

Review

Understanding cognitive processes across spatial scales of the brain

Hayoung Song^{1,2,5,*}, JeongJun Park^{2,5,*}, and Monica D. Rosenberg^{1,3,4,*}

Cognition arises from neural operations at multiple spatial scales, from individual neurons to large-scale networks. Despite extensive research on coding principles and emergent cognitive processes across brain areas, investigation across scales has been limited. Here, we propose ways to test the idea that different cognitive processes emerge from distinct information coding principles at various scales, which collectively give rise to complex behavior. This approach involves comparing brain–behavior associations and the underlying neural geometry across scales, alongside an investigation of global and local scale interactions. Bridging findings across species and techniques through open science and collaborations is essential to comprehensively understand the multiscale brain and its functions.

Cognition emerges from neural operations at multiple spatial scales

The brain operates as a multiscale system, with cognitive processes represented across various spatial scales, from individual neurons and synapses to regions and large-scale networks [1]. For instance, episodic memory is encoded not only in the electrical and chemical signals at neuronal synapses [2], but also in the firing of single neurons in the hippocampus [3], patterns of hippocampal neuronal populations [4], the communication between the hippocampus and neighboring cortical areas, such as the entorhinal cortex [5–7], and larger cortical networks, including the default mode network [8,9] (Figure 1A). This raises the question of how to determine the neural scale (or scales) that best explains a cognitive process of interest. For example, observations that a particular brain area represents a cognitive process could be the result of brain-wide state changes only partially captured by focusing on an isolated brain region. Similarly, identifying a brain-wide **representation** (see Glossary) of a cognitive process does not preclude the possibility of a specific area or circuit being the causal source.

Scientists across fields have argued that phenomena at different scales are governed by different principles [10]. In neuroscience, this is reflected in the view that different principles of neural coding operate at various spatial scales of the brain, leading to different cognitive functions emerging from each scale [11]. Parallel processes at different brain scales are thought to collectively give rise to complex behavior [12–14]. In practice, however, direct empirical tests of this perspective have been limited because research in each subfield typically associates each cognitive process with a specific neural scale. This limitation arises because neuroscience studies must choose model organisms and techniques, which inherently constrains the spatiotemporal scale of the investigation. For example, human neuroimaging captures indirect, slow-fluctuating signals of the whole brain, whereas nonhuman animal electrophysiology records fast spiking activity of neurons in a local brain area. This disparity has made it challenging to integrate findings across scales.

Decades of research that investigated the functional roles of the brain and its parts have given us preliminary ideas about which cognitive processes are associated with which neural scales. As examples, sensation and motor actions are relatively localized to unimodal areas, with individual

Highlights

Different neural coding principles operate at various spatial scales of the brain, contributing to different cognitive functions and collectively generating complex behavior.

The association between neural scales and cognitive processes can be assessed using encoding and decoding models, and model performance can be compared across scales.

The geometry of neural representations and dynamics reveals information coding principles at each scale, which can be compared across scales.

How global and local neural scales and their interaction generate cognitive processes can be understood both biologically and computationally.

A comprehensive understanding of neural scales and their functions comes from theoretically bridging findings across model species and techniques, together with open science practices and cross-lab collaborations.

¹Department of Psychology, University of Chicago, Chicago, IL, USA

²Department of Neuroscience, Washington University School of Medicine, St Louis, MO, USA

³Neuroscience Institute, University of Chicago, Chicago, IL, USA

⁴Institute for Mind and Biology, University of Chicago, Chicago, IL, USA

⁵These authors contributed equally

*Correspondence: hayoung@wustl.edu (H. Song), jeongjun@wustl.edu (J. Park), and mdrosenberg@uchicago.edu (M.D. Rosenberg).

neurons selectively tuned to specific features [15–17]. By contrast, vigilance and arousal are reflected in large-scale **brain state** changes [18–22], instantiated by ascending neuromodulatory signals [23,24]. However, even for these exemplar processes that appear to lie at the ends of the local-to-global scale spectrum, the answer is not so simple when we consider various behavioral contexts and interactions with other cognitive processes.

Thus, there remains a need to directly investigate neural operations and compare their functions across spatial scales. This is timely due to the more precise and larger-scale neuroimaging and electrophysiological recording techniques that have become feasible in recent years (e.g., 7-Tesla neuroimaging in humans [25], cellular resolution brain-wide calcium imaging [26], and simultaneous multi-site recordings using Neuropixel probes [27,28]). Technological advances have been accompanied by progress in computational neuroscience, which has examined multiscale interaction through models and simulations [29,30] and introduced new analytical approaches to investigating neural codes at each scale [31–33]. The neuroscience field has further benefited from increased data sharing through open science practices [34–36] and cross-lab collaborations [37], which allow researchers to work with multimodal data spanning diverse scales.

Here, we address ways in which the field can study the brain's cognitive functions at multiple scales (Figure 1). As a first step, we can identify neural scales relevant to cognitive processes by associating each process with neural features extracted from various spatial scales using **encoding or decoding models** and systematically comparing model performance across scales. Second, investigating the geometry of neural representations and dynamics at each scale, and comparing them across scales, will provide a more mechanistic understanding of neural coding principles. Third, exploring how multiple scales interact, including how circuits operate within larger systems and how their operations emerge from collections of neurons and synapses, will inform understanding of the multiscale brain. Finally, bridging findings across species and techniques and fostering open science and cross-lab collaborations are critical for theoretical advances. Interpretation of findings should be confined to the probed scale, and generalization to other scales should be cautioned unless directly tested. We use evidence from neuroimaging and electrophysiological studies in behaving humans, nonhuman primates, and rodents to illustrate how such cross-scale research can be conducted.

Comparison of brain–behavior associations across scales

Evidence that different cognitive processes are associated with different neural scales emerged from predictive modeling, including encoding and decoding approaches. A recent cross-lab collaboration recorded activity from hundreds of thousands of neurons across 267 brain areas as mice performed a visual decision-making task [37]. The study found that neurons in nearly all regions responded nonspecifically and simultaneously when mice initiated actions, received rewards, and engaged in a task, paralleling distributed representation found in human fMRI studies [38–41]. Similarly, visual stimuli and upcoming choices were represented in distributed areas of the brain. However, these processes were more strongly encoded in specific areas and propagated to other areas with temporal delays, which was obscured in human fMRI studies due to a slower time resolution. These findings, replicated in complementary work [27,42–44], indicate that the relevant scale of neural correlate varies across cognitive processes. Some processes, such as stimulus perception and choice behavior, are primarily computed in local brain circuits and sequentially relayed to distributed parts of the brain to support complex behavior. Others, such as motor action, reward processing, and engagement, are represented simultaneously at a whole-brain scale in rodents, potentially modulated by neuromodulatory tone [45].

Glossary

Brain state: recurring pattern of brain activity or functional coupling that is relevant to physiological or cognitive processes.

Causal mechanism: explanation of specific cause-and-effect relationships between neural activity and cognitive or behavioral outcomes.

