Title:

Association of bidirectional network cores in the brain with conscious perception and cognition

Abbreviated title:

Bidirectional Cores, Consciousness, and Cognition

Authors:

Tomoya Taguchi[1], Jun Kitazono[2], Shuntaro Sasai[3], and Masafumi Oizumi[1]

Affiliations:

1: Graduate School of Arts and Sciences, The University of Tokyo,

2: Graduate School of Data Science, Yokohama City University, Kanagawa, Japan,

3: Araya, Inc.

Corresponding Authors:

Jun Kitazono (kitazono.jun.ig@yokohama-cu.ac.jp) and Masafumi Oizumi (c-oizumi@g.ecc.u-tokyo.ac.jp)

Financial Interests of Conflicts of Interests:

The authors declare no competing interests.

Acknowledgement: This study was supported by Graduate Program for Social ICT Global Creative Leaders of the University of Tokyo by MEXT to TT, JST ACT-X Grant Number JPMJAX20A6 to JK, JST Moonshot R&D Grant Number JPMJMS2012 to SS and MO, JST CREST Grant Number JPMJCR1864 to MO, and JSPS KAKENHI Grant Numbers 18H02713 and 20H05712 to MO. Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. We acknowledge the use of ChatGPT 4 (Open AI, https://chat.openai.com) to proofread our draft. We would like to thank Editage (www.editage.jp) for English language editing.

Association of bidirectional network cores in the brain with conscious perception and cognition

Tomoya Taguchi[1], Jun Kitazono[2*], Shuntaro Sasai[3], and Masafumi Oizumi[1*]

[1]Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan

[2]Graduate School of Data Science, Yokohama City University, Kanagawa, Japan

[3]Araya Inc., Tokyo, Japan

[*]kitazono.jun.ig@yokohama-cu.ac.jp, c-oizumi@g.ecc.u-tokyo.ac.jp

Abstract

The brain comprises a complex network of interacting regions. To understand the roles and mechanisms of this complex network, its structural features related to specific cognitive functions need to be elucidated. Among such relationships, recent developments in neuroscience highlight the link between network bidirectionality and conscious perception. Given the essential roles of both feedforward and feedback signals in conscious perception, it is surmised that subnetworks with bidirectional interactions are critical. However, the link between such subnetworks and conscious perception remains unclear due to the network's complexity. In this study, we propose a framework for extracting subnetworks with strong bidirectional interactions—termed the "cores" of a network—from brain activity. We applied this framework to resting-state and task-based fMRI data to identify regions forming strongly bidirectional cores. We then explored the association of these cores with conscious perception and cognitive functions. The central cores predominantly included cerebral cortical regions, which are crucial for conscious perception, rather than subcortical regions. Furthermore, the cores were composed of previously reported regions in which electrical stimulation altered conscious perception. These results suggest a link between the bidirectional cores and conscious perception. A meta-analysis and comparison of the core structure with a cortical functional connectivity gradient suggested that the central cores were related to lower-order sensorimotor functions. An ablation study emphasized the importance of incorporating bidirectionality, not merely interaction strength for these outcomes. The proposed framework provides novel insight into the roles of network cores with strong bidirectional interactions in conscious perception and lower-order sensorimotor functions.

Significance Statement

To understand the brain's network, we must decipher its structural features linked to cognitive functions. Recent studies suggest the importance of subnetworks with bidirectional interactions for conscious perception, but their exact relationship remains unclear due to the brain's complexity. Here we propose a framework for extracting subnetworks with strong bidirectional interactions, or network "cores." We applied it to fMRI data and explored the association of the extracted cores with conscious perception and cognitive functions. The central cores predominantly included cerebral cortical regions rather than subcortical ones, and comprised previously reported regions wherein electrical stimulation altered conscious perception, suggesting the importance of strongly bidirectional cores for conscious perception. Additionally, further analysis including meta-analysis revealed the cores' relation to lower-order sensorimotor functions.

Introduction

Complex networks formed by interactions between brain regions are involved in cognitive functions. Analyzing these networks using network theory could crucially enable the identification of network features relevant to specific cognitive functions, which may further our understanding of the mechanisms that underlie cognitive functions [1–3]. For example, studies have shown that the manifestation of module structures (i.e., modularity) predicts high performance in working memory tasks [4–8]. Moreover, research on small-world properties has demonstrated that the efficiency of signal propagation between brain regions (i.e., network efficiency) can predict an individual's intellectual capacity [9].

One relationship between network features and cognitive functions that has attracted considerable attention is the link between the bidirectionality of networks and the subjective experience of consciousness [10–13]. Decades of research on the minimal neural correlates of consciousness (NCC) suggest that the bidirectionality of networks is a key to understanding consciousness [10–17]. For example, empirical research has shown that conscious perception of a sensory stimulus requires bidirectional signaling that involves both feedforward and feedback signal propagation [18–26]. Furthermore, the importance of bidirectional signal propagation for consciousness is suggested by its absence not only during the loss of specific contents of consciousness in awake states but also during unconscious states, where conscious experience is generally lost, such as general anesthesia [19, 27–29], sleep [30, 31], and vegetative states [32]. Moreover, the importance of bidirectional interactions in consciousness is independent of sensory modality [33] (vision [19–21, 24, 25], somatosensation [18, 22, 23, 26], and audition [31, 34–37]) and extends across species (humans [22, 24, 28, 32], monkeys [18–21, 25, 30], rodents [23, 26], birds

[38], and flies [29]). These experimental findings underlie the hypotheses of many theories of consciousness that suggest that the locus of consciousness within the brain network includes subnetworks with bidirectional interactions [10–12, 14–17, 39–44]. For example, the integrated information theory specifically hypothesizes that areas with strong bidirectional causal interactions constitute the locus of consciousness [14, 40, 45, 46].

Based on these experimental findings and theoretical insights, to understand the relationship between the brain network and conscious perception, the identification of brain subnetworks that have strong bidirectional interactions and ascertaining how these subnetworks relate to conscious perception is pivotal. However, the identification of such subnetworks has been challenging. To date, methods for extracting subnetworks with strong connections have been applied to brain functional networks (e.g., s-core decomposition [47–50], network hubs [51, 52], rich clubs [48, 53, 54], and modularity [55, 56]). However, the abovementioned methods do not consider the direction of influence, particularly the bidirectionality of interactions. Consequently, the locations of the subnetworks with pronounced bidirectionality and their links to cognitive functions, including conscious perception, remain elusive.

To address this gap, we propose a novel framework for extracting subnetworks with strong bidirectional interactions from brain activity that were designated as "cores." We applied this framework to functional magnetic resonance imaging (fMRI) data to identify the regions that constitute strongly bidirectional cores and subsequently performed two analyses on the identified regions: one that examined the relationship of the core regions with conscious perception and the other that explored the association of the core regions with cognitive functions. The following section provides an overview of this analysis.

First, we investigated the brain regions that were more likely to be components of strongly bidirectional cores. If certain brain regions consistently appear in the cores under various conditions, this would suggest their general importance in conscious perception and a broad range of cognitive functions. Therefore, using data from the Human Connectome Project (HCP) [57], we extracted the cores at rest and during seven cognitive tasks. Our objective was to determine the brain regions that consistently form strongly bidirectional cores under resting-state and task conditions.

We subsequently investigated the association of the core regions with conscious perception. We focused on a previous study wherein electrical brain stimulation was used to assess the association of brain regions with conscious perception [58]. We posited that, if regions within strongly bidirectional cores are integral to conscious perception, then, stimulation of these regions would propagate effects throughout the cores, and thereby result in alterations of conscious perception. Therefore, in this study, we compared the likelihood of each region of interest (ROI) being included in strongly bidirectional cores and the rate at which the intracranial electrical stimulation

(iES) [58] of each ROI elicited a change in conscious perception. iES involves direct electrical stimulation of the brain via intracranially placed electrodes [59] that enable causal modulation of neural activity. Prior research has quantitatively evaluated the extent to which the iES of each brain region induces reportable changes in conscious perceptual experiences across the entire cerebral cortex [58]. We investigated whether the rates of perceptual change elicited by iES (elicitation rates) were associated with the probability of being included in strongly bidirectional cores. Higher elicitation rates of regions that are more likely to be part of strong cores would suggest a significant association between strongly bidirectional cores and conscious perception.

Then, to investigate the association of the cores with cognitive functions more broadly, we performed a hypothesis-free meta-analysis using NeuroSynth [60]. NeuroSynth is a platform for examining associations between brain regions and specific psychological and neurological terms based on a large database of fMRI studies. Using this tool, thousands of fMRI data can be aggregated to identify statistical associations between ROIs and specific terms, such as "memory" and "attention." This approach facilitated the identification of differences in the associated terms between ROIs with a strong or weak tendency to be included in the cores. Finally, to further assess the association of the cores with cognitive function, we examined the relationship between the cores to a functional connectivity gradient [61]. The gradient is known to correspond to a spatial gradient over the cerebral cortex, ranging from lower- to higher-order regions.

Materials and Methods

fMRI Data Acquisition and Preprocessing

The 3T-fMRI data were obtained from the Washington University-Minnesota Consortium Human Connectome Project (HCP, [57]). We started with minimally preprocessed fMRI data at rest and during seven cognitive tasks (emotion, gambling, language, motor, relational, social, and working memory). From the HCP 1200 Subjects Data Release [57] (`www.humanconnectome.org/`), we used 352 low-motion subjects in accordance with the criteria proposed in ref. [62]. Next, we performed denoising by estimating nuisance regressors and subtracting them from the signal at every vertex [63]. Accordingly, we used 36 nuisance and spike regressors, developed by ref. [63], that comprised: (1–6) six motion parameters, (7) white-matter time-series, (8) cerebrospinal fluid time-series, (9) global signal time-series, (10–18) temporal derivatives of (1–9), and (19–36) quadratic terms for (1–18). The spike regressors were computed as described previously in ref. [63], with 1.5-mm movement used as a spike-identification threshold. These fMRI data have a temporal resolution (TR) of 0.72 seconds. The number of frames for rest and each task state is as follows: emotion, 176; gambling, 253; language, 316; motor, 284;

relational, 232; social, 274; working memory, 405; and rest, 1200 frames. There were two resting-state sessions in the HCP data: Rest1 and Rest2. As Rest1 alone provided a larger number of frames than the other task states, in this study, only Rest1 was used.

Using the methodology proposed in ref. [64], we parcellated the cerebral cortex and subcortex into 200 and 32 ROIs, based on Schaefer's parcellation [65] and Tian's parcellation [66], respectively, that generated a total of 232 ROIs. Functional and structural networks obtained using this parcellation are close to the average of those derived from other parcellations, which makes them representative networks [64] and, accordingly, we employed this parcellation in this study. We normalized the time-series data for each session to z-scores and then applied a 0.008–0.08 Hz second-order Butterworth band-pass filter.

