes under both single-task and multi-task setups. For single-task learning, we only assigned a single task to each network configuration and assessed the average performance on the set of 20 tasks after training (Fig. 2a). For multi-task learning, we trained RNN models with tasks randomly interleaved at each iteration to perform all tasks together (Fig. 2b). First of all, it is noticeable that for both learning paradigms, the increase in the number of hidden neurons enhances the model’s task performance in both learning speed and final accuracy (# Hidden Neurons: 8, 16, 32, 64 as illustrated in Fig. 2a, b). In addition, models trained on multi-task learning perform generally worse than models trained on single-task learning (FDR $q < 0.05$ after Bonferroni’s correction for different setups of # Hidden Neurons). This discrepancy suggests that the networks make concessions in the representation of each individual task in every updating step to manage and accommodate multiple tasks concurrently. As a cost, the overall accuracy at a given time step drops even though the final performance might be as high once the overall capacity is sufficient. Consistent with [9], we observe the similar phenomenon of co-embedding functionally similar tasks for reducing the demand of hidden nodes. Indeed, a single model with a sufficient number of hidden units (e.g., # Hidden Neurons = 32) trained under multi-task learning (Fig. 2b) achieves comparable average performance to separate models, each with a small number of hidden units (e.g., # Hidden Neurons = 8 for 20 tasks, which means 160 units in total) trained under single-task learning (Fig. 2a). Overall, such observations verify the emergence of task co-embedding when the network is trained for learning multiple tasks with limited capacity.

Modularization as the Structural Response to Task Co-embedding

Our working hypothesis is that brains display modular structure to promote adaptation and increase flexibility in response to a changing environment [23]. How does this happen? If we adopt the dogma from Donald Hebb that “*Neurons that fire together, wire together*” [24], here we can hypothesize that the co-embedding of tasks leads to the modularization of the functional covariance of the hidden layer, which further facilitates the structural wiring. To validate this hypothesis, we applied a widely used community detection technique [25] to the recurrent layer of the RNN model and examined the modular structure through model training. Specifically, hidden neurons in the recurrent layer were grouped into communities, exhibiting dense intra-community connections and sparse inter-community connections, based on the absolute values of the

recurrent weights (Fig. 2c). To assess how task learning shapes the network, we computed the modularity of the RNNs in both single-task (Mode I) and multi-task (Mode II) learning (Fig. 1 b). We find that for the RNN with a fair number (e.g., ≥ 16) of hidden neurons (Fig. 2d), multi-task learning induces significantly higher levels of modularization compared to single-task learning by a substantial margin ($p < 0.05$ for a two-sample t -test at any selected iteration after 10000; see Fig. A1 for more detailed results). In addition to a more modularized structure, multi-task learning also facilitates a more enduring process of modularization compared to single-task learning. In the initial stage of learning, both modes displayed a rapid increase of modularity, probably due to the embedding of input features. In the subsequent learning phase, however, the modularity of the multi-task learning model exhibits a sustained upward trajectory, while the modularity of the single-task learning model reaches a plateau. This difference may be a result of additional demands of task embedding together with the feature embedding, underscoring the differential durations of adaptation made by the model in response to the task contents and types.

To obtain a more comprehensive understanding of modularization under different balances of model capacity and task load, we varied the parameters of task number and hidden neuron number at a finer granularity, as illustrated in Fig. 2 d, e, f, and g (for details of the four task sets, see the Supplementary Methods). When the numbers of hidden neurons are sufficiently large (e.g., ≥ 32 for the current setup), it is almost possible to represent three tasks in separate, non-overlapping populations. Thus, incorporating more tasks may not result in a more significant modularization given Eqn. 2. With relatively few hidden units (e.g., 16 and 8), the task embedding will make a difference and lead to a higher level of modularity. Such findings raise an intriguing possibility that the emergence of modular structures in the brain may serve as a structural cause to address the limitations of model capacity while supporting multiple tasks.

A Causal Explanation for the Emergence of Modularization

To further elucidate how the task co-embedding in the hidden layer representation causes the emergence of modularization, we turned to an examination of the relationship among tasks in embedding and performed the representational similarity analysis on task representations.

Co-embedding of Tasks into the Neuronal Representation

For the relationship among tasks regarding their interaction in the embedding, we illustrate three typical phenotypes based on the overlap of embedding neurons in the hidden layer (Fig. 3a). For each task pair, we compared single-task learning (blue and orange) and interleaved learning (green), and determined the minimal number of hidden neurons required to achieve $\geq 95\%$ accuracy. Left-column subplots (1, 3, 5) show