Descriptive mechanism: explanation of how information is processed in the brain to produce cognitive or behavioral outcomes, without necessarily identifying the underlying reasons.

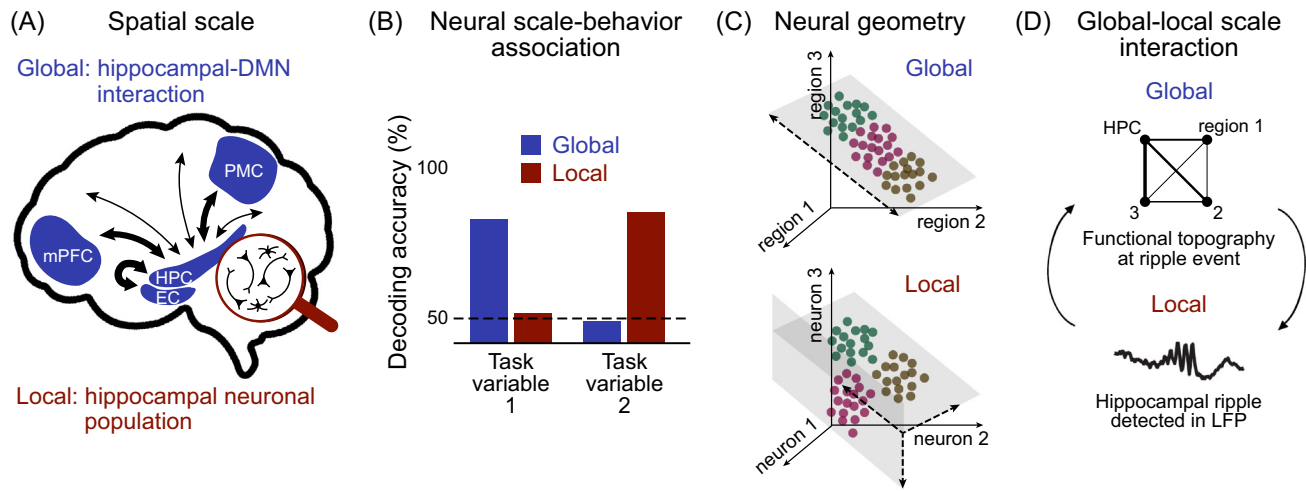
Decoding model: cross-validated model that predicts task-related variables, cognitive states, or behavior based on patterns of neural activity.

Encoding model: cross-validated model that predicts brain activity based on given stimuli or input.

High-dimensional codes: representations that rely on relatively many latent variables due to less correlated and more variable neural population activity.

Low-dimensional codes: representations that rely on relatively few latent variables due to highly correlated and redundant neural population activity.

Representation: neural signal that carries meaningful information about internal states or the external world to guide behavior.



Trends in Cognitive Sciences

Figure 1. Schematic overview of multiscale investigation. (A) Illustration of hippocampus (HPC) function across spatial scales. The function of the HPC can be studied at the scales of individual neurons, synaptic interactions between pairs of neurons, neuronal population activity, communication within hippocampal subfields (e.g., CA1–CA3–dentate gyrus; anterior-to-posterior gradient), communication within hippocampal subfields (e.g., CA1–CA3–dentate gyrus; anterior-to-posterior gradient), communication with the entorhinal cortex (EC) or broader default mode network (DMN) areas, including the medial prefrontal cortex (mPFC) and the posterior medial cortex (PMC), and even interaction with macroscale brain-wide states. Among this wide range of spatial scales, example global and local scale neural signals are noted in blue and red, respectively. Arrows indicate bidirectional interactions centered around the HPC, with the thickness representing the relative strength of these interactions. (B) Comparisons of decoding model performance of global and local neural signals with respect to two example task variables. The bar graph illustrates that the global signal is associated with task variable 1 but not 2, whereas the local signal is associated with task variable 2 but not 1. (C) Understanding neural processes at each scale by estimating latent geometry and dynamics across task conditions. The figure shows a state space represented by the activity of three regions (global; top) or neurons (local; bottom). Each dot represents population activity at a specified moment (e.g., a time point or a trial), with colors representing task conditions of different sensory, behavioral, or contextual variables. The top figure illustrates low-dimensional representation, where variance in neural data is explained by a 2D subspace (gray plane), with a single coding dimension to distinguish three conditions (dotted arrow). The bottom figure illustrates high-dimensional representation, where three coding dimensions are necessary to explain variances in neural data across task conditions. (D) Understanding the relationship between global and local neural signals. The bottom figure illustrates hippocampal ripple activity detected by the local field potentials (LFP) in the HPC. The global functional topography, or the coupling of the HPC with nearby DMN areas, during these ripple events can be statistically compared to the topography during periods when ripple event is not detected. Circles denote four regions, and lines connecting the circles denote functional interactions with the thickness indicating the degree of interaction.

These results clearly show that the relevant scales and mechanisms differ across cognitive processes. Nevertheless, a disproportionately large body of work has focused on testing the sensitivity of each neural feature and scale to a cognitive process of interest, rather than the specificity. That is, while many neural features may have been tested for their association with a cognitive process, whether this association is specific to that scale or generalizable to other scales has been under-explored. By applying predictive modeling iteratively across neural features extracted from various scales, studies can formally compare how well each scale explains and/or predicts the cognitive process of interest. For example, a monkey electrophysiology study directly compared three known neural signatures of spatial attention and found that the variability in spike counts of V4 neuronal pairs explained 79% of variance in attentional modulation, compared with single neuronal metrics, such as firing rates and spike count variability (i.e., Fano Factor), which accounted for only 9% and 4% of variance, respectively [46]. This population account was supported by a model fitted to the same data, which showed that attention reduces variance in the shared gain of V4 populations [47].

Although studies have had the capacity to probe multiple spatial scales simultaneously, tests for scale specificity and generalizability have been surprisingly scarce. Instead, a more common research practice has been to compare decoding accuracies of neural features that differ in the order of statistics, extracted from the same spatial scale. A human fMRI study showed that higher-order correlations in regional activity reflected higher-order cognitive processing during

story listening, whereas first-order correlations and noncorrelational activity patterns reflected lower-level perceptual and cognitive processes [48]. Moreover, patterns of pairwise regional interaction, but not regional or voxelwise activations, predicted changes in comprehension [49], attentional engagement [50], and emotional arousal [51] during narratives. These findings suggest that features with different order statistics and levels of granularity, even when extracted from the same neural scale, are not simply redundant but implicate different functions.

This approach can further be used to identify the ‘building blocks’ of cognitive processes based on their associations with different neural scales. Consider a scenario where a researcher extracts two types of neural feature: one at the microscale and another at the macroscale. The researcher also has two attention metrics linked to these data: one representing the severity of an attention disorder and the other reflecting behavioral performance in an attention task. By applying a decoding method, the researcher discovers that the microscale neural feature better predicts attention task performance, while the macroscale neural feature better predicts the severity of the attention disorder (Figure 1B). Although data types and preprocessing methods may vary across scales, systematic comparison is possible by normalizing model performance against the chance distribution generated for each scale. This hypothetical distinction suggests not only different brain–behavior associations across scales, but also that the attention task and the attention disorder involve different aspects of attention. This insight can prompt investigation into how attention is operationalized differently in these two contexts.