Estimation of directed functional network

To extract cores with strong bidirectional interactions, it is first necessary to quantify the causal strengths between brain regions and construct a directed network. For this purpose, we computed the normalized directed transfer entropy (NDTE, [67]) for each pair of ROIs from the BOLD signals: We fitted a bivariate VAR model to the BOLD signals X and Y of each pair of ROIs and set the lag order T of the VAR model to 10. The estimation of the VAR model was performed using locally weighted regression, using data from the RL and LR sessions of 352 subjects as trials. The Multivariate Granger Causality Toolbox [68] was employed for these computations. Subsequently, the transfer entropy (TE) for each pair of ROIs was calculated using the parameters of this estimated VAR model. TE T_{XY} from the ROI X to ROI Y quantifies how much the past activity of X contributes to predicting the future of Y [67]. Mathematically, T_{XY} is conditional mutual information $I\left(Y_{i+1}; X^i \mid Y^i\right)$ that can be written using entropy as

$$T_{\mathrm{XY}} = I\left(Y_{i+1}; X^i \mid Y^i\right) = H\left(Y_{i+1} \mid Y^i\right) - H\left(Y_{i+1} \mid X^i, Y^i\right) \tag{1}$$

where Y_{i+1} denotes the activity of the ROI Y at time $i+1$, and X^i and Y^i denote the history of activity of X and Y ($X^i \coloneqq \left(X_i, X_{i-1}, \ldots, X_{i-(T-1)}\right)$ and $Y^i \coloneqq \left(Y_i, Y_{i-1}, \ldots, Y_{i-(T-1)}\right)$, respectively). Thus, TE quantifies the reduction in uncertainty of Y_{i+1} when Y^i and X^i are known, compared to when only Y^i is known. Then, aiming to add and compare the TE between different ROI pairs, we normalized TE by mutual information and obtained NDTE F_{XY} [67]:

$$F_{\mathrm{XY}} = T_{\mathrm{XY}} / I\left(Y_{i+1}; X^i, Y^i\right) = I\left(Y_{i+1}; X^i \mid Y^i\right) / I\left(Y_{i+1}; X^i, Y^i\right). \tag{2}$$

The normalization factor $I\left(Y_{i+1}; X^i, Y^i\right)$ is the mutual information between Y_{i+1} and $\{X^i, Y^i\}$, and can be decomposed as

$$I\left(Y_{i+1}; X^i, Y^i\right) = I\left(Y_{i+1}; X^i \mid Y^i\right) + I\left(Y_{i+1}; Y^i\right). \tag{3}$$

This decomposition indicates that the normalization compares the predictability of Y_{i+1} by $X^i \mid Y^i$ (i.e., $I\left(Y_{i+1}; X^i \mid Y^i\right)$) with the internal predictability of Y_{i+1} by Y^i (i.e., $I\left(Y_{i+1}; Y^i\right)$).

Generally, weak functional or causal connections estimated from brain activity can be spurious connections owing to noise [69]; thus, edge filtering is often applied [64, 70]. Therefore, we also filtered the edges so that the density of edges was 10%.

Extraction of bidirectionally interacting cores

From the directed network constructed using NDTE, we extracted subnetworks with particularly strong bidirectional connections—the network "cores." To identify such cores, our proposed framework uses the algorithm proposed by Kitazono et al. [13]. This algorithm enables the hierarchical decomposition of the entire network based on the strength of bidirectional connections.

To demonstrate how the algorithm works, we present its application to a simple toy network in Supplementary Fig. 1. In the toy network, the node set $\{B, E, F, I, J\}$ is connected bidirectionally whereas the node set $\{A, C, D, G, H\}$ is connected in a feedforward manner. Extracting cores from this network according to the method of Kitazono et al. [13] reveals the network cores, as indicated by various shades of color. The subnetwork $\{E, F, I, J\}$, highlighted in orange, represents the core with the strongest bidirectional connections, followed by the subnetwork $\{B, E, F, I, J\}$, shown in blue. As shown in this example, cores with stronger bidirectional connections are nested within those with weaker bidirectional connections and thus generally form a unimodal or multimodal hierarchical structure.

Below, we outline the definition of cores according to the method of ref. [13]. A core is defined as a subnetwork that cannot be separated because of its strong bidirectional connections. To define a core, we first introduce the definition of the strength of bidirectional connections. Based on the definition of the strength, we introduce a minimum cut that quantifies the inseparability of a network, which is then used to define cores. Subsequently, we introduce "coreness"—an index, which is defined for each node, that quantifies the strength of the core in which each node is included.

Strength of bidirectional connections

In this subsection, we introduce a measure that quantifies the strength of bidirectional connections between two parts of a network. Let us consider a directed graph $G(V, E)$ where V denotes the node set and E the edge set. For a bi-partition of $(V_{\mathrm{L}}, V_{\mathrm{R}})$ of the node set V, there are two types of edges connecting V_{L} and V_{R}, determined by their direction. One is the set of edges that outgoing from V_{L} to V_{R}:

$$E(V_{\mathrm{L}} \to V_{\mathrm{R}}) = \{(u, v) \in E | u \in V_{\mathrm{L}}, v \in V_{\mathrm{R}}\}. \tag{4}$$

The other is in the opposite direction:

$$E(V_{\mathrm{R}} \to V_{\mathrm{L}}) = \{(u, v) \in E | u \in V_{\mathrm{R}}, v \in V_{\mathrm{L}}\}. \tag{5}$$

To evaluate the strength of bidirectional connections between V_{L} and V_{R}, we first sum up the weight of the edges in each direction,

$$w(V_{\mathrm{L}} \to V_{\mathrm{R}}) = \sum_{e \in E(V_{\mathrm{L}} \to V_{\mathrm{R}})} w_e, \tag{6}$$

$$w(V_{\mathrm{R}} \to V_{\mathrm{L}}) = \sum_{e \in E(V_{\mathrm{R}} \to V_{\mathrm{L}})} w_e. \tag{7}$$

We then define the strength of bidirectional connections as the minimum of the two:

$$w(V_{\mathrm{L}}; V_{\mathrm{R}}) := \min\left(w(V_{\mathrm{L}} \to V_{\mathrm{R}}), w(V_{\mathrm{R}} \to V_{\mathrm{L}})\right). \tag{8}$$

With this definition, when two segments of a network are linked only unidirectionally, the strength of the bidirectional connections $w(V_{\mathrm{L}}; V_{\mathrm{R}})$ equals zero (Supplementary Fig. 2**a**); this indicates that the segments are bidirectionally "disconnected." In Supplementary Fig. 2**b**, the connection from one segment to the other is strong ($w(V_{\mathrm{L}} \to V_{\mathrm{R}}) = 3$), whereas the connection in the opposite direction is weak ($w(V_{\mathrm{R}} \to V_{\mathrm{L}}) = 1$). Hence, the bidirectional connections are weak ($w(V_{\mathrm{L}}; V_{\mathrm{R}}) = 1$). In Supplementary Fig. 2**c**, the connections in both directions are strong ($w(V_{\mathrm{L}} \to V_{\mathrm{R}}) = w(V_{\mathrm{R}} \to V_{\mathrm{L}}) = 3$), resulting in strong bidirectional connections ($w(V_{\mathrm{L}}; V_{\mathrm{R}}) = 3$).

Measuring inseparability by minimum cut

Using the strength of the bidirectional connections between the two parts of a network defined above, we define a minimum cut (min-cut). As mentioned earlier, a core is a subnetwork that cannot be separated because of its strong bidirectional connections. In other words, a core cannot be cut into two parts without losing many strong edges regardless of how it is cut. To measure such inseparability, we consider the bipartitioning of a network for which the strength of bidirectional connections is the minimum among those for all possible bipartitions—the min-cut. Mathematically, a min-cut $(V_{\mathrm{L}}^{\mathrm{mc}}, V_{\mathrm{R}}^{\mathrm{mc}})$ is defined as

$$(V_{\mathrm{L}}^{\mathrm{mc}}, V_{\mathrm{R}}^{\mathrm{mc}}) = \underset{(V_{\mathrm{L}}, V_{\mathrm{R}}) \in \mathcal{P}_V}{\arg\min}\ w(V_{\mathrm{L}}; V_{\mathrm{R}}), \tag{9}$$

where $\mathcal{P}_V$ denotes the set of all bipartitions of V. We refer to the strength of bidirectional connections for a min-cut as "min-cut weight." We denote the min-cut weight of a min-cut $(V_{\mathrm{L}}^{\mathrm{mc}}, V_{\mathrm{R}}^{\mathrm{mc}})$ of graph G as

$$w_G^{\mathrm{mc}} := w(V_{\mathrm{L}}^{\mathrm{mc}}; V_{\mathrm{R}}^{\mathrm{mc}}). \tag{10}$$

Based on the definition of a min-cut weight, because two parts of a network G given by any bipartition are connected with a strength greater than or equal to its min-cut weight w_G^{mc}, w_G^{mc} can be taken to represent the network's inseparability.

Definition of cores: Complexes

Finally, we introduce the definition of a core [13,71]. To formally define the cores, we must introduce the concept of an induced subgraph. Let G be a graph consisting of a node set V and an edge set E and let $S \subseteq V$ be a subset of nodes. Then, an induced subgraph $G[S]$ is a graph that consists of all the nodes in S and all the edges that connect the nodes in S. The weight of the minimum cut of $G[S]$ is denoted by $w_{G[S]}^{\mathrm{mc}}$. We are now ready to define the complexes as follows.

Definition 1 (Complex)**.** *An induced subgraph $G[S]$ ($S \subseteq V$) is called a complex if it satisfies $w_{G[S]}^{\mathrm{mc}} > 0$ and $w_{G[S]}^{\mathrm{mc}} > w_{G[T]}^{\mathrm{mc}}$ for any subset T that is a superset of S ($T \supset S$ and $T \subseteq V$).*

The definition of complexes is shown schematically in Supplementary Fig. 1, wherein we consider the induced subgraphs of graph G that comprises ten nodes $\{A, B, \ldots, J\}$. An induced subgraph $G[\{E, F, I, J\}]$ is a complex because its w^{mc} is greater than that of any induced subgraph of G that is its supergraph (e.g., $G[\{B, E, F, I, J\}]$ and $G[\{D, E, F, H, I, J\}]$). The whole graph G is a complex if it satisfies $w_G^{\mathrm{mc}} > 0$ by definition. We define

$w^{\mathrm{mc}} = 0$ for single nodes because we cannot consider partitions of a single node. Therefore, single nodes cannot be complexes.

If we identify complexes by exhaustive search, the computational time increases exponentially with the number of nodes. However, the algorithm proposed by Kitazono et al. [13] reduces the time complexity of the search to polynomial order [13, 71, 72], which allows complexes to be extracted even from large networks.

Coreness

To quantify the strength of the bidirectional connections of the cores wherein each node is included, we measured the "coreness" of each node that we defined using the complexes and their w^{mc}. A node that is included in complexes with a high w^{mc} has high coreness; conversely, a node that is included only in complexes with low w^{mc} has low coreness. Specifically, we defined the coreness of a node v as k_v if the node v is included in a complex with $w^{\mathrm{mc}} = k_v$ but not included in any complex with $w^{\mathrm{mc}} > k_v$. Equivalently, we can define the coreness of a node v as the largest of the w^{mc} of all complexes containing the node v:

$$k_v = \max_{C \in \mathcal{G}_{\mathrm{complex}} | v \in V(C)} w_C^{\mathrm{mc}}, \tag{11}$$

where $\mathcal{G}_{\mathrm{complex}}$ denotes the set of all complexes in the graph G and $V(C)$ denotes the set of all nodes in the complex C. The coreness Eq. (11) is equal to the largest of the w^{mc} of all subnetworks containing node v [13]

$$k_v = \max_{S | v \in V(S)} w_{G[S]}^{\mathrm{mc}}. \tag{12}$$

Herein, to investigate the common tendencies of cores across the conditions of the resting state and seven tasks, we normalized coreness by dividing it by the maximum coreness among all nodes for each brain state; then, we averaged these values across all brain states. In the remainder of this paper, "coreness" refers to this averaged value across the conditions.

Cortical and subcortical rendering

For visualization, the coreness of each ROI was assigned to the cortical surface and the subcortical volume map as follows. First, the coreness of the cerebral cortex was assigned to the fsLR-32k CIFTI space using the parcellation label defined by ref. [65]. The resulting map was then displayed on the "fsaverage" inflated cortical surface by Connectome Workbench [73]. For the subcortex, the coreness was assigned to the MNI152 nonlinear 6th-generation space by using the parcellation label from ref. [66]. The resulting map was then plotted using

nilearn (https://nilearn.github.io/).

Cortical and subcortical major divisions

To analyze the variability in the coreness of ROIs in the cerebral cortex from a cognitive functional perspective, we used the brain atlas developed by Yeo et al. [74], which divides a functional brain network within the cerebral cortex into seven major subnetworks. Each ROI was assigned to one of the seven subnetworks [65], specifically as: default mode, control, limbic, salience/ventral attention, dorsal attention, somatomotor, and visual networks.

Similarly, the subcortical ROIs were classified according to the Melbourne subcortical atlas [66]. Each ROI was assigned to one of the seven areas: the hippocampus, amygdala, thalamus, nucleus accumbens, globus pallidus, putamen, and caudate nucleus.