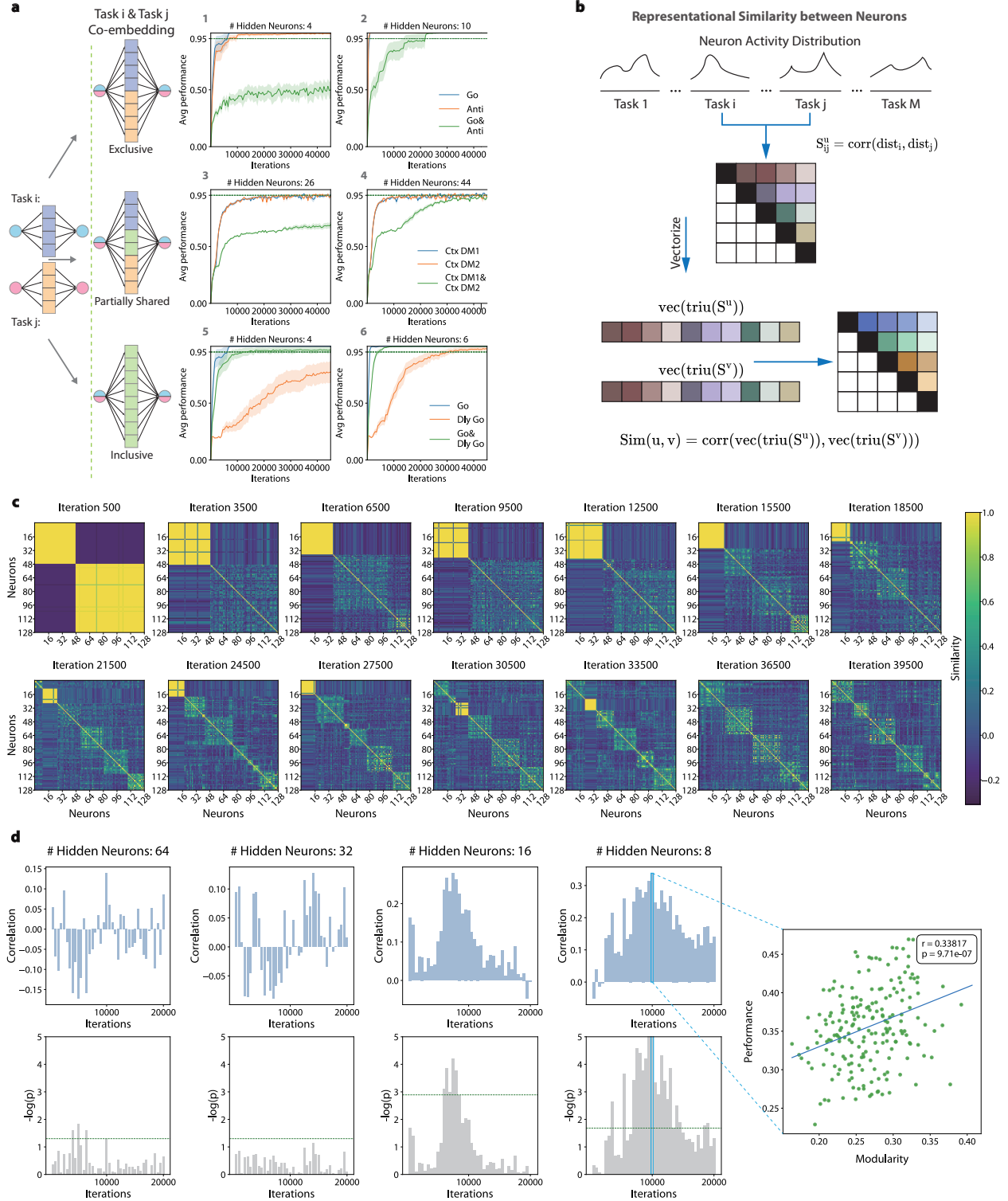


Fig. 3: Emergence of Modular Structure over Task Learning. **a**, Three examples illustrating the typical relationship between tasks in terms of their neuronal embedding. The dark curve represents the mean of 10 independent runs, with the shaded area indicating the standard error. **b**, A representational similarity analysis method is adapted and tailored to quantify the neuronal similarity with respect to their response to different tasks. **c**, Building on panel **b**, neurons are clustered into modules that evolve throughout training. **d**, A significant positive correlation between modularity and task performance is observed in capacity-limited models. Each green dot represents one of 200 model instances under different seeds. The green dashed line denotes the False Discovery Rate-adjusted p -value threshold of 0.05, on a log scale.

single-task learning with minimal sufficient capacity, while right-column subplots (2, 4, 6) show interleaved learning with sufficient capacity for both tasks. Our results reveal three types of task relationships:

Exclusive Representation (Subplot 1 & 2): Tasks *fdgo* and *fdanti* each require only 4 hidden neurons when trained independently. However, when trained jointly in an interleaved fashion, the network requires 10 neurons to reach the same performance. This sharp increase indicates that the tasks use non-overlapping neuronal populations, consistent with their contradictory objectives.

Partially Shared Representation (Subplots 3 & 4): Tasks *contextdm1* and *contextdm2* each require 26 neurons when learned independently, while 44 neurons are sufficient for interleaved learning. The difference suggests that the network allocates some shared neurons alongside private subpopulations, similar to [9].

Inclusive Representation (Subplots 5 & 6): Task *fdgo* is a simplified variant of *delaygo*. The minimal network required for learning them jointly is nearly identical to that for *delaygo* alone, indicating that *fdgo* can be fully embedded within the representational space of *delaygo* without the need for additional capacity.

When multiple tasks are learned together, the network can exploit shared neural representations, resulting in only a modest increase in capacity relative to single-task learning, suggesting overlapping neural substrates. We identify three representative task-pairing regimes—exclusive, partially shared, and fully inclusive—that illustrate distinct patterns of representational modularity. In the case of exclusive tasks (e.g., *fdgo* and *fdanti*), the network forms separate modules, as their contradictory objectives require task-specific representations, reflecting the need for non-overlapping representations due to conflicting task objectives. In contrast, partially shared (e.g., *contextdm1* and *contextdm2*) and fully inclusive (e.g., *fdgo* and *delaygo*) tasks promote the reuse and reinforcement of shared modules, thereby enhancing representational robustness and integration. Together, these patterns demonstrate how task interactions can shape the modular organization of neural representations, either by driving the emergence of distinct functional modules or by strengthening shared, adaptable ones.

Neuronal Similarity Analysis of Task Involvement

Building on the observation that task similarity influences co-embedding patterns, we further analyze how individual neurons are engaged across tasks. Inspired by representational similarity analysis (RSA) [26], we introduce a metric to quantify neuronal similarity based on activation patterns during task execution, thereby enabling a fine-grained assessment of shared versus task-specific neural involvement. Specifically, for each neuron u , we extracted its activity distributions across tasks and computed the pairwise correlation among these distributions to form a task representation similarity matrix (Fig. 3b). We then vectorized the upper triangular portion of each matrix and used it as a feature vector to quantify task activation similarity between

neurons. By clustering the neuronal representational similarity matrix defined in Fig. 3b, we grouped neurons with similar task representations into distinct communities (Fig. 3c). Early in training ($\leq$ iteration 500), a single dominant community emerged, indicating a homogeneous embedding of all tasks across neurons. As training progressed, this community gradually split into multiple modules (iterations 3500–15500), which further sharpened in contrast to non-community regions (iterations 18500–39500), eventually stabilizing into a consistent modular structure. These dynamics suggest that co-embedding of related tasks progressively drives modularization in the neuronal representation space, as measured by task activation similarity.