Geometry of neural representations and dynamics

Encoding and decoding models provide statistical associations of brain and behavior across scales. However, even when decoding accuracies are similar across these scales, the underlying neural geometry may differ. In this section, we propose examining the geometry of neural representations and dynamics, which offers mechanistic insights into how cognition emerges from neural activity at multiple scales (we focus here on **descriptive** rather than **causal mechanisms**).

Computational neuroscience work has adopted a dynamical systems approach to explore the geometries of neural population activity in relation to cognition (Figure 2; Box 1). In this framework, neural activity at each time segment is represented as a point and neural dynamics as a trajectory in a state space. The state space is defined from a specified neural scale, with each axis representing the activity of one neural unit, whether it is the firing rate of a neuron, signal in a pixel or voxel, or mean activity level of a brain region. The neural geometry reveals the underlying latent structure of population activity [31–33,52]. Given that neurons or regions within a population display tightly correlated responses, population activity follows stereotypical patterns and visits only confined surfaces within the state space, known as neural manifolds or subspaces [53]. Knowing the shape of this manifold and underlying dynamical systems is like knowing the likely routes that brain activity can take. Knowing how task variables are represented in the manifold and how they change across conditions, contexts, or task demands provides insight into how information is processed by the neural population. Thus, this framework allows us to penetrate the black box of a brain–behavior relationship and understand how information is internally organized and processed amid seemingly noisy and heterogeneous neural activity.

Global scale neural geometry

At a global scale, neuroimaging studies have investigated whole-brain activity and connectivity to understand the geometry of neural dynamics. The axes comprising this state space correspond to the activities of parcellated whole-brain areas. Studies have found that macroscale neural dynamics are low dimensional, constrained by the anatomical and functional organization of the brain [54–56]. Given that regional activity covaries across the brain according to its modular

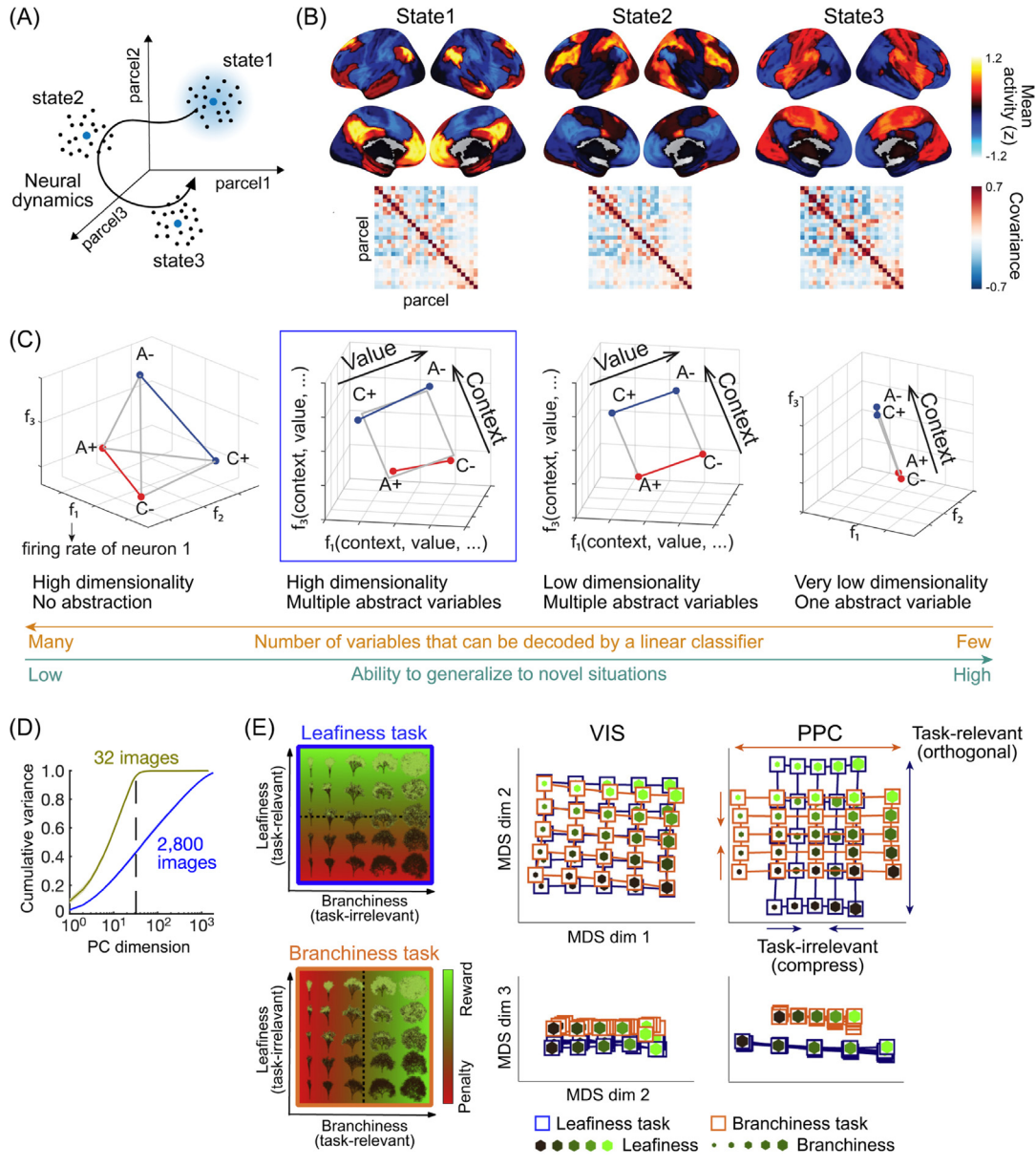


Figure 2. Geometry of global and local scale neural population activity. (A) Macroscale neural dynamics characterized as trajectories within a state space, where axes represent activities of parcellated brain areas. The black dot represents population activity at each time step. A latent brain state is characterized by unique patterns of mean activity (blue circle) and covariance between parcels (blue area) [22]. (B) Three example brain states identified using the hidden Markov model fits to the human fMRI data [22]. The mean activity (top) and pairwise covariance (bottom) of the 25 parcels' time series are shown for each state. (C) Illustration of the dimensionalities of neural codes and the orthogonal representations of task variables [73]. Dots indicate representations of the four stimulus–outcome conditions in the state space (stimulus identity: A and C; reward value: + and –), with same-colored lines indicating the same context. Empirical data recorded from the hippocampus and prefrontal cortex supported the second account (blue contour), which exhibited high-dimensional codes while maintaining orthogonal coding of different task variables. (D) Evidence supporting high-dimensional neural coding in the mouse visual cortex [77]. Cumulative variance explained relative to the number of neural dimensions for cross-validated principal component analyses conducted with either 2800 images (blue) or 32 images (green). The dashed line denotes 32 dimensions. All 32 dimensions were required to represent 32 images, whereas the required neural dimensions did not saturate (when tested up to 1000 dimensions) when representing 2800 images. (E) Evidence supporting orthogonal neural coding of multiple task variables in the human posterior parietal cortex (PPC) [78]. (Left) Twenty-five stimuli associated with reward or penalty in the two contexts. (Right) Multidimensional scaling (MDS) projections of fMRI data from the visual cortex (VIS) and PPC, reconstructed from coefficients of the best-fitting, hypothesis-driven models. In PPC, representations of task-irrelevant features were compressed, while representations of the two task-relevant features in the two contexts were orthogonalized.