Comparison of coreness and mean response rates for intracranial electrical stimulation

In this study, we used the mean response rate (MRR) for intracranial electrical stimulation (iES) [58], which was obtained as follows: Electrical stimulation was administered to each stimulation site via electrodes. Subjects were then asked to report whether they felt any change in their perceptions, including tactile, visual, emotional perceptions, or motor movement. Electrodes were classified as either "responsive" or "silent" based on the subject's feedback. Data from 119 subjects were aggregated, and the MRR was calculated for each of the subnetworks of Yeo-17 (or -7) network atlas [74] as the proportion of responsive electrodes compared to all electrodes within each subnetwork. We rendered the MRR for the 17 subnetworks [58] on the brain surface in Fig. 3**a** and that for the seven networks in Supplementary Fig. 4. For a comprehensive explanation of the methodology, refer to ref. [58].

To compare coreness with the MRR (Fig. 3**c** and Supplementary Fig. 4**c**), we first calculated the average coreness of the ROIs within each subnetwork of the Yeo-17 network atlas, according to the ROI assignment to each subnetwork by Schaefer et al. [65], and compared the two metrics. However, prior to this comparison, a minor adjustment to the MRR was necessary owing to the slight differences in voxel allocation to subnetworks between the original Yeo-17 network atlas and Schaefer's assignment. We projected the MRR onto the brain surface voxels as per the original Yeo-17 network atlas and then recalculated the averages according to Schaefer's assignment.

NeuroSynth term-based meta-analysis

To investigate whether the functions of ROIs with high and low coreness in the cerebral cortex differ, we performed a meta-analysis using NeuroSynth—a platform for large-scale automated meta-analysis of fMRI data (www.neurosynth.org) [60]. We investigated how the degree of association between ROIs and topics that represent cognitive functions vary depending on their coreness as follows. First, we divided the 200 cerebral cortical ROIs into 20 groups according to coreness intervals of 0.05 (0–0.05 to 0.95–1) and, based on which interval respective ROIs belonged to, we classified them into one of these 20 divisions.

Next, for ROIs in each interval, we calculated the association with specific terms as follows [60]. We first calculated the average value of the "association test" meta-analytic maps provided by Neurosynth [60] for the ROIs in each interval. These "association test" maps comprise z-scores obtained from a two-way ANOVA that tests for non-zero associations between voxel activation and the use of a specific term in an article. For example, a large positive z-score for a voxel i in the association test map for the term "reward" implies that compared to studies without the term, those with the term in the title or abstract are more likely to report activation of the voxel i. By calculating the average of these z-scores for all voxels of ROIs in each interval for every term, we obtain a vector X with the dimension of the number of terms that represents the association between the ROIs in the interval and the terms. The voxels were assigned to ROIs by the NIFTI format parcellation label defined in the FSL MNI152 2mm space by Schaefer et al. [65].

Finally, we evaluated the association of the ROIs in each interval with a specific topic by calculating the correlation r between this vector X and a vector Y, which is a binary vector assigned for each topic with the dimension of the number of terms, and indicates whether each topic involves the terms. The list of topics and terms, and which term each topic involves, were based on the data available at https://github.com/NeuroanatomyAndConnectivity/gradient_analysis/blob/master/gradient_data/neurosynth/v3-topics-50-keys.txt. The number of topics was 50, of which we used 44 topics for analysis, after excluding six topics that did not capture any consistent cognitive functions following ref. [61]. The topic terms (e.g., "eye movements," "cued attention," and "emotion") that represent the topics were set according to Margulies et al. [61]. Fisher's z-score, obtained by Fisher's z-transformation of the correlation r between the vectors X and Y, was used as the degree of association between each topic and ROIs in the interval. For more information see https://neurosynth.org and ref. [60].

We calculated Fisher's z-score for all intervals and all topics. Subsequently, for visualization, we sorted the topics based on the weighted average of the coreness of intervals by using Fisher's z-score as the weight, wherein topics related to high- and low-coreness ROIs were placed at the top and bottom, respectively. Any

topic that did not reach a significant threshold of $z > 3.1$ in any interval was excluded; thus, only the remaining 24 topics were displayed (Fig. 4). The analysis was performed using a modified code available at https://github.com/NeuroanatomyAndConnectivity/gradient_analysis.

Comparison of the functional connectivity gradient and coreness

A functional connectivity gradient refers to the spatial variation in functional connectivity patterns across different brain regions [61, 75–77]. Mathematically, a functional connectivity gradient is obtained through an embedding of a functional network into a low-dimensional space. Previous studies have shown that functional connectivity gradients are associated with numerous neuroscientific features [61, 75–77], including molecular, cellular, anatomical, and functional aspects, indicating that these gradients capture essential properties of the brain's functional organization. Among those functional connectivity gradients, in this study, we utilized the one for the entire human cerebral cortex developed by Margulies et al. [61], which was obtained by the following procedures: A functional connectivity matrix was first obtained by calculating the correlation between all pairs of gray coordinates from resting-state fMRI data. Next, a nonlinear dimensionality-reduction technique called diffusion embedding [78] was applied to this connectivity matrix. The functional connectivity gradient was defined as the first component in the embedding space, which accounts for the greatest variance in the connectivity patterns. Margulies et al. have demonstrated that the functional connectivity gradient corresponds to a cortical spatial gradient of the degree of abstraction and integration in processing, ranging from the primary sensory/motor cortex to transmodal areas [61]. The detailed methodology for calculating the functional connectivity gradient has been described in ref. [61]. Data of the functional connectivity gradient are publicly available at https://github.com/NeuroanatomyAndConnectivity/gradient_analysis. For our analysis, we averaged the functional connectivity gradient for the voxels in each ROI, and these averaged values were then used for the scatterplot in Fig. 5**c**.

Extraction of cores regardless of bidirectionality

To assess the impact of incorporating the bidirectionality of connections on the results, we conducted an ablation study. Instead of using Eq. (8), which defines the measure of bidirectional connection strength, we used a simpler measure: we summed the weights of all edges that connect two parts, regardless of their directions:

$$w_{\text{simple sum}}\ (V_{\mathrm{L}}; V_{\mathrm{R}}) = \frac{1}{2} \sum_{e \in E(V_{\mathrm{L}} \to V_{\mathrm{R}}) \cup E(V_{\mathrm{R}} \to V_{\mathrm{L}})} w_e \tag{13}$$

Here, the factor of 2 in the denominator is to maintain consistency with the case wherein bidirectionality is considered. The use of this connection strength Eq. (13) is equivalent to applying the bidirectional connection strength Eq. (8) to an undirected network, which is obtained by ignoring edge directions and is equivalently achieved by setting its connection matrix to $W' = (W + W^\top)/2$, where W is the connection matrix of the original directed network [13].

Degree of nodes

The degree $\deg(v)$ of a node v is defined as the sum of the weights of all edges connected to the node v, regardless of the edge directions:

$$\deg(v) = \sum_{e \in E(\{v\} \to V) \cup E(V \to \{v\})} w_e. \tag{14}$$

To investigate the common tendencies of node degree across the conditions of the resting state and seven tasks, we normalized node degree in the same way as we normalized coreness: by dividing it by the maximum degree among all nodes for each brain state and then averaging these values across all brain states. In the remainder of this paper, "degree" refers to this averaged value across the conditions.

Results

Framework for extracting bidirectionally interacting cores from a directed functional network

This section outlines a novel framework that has been proposed in this paper for extracting the cores of brain networks, with strong bidirectional interactions from brain activity data. The framework consists of two steps. First, the strength of causal influence between brain regions is estimated from brain activity to construct a whole-brain-directed functional network. Second, strongly bidirectional cores are extracted from the network.

In the first step, we quantify the strength of causal influence from one ROI to another using fMRI data (Fig. 1**a**). This quantification is performed for every possible combination of ROIs, enabling us to construct a whole-brain directed network (Figure 1**b**) that illustrates the strength of causal influences among brain ROIs. To measure these strengths, we used normalized directed transfer entropy (NDTE) [67]—a statistical method that is designed to estimate the strengths of the causal influence between two sets of time-series data. The direction of the arrow in the schematic (Figure 1**b**) indicates the direction of causal influence between ROIs, whereas the thickness of these arrows reflects the magnitude of the influence. The methodology for this calculation has been

described in the Methods section.

In the second step, from the constructed directed network, we extract subnetworks with strong bidirectional interactions, or the "cores" of a network that are identified using the method proposed by Kitazono et al. [13]. In their study, a core is termed a "complex" [13], which is defined as a subnetwork that is composed of stronger bidirectional connections than other subnetworks that include it. Using the method, the directed network can be hierarchically decomposed into complexes based on the strength of bidirectional connections. The complexes extracted from the network in Fig. 1**b** are shown in Fig. 1**c**. The subnetwork highlighted with a yellow background is the complex with the strongest bidirectional connections; the subnetworks distinguished by blue and purple backgrounds represent the second-strongest and the weakest complexes, respectively. Generally, complexes exhibit a nested structure wherein stronger bidirectional complexes are contained in weaker ones to form either unimodal or multimodal hierarchical structures.

To quantify the strength of the bidirectional connections of the cores wherein each node is included, we use a measure called the "coreness." We defined the coreness of a node v as the largest of the w^{mc} of all complexes containing the node v (see Methods for details.)

Bidirectional cores in the functional network of the human brain

In this subsection, we present the results of our analysis of fMRI data from the human connectome project (HCP) [57] using our proposed framework. We assume that if a brain region is consistently included in the strongly bidirectional, central cores during the execution of various tasks, those regions are crucial for diverse cognitive functions in general. Therefore, to determine which regions are consistently included in central cores, we first extracted cores at rest and during seven tasks in the HCP. Next, to explore how regions consistently included in the central cores relate to conscious perception, we compared coreness with the rates of iES-induced perceptual elicitation. Additionally, to further characterize the regions consistently included in the central cores, we performed a meta-analysis using NeuroSynth, and compared the cores with the functional connectivity gradient.

Brain regions frequently included in the central cores

We applied our proposed framework to HCP fMRI data to analyze which brain regions were more likely, on average, to be included in the central cores. First, we extracted cores for the resting state and seven tasks (Supplementary Fig. 3). Subsequently, we calculated the average coreness over these eight conditions (see Methods for details). This average is simply referred to as coreness, hereafter.