Modularity as an indicator of Model Performance

Previous analyses demonstrated that learning multiple tasks induces modularization in task-related neuronal representations, consistent with evidence linking modular organization to cognitive capacity in the brain [27]. Here, we investigate whether modularity can serve as a performance indicator, analogous to its role in explaining individual differences in human cognition on brain networks. Specifically, we calculated the correlation coefficients between the network modularity and model accuracy across 200 model instances along 20,000 training iterations (Fig. 3d). A significant positive correlation emerges during the formation of modular structures but tends to diminish in later stages. The strength of this relationship depends on the effectiveness of multi-task co-embedding, which is modulated by both task demands and representational capacity. Such variation in co-embedding effectiveness leads to clear differences between smaller models (e.g., 8 or 16 hidden neurons), where the correlation is more pronounced, and larger models (e.g., 32 or 64 hidden neurons), where the correlation weakens. For example, at the peak iteration for the 8-neuron models, we observed a significant positive correlation ($r = 0.34$, $p < 1.0 \times 10^{-6}$). Together with the findings in Fig. 2, these results suggest that when representational capacity is constrained, modularization emerges as a compensatory procedure to improve the neural efficiency in representation. In contrast, when the model’s capacity is sufficiently large, the model exhibits a certain degree of over-parameterization, endowing it with inherent robustness. In the late iterations of training, further perturbations in the weights have minimal impact on task performance, leading to a weakened correlation between modularity and performance.

Further Modularization from Connection-incremental Learning

The brain reaches its structural complexity through a gradual evolutionary trajectory rather than beginning with all its current neurons and a fully connected architecture. To investigate the impact of this developmental paradigm on modularization, we designed a connection-incremental learning strategy in which RNN models start with no recurrent connections—functionally equivalent to feed-forward networks (Fig. 4a top left)—and recurrent connections are progressively added as new tasks are introduced (Fig. 4b left; each colored line

denotes a task). Over time, the networks eventually result in sparse RNNs that resembles the structural sparsity observed in biological brain networks (Fig.4a top right). As a control, we also trained standard RNN models with the full set of recurrent connections initialized from the outset, matched in configuration to their incremental counterparts (Fig.4a bottom).

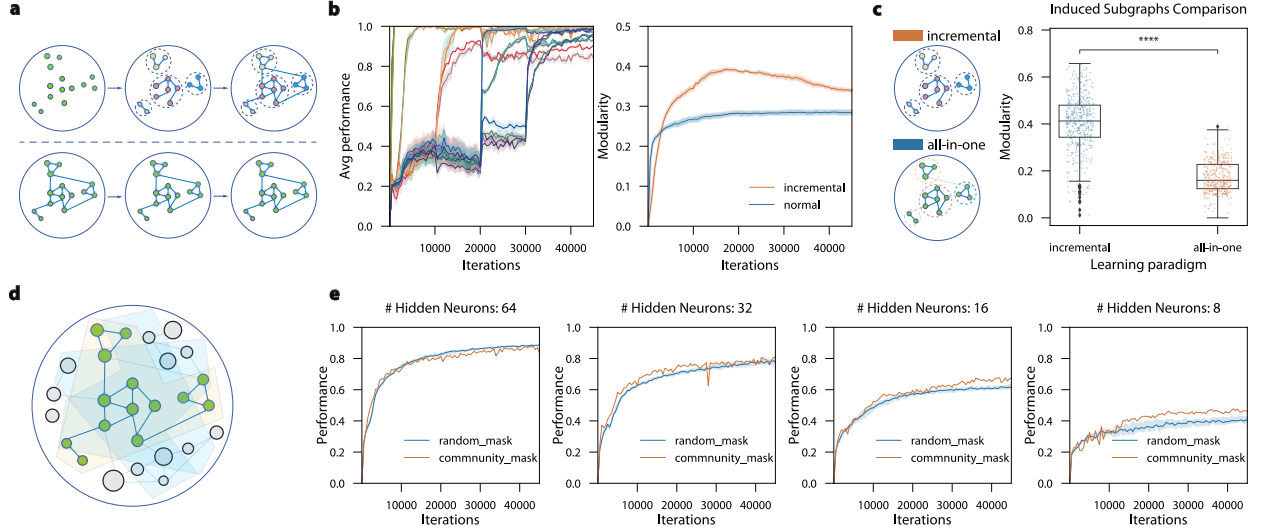


Fig. 4: Sequential introduction of tasks and incremental addition of connections during learning leads to higher modular structures and the lottery hypothesis for modular structures. **a**, Schematic of incremental addition of connections vs. all connections that exist at first. **b**, Tasks are introduced sequentially with an increasing number of connections. Each line with different color represents a task (left). The incremental models show a higher modularity than those trained all-at-once (right). **c**, In the process of dissecting the connected components emerged during the growing process of the incremental models, we found notably high modular subgraphs. **d**, Schematic of the masks of modular structures (orange) and of random structures (blue). **e**, Connections masked by the emerging modular structures after training are the *winning tickets* that refer to the subnetworks being able to reach comparable accuracy when trained in isolation by the *Lottery Ticket Hypothesis*. They can enhance the learning compared to connections masked randomly. Solid lines represent the median over 20 i.i.d. random masks, and shaded areas show the interquartile range (25th to 75th percentile).

The modularity of connection-incremental models consistently exceeds that of all-at-once training by a significant margin (Fig.4b right; $p < 0.05$ via a two-sample t -test at all selected iterations after 10000; see Fig. A1 for extended results). To examine the underlying structural change, we partitioned the incrementally trained networks into their evolving connected components and extracted their corresponding induced subgraphs from the all-at-once models for comparison (Fig. 4c). As shown in Fig. 4b, c, with each successive addition of connections, the gradual addition of recurrent connections leads to the emergence of distinct, isolated communities of hidden neurons, each exhibiting strong modular organization to accommodate the growing complexity of multi-task demands. This results in a more modular recurrent structure than that produced by all-at-once learning. These findings suggest that modularization is shaped not only by task co-embedding but also by the learning dynamics imposed by the connection-incremental paradigm.

Given the observed emergence of modularization in the multi-task learning process, a natural yet unanswered question is whether the induced modularized structure, in turn, computationally facilitates more efficient task embedding. Drawing inspiration from the so-called *lottery ticket hypothesis* in network pruning [28], we investigated this question by identifying the emergent communities after training (Fig. 4d, orange) and treating them as sub-networks for reinitialization and retraining. Specifically, we defined a *community mask* by preserving all the intra-community connections and the top 15% high-magnitude inter-community connections, thereby maintaining the interactions between different modules. As a control, we trained models with an equal number of randomly selected connections (*random mask*, Fig. 4d blue). The network initialized with the *community mask* exhibited superior performance over those marked by the *random mask* (Fig. 4e). Notably, this performance advantage was especially pronounced in models with smaller capacity, consistent with earlier results shown in Fig. 1 and Fig. 2. These findings support the functional utility of modularized structure under capacity constraints, while leaving an open question about their role in very large models with extensive task loadings.