functional organization [57,58], neural dynamics are not a randomized flow in the state space of the brain (Figure 2A). Instead, dynamics systematically traverse a relatively small number of brain states, characterized by distinct patterns of activity and interaction [59–65] (Figure 2B). Recent work found that these brain states commonly recur across various contexts, including tasks, rest, and movie watching [22], and span the canonical gradients of functional brain organization [22,66]. Importantly, these low-dimensional neural state dynamics reflect cognitive state dynamics in humans during controlled task performance [67–69] and naturalistic movie watching [49,70], with complementary evidence emerging from work in nonhuman animals [64].

Local scale neural geometry

At a local scale, research has investigated the geometry of neural representations, that is, how task variables are encoded by the neural population. A key topic of debate has been whether the neural representation is governed by **high-** or **low-dimensional codes** (i.e., uncorrelated and variable versus correlated and redundant population activity), which is associated with the efficiency versus generalizability of the representation [53,71–73] (Figure 2C).

High-dimensional neural codes are efficient in distinguishing combinations of multiple variables. For example, a study showed that the nonlinear mixed selectivity of macaque prefrontal cortical

Box 1. Dimensionality in neuroscience

The term ‘dimensionality’ mathematically refers to the number of latent variables required to account for a specific proportion of variance in data. In linear dimensionality reduction, this corresponds to the number of principal components (PCs) that explain a certain amount of variance in N -dimensional data. For low-dimensional data, the variance is concentrated in the first few eigenmodes with large eigenvalues. For high-dimensional data, the variance is distributed across eigenmodes with similar eigenvalues.

In neuroscience, a state space is defined by each axis representing the activity of a neural unit, such as the firing rate of a neuron, signal in a pixel or voxel, or mean activity level of a brain region. A pattern of population activity at a specified moment is geometrically conceptualized as a point in this state space. The dimensionality of neural data indicates the number of latent axes required to explain variance in these patterns. Dimensionality, in this sense, is estimated by applying a dimensionality reduction algorithm directly to neural data without associating it with task or behavioral variables.

A primary goal of characterizing neural geometry, however, is to understand how information is encoded in neural activity patterns. Neural dimensionality can be associated with task-related variables, including sensory, behavioral, and contextual variables chosen by the experimenter. This is referred to as neural coding dimensionality.

Neural coding dimensionality, also known as ‘representation dimensionality’ [71,72] or ‘embedding dimensionality’ [32,53], refers to the number of latent dimensions of neural activity required to represent and discriminate between task variables. In practice, neural coding dimensionality is estimated in two primary ways. The first method estimates the latent dimensions of neural data based on explained variance, then determines the number of latent dimensions needed to decode task variables (Figure 1A). For instance, PCs extracted from neural data can be used as features in a decoding model to distinguish between task variables. Researchers can ask whether the top-three PCs from a ten-dimensional state space are sufficient to discriminate task conditions (low-dimensional coding) or whether all ten PCs are required (high-dimensional coding). These components are derived directly from neural data, independent of task features, with the link to task variables established in subsequent analysis.

The second way of estimating neural coding dimensionality is by directly projecting neural data onto task-related subspaces that best discriminate task variables (Figure 1B). This analysis identifies hyperplanes that optimally separate neural activity patterns of different task conditions. For example, hyperplanes can be estimated with linear regression applied to the data with respect to task variables [80]. A neural representation is considered low dimensional if a small number of latent components is enough to discern task conditions.

Whether the neural code is low or high dimensional is determined relative to the dimensionality of the original state space and the dimensionality dictated by the task variables. Estimated neural coding dimensionality is mathematically equivalent to or lower than the original dimensionality of the neural data. The dimensionality required for successful neural coding is the same as or higher than the dimensions of task variables [32].

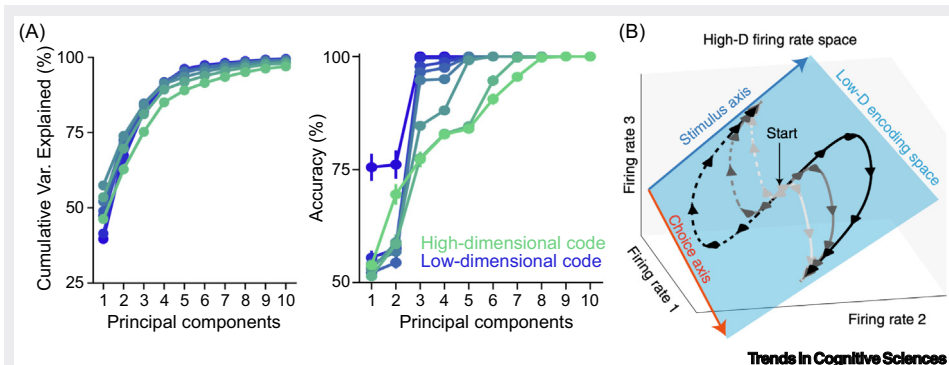


Figure 1. Neural coding dimensionality. (A) Illustration of the dimensionality reduction algorithm applied to the neural data, followed by subsequent association with task variables [78]. (Left) The latent dimensionality of the neural data is estimated using principal component (PC) analysis. Values denote how much variance in the neural data is explained upon retention of one to ten PCs. (Right) Decoding accuracy for task-related variables is assessed based on the number of PCs retained. If a small number of PCs is sufficient to represent task variables, neural coding dimensionality is low (blue lines). If a higher number of PCs is needed for representation, neural coding dimensionality is high (light-green lines). (B) Illustration of targeted dimensionality reduction, where neural data are projected onto a hyperplane defined by orthogonal stimulus- and choice-coding axes [81]. Modulations of neuronal firing rates by the task variables span a 2D subspace relative to the 3D state space of firing rates. The dimensions of this subspace include a stimulus axis (blue arrow), which captures the main variance in neural activity due to differences in stimulus values, and a choice axis (red arrow), which captures the main variance in neural activity due to differences in decision outcomes. Lines denote task conditions: stimulus intensity from +1 (light gray) and +2 (gray) to +3 (black), and choice of either left (dotted) or right (solid).

(PFC) neurons confers high-dimensional representations, allowing binary classifications of all combinations of task-related variables to be performed by a linear readout [74]. Indeed, increased task complexity correlated with increased coding dimensions in the PFC [75,76]. Research that estimated the stimulus-related variance in mouse visual cortical neurons found that the explained variance continued to increase with the inclusion of additional neural dimensions [77] (see also [26], which found similar results on a more global scale; Figure 2D). This suggests that neural responses need to be high dimensional and uncorrelated to be able to code multiple variables with distinct representation patterns.