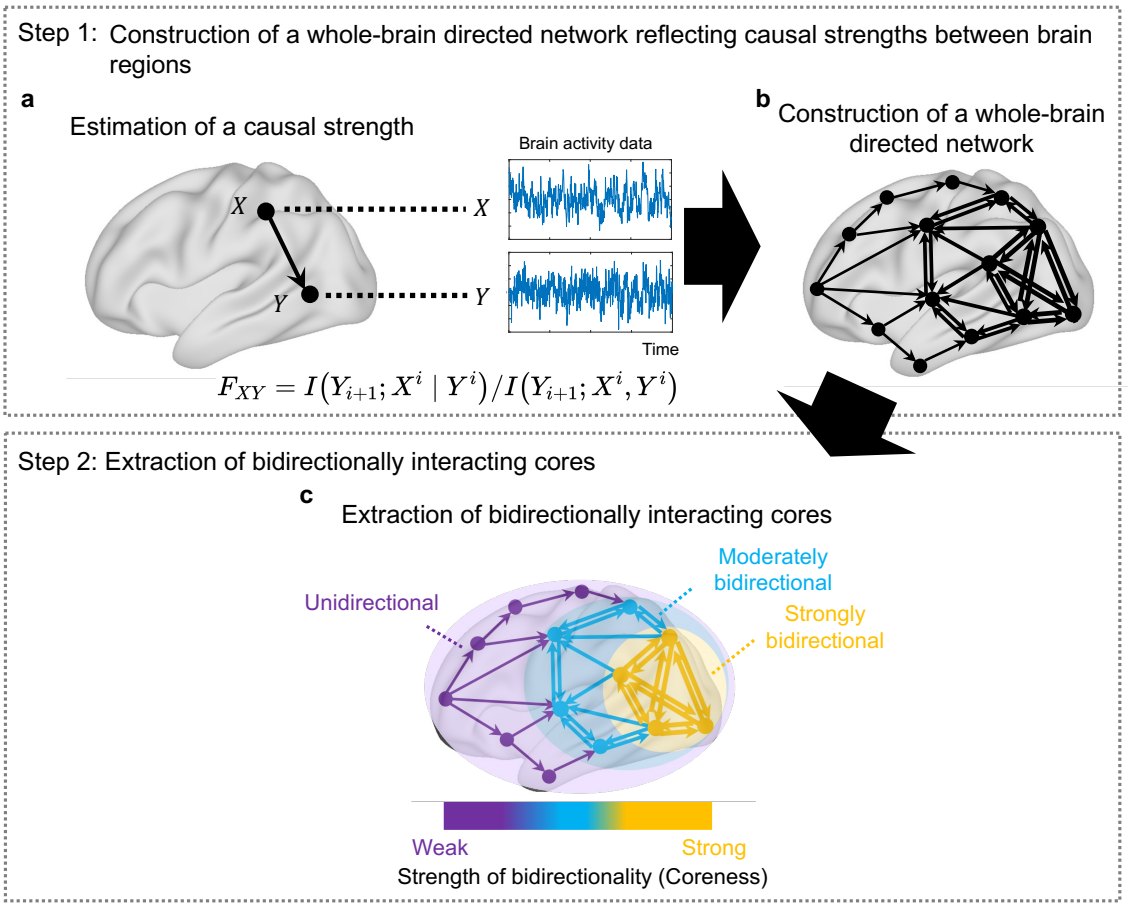


Figure 1: **Schematic of our framework for extracting bidirectionally interacting cores from a whole-brain directed network.** Our proposed framework in this study comprises two steps. In the first step, based on brain activity, we estimate the strength of the causal influence between brain regions. **a**, Specifically, we calculate the NDTE F_{XY} [67] from the brain activities of regions X and Y. NDTE F_{XY} represents the strength of causal influence from the source brain region X to the target brain region Y (See Methods for details). **b**, We then apply this approach to all pairs of brain regions across the whole brain and thereby construct a whole-brain-directed network. **c**, In the second step, from the constructed directed network, we extract cores with strong bidirectional connections. To extract cores, we use the algorithm by Kitazono et al. [13]. This algorithm hierarchically decomposes a network into the cores with the strongest, second strongest, and third strongest bidirectional connections, and so on. In this figure, the subnetwork in yellow represents the core with the strongest bidirectional connections, whereas that in blue represents the second-strongest core. The entire network shown in purple is the weakest core and has parts with completely unidirectional connections.

The results revealed that compared to the subcortical regions, cerebral cortical regions tend to have higher coreness (Figs. 2**a** and 2**b**). Many ROIs in the cerebral cortex display high coreness, whereas all subcortical regions show lower coreness, although they are from diverse areas, such as the hippocampus, amygdala, and thalamus. The results, in conjunction with the argument supported by experimental evidence and theoretical frameworks—that the cerebral cortex is crucial for conscious perception, whereas subcortical areas may not be—imply that bidirectional interactions play a significant role in conscious perception, and this interpretation is

further explored in the Discussion section.

A more detailed examination within the cerebral cortex revealed that not all ROIs possess high coreness, indicating variability in coreness among different regions (Fig. 2**a**). To understand this variability in coreness from a cognitive functional perspective, we divided ROIs into seven functional subnetworks according to Yeo's 7 network atlas [74] (Fig. 2**b**). The results showed that ROIs with high coreness were particularly located in the salience/ventral attention, dorsal attention, somatomotor, and visual networks. Conversely, ROIs in the limbic network exhibited low coreness, and the default mode network and control network included ROIs with low coreness.

Next, to explore the spatial distribution trend of ROIs with particularly high or low coreness within the cerebral cortex, we visualized ROIs with a coreness greater than 0.7 and less than 0.3 on the surface of the cerebral cortex (Fig. 2**c**). This visualization used color coding based on the Yeo-7 network atlas. The results showed that ROIs with high coreness were mainly located in the somatosensory and motor areas surrounding the central sulcus, which are part of the somatomotor and dorsal attention networks, and in the occipital areas, which belong to the visual network. Additionally, high-coreness ROIs were identified in the salience network, which were relatively dispersed in their locations. On the other hand, low-coreness ROIs were scattered across various areas, including the lateral temporal, medial temporal, orbitofrontal, and cingulate cortices.

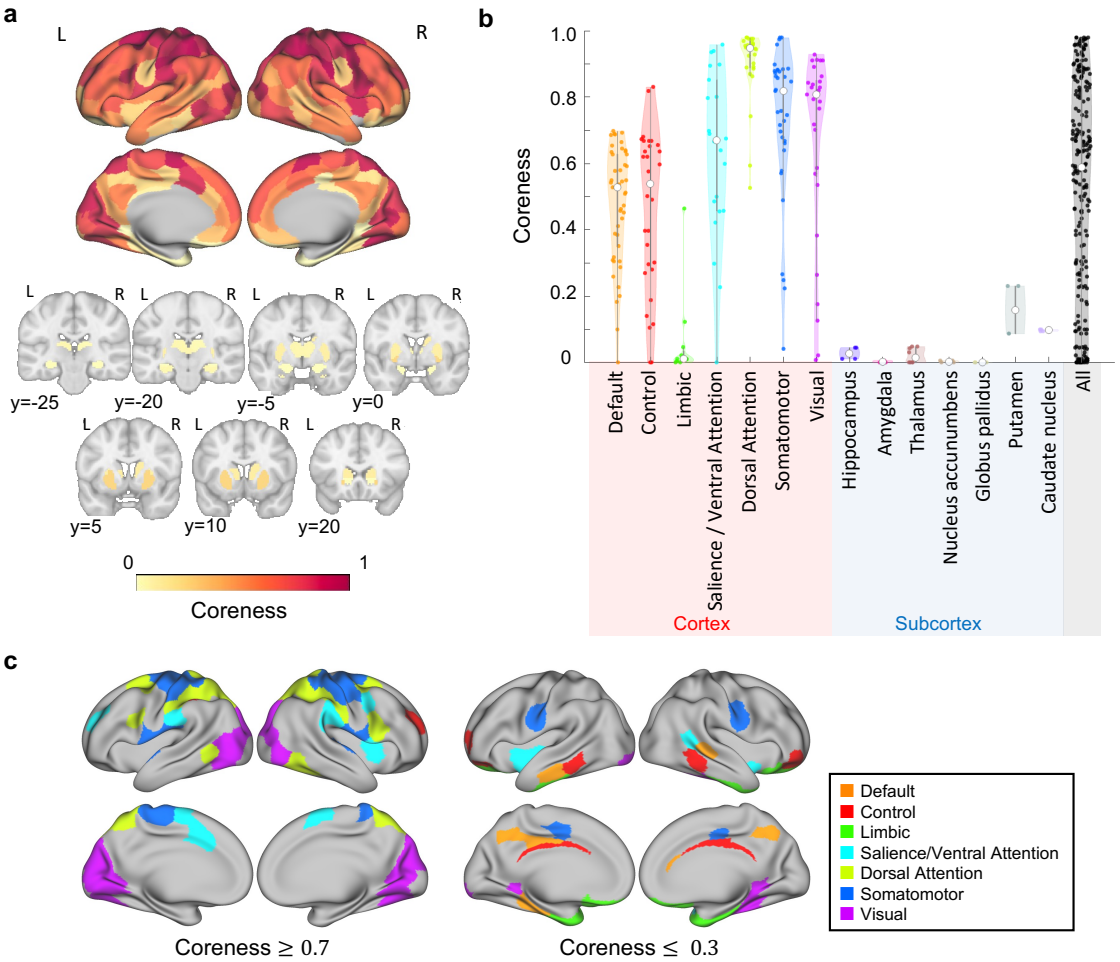


Figure 2: **Extracted bidirectionally interacting cores.** Compared to subcortical regions, cerebral cortical regions tend to have higher coreness. Furthermore, within the cerebral cortex, ROIs with high coreness are particularly found in the salience/ventral attention, dorsal attention, somatomotor, and visual networks. In contrast, ROIs in the limbic network show low coreness, and the default mode network and frontoparietal control network include ROIs with low coreness. **a**, Coreness in the cerebral cortex (top) and the subcortex shown in seven coronal slices (bottom). Coordinates of the slices were given in the MNI (Montreal Neurological Institute) space. **b**, Violin plots of coreness according to the divisions of the Yeo-7 network atlas and major divisions of the subcortex. Each violin plot represents the probability density of coreness for each division, and the colored dots inside the plots represent ROIs. The white dot represents the median, and the thick line inside indicates the interquartile range. **c**, ROIs with coreness > 0.7 or < 0.3 are colored as per the Yeo-7 network atlas.

Comparison of the core structure with the mean response rate for iES

In the preceding subsection, we observed that, compared to subcortical ROIs, the cerebral cortical ROIs tended to have higher coreness and that there was a variation in coreness within the cerebral cortex. Next, we investigated whether the coreness of cerebral cortical ROIs, whether high or low, is associated with their importance for conscious perception. Specifically, we compared the coreness with the rates of iES-induced perceptual elicitation [58].

Figure 3**a** shows the cortical map representing the mean response rate (MRR) for iES. By comparing this map of the MRR with the map of coreness (Fig. 3**b**, the same figure as Fig. 2**a** is reprinted), similar structures can be observed in the two maps—high MRR and coreness are commonly found in the areas such as the somatosensory and motor areas surrounding the central sulcus, and in the visual cortex in the occipital region, whereas low MRR and coreness are visible in the areas such as the prefrontal, lateral temporal, posterior parietal, and posterior cingulate cortices. The relationship between the coreness and the MRR is directly observable from the scatterplot (Fig. 3**c**), showing a positive correlation between the two metrics ($r = 0.46$). This relationship was confirmed not only in the comparison using the Yeo-17 network atlas (Fig. 3**c**) but also in the comparison using the Yeo-7 network atlas (with $r = 0.57$, Supplementary Figs. 4**a** and 4**b**).