Task-driven Adaption Enhances More Brain-like Structural Organization

The formation of brain network structure has been extensively studied, ranging from graph-based approaches [8] to the RNN-based spatial embedding [6]. In the previous sections, we demonstrated that, beyond spatial constraints, functional adaptation to multi-task learning can also drive the emergence of modularized structure within networks. Thus, there arises a natural question: does task learning explain additional aspects of brain-like connectivity beyond those accounted for by spatial embedding alone? To answer this question, we employed generative network modeling [8] to reconstruct network connectivity based on unsupervised wiring rules and assessed whether the task loads induced more brain-like structural features. Specifically, here we compared four groups of generative models (Fig. 5b), each comprising 200 model instances. The ‘*random*’ models connect node pairs with equal probability. The ‘*spatial*’ models encode the biological wiring costs based on the Euclidean distances, serving as the baseline models. The ‘*task*’ models disregard the spatial constraints and define the feature kernel $K(u, v)$ using the functional connectivity matrices of the trained RNNs. The ‘*spatial+task*’ models integrate both spatial constraints and functional connectivity in the kernel.

As discussed in prior work [6, 8], the ‘*spatial*’ models, improve the fitness to empirical brain connectivity by favoring the short-range connections to enhance the local clustering following the nodal degree distribution (Fig. 5c). When applied separately, the ‘*task*’ models similarly outperform random baselines, indicating that task-induced functional organization also contributes to brain-like structure (Fig. 5c). To evaluate structural fidelity more precisely, we assessed four key topological properties: node degree (KS_k), clustering coefficient

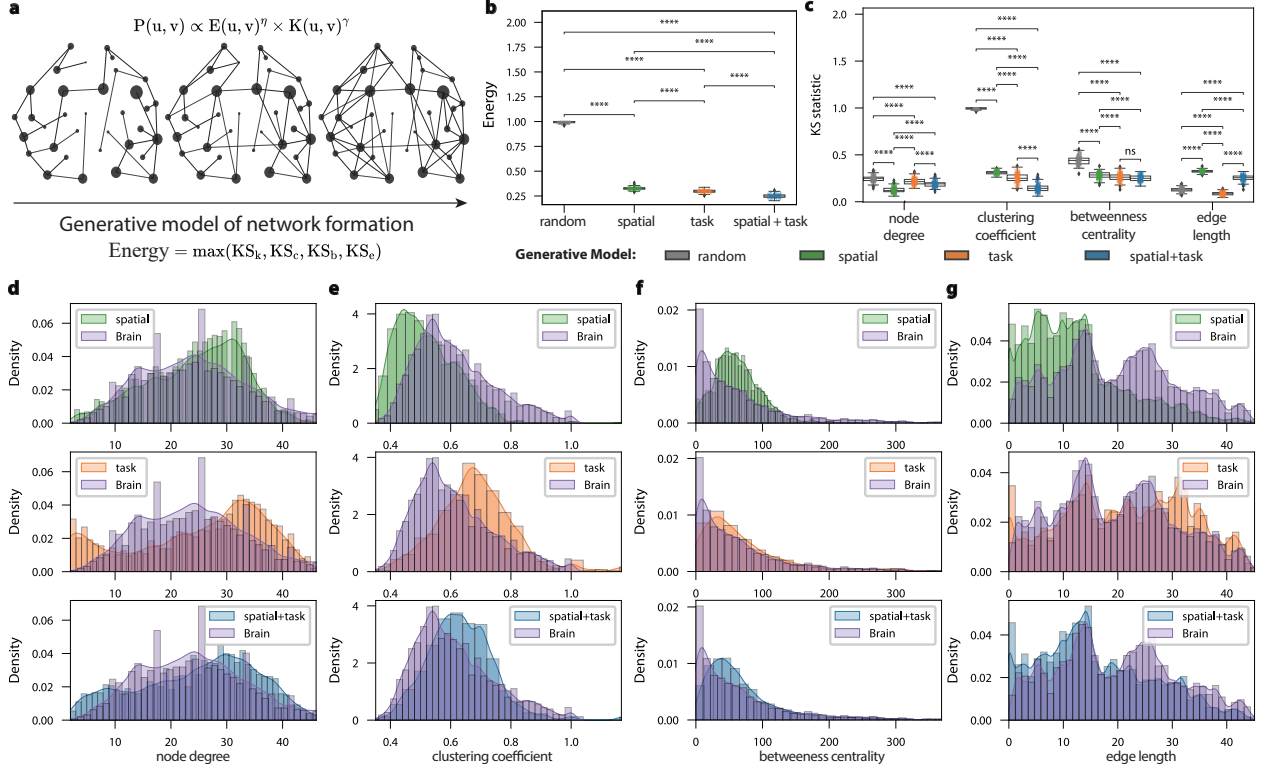


Fig. 5: Tasks induce more brain-like structures. **a**, The generative model characterizes the wiring of networks based on the Euclidean distance $E(u, v)$ and feature kernel $K(u, v)$. The energy function quantifies the dissimilarity between simulated and observed networks with Kolmogorov–Smirnov (KS) statistics. **b**, Energy distributions of 4 generative models. "Random" means the connections are randomly determined. "Spatial" means that the connections follow the distribution determined by the spatial constraints. "Task" means that the feature kernel $K(u, v)$ adopts the functional connectivity matrix during the inference of the tasks by the trained RNNs. Each box denotes the distribution of the lowest energy of synthetic networks. Each model generates 200 model instances. **c** The detailed distribution of Kolmogorov–Smirnov statistics for each argument in the energy form defined in panel **a**. **d–f**, Detailed comparisons of the distributions with respect to the graph measurements in panel **c**. As a consequence, the task-induced generative models provide better predictions of the long-range edges and high-clustered structures.