However, efficient coding alone is not enough. Neural codes need to be structured and abstract to generalize to novel situations or contexts. This is achieved by low-dimensional and orthogonal representation of task variables [53,71–73]. Low-dimensional neural codes confine the complexity of neural representations to only a few dimensions, which increases robustness to variability and allows selective coding of task-relevant features while ignoring the irrelevant ones. Encoding task variables in subspaces orthogonal to one another using independent neural patterns prevents interference between the codes, allowing generalization to other contexts (Figure 2C).

This coding scheme was evidenced in a study that examined whole-brain human fMRI data and single-neuron recordings in the macaque frontal eye fields (FEF) as humans and monkeys performed a context-dependent categorization task [78] (Figure 2E). In the task, the correct choice depended on one stimulus feature in context A and another feature in context B. The early visual cortex in humans encoded both feature dimensions irrespective of context, deploying high-dimensional and context-agnostic representations. By contrast, in the frontoparietal regions of both humans and monkeys, irrelevant features in each context were compressed, allowing relevant features in the two contexts to be structured in low-dimensional, orthogonal subspaces.

Additionally, a human fMRI study observed low-dimensional orthogonal coding of target and distractor stimuli in the intraparietal sulcus during a cognitive control task [79]. Similar results were found with monkey electrophysiology [80,81]. In an oculomotor task, prior expectations about the visual target were encoded orthogonally to the sensory coding axis within a low-dimensional subspace of the middle temporal visual area (MT) activity before the target appeared [82]. Another study demonstrated that, despite heterogeneous temporal dynamics of PFC neurons, their population coding of mnemonic stimuli was stable and robust across time during working memory maintenance when projected onto a subspace orthogonal to the time-coding axis [83].

Research has also shown that even one task variable can be represented in orthogonal subspaces depending on task demands [84,85]. For example, motor neurons encode population activity in two orthogonal subspaces during preparation and execution [84]. If neuronal population activity is confined to a dimension that is orthogonal to a readout dimension, an animal can withhold muscle movement despite preparatory motor neuron firing. Together, these findings suggest that task variables are represented through a balance between efficient (high-dimensional) and structured (low-dimensional) neural codes [73], which vary across regions and potentially across scales.

Integrating neural geometries across scales

Evidence points to relatively lower-dimensional codes at the macroscale and higher-dimensional codes at the microscale. This variation in neural representation geometries and dynamics across scales suggests different mechanisms of information processing. A recent preprint examined how these opposing neural coding principles coexist across scales in a biological system [86]. The study analyzed an extensive set of openly available calcium imaging data collected from zebrafish, nematodes, flies, mice, and macaques. Calculating pairwise neuronal correlations from micro-, meso-, to macroscales, the study found that absolute correlation increased with scale, recapitulating the high-dimensional and efficient microscale activity and low-dimensional and redundant macroscale activity that were conserved across species. Network simulations found that this efficiency–redundancy balance facilitates information flow, increases susceptibility to a wide range of stimuli, and allows flexible adaptation to behavioral demands [86].

Many studies have characterized regional differences in neural geometries [78,87]. The logic of comparing across regions can be extended to comparing across scales. Researchers can characterize neural geometries at multiple neural scales (e.g., finding low-dimensional subspaces using multidimensional scaling [87]) and compare how information is processed differently.

Global and local scale interaction

How do brain-wide state dynamics interact with local information processing [88]? To understand cognitive functions associated with different neural scales, we must understand how these scales are biologically intertwined.

Empirical understanding

Work has begun to probe multiple spatial scales simultaneously in a single study [31]. For example, a recent study conducted simultaneous local electrophysiological recordings of either V1 or retrosplenial cortex (RSP) and wide-field calcium imaging of the dorsal cortex of head-fixed mice as the mice were free to run on a rotating wheel [89]. Investigating how individual neurons couple to the broader cortex, the study found that V1 and RSP neurons that were less coupled to nearby neurons had more global connections with other cortical areas (Figure 3A). Another study analyzed publicly available datasets of large-scale extracellular recordings of head-fixed mice [90]. Hippocampal sharp wave ripples not only modulated intrahippocampal activity, but also

modulated brain-wide firing rates and functional network topology (Figure 3B). Using fMRI, researchers found that discrete states of the ventromedial prefrontal cortex (vmPFC), detected with a hidden Markov model, mapped not only onto participants' affective experiences during movie watching, but also to systematic changes in whole-brain univariate activity [91]. This shows that affective state dynamics are not only represented in the vmPFC, but also that vmPFC activity states coincide with characteristic changes in brain-wide activity (Figure 3C). Other fMRI work showed that the large-scale functional topography of the cortex, which is modulated by arousal and cognitive states, is associated with the activity of the thalamus [92] and recapitulated by thalamocortical connectivity [93]. Together, these studies demonstrate an exciting new avenue of research where neural activities at different scales can be probed simultaneously. This allows us to investigate how neural activity at local and global scales interacts to give rise to cognitive processes. Future work could also ask whether functions arise from the global–local interaction itself in addition to the isolated scale alone.

Computational understanding

Empirical work gives us biological insights into global–local scale interactions and how they are modulated by cognitive states and behavioral contexts. Computational neuroscience has further explained these interactions by developing generative models that simulate neurophysiological properties of the brain [29,30,94]. For example, large-scale neural mass models characterize brain activity as an interconnected network of local nodes [95,96]. In one implementation of this model, the network is constrained by structural connectivity (i.e., white matter tracts measured with diffusion tensor imaging), with each node (representing a parcellated cortical area) modeled by interconnected excitatory and inhibitory local neural mass models [96]. By simulating brain activity (e.g., measured with fMRI, EEG, or MEG) while controlling the parameters of this network, researchers can determine which scales and parameters are causally responsible for specific