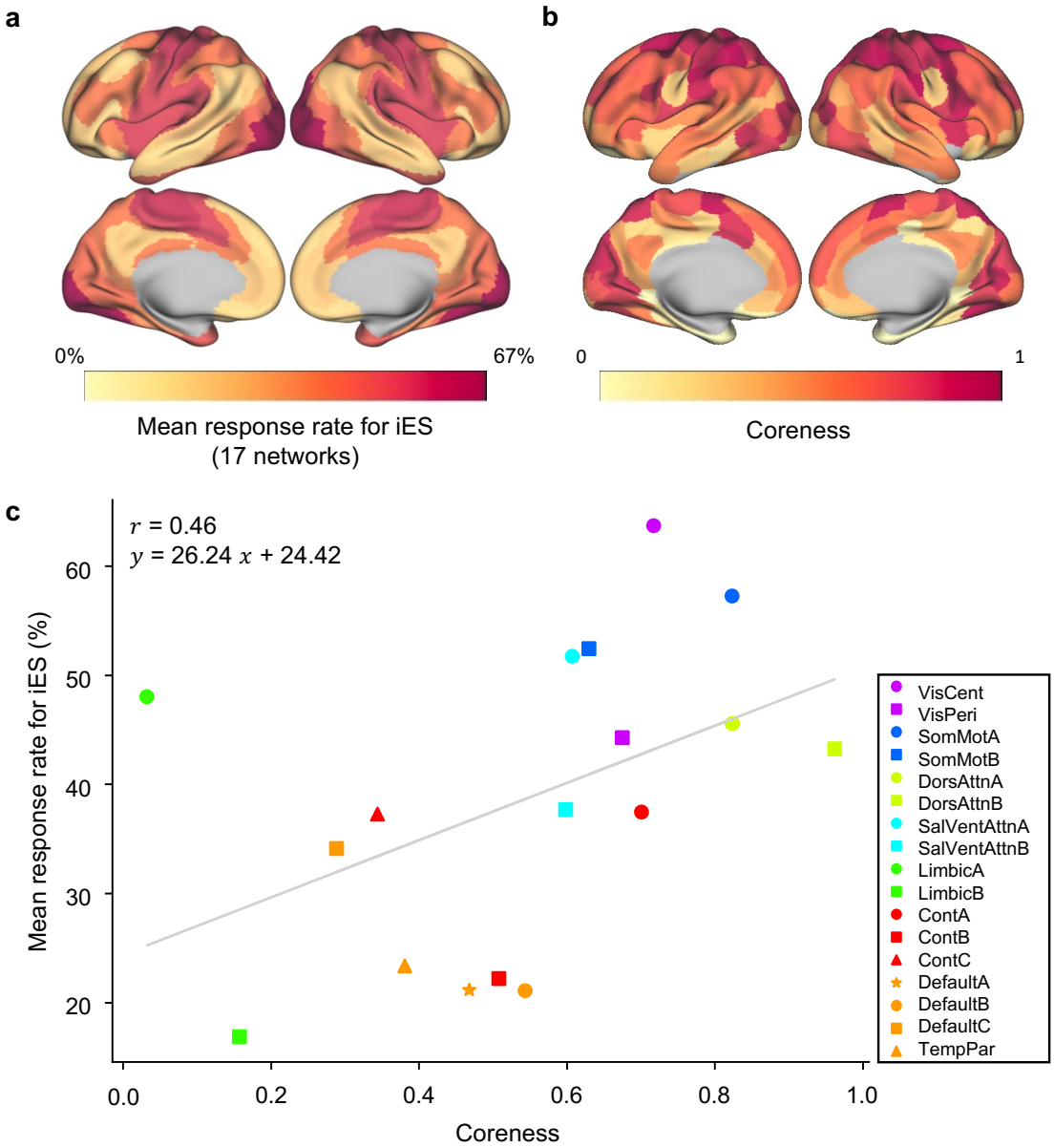


Figure 3: **Comparison of the core structure with the mean response rate for iES.** There is a correlation between coreness and the mean response rate (MRR) for iES. **a, b**, A cortical surface rendering of the MRR for iES [58] (**a**) and that of coreness (**b**, same as Fig. 2**a**). **c**, A scatter plot showing the average coreness for each division of the Yeo-17 network atlas (horizontal axis) and the MRR (vertical axis). A positive correlation exists between the coreness and the MRR ($r = 0.46$). The solid line in the figure represents the least-squares line. Each point is color-coded according to the Yeo-7 network atlas and then assigned marker shapes according to the Yeo-17 network atlas.

Term-based meta-analysis of the core structure using NeuroSynth

In the previous subsection, we observed a correlation between the coreness and the mean response rate for iES in the cerebral cortex, indicating that high-coreness ROIs are associated with conscious perception. To further

explore the cognitive functions linked to ROIs with varying levels of coreness, we conducted a term-based meta-analysis using NeuroSynth [60], which statistically evaluates the association between each ROI and cognitive functions in the literature by analyzing information from thousands of published fMRI-based studies.

The results indicated that high-coreness ROIs are more likely to be associated with lower-order sensorimotor cognitive functions, such as "eye movements," "visual perception," and "motor" (Fig. 4). In contrast, low-coreness ROIs are more likely to be associated with terms related to complex, abstract processes, including "emotion," "declarative memory," "reward-based decision making," and "autobiographical memory" (Fig. 4).

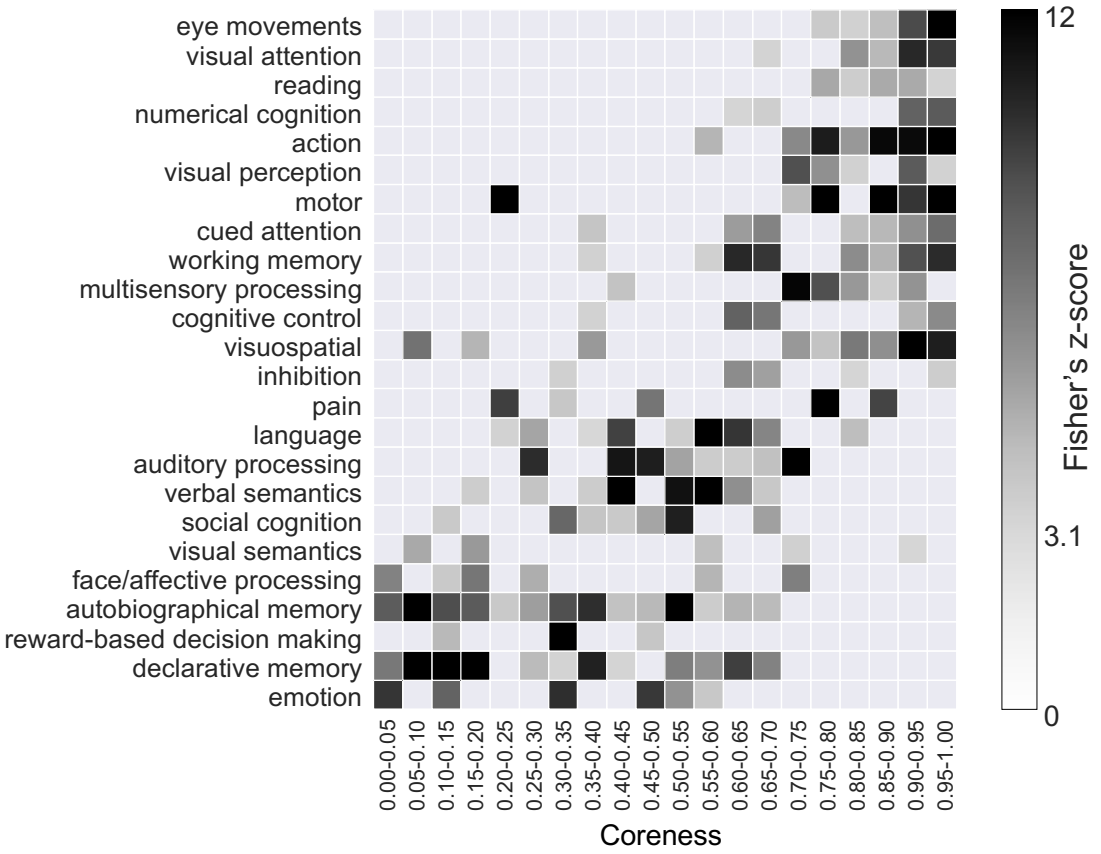


Figure 4: **Term-based meta-analysis of the core structure using NeuroSynth.** High-coreness ROIs are related to lower-order sensorimotor functions whereas low-coreness ROIs are related to higher-order cognitive functions. This figure shows a term-based meta-analysis using NeuroSynth applied to the coreness of the cerebral cortex. The columns represent coreness, at intervals of 0.05 from 0–0.05 to 0.95–1. The rows represent the topic terms used in the meta-analysis. The grayscale of each cell indicates the Fisher's z-score representing the association strength between ROIs in each division of coreness and topic terms, as obtained from the meta-analysis. Only components that reached a significant threshold of $z > 3.1$ are colored. For visualization, topic terms are arranged by the weighted mean of the coreness of intervals with Fisher's z-score as weights, by placing terms related to high-coreness regions at the top and those related to low-coreness regions at the bottom. Only the cerebral cortical ROIs were used for this analysis (see Methods for details).

Comparison of the core structure with the functional connectivity gradient

A meta-analysis using NeuroSynth indicated that regions with high and low coreness are associated with lower-order and higher-order cognitive functions, respectively. To further assess the relationship between coreness and lower-order or higher-order functions in the cerebral cortex, we compared coreness with the functional connectivity gradient [61], which is a low-dimensional embedding of the functional connectivity at rest and accounts for the greatest variance in the connectivity patterns. The gradient is known to correspond to a spectrum of the degree of abstraction and integration in processing, where its upper end is associated with lower-order sensory processing and the lower end with higher-order (abstract and integrative) cognitive functions [61].

Therefore, if there exists a relationship where high-coreness and low-coreness ROIs are associated with lower-order and higher-order functions, respectively, we would expect the coreness to correlate with the functional connectivity gradient. This expectation was confirmed by the comparison of the two (see Fig. 5**a** and Fig. 5**b**, Fig. 2**a** reprinted). Regions at the upper end of the gradient comprised visual, somatosensory/motor, and auditory areas, demonstrating partial overlap with regions of high coreness. Similarly, regions at the lower end of the gradient predominantly comprise default-mode network areas, which show partial overlap with low-coreness regions. This relationship is further evidenced in the scatterplot for comparing the coreness with the functional connectivity gradient (Fig. 5**c**), indicating a positive correlation between them ($r = 0.40$).

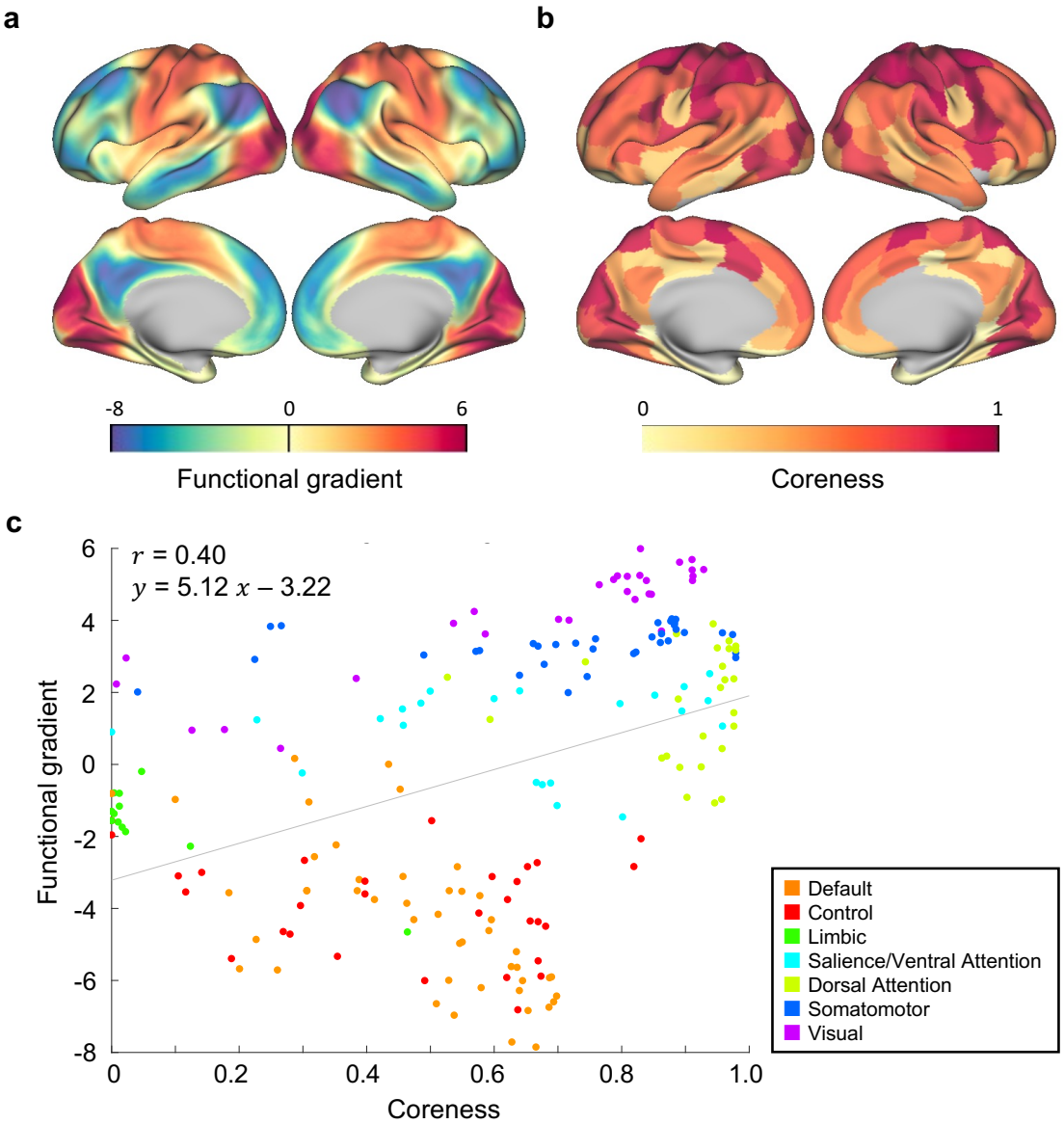


Figure 5: **Comparison of the core structure with the functional connectivity gradient. a, b**, The functional connectivity gradient [61] and coreness are correlated. **a, b**, A cortical surface rendering of the cortical functional connectivity gradient [61] (**a**) and that of coreness (**b**, same as Fig. 2**a**). **c**, A scatterplot of coreness and the functional connectivity gradient. A positive correlation exists between coreness and the functional connectivity gradient ($r = 0.40$). The solid line represents the least-squares line. Each point represents an ROI and is color-coded according to the Yeo-7 network atlas (See Methods for details).