259 (KS_c), betweenness centrality (KS_b) and edge length (KS_e). For node degree, the 'spatial' models achieved
 260 strong alignment with empirical data. For clustering, 'spatial' models generated networks with slightly lower
 261 clustering than the real brain, while task-induced networks shifted higher, likely reflecting the increased local
 262 connectivity required to support task co-embedding, as previously discussed. For betweenness centrality, both
 263 'spatial' and 'task' models produced positively biased fits. For the edge length, spatial models exhibited a
 264 negative bias due to their preference for short connections, whereas task-based models captured more long-
 265 range connectivity and even reproduced the brain's characteristic multi-peak distribution. This pattern of
 266 findings suggests that long-range connections in the brain might support not only small-world organization
 267 [29] but also functional integration demanded by task load [30, 31]. Taken together, these findings indicate
 268 that combining spatial constraints with task-driven connectivity in the 'spatial+task' model can result in
 269 superior structural fidelity. Indeed, 'spatial+task' models significantly outperformed spatial-only models

($t = 47.04$, $p < 1 \times 10^{-163}$, Fig. 5b) and task-only models ($t = 29.20$, $p < 1 \times 10^{-99}$, Fig. 5b). These results support a complementary view of the brain: not only as a dynamical system shaped by physical constraints, but also as an information-processing system optimized for functional demands, as exemplified by RNNs trained on multi-task learning.

Crucially, they suggest that modularization in artificial networks, emerging through task-driven adaptation, may reflect a computational principle underlying the formation of brain-like connectivity, offering valuable insights for both neuroscience and neuroAI system design.

Discussion

In this work, we investigated the emergence of modularization in recurrent neural networks (RNNs) during multi-task learning. We first showed that modularization is a general phenomenon associated with the embedding of input features into the network’s internal representation. Beyond feature-level organization, the demand to encode multiple tasks introduces additional representational constraints, which further increases the modularity of the trained network. Furthermore, introducing tasks gradually via a staged unfreezing of connections promotes even stronger modular structure, highlighting the influence of learning dynamics on architectural organization. Notably, networks shaped by multi-task-induced constraints exhibit greater structural similarity to empirical brain networks than those trained under static structural constraints alone.

It has been an emerging trend in both neuroscience and AI to model the brain’s computational and cognitive function using recurrent neural networks (RNNs). The existing research approaches this bidirectionally. On one hand, RNNs are used to replicate the brain system to empower the modeling of internal neural dynamics [32–34] and the generation of external behavior aligned with experimental observations [9, 35, 36]. On the other hand, RNNs can also serve as virtual experimental platforms to assess whether specific brain mechanisms facilitate task learning [37, 38] or emerge from the physical and structural constraints [6, 36]. Our work spans both perspectives: it models internal dynamics while also demonstrating how adaptation to external task demands induces modular organization in network connectivity. While prior studies on image classification with network architecture search suggest that the feature representation alone may not give rise to the brain-like topology [39], our findings, alongside recent ones on familiarity detection [13], functional specialization [37], and spatial modularization [40], support a task-driven view of brain-like organization. Our findings also align with recent perspectives in cognitive neuroscience that emphasize the role of task demands and generalization [35], suggesting that the modularization and other topological features in the brain may emerge to support complex task environments, including multitasking and continual learning, over the long-term evolutionary scales.

301 Network neuroscience traditionally views the brain as a dynamical system governed by physical con-
 302 straints [10]. Combined with brain imaging techniques, this methodology enables researchers to infer brain
 303 properties and map external behaviors onto internal network topologies [41–43]. However, such investigations
 304 are often correlational rather than causal in nature [44] as there is a fundamental gap in the methodolo-
 305 gies between the dynamical system and the information processing system [45]. Specifically, the dynamical
 306 system pays attention to the physical constraints like spatial embedding and metabolic cost [46], while the
 307 information processing system focuses on the representational efficiency and computational utility [40, 47].
 308 This distinction raises important questions about the limitations of correlational analyses in revealing how
 309 neural structures emerge or adapt. By contrast, computational models such as RNNs offer a tractable testbed
 310 for probing how functional demands shape network organization, enabling causal insights that complement
 311 empirical observations. In this sense, our work contributes to bridging the theoretical gap by demonstrating
 312 how modularity can arise as an emergent property of task optimization under representational constraints.
 313 In the context of multi-task learning, both cognitive neuroscience [48, 49] and RNN-based task modeling
 314 [9] have recognized the compositional nature of task representation. This compositionality gives rise to the
 315 hierarchical modular representation in the task space, which may in turn induce the modularization in brain
 316 networks, a pattern that echoes the task-cortical mappings analogous to the well-studied retino-cortical map
 317 [50, 51]. Such a view, along with its underlying rationale, provides a causal explanation for why modular-
 318 ity and its variation across learning can be informative for understanding both the acquisition of cognitive
 319 functions [13, 52] and their changing trajectory through neurodevelopment [3, 4].

320 Despite providing a causal explanation of how multitask learning induces modularization in network
 321 topology, some limitations remain, both theoretically and practically. Although we compare the model per-
 322 formances with regard to the generative model [7, 53], the proposed protocol here is not fully biologically
 323 validated and remains a simplified abstraction. The use of RNNs enables tractable analysis of modularity
 324 emergence, but both the human brain and the modern foundation models [54] are far more complex, wherein
 325 there may be no single principle to explain the generation of intelligence on both sides. Thus, it remains to
 326 be explored whether such investigation can be extended to explain other characteristics of brain function
 327 and yield generally efficient machine learning algorithms.

328 In conclusion, in spite of these limitations, our work contributes a function-to-structure perspective,
 329 complementing traditional structure-to-function approaches, and offers a pathway for understanding how the
 330 brain’s specialized topology may emerge through task-driven adaptation. Future progress along this direction
 331 may help illuminate why the brain is organized as it is and how its principles can inform the design of more
 332 effective computational systems.

333 Materials and Methods

334 Recurrent Neural Networks and Cognitive Tasks

335 We adopt a recurrent neural network whose network activities $\mathbf{r}$ follow a continuous dynamical equation:

$$\tau \frac{d\mathbf{r}}{dt} = -\mathbf{r} + f(\mathbf{W}^{rec}\mathbf{r} + \mathbf{W}^{in}\mathbf{u} + \mathbf{b} + \sqrt{2\tau\sigma_{rec}^2}\boldsymbol{\epsilon}) \quad (1)$$

336 where τ is a time constant(100 ms), $\mathbf{W}^{rec}$ are the weights of the recurrent layer, $\mathbf{W}^{in}$ are the weights of the
337 input layer, $\mathbf{u}$ is the input to the model, $\mathbf{b}$ is the bias, $f(\cdot)$ is the nonlinearity activation (e.g., ReLU function
338 [55]), and $\sqrt{2\tau\sigma_{rec}^2}\boldsymbol{\epsilon}$ is the Gaussian white noise. The neurons of the output layer read out the network
339 activity $\mathbf{r}$ nonlinearly as $\mathbf{o} = g(\mathbf{W}^{out}\mathbf{r})$, where $g(\cdot)$ is the sigmoid function, and $\mathbf{W}^{out}$ are the weights of the
340 output layer. For each task, the RNN receives inputs of three types: fixation, stimulus, and rule (Fig.1a).
341 The fixation input instructs the network to ‘fixate’ or respond (e.g., ‘saccade’). When the fixation input is
342 turned off, the network receives a ‘go’ signal. The stimulus inputs consist of two modalities, each represented
343 by a ring of input units that encode a one-dimensional circular variable that indicates the motion direction of
344 color on a color wheel [56]. Task identity is provided via a one-hot encoded rule input, in which a single unit
345 is activated per trial to instruct the network on which task to perform. The network is trained to generate
346 output through a fixation unit and a set of motor output units arranged in a circular array, representing
347 response direction. For more details on the task RNNs and the full task set implementation, please refer to
348 the prior work [9].