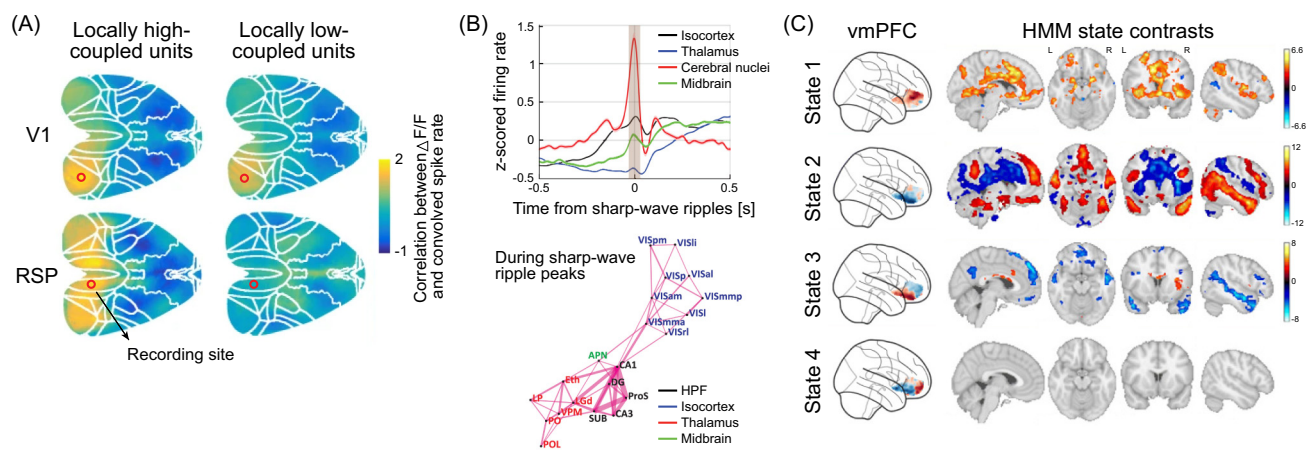


Figure 3. Interactions between global and local neural processes. (A) Simultaneous imaging of the mouse dorsal cortex and recording of the spiking neuronal population [89]. Units recorded in V1 and retrosplenial cortex (RSP) were grouped based on the degree of local population coupling, which is the correlation between the spiking of a neuron and the sum of spikes from all other simultaneously recorded nearby neurons (locally high-coupled units: above the 80th percentile; locally low-coupled units: below the 20th percentile). The convolved spike trains of these units were correlated with changes in calcium signals (F/F) of every pixel across the dorsal cortex measured with wide-field imaging, with the color indicating z-scored correlation coefficients. Red circles denote recording sites. (B) Changes in brain-wide activity and functional topography at moments when hippocampal ripple activity is detected [90]. (Top) Average firing rates of the four major extrahippocampal areas, time-aligned to hippocampal sharp-wave ripples. (Bottom) The undirected graph shows mutual information between pairs of brain regions during hippocampal sharp-wave ripples. Abbreviation: HPF, hippocampal formation. (C) Brain-wide activity time-aligned to states defined by the activity patterns of a local brain area [91]. (Left) An example participant's ventromedial prefrontal cortex (vmPFC) states inferred from a hidden Markov model (HMM). (Right) Whole-brain univariate contrasts computed based on the HMM state occurrences.

brain functions. Local circuit models, which simulate the neural activity of local areas, have advanced to a level of simulating behavior in different task contexts and cognitive states [52]. Although current large-scale models have simulated neural activity at rest [95–97], work is heading toward simulating task-based neural data to understand how cognition modulates global–local interaction and neural topography at various scales [98].

Concluding remarks

Neuroscience has long focused on dissecting the brain into its parts and understanding the cognitive functions associated with each part. The spatial scales at which these studies were conducted received less attention, often dictated by the choice of model species and neuroscientific techniques. Consequently, theories and empirical tests of how functions emerge from each scale, and from interactions across scales, have been scarce. However, we now have the tools to test how cognitive functions emerge from different scales.

To achieve this, we need to expand our current approaches to encompass multiple scales. Brain–behavior associations identified through predictive modeling and neural geometries identified through dynamical systems approaches should be tested across scales for empirical comparisons. Interactions between global and local scales should be characterized both at empirical (descriptive) and computational (generative) levels. However, the range of scales that any single study or lab can probe is limited by expertise, time, cost, and other constraints. Identifying functions associated with various scales is not a task that can be easily achieved by a single lab. A collective effort is needed, requiring us to theoretically bridge findings from different animal models and techniques uniquely suited to probe different scales of neural processing. Open science practices, particularly data, code, and model sharing, and cross-lab collaborations can foster the use of diverse datasets and computational approaches to investigate multiple neural scales (see [Outstanding questions](#)).

Finally, although this may sound obvious to some readers, we cannot help but emphasize one last point. Successful crosstalk starts from researchers' awareness that the scale they probe is the tip of an iceberg and their findings are not guaranteed to generalize across scales. For example, if a researcher uses fMRI to investigate decision-making, they should acknowledge that they are looking at a specific representation that occurs at a temporal scale of seconds and a spatial scale of $\sim 2 \text{ mm}^3$ voxels, which is a blood-flow proxy for an averaged activity of a million neurons and billions of synapses. If a researcher records the spiking activity of PFC neurons to investigate decision-making, they should acknowledge that the evidence is a part of the whole; their findings may be specific to PFC neurons, but not others, and the findings may have different computational roles when interaction with other regions or brain-wide state modulations are taken into account. Even in studies that argue for causal effects by directly manipulating neural activity using microstimulation, lesions, or optogenetics, observed effects may not indicate that the targeted area is a single source of the cognitive function but only that it contributes as a part of a larger processing network. Acknowledging the limited scale that each study probes will explicitly position the findings along the continuum of all possible scales, allowing us to connect the dots across scales and identify missing pieces ([Box 2](#)).

In summary, cognition and behavior emerge from neural operations at various spatial scales. Isolating a single neural scale provides a limited understanding of complex behavior. We highlight the need to probe multiple scales to assess and compare their relationships to cognition and behavior. The principles of information coding should be identified within each scale and compared across scales, complemented by empirical and computational understanding of global and local scale interactions. While recent work has started to investigate multiple scales simultaneously, the technology afforded by each subfield unavoidably confines the spatial scale that is

Outstanding questions

Cognitive processes are represented across different temporal scales of the brain, ranging from fast (e.g., milliseconds and seconds) to slow (e.g., minutes and hours) time-scales. How can we investigate cognitive functions of neural processes at each timescale, and their relationships across scales?

Empirical cross-scale comparisons are frequently confounded by variations in model species, recording techniques, and subsequent preprocessing pipelines tailored to each data type. How can we harmonize these diverse data types to make fair comparisons across scales?

Causal mechanism studies can test whether the effects of perturbation, stimulation, or lesion are more pronounced at global or local scales. How can we assess the causal effects of global–local scale interactions?

Low-dimensional state spaces have been shown to encode features of controlled tasks and associated behavior. Does the low-dimensional coding principle similarly explain cognitive processes in less controlled, more naturalistic contexts?

Box 2. Range of spatiotemporal scales in neuroscience research

Lets do a fun exercise. Think about the most recent neuroscience paper you wrote or read. What spatial and temporal scales of the brain were investigated? Feel free to check all scales that apply. Now think about the cognitive or behavioral variable that was targeted in the paper. Do you think that the spatiotemporal scales you checked well capture the scope of the behavioral construct? Do the given neural observations tile all possible explanatory scales? If not, what scales are missing? What subfields in neuroscience (using which model species, recording/imaging techniques, and computational analyses) could complement the missing pieces? We imagine an interactive repository where researchers can share their chosen scales for each of their scientific papers, tagged with the associated cognitive process or behavior and techniques used to collect the data.

	Molecular	Cellular	Inter-neuronal	Neuronal population	Area	Inter-areal	Circuit/network	Whole brain	Brain-body	Brain-brain	Brain-environment
Spatial	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	~1 ms	~10 ms	~100 ms	~1 s	~10 s	Minutes	Hours	Days	Months	Years	Decades
Temporal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Trends in Cognitive Sciences

Figure 1. Spatiotemporal scales in neuroscience research.

probed. Therefore, we argue for the fostering of open science and cross-lab collaborations and emphasize that findings from a specified scale are not guaranteed to generalize across scales.