Comparison with the complexes when bidirectionality is ignored

So far, we have observed that compared to subcortical regions, the cortical regions tend to exhibit higher coreness. Additionally, within the cerebral cortex, higher coreness is associated with higher rates of iES-induced perceptual elicitation and lower-order cognitive functions. In this subsection, we explore the importance of considering bidirectionality in these findings. To do so, we examined how cores would change if we only considered the strength of interactions while ignoring bidirectionality. Not accounting for bidirectionality equates to the symmetrizing of a network and treating it as an undirected network (see Methods for details).

First, on comparing the broad categories—cortical and subcortical—similar to the observations made when considering bidirectionality (Figs. 2**a** and 2**b**), cortical regions generally exhibited higher coreness than subcortical regions, even in the absence of consideration of bidirectionality (Figs. 6**a** and 6**b**). Next, we examined the differences within the cerebral cortex when considering, versus ignoring, bidirectionality. The results showed that although coreness was correlated between the two cases, large differences were observed in certain regions (Fig. 6**c**); specifically, in regions with low or medium coreness when considering bidirectionality, coreness was largely increased by ignoring bidirectionality (see Fig. 6**c**, highlighted area, and Fig. 6**d**). Thus, there was a change that counteracted the trend in coreness that existed when bidirectionality was considered. This suggests that the correlation of coreness with the MRR for iES, as well as with the functional connectivity gradient—both observed when considering bidirectionality— weakened when bidirectionality was ignored. Indeed, in regions with smaller MRR and the functional connectivity gradient tended to show a greater increase in coreness if bidirectionality was ignored (Supplementary Figs. 5**a** and 5**b**), resulting in a weaker correlation of coreness with the MRR and with the functional connectivity gradient (Supplementary Figs. 6**a** and 6**b**).

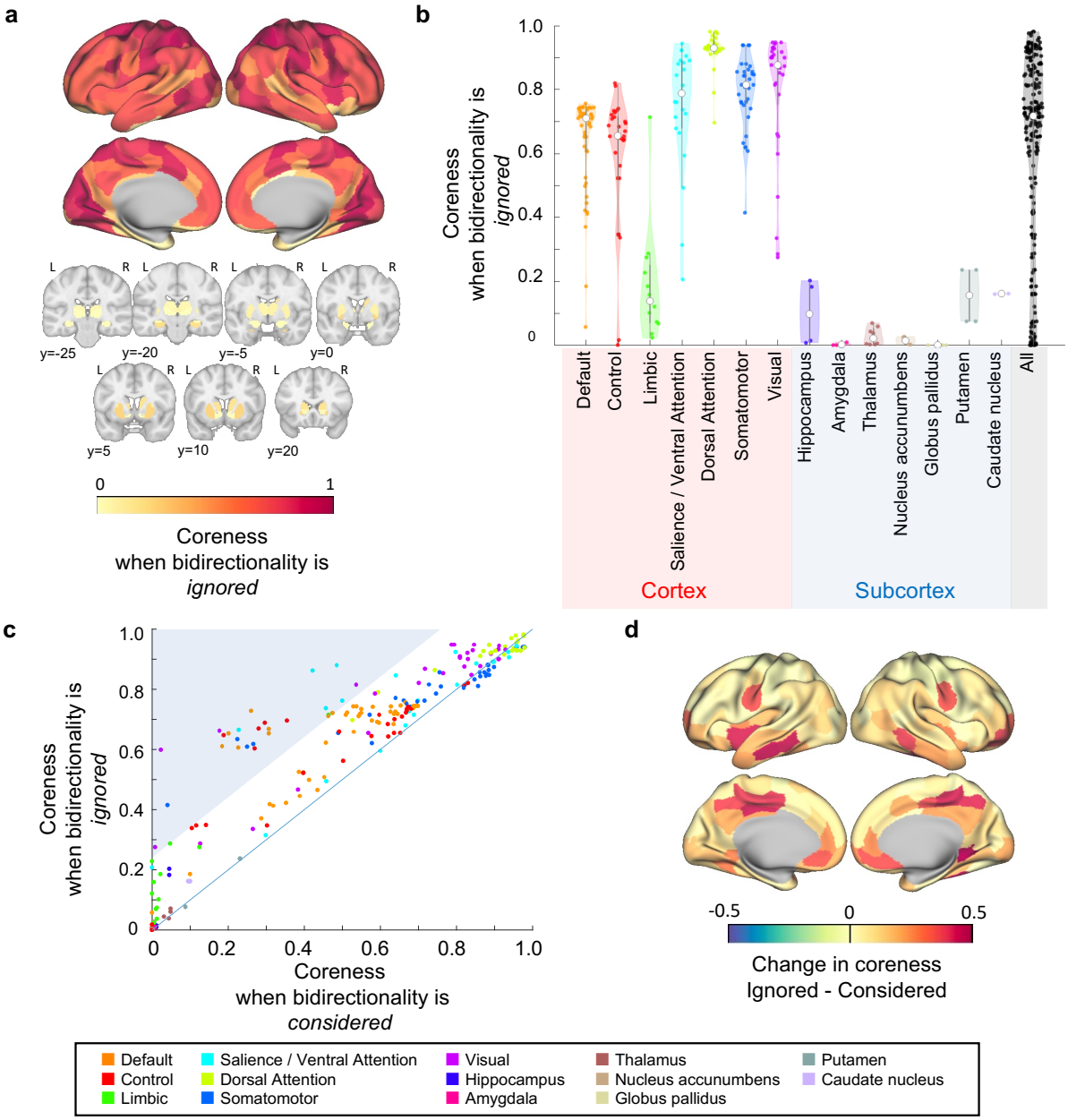


Figure 6: **Comparison of the core structure with the case when bidirectionality is ignored.** **a** and **b**, When bidirectionality is ignored, the cortical regions tend to have higher coreness compared to the subcortical regions, as in the case when bidirectionality is considered. **a**, A cortical surface rendering of coreness when bidirectionality is ignored in the cerebral cortex (top) and that in the subcortex shown in seven coronal slices (bottom). The coordinates of the slices were given in the MNI space. **b**, Violin plots of coreness when bidirectionality is ignored according to the divisions of the Yeo-7 network atlas and major divisions of the subcortex. **c**, A comparison of coreness between considering and ignoring bidirectionality. The difference between the two cases is small and large for regions with high and low coreness when bidirectionality is considered, respectively. Each point is an ROI and is color-coded according to the Yeo-7 network atlas and subcortical divisions. The solid line represents the identity line ($y = x$). The blue background highlights regions with relatively large changes in coreness between considering and ignoring bidirectionality. **d**, A cortical surface rendering of the changes in coreness.

Comparison with other existing methods for extracting cores

To further assess the importance of considering bidirectionality, we compared our core extraction method using complexes with other existing methods that do not consider bidirectionality. Specifically, we conducted comparisons with s-core decomposition [47–50] and network hubs [51, 52].

First, we undertook a comparison with the s-core decomposition. The s-core decomposition is frequently used for extracting subnetworks with strong connections, which are called s-cores, and does not consider the bidirectionality of connections [47–50]. The comparison revealed that the coreness for the complexes when bidirectionality is considered and the coreness for s-core decomposition (where coreness can be defined in the same way as the complexes. See Methods for details.) were not necessarily equal (Fig. 7**a**). This indicates that, by considering bidirectionality, we can extract core structures that are not identifiable by s-core decomposition. Additionally, the coreness for s-cores and that for complexes when bidirectionality is ignored were approximately equal (Fig. 7**b**). This means that the coreness for s-core decomposition demonstrated a weaker correlation with the MRR for iES or the functional connectivity gradient than did the coreness for the complexes when bidirectionality was considered.

Next, we compared with network hubs, which are nodes with a high degree. Here, the degree of a node is defined as the sum of the weights of edges that are connected to the node, regardless of their directions (Eq. (14)). If strong complexes when bidirectionality is considered basically comprise hubs, this means that bidirectionality does not matter for the extraction of complexes. The results show that the coreness when bidirectionality was considered did not necessarily comprise hubs. Figure 7**c** shows that strong complexes include not only hubs but also medium-degree nodes, although the degree and the coreness were correlated. In contrast, the degree was strongly correlated with the coreness when bidirectionality was ignored, particularly for regions with lower degrees (Fig. 7**d**). Additionally, the degree demonstrated a weaker correlation with the MRR and the functional connectivity gradient than did the coreness when bidirectionality was considered (Supplementary Figs. 7**a** and 7**b**).

These results demonstrate that considering bidirectionality for extracting cores can reveal core structures of the human functional network that are unidentifiable when analyzed using simple methods that do not consider bidirectionality.

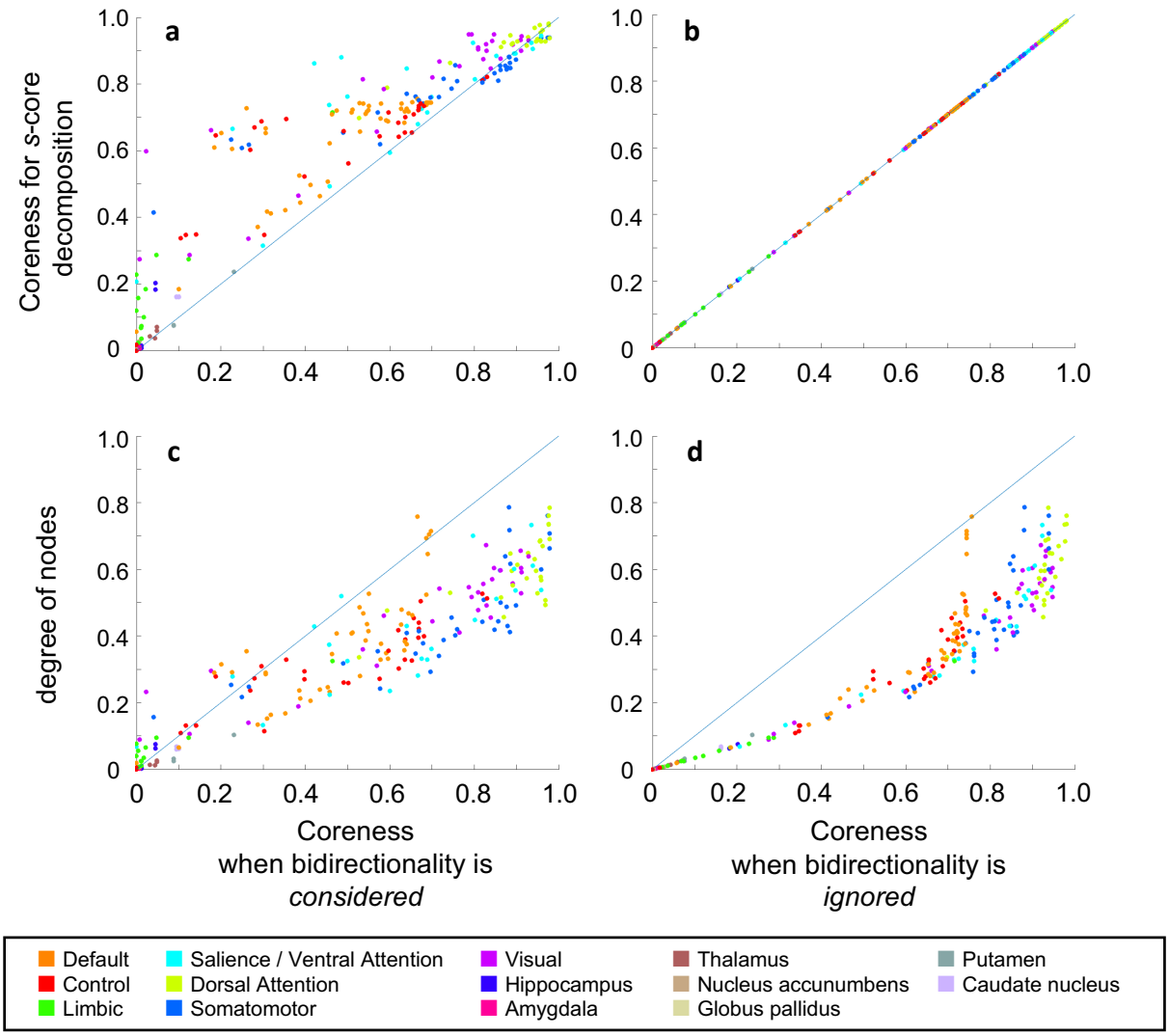


Figure 7: **Comparison with other methods to extract strongly connected cores.** Each metric is the average value across the resting state and seven tasks, after being normalized at the maximum in each brain state. Each point is color-coded according to the Yeo-7 network atlas and subcortical divisions. The solid line represents the identity line ($y = x$).