349 Modularization

350 The modularity statistic Q quantifies the extent to which the network can be subdivided into clearly delin-
351 eated groups. Here, we employ a widely used community discovery technique on the recurrent weight matrix
352 to calculate its modularity value [25], which can be written as

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i^{in} k_j^{out}}{2m} \right] \delta(c_i, c_j) \quad (2)$$

353 where A_{ij} denotes the absolute value of W_{ij} (an element of the recurrent weight matrix of the RNN model),
354 $2m = \sum_i (k_i^{in} + k_j^{out})$ denotes the overall weight, δ is the Kronecker delta symbol, and c_i is the label of the
355 community to which vertex i is assigned.

356 Representational Similarity between Neurons

357 In order to characterize the representational similarity of neurons, we conduct a tailored two-level represen-
358 tational similarity analysis [26] during the training process. On the first level, we calculate the similarity
359 between tasks based on the activation distribution, following a similar approach in [9]. On the second level,
360 we further model the similarity of neurons based on their similarity matrices over different tasks. Specifi-
361 cally, we employ an RNN model to perform the 20 tasks and partition the task-specific activity distribution
362 of each neuron into 50 bins as $\mathbf{dist}_i^u = \text{histogram}(\mathbf{A}_i^u, 50)$, where $\mathbf{A}_i^u = [a_i^u(1), \dots, a_i^u(t), \dots, a_i^u(T)]$ and $a_i^u(t)$
363 denotes the activation value of neuron u at the t -th time step of task i , and $\text{histogram}(\cdot)$ is a function that
364 counts how many activation values fall into each of the 50 bins. Therefore, the $\mathbf{dist}_i^u$ is a 50-element vector
365 representing the activity pattern of neuron u during the execution of the i -th task. Next, for each neu-
366 ron u , we compute the task similarity matrix as the correlation matrix $\mathbf{S}^u$, where each element represents
367 the Pearson’s correlation coefficient between two task-specific activity patterns, i.e. $S_{ij}^u = \text{corr}(\mathbf{dist}_i^u, \mathbf{dist}_j^u)$.
368 This resulting correlation matrix $\mathbf{S}^u$ provides structural insights into the specialization of neuron u across
369 tasks, capturing its neural relationships for different task pairs. We vectorize its upper triangle to serve as
370 the signature vector of neuron u . Building on this approach, we consequently define the similarity between
371 two neurons as the Pearson correlation coefficient between their signature vectors,

$$\text{Sim}(u, v) = \text{corr}(\text{vectorize}(\text{triu}(\mathbf{S}^u)), \text{vectorize}(\text{triu}(\mathbf{S}^v))) \quad (3)$$

372 where $\text{triu}(\mathbf{S}^u)$ denotes the upper triangle of matrix $\mathbf{S}^u$, and $\text{vectorize}(\cdot)$ is a function that flattens the matrix
373 into a vector. The element $\text{Sim}(u, v)$ then captures the similarity between neurons u and v with respect
374 to their neuronal signatures. By applying a community discovery technique like Eqn. 2, we can group the
375 neurons into different communities based on their similarities and analyze the entire training process.

376 Connection-incremental Learning

377 To implement the connection-incremental learning shown in Fig. 4a, we first generate a series of binary masks,
378 each containing an increasing number of connections, with 25 new connections added every 500 iterations,
379 culminating in a total of 2000 connections. For the regular models, we apply the final mask from the series to
380 the recurrent layer, ensuring that the set of connections matches the final set obtained from the connections-
381 incremental learning. In line with the setting outlined in 7, the weights of nonexistent connections are
382 scaled down to 0.1 of their initial values to serve as background, thereby preventing numerical issues when

calculating modularity. During forward propagation, the aforementioned mask is applied to the recurrent connection weights, allowing only the existing connections to facilitate information flow.

Generative Network Modeling

Generative network models have been employed in recent years to validate whether a wiring rule can be used to simulate an empirical network topology. The idea is to start from an empty network and at each step we allow for the possibility that any pair of unconnected nodes, u and v , will be connected, where the relative probability of connection formation is given by

$$P(u, v) = E(u, v)^\eta \times K(u, v)^\gamma \quad (4)$$

where $E(u, v)$ denotes the Euclidean distance and $K(u, v)$ denotes a feature kernel characterizing the non-geometric relationship between nodes u and v , the exponent η controls the characteristic connection length, and the value of γ scales its relative importance. In this work, we include four generative models. For the ‘*random*’ models, we set both $E(u, v)$ and $K(u, v)$ to be the identity matrices. For the ‘*spatial*’ models, $E(u, v)$ is assigned as the Euclidean distance between the centers of different cortex regions, while $K(u, v)$ is set to the identity matrix. In the ‘*task*’ models, $E(u, v)$ is set to the identity matrix and $K(u, v)$ is assigned as the functional connectivity of the trained RNN’s recurrent layer. For the ‘*spatial+task*’ models, we set $E(u, v)$ as the Euclidean distance between the centers of different cortex regions and $K(u, v)$ as the functional connectivity of the trained RNN’s recurrent layer.