Acknowledgments

We thank Wilma Bainbridge, Jin Ke, Jungwoo Kim, Mina Kwon, Yuan Chang Leong, Jady Park, Lawrence Snyder, and Ziwei Zhang for their feedback on the initial version of the manuscript (ordered alphabetically). We thank the four anonymous reviewers for their constructive feedback. The research was supported by National Science Foundation BCS-2043740 and Alfred P. Sloan Research Fellowship in Neuroscience FG-2022-19032 (M.D.R.).

Declaration of interests

The authors declare no competing interests.

References

- Betzell, R.F. and Bassett, D.S. (2017) Multi-scale brain networks. *Neuroimage* 160, 73–83
- Martin, S.J. *et al.* (2000) Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* 23, 649–711
- Kolb, L.D. *et al.* (2023) Hippocampal neurons code individual episodic memories in humans. *Nat. Hum. Behav.* 7, 1968–1979
- Brune, I.K. *et al.* (2020) Integration and differentiation of hippocampal memory traces. *Neurosci. Biobehav. Rev.* 118, 196–208
- Kitamura, T. *et al.* (2015) Entorhinal-hippocampal neuronal circuits bridge temporally discontinuous events. *Learn. Mem.* 22, 438–443
- Eichenbaum, H. (2017) On the integration of space, time, and memory. *Neuron* 95, 1007–1018
- Buzsáki, G. and Moser, E.I. (2013) Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nat. Neurosci.* 16, 130–138
- Ranganath, C. and Ritchey, M. (2012) Two cortical systems for memory-guided behaviour. *Nat. Rev. Neurosci.* 13, 713–726
- Chen, J. *et al.* (2017) Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* 20, 115–125
- Andersen, P.W. (1972) More is different. *Science* 177, 393–396
- Gómez-Ramírez, J. (2013) Towards a Theory of Brain Structure and Function. In *A New Foundation for Representation in Cognitive and Brain Science*, Springer
- Goldman-Rakic, P.S. (1988) Topography of cognition: parallel distributed networks in primate association cortex. *Annu. Rev. Neurosci.* 11, 137–156
- Buzsáki, G. and Draguhn, A. (2004) Neuronal oscillations in cortical networks. *Science* 304, 1926–1929
- Pessoa, L. (2022) *The Entangled Brain: How Perception, Cognition, and Emotion Are Woven Together*, MIT Press
- Hubel, D.H. and Wiesel, T.N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154
- Georgopoulos, A.P. *et al.* (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* 2, 1527–1537
- Kanwisher, N. *et al.* (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311
- Chang, C. *et al.* (2016) Tracking brain arousal fluctuations with fMRI. *Proc. Natl. Acad. Sci. U. S. A.* 113, 4518–4523
- Goodale, S.E. *et al.* (2021) Fmri-based detection of alertness predicts behavioral response variability. *Elife* 10, e62376
- Kucyi, A. *et al.* (2020) Electrophysiological dynamics of antagonistic brain networks reflect attentional fluctuations. *Nat. Commun.* 11, 1–14
- Rosenberg, M.D. *et al.* (2020) Functional connectivity predicts changes in attention observed across minutes, days, and months. *Proc. Natl. Acad. Sci. U. S. A.* 117, 3797–3807
- Song, H. *et al.* (2023) Large-scale neural dynamics in a shared low-dimensional state space reflect cognitive and attentional dynamics. *Elife* 12, e85487