Discussion

In this study, we investigated the subnetworks of the brain that exhibit strong or weak bidirectionality, and examined their relationship with conscious perception and other cognitive functions. To achieve this, we proposed a novel framework for extracting cores (complexes) with strong bidirectional interactions from brain activity (Fig. 1). We applied this framework to HCP fMRI data of the resting state and seven cognitive tasks and extracted complexes. The analysis revealed the average tendencies of the regions included in the cores with strong and weak bidirectionality across different brain states, and how these are characterized in relation to conscious

perception and other cognitive functions (Figs. 2–5). Regarding the relationship with conscious perception, cerebral cortical regions were more likely to be included in the strongly bidirectional cores whereas subcortical regions were less likely to be included (Fig. 2). Specifically within the cerebral cortex, there was a correlation between iES-induced perceptual elicitation rates and the likelihood of being included in the strongly bidirectional cores (Fig. 3). Regarding the relationship with other cognitive functions, a meta-analysis and a comparison with the functional connectivity gradient revealed that cores with stronger bidirectionality tended to be associated with lower-order, rather than higher-order, cognitive functions (Figs. 4 and 5).

In the following subsections, we first examine the relationship of cores that exhibit strong bidirectional interactions with conscious perception. Based on the results of this study, we argue that the regions with strong bidirectional interactions broadly correspond to regions considered important for conscious perception. We then discuss how the implications of this correspondence relate to previously proposed theories of consciousness. In particular, we delve into the underlying mechanisms of the observed correspondence between iES-induced perceptual elicitation rates and the strength of bidirectional interactions. We focus on interpretations from the perspective of the integrated information theory (IIT), given its significant relevance to our framework among various theories of consciousness. Next, we discuss the interpretation of our findings suggesting that regions associated with lower-order sensorimotor processing belong more to strongly bidirectional cores compared to regions in the association cortices, which are known for their role in higher cognitive functions. Additionally, we contextualize our results by comparing the extracted cores with existing core extraction methods from a network analysis perspective. Finally, we discuss future prospects for applications of our proposed framework.

Correspondence between the strength of bidirectional interactions and the significance for conscious perception

In this subsection, we discuss a link between a region's inclusion in the central cores and its importance for conscious perception. This argument is supported by two key findings: firstly, cortical regions are identified as more likely to be included in central cores than subcortical ones (Fig. 2). Secondly, the coreness is correlated with the iES-induced perceptual elicitation rates (Fig. 3).

First, we address the observation that cortical regions tend to be included in the central cores (i.e., exhibit higher coreness) compared to subcortical regions. The cerebral cortex plays a crucial role in conscious perception [41, 43, 79–81], whereas evidence suggests and theories propose that subcortical regions have a less direct role [41,43,79]. Our findings hint at a tentative correspondence between the coreness and its significance for conscious perception when considering the distinction between the cerebral cortex and the subcortex. It should be noted,

however, that some studies have pointed to a link between subcortical areas and conscious perception [82–85], and therefore, continued careful research is needed.

Second, we address the positive correlation between the coreness and the iES-induced perceptual elicitation rates (see Fig. 3). The use of iES allows for causal modulation of neural activity, thereby suggesting that if stimulation of a specific region results in changes in conscious perception, then that region is directly involved in that perception [86]. This implies that, within the cerebral cortex, the inclusion of regions in the central cores broadly corresponds to their importance for conscious perception. This underscores the importance of considering bidirectionality in identifying regions critical for conscious perception. However, iES can induce motor responses, making it uncertain whether conscious perception results directly from the stimulation or from awareness of the induced movement. Therefore, the identified core regions might be only indirectly related to conscious perception.

Overall, although further cautious investigation is needed, it can be said that areas strongly associated with conscious perception have been identified as strongly bidirectional cores. This underscores the significance of bidirectional interactions in conscious perception.

Relationship with theories of consciousness

This study suggests that cores with strong bidirectional interactions are important for conscious perception. Additionally, the findings reveal that cores with strong bidirectional interactions tend to comprise lower-order sensorimotor areas. In this subsection, we discuss the similarities and differences between these results and predictions made by existing theories of consciousness. Specifically, we focus on two key aspects: (1) the relationship each theory hypothesized between bidirectional interactions and conscious experiences, including how this compares and contrasts with our study's concept of "cores with strong bidirectional interactions," and (2) the identification of commonalities and differences between the regions deemed crucial for conscious experience in each theory and those included within the cores with strong bidirectional interactions in our study. First, we discuss the predictions of the IIT among theories of consciousness. As will be explained, the concept of cores with strong bidirectional interactions in this study originates from IIT. Thus, among theories of consciousness, IIT has many commonalities with the conceptual tenets of our study. We then discuss two representative theories that indicate the connection between bidirectional interactions and conscious perception: recurrent processing theory (RPT) and global workspace theory (GWT).

IIT [12, 14, 40, 45, 46, 87] posits that conscious perception crucially depends on bidirectional exchange and integration of information among brain regions. Specifically, IIT identifies a "complex"—a group of brain regions

marked by strong bidirectional information integration—as the place of consciousness. Although our study's definition of "complexes" diverges from IIT's, the fundamental concept stems from IIT [13,71]. Both approaches define complexes as subnetworks that, compared to other regions, are distinguished by stronger bidirectionality. Our findings indicate that cores with strong bidirectional interactions are associated with conscious perception, which broadly aligns with the predictions of IIT. Regarding which regions of the brain, such as the visual or prefrontal cortex, are integral to conscious perception, IIT does not pre-specify any particular region. Whether a region contributes to consciousness is determined solely based on the degree of information integration. Thus, a region's contribution to consciousness depends on whether it is part of the core. Furthermore, in principle, the regions included in the core may, depending on the brain state, vary from time to time. However, some advocates of IIT, including its originator Tononi, speculate that the posterior hot zone of the brain, including areas of the occipital, parietal, and temporal lobes, may be the place of consciousness [43,88]. This speculation is consistent with our findings that strongly bidirectional cores do not contain the frontal parts of the brain. However, there are differences as, for example, some posterior areas of the cerebral cortex are not included in the strongly bidirectional cores that were identified in our study.

Next, we discuss the similarities and differences of our results with two other theories of consciousness, RPT and GWT. RPT [11,41,89] predicts that bidirectional signal propagation between brain regions constitutes the neural basis of conscious perception. Although RPT does not specifically mention the "strength" of bidirectional signal propagation, its hypothesis is aligned with this study's results in that bidirectional signal propagation between brain regions is important for conscious perception. Furthermore, RPT emphasizes the importance of sensory pathways, including both lower and higher-order areas, for conscious perception. In contrast, our study found that higher visual areas, such as the inferior temporal cortex which is located in the temporal lobe, are less likely to be included in the strong cores compared to lower-order areas such as the somatosensory and lower visual areas (see Fig. 2**a**). Moreover, while RPT discusses the involvement of motor areas in reporting conscious experiences, it does not extensively discuss the relationship between motor areas and conscious experiences, although our study found that motor areas are also included in the strong cores. GWT [15,43,90] predicts that long-distance bidirectional interactions between higher-order areas within the frontal and parietal lobes and lower sensory areas constitute the neural basis of conscious perception. GWT particularly identifies high-order areas (within the parietal and frontal lobes, especially the prefrontal cortex) as central to the locus of consciousness. In contrast, the strongly bidirectional cores identified in this study did not include many regions of the prefrontal cortex, though they included some regions within the dorsal attention network of the frontal and parietal lobes.

In summary, the results of this study have both similarities and differences with the predictions of any of the

Theory	Relation of bidirectional interactions to conscious experience	Brain regions particularly important for conscious perception
Integrated information [12, 14, 40, 45, 46, 87]	Bidirectional exchange and integration of information among brain regions establish conscious perception	Cores of brain networks where bidirectional integration of information is strong. Some IIT proponents speculate that the posterior hot zone, including the occipital, parietal, and temporal lobes, may be the locus of consciousness
Recurrent processing [11, 41, 89]	Bidirectional signal propagation in sensory pathways constitute the neural basis of conscious perception	Sensory pathways including both lower-order and higher-order areas
Global workspace [15, 43, 90]	Long-distance bidirectional interactions between higher-order areas in the frontal and parietal lobes and lower sensory areas constitute the neural basis of conscious perception	Areas within the frontal and parietal lobes, especially the prefrontal cortex

Table 1: Predictions of Theories of Consciousness

three theories. Of course, the differences are not surprising because the theories differ fundamentally in some respects from our study. Additionally, it is crucial to emphasize that our results represent average tendencies for limited brain states (resting and seven task states). Further analysis is required because other brain regions not included in the strong core, such as higher visual areas and the prefrontal cortex, may also be included in the cores during various brain states.

Interpretation of the mechanism behind the high elicitation rates induced by iES for regions within cores

Our analysis revealed a tendency for high elicitation rates induced by iES for regions within strongly bidirectional cores. What underlying mechanisms might explain this? Below, we provide an interpretation from the perspective of IIT, which is closely related to our study [12, 40, 46].

IIT hypothesizes that conscious perception is supported by parts of the brain where bidirectional causal interactions are strong, namely the complexes. It should be noted, however, that the definition of "complexes" in IIT is not exactly the same as the complexes defined in our study [13]. IIT posits that causal interactions between areas within a complex, which integrates various types of information (e.g., color, shape, smell, sound, etc.), are essential for our unified perceptual experiences. Whether a region is included in a strongly connected part (i.e., a complex) or outside the complex determines if it is involved in conscious perception at a given moment. As regions within a core are strongly connected to other regions within the core, it is expected that stimulating a

region within the core will propagate its effects throughout the core, and thereby result in a change in conscious perception. Our results do not contradict this expected mechanism behind the high elicitation rates induced by iES for regions within cores.

On the other hand, it is also expected that even when a region outside a core is stimulated, if the region provides input to the core, the effect of the stimulus could propagate to the core, and thus result in a change in perception. However, our results showed that the perceptual elicitation rates in regions not included in the strongly bidirectional core were generally low (Fig. 3, Supplementary Figs. 5 and 6). This indicates that when a region that causally affects the core, that is, provides input to the core, is stimulated, there tends to be a lower tendency to induce changes in perception. Whether a perceptual change occurs when a region outside the core is stimulated depends on factors such as the intensity of the stimulation. Even when stimulating non-core regions that provide input to the core, perceptual change could occur if the stimulus intensity is strong enough.

To validate this interpretation in the future, it will be crucial to monitor the extent and manner in which the signal propagates bidirectionally upon stimulation. Additionally, it will be important to determine whether the resulting perception aligns with the perception associated with the region to which the signal has propagated. This will involve adjusting the intensity of the stimulation and correlating it with the outcomes of extracted cores. Such analysis requires measurement techniques with high temporal resolution that are capable of capturing signal propagation and perceptual changes within a timeframe of milliseconds to a few seconds. However, the temporal resolution of fMRI exceeds this timeframe, being longer than a few seconds [91]. Therefore, incorporating EEG or ECoG data in future studies could be beneficial, offering higher temporal resolution than fMRI. These methods enable the capturing of rapid signal dynamics and perception changes, effectively complementing fMRI's spatial insights.