Functional Connectivity of Network Activity

Analogous to brain networks, to obtain functional connectivity as demonstrated in previous work [13], we collected the activation from the recurrent layer in response to task inference. This process allows us to construct functional networks by analyzing correlations within time windows. To be specific, we first train an RNN while considering the communication cost, where the optimization term simulates the cost of information flow between neurons spatially embedded in a 3D Euclidean space:

$$L = L_{task} + \gamma \sum_{t=1}^T \sum_{u=1}^N \sum_{v=1}^N (\text{Distance}(u, v) \cdot W^{rec}(u, v) \cdot a^u(t)) \quad (5)$$

where γ is set to 1×10^{-6} , and u, v denote two neurons separated by a distance $\text{Distance}(u, v)$, and $a^u(t)$ denotes the activation of neuron u at the t -th time step. After training, we utilize the obtained RNN to

perform the tasks and collect its neural activity during inference. Specifically, we record the activity of the hidden layer comprising N neurons over T time steps for a specific task, represented as $\mathbf{A}_H = \{a^u(t)\}$ with dimension $N \times T$. The functional connectivity is then defined as the Pearson correlation coefficient for each row in the $\mathbf{A}_H$, resulting a connection matrix of dimension $N \times N$. Given that the correlation values fall within the interval $[-1, 1]$, we adapted the matrix as a feature kernel in Eqn. 4 by replacing all negative values with zeros. This adjusted functional connectivity enables us to infer functional interactions among neurons embedded across space, mimicking the interplay observed in various regions of the cortex. Building on this approach, we derived the ‘*task*’ and ‘*spatial+task*’ models shown in Fig. 5.

Data and Code Availability

The structural connectivity data used in this work were derived from the Human Connectome Project’s (HCP) [57]. We selected only 84 cortical areas and obtained the mean distance matrix between these regions, as well as the structural connectivity matrix. The code for the analysis is currently available at <https://github.com/brain-intelligence-lab/modular-brain-rnn> and will be moved to a permanent address upon acceptance.

Acknowledgements

Funding: S.G. is supported by NSFC Key Program 62236009, General Program 61876032, and Fundamental Research Funds for the Central Universities. G.P. is supported by the National Key R&D Plan of China (2022YFB4500100), Natural Science Foundation of China (61925603), Major Program of Natural Science Foundation of Zhejiang Province in China (LDQ23F040001).

Competing interests: The authors declare that they have no competing interests.

Author contributions: Conceptualization, methodology, investigation, and visualization: S.G., G.P., Y.H.W. and Y.F.W. Writing—original draft: S.G., Y.H.W., Y.F.W., and M.G.M. Writing—review and editing: S.G., D.S.B., Y.H.W., Y.F.W., M.G.M., S.D., K.D., H.T., and G.P.

References

- [1] Sporns, O. & Betzel, R. F. Modular brain networks. *Annual review of psychology* **67**, 613–640 (2016).
- [2] Power, J. D. *et al.* Functional network organization of the human brain. *Neuron* **72**, 665–678 (2011).
- [3] Baum, G. L. *et al.* Modular segregation of structural brain networks supports the development of executive function in youth. *Current Biology* **27**, 1561–1572 (2017).
- [4] Gu, S. *et al.* Emergence of system roles in normative neurodevelopment. *Proceedings of the National Academy of Sciences* **112**, 13681–13686 (2015).
- [5] Braun, U. *et al.* Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences* **112**, 11678–11683 (2015).
- [6] Achterberg, J., Akarca, D., Strouse, D., Duncan, J. & Astle, D. E. Spatially embedded recurrent neural networks reveal widespread links between structural and functional neuroscience findings. *Nature Machine Intelligence* **5**, 1369–1381 (2023).
- [7] Akarca, D., Vértes, P. E., Bullmore, E. T. & Astle, D. E. A generative network model of neurodevelopmental diversity in structural brain organization. *Nature communications* **12**, 4216 (2021).
- [8] Betzel, R. F. *et al.* Generative models of the human connectome. *Neuroimage* **124**, 1054–1064 (2016).
- [9] Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T. & Wang, X.-J. Task representations in neural networks trained to perform many cognitive tasks. *Nature neuroscience* **22**, 297–306 (2019).
- [10] Bassett, D. S. & Sporns, O. Network neuroscience. *Nature neuroscience* **20**, 353–364 (2017).
- [11] Pagani, M., Gutierrez-Barragan, D., de Guzman, A. E., Xu, T. & Gozzi, A. Mapping and comparing fmri connectivity networks across species. *Communications Biology* **6**, 1238 (2023).
- [12] Brynildsen, J. K., Rajan, K., Henderson, M. X. & Bassett, D. S. Network models to enhance the translational impact of cross-species studies. *Nature Reviews Neuroscience* **24**, 575–588 (2023).
- [13] Gu, S., Mattar, M. G., Tang, H. & Pan, G. Emergence and reconfiguration of modular structure for artificial neural networks during continual familiarity detection. *Science Advances* **10**, eadm8430 (2024).

- 456 [14] Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's
457 dorsolateral prefrontal cortex. *Journal of neurophysiology* **61**, 331–349 (1989).
- 458 [15] Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574
459 (2007).
- 460 [16] Siegel, M., Buschman, T. J. & Miller, E. K. Cortical information flow during flexible sensorimotor
461 decisions. *Science* **348**, 1352–1355 (2015).
- 462 [17] Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent
463 dynamics in prefrontal cortex. *nature* **503**, 78–84 (2013).
- 464 [18] Raposo, D., Kaufman, M. T. & Churchland, A. K. A category-free neural population supports evolving
465 demands during decision-making. *Nature neuroscience* **17**, 1784–1792 (2014).
- 466 [19] Romo, R., Brody, C. D., Hernández, A. & Lemus, L. Neuronal correlates of parametric working memory
467 in the prefrontal cortex. *Nature* **399**, 470–473 (1999).
- 468 [20] Munoz, D. P. & Everling, S. Look away: the anti-saccade task and the voluntary control of eye movement.
469 *Nature Reviews Neuroscience* **5**, 218–228 (2004).
- 470 [21] Miller, E. K., Erickson, C. A. & Desimone, R. Neural mechanisms of visual working memory in prefrontal
471 cortex of the macaque. *Journal of neuroscience* **16**, 5154–5167 (1996).
- 472 [22] Freedman, D. J. & Assad, J. A. Neuronal mechanisms of visual categorization: an abstract view on
473 decision making. *Annual review of neuroscience* **39**, 129–147 (2016).
- 474 [23] Kashtan, N. & Alon, U. Spontaneous evolution of modularity and network motifs. *Proceedings of the*
475 *National Academy of Sciences* **102**, 13773–13778 (2005).
- 476 [24] Löwel, S. & Singer, W. Selection of intrinsic horizontal connections in the visual cortex by correlated
477 neuronal activity. *Science* **255**, 209–212 (1992).
- 478 [25] Leicht, E. A. & Newman, M. E. Community structure in directed networks. *Physical review letters*
479 **100**, 118703 (2008).
- 480 [26] Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis-connecting the
481 branches of systems neuroscience. *Frontiers in systems neuroscience* **2**, 249 (2008).