23. Aston-Jones, G. and Cohen, J.D. (2005) An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450
24. Munn, B.R. *et al.* (2021) The ascending arousal system shapes neural dynamics to mediate awareness of cognitive states. *Nat. Commun.* 12, 1–9
25. Allen, E.J. *et al.* (2022) A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat. Neurosci.* 25, 116–126
26. Manley, J. *et al.* (2024) Simultaneous, cortex-wide dynamics of up to 1 million neurons reveal unbounded scaling of dimensionality with neuron number. *Neuron* 112, 1694–1709.e5
27. Steinmetz, N.A. *et al.* (2019) Distributed coding of choice, action and engagement across the mouse brain. *Nature* 576, 266–273
28. Siegle, J.H. *et al.* (2021) Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* 592, 86–92
29. Kringsbach, M.L. and Deco, G. (2020) Brain states and transitions: insights from computational neuroscience. *Cell Rep.* 32, 108128
30. Breakspear, M. (2017) Dynamic models of large-scale brain activity. *Nat. Neurosci.* 20, 340–352
31. Engel, T.A. and Steinmetz, N.A. (2019) New perspectives on dimensionality and variability from large-scale cortical dynamics. *Curr. Opin. Neurobiol.* 58, 181–190
32. Ostojic, S. and Fusi, S. (2024) Computational role of structure in neural activity and connectivity. *Trends Cogn. Sci.* 28, 677–690
33. Ebitz, R.B. and Hayden, B.Y. (2021) The population doctrine in cognitive neuroscience. *Neuron* 109, 3055–3068
34. Markiewicz, C.J. *et al.* (2021) The Openneuro resource for sharing of neuroscience data. *Elife* 10, e71774
35. Milham, M.P. *et al.* (2018) An open resource for non-human primate imaging. *Neuron* 100, 61–74.e2
36. Rübél, O. *et al.* (2022) The Neurodata Without Borders ecosystem for neurophysiological data science. *Elife* 11, e78362
37. International Brain Laboratory *et al.* (2023) A brain-wide map of neural activity during complex behaviour. *bioRxiv*, Published online July 4, 2023. <https://doi.org/10.1101/2023.07.04.547681>
38. Huth, A.G. *et al.* (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532, 453–458
39. Vickery, T.J. *et al.* (2011) Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron* 72, 166–177
40. Chang, L.J. *et al.* (2015) A sensitive and specific neural signature for picture-induced negative affect. *PLoS Biol.* 13, e1002180
41. Rissman, J. and Wagner, A.D. (2012) Distributed representations in memory: insights from functional brain imaging. *Annu. Rev. Psychol.* 63, 101–128
42. Chen, S. *et al.* (2024) Brain-wide neural activity underlying memory-guided movement. *Cell* 187, 676–691
43. Stringer, C. *et al.* (2019) Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364, eaav7893
44. Musall, S. *et al.* (2019) Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* 22, 1677–1686
45. McCormick, D.A. *et al.* (2020) Neuromodulation of brain state and behavior. *Annu. Rev. Neurosci.* 43, 391–415
46. Cohen, M.R. and Maunsell, J.H.R. (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* 12, 1594–1600
47. Rabinowitz, N.C. *et al.* (2015) Attention stabilizes the shared gain of V4 populations. *Elife* 4, e08998
48. Owen, L.L.W. *et al.* (2021) High-level cognition during story listening is reflected in high-order dynamic correlations in neural activity patterns. *Nat. Commun.* 12, 1–14
49. Song, H. *et al.* (2021) Cognitive and neural state dynamics of narrative comprehension. *J. Neurosci.* 41, 8972–8990
50. Song, H. *et al.* (2021) Neural signatures of attentional engagement during narratives and its consequences for event memory. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2021905118
51. Ke, J. *et al.* (2023) Dynamic functional connectivity encodes generalizable representations of emotional arousal but not valence. *bioRxiv*, Published online November 16, 2023. <https://doi.org/10.1101/2023.11.14.566767>
52. Langdon, C. *et al.* (2023) A unifying perspective on neural manifolds and circuits for cognition. *Nat. Rev. Neurosci.* 24, 363–377
53. Jazayeri, M. and Ostojic, S. (2021) Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Curr. Opin. Neurobiol.* 70, 113–120
54. Pang, J.C. *et al.* (2023) Geometric constraints on human brain function. *Nature* 618, 566–574
55. Power, J.D. *et al.* (2011) Functional network organization of the human brain. *Neuron* 72, 665–678
56. Mesulam, M.M. (1998) From sensation to cognition. *Brain* 121, 1013–1052
57. Yeo, B.T.T. *et al.* (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 1125–1165
58. Fox, M.D. *et al.* (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U. S. A.* 102, 9673–9678
59. Vidaurre, D. *et al.* (2017) Brain network dynamics are hierarchically organized in time. *Proc. Natl. Acad. Sci. U. S. A.* 114, 12827–12832
60. Baker, A.P. *et al.* (2014) Fast transient networks in spontaneous human brain activity. *Elife* 3, e01867
61. Chen, S. *et al.* (2016) Spatiotemporal modeling of brain dynamics using resting-state functional magnetic resonance imaging with Gaussian hidden Markov model. *Brain Connect.* 6, 326–334
62. Abbas, A. *et al.* (2019) Quasi-periodic patterns contribute to functional connectivity in the brain. *Neuroimage* 191, 193–204
63. Bolt, T. *et al.* (2022) A parsimonious description of global functional brain organization in three spatiotemporal patterns. *Nat. Neurosci.* 25, 1093–1103
64. MacDowell, C.J. and Buschman, T.J. (2020) Low-dimensional spatiotemporal dynamics underlie cortex-wide neural activity. *Curr. Biol.* 30, 2665–2680
65. Greene, A.S. *et al.* (2023) Why is everyone talking about brain state? *Trends Neurosci.* 46, 508–524
66. Shine, J.M. *et al.* (2019) Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nat. Neurosci.* 22, 289–296
67. Cornblath, E.J. *et al.* (2020) Temporal sequences of brain activity at rest are constrained by white matter structure and modulated by cognitive demands. *Commun. Biol.* 3, 1–12
68. Taghia, J. *et al.* (2018) Uncovering hidden brain state dynamics that regulate performance and decision-making during cognition. *Nat. Commun.* 9, 2505
69. Yamashita, A. *et al.* (2021) Brain state-based detection of attentional fluctuations and their modulation. *Neuroimage* 236, 118072
70. van der Meer, J.N. *et al.* (2020) Movie viewing elicits rich and reliable brain state dynamics. *Nat. Commun.* 11, 5004
71. Badre, D. *et al.* (2021) The dimensionality of neural representations for control. *Curr. Opin. Behav. Sci.* 38, 20–28
72. Fusi, S. *et al.* (2016) Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* 37, 66–74
73. Bernardi, S. *et al.* (2020) The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell* 183, 954–967
74. Rigotti, M. *et al.* (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590
75. Spaak, E. *et al.* (2017) Stable and dynamic coding for working memory in primate prefrontal cortex. *J. Neurosci.* 37, 6503–6516
76. Bartolo, R. *et al.* (2020) Dimensionality, information and learning in prefrontal cortex. *PLoS Comput. Biol.* 16, e1007514
77. Stringer, C. *et al.* (2019) High-dimensional geometry of population responses in visual cortex. *Nature* 571, 361–365
78. Flesch, T. *et al.* (2022) Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron* 110, 1258–1270
79. Ritz, H. and Shenhav, A. (2024) Orthogonal neural encoding of targets and distractors supports multivariate cognitive control. *Nat. Hum. Behav.* 8, 945–961
80. Mante, V. *et al.* (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84
81. Aoi, M.C. *et al.* (2020) Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. *Nat. Neurosci.* 23, 1410–1420
82. Park, J. *et al.* (2023) Prior expectation enhances sensorimotor behavior by modulating population tuning and subspace activity in sensory cortex. *Sci. Adv.* 9, eadg4156

83. Murray, J.D. *et al.* (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 114, 394–399
84. Kaufman, M.T. *et al.* (2014) Cortical activity in the null space: permitting preparation without movement. *Nat. Neurosci.* 17, 440–448
85. Yoo, S.B.M. and Hayden, B.Y. (2020) The transition from evaluation to selection involves neural subspace reorganization in core reward regions. *Neuron* 105, 712–724.e4
86. Munn, B.R. *et al.* (2024) Phylogenetically-preserved multiscale neuronal activity: iterative coarse-graining reconciles scale-dependent theories of brain function. *bioRxiv*, Published online June 27, 2024. <https://doi.org/10.1101/2024.06.22.600219>
87. Muhle-Karbe, P.S. *et al.* (2023) Goal-seeking compresses neural codes for space in the human hippocampus and orbitofrontal cortex. *Neuron* 111, 3885–3899.e6
88. MacDowell, C.J. *et al.* (2022) A Goldilocks theory of cognitive control: balancing precision and efficiency with low-dimensional control states. *Curr. Opin. Neurobiol.* 76, 102606
89. Clancy, K.B. *et al.* (2019) Locomotion-dependent remapping of distributed cortical networks. *Nat. Neurosci.* 22, 773–786
90. Nitzan, N. *et al.* (2022) Brain-wide interactions during hippocampal sharp wave ripples. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2200931119
91. Chang, L.J. *et al.* (2021) Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Sci. Adv.* 7, eabf7129
92. Shine, J. *et al.* (2019) The low-dimensional neural architecture of cognitive complexity is related to activity in medial thalamic nuclei. *Neuron* 104, 849–855
93. Müller, E.J. *et al.* (2020) Core and matrix thalamic subpopulations relate to spatio-temporal cortical connectivity gradients. *Neuroimage* 222, 117224
94. D'Angelo, E. and Jirsa, V. (2022) The quest for multiscale brain modeling. *Trends Neurosci.* 45, 777–790
95. Demirtaş, M. *et al.* (2019) Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. *Neuron* 101, 1181–1194
96. Schirner, M. *et al.* (2018) Inferring multi-scale neural mechanisms with brain network modelling. *Elife* 7, e28927
97. Singh, M.F. *et al.* (2020) Estimation and validation of individualized dynamic brain models with resting state fMRI. *Neuroimage* 221, 117046
98. John, Y.J. *et al.* (2022) It's about time: linking dynamical systems with human neuroimaging to understand the brain. *Netw. Neurosci.* 6, 960–979