482 [27] Wig, G. S. Segregated systems of human brain networks. *Trends in cognitive sciences* **21**, 981–996
483 (2017).

484 [28] Frankle, J. & Carbin, M. The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv*
485 *preprint arXiv:1803.03635* (2018).

486 [29] Bassett, D. S. & Bullmore, E. T. Small-world brain networks revisited. *The Neuroscientist* **23**, 499–516
487 (2017).

488 [30] Vatansever, D., Menon, D. K., Manktelow, A. E., Sahakian, B. J. & Stamatakis, E. A. Default mode
489 dynamics for global functional integration. *Journal of Neuroscience* **35**, 15254–15262 (2015).

490 [31] Nour, M. M. *et al.* Task-induced functional brain connectivity mediates the relationship between striatal
491 d2/3 receptors and working memory. *Elife* **8**, e45045 (2019).

492 [32] Kim, J. Z., Lu, Z., Nozari, E., Pappas, G. J. & Bassett, D. S. Teaching recurrent neural networks to
493 infer global temporal structure from local examples. *Nature Machine Intelligence* **3**, 316–323 (2021).

494 [33] Durstewitz, D., Koppe, G. & Thurm, M. I. Reconstructing computational system dynamics from neural
495 data with recurrent neural networks. *Nature Reviews Neuroscience* **24**, 693–710 (2023).

496 [34] Driscoll, L. N., Shenoy, K. & Sussillo, D. Flexible multitask computation in recurrent networks utilizes
497 shared dynamical motifs. *Nature Neuroscience* **27**, 1349–1363 (2024).

498 [35] Nau, M., Schmid, A. C., Kaplan, S. M., Baker, C. I. & Kravitz, D. J. Centering cognitive neuroscience
499 on task demands and generalization. *Nature Neuroscience* 1–12 (2024).

500 [36] Fascianelli, V. *et al.* Neural representational geometries reflect behavioral differences in monkeys and
501 recurrent neural networks. *Nature Communications* **15**, 6479 (2024).

502 [37] Dobs, K., Martinez, J., Kell, A. J. & Kanwisher, N. Brain-like functional specialization emerges
503 spontaneously in deep neural networks. *Science advances* **8**, eabl8913 (2022).

504 [38] Rajalingham, R., Piccato, A. & Jazayeri, M. Recurrent neural networks with explicit representa-
505 tion of dynamic latent variables can mimic behavioral patterns in a physical inference task. *Nature*
506 *Communications* **13**, 5865 (2022).

507 [39] Xie, S., Kirillov, A., Girshick, R. & He, K. Exploring randomly wired neural networks for image
508 recognition. *Proceedings of the IEEE/CVF International Conference on Computer Vision* 1284–1293
509 (2019).

510 [40] Pfeiffer, J., Ruder, S., Vulić, I. & Ponti, E. M. Modular deep learning. *Transactions on Machine*
511 *Learning Research* (2024).

512 [41] Medaglia, J. D., Lynall, M.-E. & Bassett, D. S. Cognitive network neuroscience. *Journal of cognitive*
513 *neuroscience* **27**, 1471–1491 (2015).

514 [42] Bassett, D. S., Zurn, P. & Gold, J. I. On the nature and use of models in network neuroscience. *Nature*
515 *Reviews Neuroscience* **19**, 566–578 (2018).

516 [43] Barbey, A. K. Network neuroscience theory of human intelligence. *Trends in cognitive sciences* **22**,
517 8–20 (2018).

518 [44] Bailey, D. H. *et al.* Causal inference on human behaviour. *Nature Human Behaviour* **8**, 1448–1459
519 (2024).

520 [45] Man, D. & Vision, A. A computational investigation into the human representation and processing of
521 visual information. *WH San Francisco: Freeman and Company, San Francisco* **1**, 1 (1982).

522 [46] Lynn, C. W. & Bassett, D. S. The physics of brain network structure, function and control. *Nature*
523 *Reviews Physics* **1**, 318–332 (2019).

524 [47] Kanwisher, N., Khosla, M. & Dobs, K. Using artificial neural networks to ask ‘why’ questions of minds
525 and brains. *Trends in Neurosciences* **46**, 240–254 (2023).

526 [48] Cole, M. W., Etzel, J. A., Zacks, J. M., Schneider, W. & Braver, T. S. Rapid transfer of abstract rules
527 to novel contexts in human lateral prefrontal cortex. *Frontiers in human neuroscience* **5**, 142 (2011).

528 [49] Cole, M. W. *et al.* Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature*
529 *neuroscience* **16**, 1348–1355 (2013).

530 [50] St-Yves, G., Allen, E. J., Wu, Y., Kay, K. & Naselaris, T. Brain-optimized deep neural network models
531 of human visual areas learn non-hierarchical representations. *Nature communications* **14**, 3329 (2023).

- 532 [51] Li, Y., Yang, H. & Gu, S. Enhancing neural encoding models for naturalistic perception with a multi-level
533 integration of deep neural networks and cortical networks. *Science Bulletin* (2024).
- 534 [52] Bassett, D. S. *et al.* Dynamic reconfiguration of human brain networks during learning. *Proceedings of*
535 *the National Academy of Sciences* **108**, 7641–7646 (2011).
- 536 [53] Betzel, R. F. & Bassett, D. S. Generative models for network neuroscience: prospects and promise.
537 *Journal of The Royal Society Interface* **14**, 20170623 (2017).
- 538 [54] Achiam, J. *et al.* Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).
- 539 [55] Agarap, A. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375* (2018).
- 540 [56] Chaisangmongkon, W., Swaminathan, S. K., Freedman, D. J. & Wang, X.-J. Computing by robust
541 transience: how the fronto-parietal network performs sequential, category-based decisions. *Neuron* **93**,
542 1504–1517 (2017).
- 543 [57] Schaefer, A. *et al.* Local-global parcellation of the human cerebral cortex from intrinsic functional
544 connectivity mri. *Cerebral cortex* **28**, 3095–3114 (2018).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement.pdf](#)