Correspondence of the strength of bidirectional interactions with other cognitive functions

In this study, a meta-analysis using NeuroSynth and a comparison with the functional connectivity gradient revealed that cores with strong bidirectional interactions tended to include regions associated with lower-order sensorimotor functions rather than regions associated with higher-order cognitive functions. This section elucidates the implications of these findings and explores potential explanations for these observations.

The results in this paper suggest that lower-order regions form cores in which they interact with each other in a bidirectional manner, whereas other regions have weak bidirectional interactions with those cores. Lower-order sensory processing regions are the first in the cortex to receive stimuli from the outside world and play a fundamental role in sending signals to higher-order regions. Lower-order regions also receive top-down signals from

higher-order regions. This suggests that lower-order regions function as the cores of multidirectional information flow by coordinating with each other.

Next, we discuss the factors that led to the result that lower-order regions form cores with strong bidirectional interactions. The simplest, albeit naïve, interpretation is that this result is due to differences in the strength of bidirectional interactions. That is, the interpretation is that the bidirectional interactions are stronger between lower-order regions, while such interactions are weaker between lower-order and higher-order regions and between higher-order regions by comparison.

Another interpretation is that the present results were obtained because bidirectional interactions universally occur in lower-order regions, but whether such interactions occur between lower- and higher-order regions depends on many factors, including attention [92, 93] and strength of sensory stimuli [94]. It should also be noted that this study analyzed average trends across resting and seven different task states [57, 95]. It is possible that in lower-order regions, task-independent bidirectional interactions occur, while in higher-order regions, they may be task-dependent.

To better understand the relationship between cognitive functions and bidirectional interactions, it will be important to analyze data that takes into account the abovementioned factors, as well as a more detailed analysis of task-specific brain states.

Comparison with other core extraction methods in terms of association with cognitive functions

In this research, we evaluated our proposed methodology against traditional core extraction techniques, namely s-core decomposition and network hubs, which do not consider bidirectionality (Fig. 7). In what follows, we compare these methods in the context of their association with cognitive functions.

Our findings revealed an association between the coreness of our method and two key metrics: the MRR of iES and the functional connectivity gradient. Conversely, the s-cores and hubs demonstrated a weaker correlation with these metrics. This suggests that our approach can unveil connections between network cores and cognitive functions that remain undetected by conventional methods due to their non-consideration of bidirectionality.

Unlike our study, previous research utilizing s-core decomposition and network hubs have identified central cores or hubs primarily identified network cores within higher-order regions [96–98]. In contrast, our study revealed that the central cores and hubs were primarily situated in lower-order regions. This divergence likely stems from our approach of employing normalized directed transfer entropy (NDTE, [67]) as an edge weighting metric. NDTE captures the directional and causal relationships within the network, unlike previous studies that

relied on correlations as edge weights, ignoring the directionality or causality of interactions. This methodological difference elucidates the distinct outcomes between our findings and those of previous research, highlighting the significance of considering directionality in network analysis.

In summary, our research underscores the value of incorporating directionality based on estimated directed influences in network analysis. Through this innovative approach, we have established a new link between network cores with strong bidirectional interactions and critical neurological metrics, namely the MRR and the functional connectivity gradient, enhancing our understanding of the relationship between the brain network structure and its functions.

Future directions

The present study targeted human subjects; nevertheless, the extension of future analyses to include non-human species is an intriguing possibility [99–102]. Experimental evidence indicates that bidirectional interactions play a pivotal role in conscious perception across various species [18–21, 23, 25, 26, 29, 30, 38]. A comparative study investigating whether the cores with strong bidirectional interactions in non-human species consist of regions analogous to those that were identified in the strong cores in this study could significantly enhance our understanding of the relationship between bidirectional interactions and conscious perception. Such an exploration could offer a broader perspective on the neural mechanisms underlying conscious perception across different species and shed light on the evolutionary aspects of consciousness. Cross-species comparison is also interesting for understanding the relationship between the cores and lower- and higher-order cognitive functions. Lower-order cognitive functions, such as sensory perception, are common to many species, whereas higher-order cognitive functions are more developed in more advanced species. By analyzing cores with bidirectional interactions and their correlation with both lower- and higher-order cognitive functions in different species, we can elucidate the fundamental impact of bidirectional interactions on cognitive functionality.

The current study delved into the association between cores with strong bidirectional interactions and conscious perception; however, further investigation into how cores correlate with different consciousness states—such as wakefulness, sleep, anesthesia, and coma—is required. Previous research has highlighted the role of bidirectional interactions not only in the generation of conscious perception but also in maintaining wakefulness [40]. Additionally, recent studies revealed that the degree of system-level integration between brain regions was associated with consciousness states [103–105]. In light of these considerations, a critical next step would involve identifying cores from datasets acquired during unconscious states, such as sleep or anesthesia, and contrasting these with observations from wakeful states. Such a comparison could provide deeper insights into the relation-

ship between bidirectional interactions and varying levels of consciousness and enhance our understanding of the neurobiological mechanisms governing consciousness and its different states.

Data and code availability

The neuroimaging data are freely available from HCP https://db.humanconnectome.org/. Parcellation labels were used from https://github.com/ThomasYeoLab/CBIG (for the cerebral cortex) and https://github.com/yetianmed/subcortex (for the subcortex). The MATLAB codes for extracting bidirectionally connected cores are available at https://github.com/JunKitazono/BidirectionallyConnectedCores. The estimation of the VAR model was performed using the MVGC toolbox (Version 1.2) https://github.com/SacklerCentre/MVGC1. Normalized directed transfer entropy (NDTE) was calculated using a modified version of the codes at https://github.com/gustavodeco/nhb-ndte. For the visualization of regions within the cerebral cortex and subcortex, the connectome workbench https://www.humanconnectome.org/software/connectome-workbench and nilearn https://github.com/nilearn/nilearn, respectively, were used. For the neurosynth meta-analysis, we used a modified version of the codes at https://github.com/NeuroanatomyAndConnectivity/gradient_analysis. Violin plots were drawn using the codes at https://github.com/bastibe/Violinplot-Matlab.

References

[1] Liao, X, Vasilakos, A. V, & He, Y. (2017) Small-world human brain networks: Perspectives and challenges. *Neurosci. Biobehav. Rev.* **77**, 286–300.

[2] Bullmore, E & Sporns, O. (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198.

[3] Sporns, O. (2018) Graph theory methods: applications in brain networks. *Dialogues Clin. Neurosci.* **20**, 111–121.

[4] Stevens, A. A, Tappon, S. C, Garg, A, & Fair, D. A. (2012) Functional brain network modularity captures inter- and intra-individual variation in working memory capacity. *PLoS One* **7**, e30468.

[5] Finc, K, Bonna, K, Lewandowska, M, Wolak, T, Nikadon, J, Dreszer, J, Duch, W, & Kühn, S. (2017) Transition of the functional brain network related to increasing cognitive demands. *Hum. Brain Mapp.* **38**, 3659–3674.

[6] Gamboa, O. L, Tagliazucchi, E, von Wegner, F, Jurcoane, A, Wahl, M, Laufs, H, & Ziemann, U. (2014) Working memory performance of early MS patients correlates inversely with modularity increases in resting state functional connectivity networks. *Neuroimage* **94**, 385–395.

[7] Liang, X, Zou, Q, He, Y, & Yang, Y. (2016) Topologically reorganized connectivity architecture of Default-Mode, Executive-Control, and salience networks across working memory task loads. *Cereb. Cortex* **26**, 1501–1511.

[8] Finc, K, Bonna, K, He, X, Lydon-Staley, D. M, Kühn, S, Duch, W, & Bassett, D. S. (2020) Dynamic reconfiguration of functional brain networks during working memory training. *Nat. Commun.* **11**, 2435.

[9] van den Heuvel, M. P, Stam, C. J, Kahn, R. S, & Hulshoff Pol, H. E. (2009) Efficiency of functional brain networks and intellectual performance. *J. Neurosci.* **29**, 7619–7624.

[10] Edelman, G. M. (2003) Naturalizing consciousness: A theoretical framework. *Proceedings of the National Academy of Sciences* **100**, 5520–5524.

[11] Lamme, V. A. F. (2006) Towards a true neural stance on consciousness. *Trends Cogn. Sci.* **10**, 494–501.

[12] Oizumi, M, Albantakis, L, & Tononi, G. (2014) From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS Comput. Biol.* **10**, e1003588.

[13] Kitazono, J, Aoki, Y, & Oizumi, M. (2023) Bidirectionally connected cores in a mouse connectome: towards extracting the brain subnetworks essential for consciousness. *Cereb. Cortex* **33**, 1383–1402.

[14] Tononi, G & Edelman, G. M. (1998) Consciousness and complexity. *Science* **282**, 1846–1851.

[15] Dehaene, S & Changeux, J.-P. (2011) Experimental and theoretical approaches to conscious processing. *Neuron* **70**, 200–227.

[16] Pennartz, C. M. A, Dora, S, Muckli, L, & Lorteije, J. A. M. (2019) Towards a unified view on pathways and functions of neural recurrent processing. *Trends Neurosci.* **42**, 589–603.

[17] Aru, J, Suzuki, M, & Larkum, M. E. (2020) Cellular mechanisms of conscious processing. *Trends Cogn. Sci.* **24**, 814–825.

[18] Cauller, L. J & Kulics, A. T. (1991) The neural basis of the behaviorally relevant N1 component of the somatosensory-evoked potential in SI cortex of awake monkeys: evidence that backward cortical projections signal conscious touch sensation. *Exp. Brain Res.* **84**, 607–619.

[19] Lamme, V. A, Zipser, K, & Spekreijse, H. (1998) Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 3263–3268.

[20] Supèr, H, Spekreijse, H, & Lamme, V. A. (2001) Two distinct modes of sensory processing observed in monkey primary visual cortex (v1). *Nat. Neurosci.* **4**, 304–310.

[21] Self, M. W, Kooijmans, R. N, Supèr, H, Lamme, V. A, & Roelfsema, P. R. (2012) Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11031–11036.

[22] Auksztulewicz, R, Spitzer, B, & Blankenburg, F. (2012) Recurrent neural processing and somatosensory awareness. *J. Neurosci.* **32**, 799–805.

[23] Sachidhanandam, S, Sreenivasan, V, Kyriakatos, A, Kremer, Y, & Petersen, C. C. H. (2013) Membrane potential correlates of sensory perception in mouse barrel cortex. *Nat. Neurosci.* **16**, 1671–1677.

[24] Tang, H, Buia, C, Madhavan, R, Crone, N. E, Madsen, J. R, Anderson, W. S, & Kreiman, G. (2014) Spatiotemporal dynamics underlying object completion in human ventral visual cortex. *Neuron* **83**, 736–748.

[25] Koivisto, M, Kastrati, G, & Revonsuo, A. (2014) Recurrent processing enhances visual awareness but is not necessary for fast categorization of natural scenes. *J. Cogn. Neurosci.* **26**, 223–231.

[26] Manita, S, Suzuki, T, Homma, C, Matsumoto, T, Odagawa, M, Yamada, K, Ota, K, Matsubara, C, Inutsuka, A, Sato, M, Ohkura, M, Yamanaka, A, Yanagawa, Y, Nakai, J, Hayashi, Y, Larkum, M. E, & Murayama, M. (2015) A Top-Down cortical circuit for accurate sensory perception. *Neuron* **86**, 1304–1316.

[27] Ku, S.-W, Lee, U, Noh, G.-J, Jun, I.-G, & Mashour, G. A. (2011) Preferential inhibition of frontal-to-parietal feedback connectivity is a neurophysiologic correlate of general anesthesia in surgical patients. *PLoS One* **6**, e25155.

[28] Boly, M, Moran, R, Murphy, M, Boveroux, P, Bruno, M.-A, Noirhomme, Q, Ledoux, D, Bonhomme, V, Brichant, J.-F, Tononi, G, Laureys, S, & Friston, K. (2012) Connectivity changes underlying spectral EEG changes during propofol-induced loss of consciousness. *J. Neurosci.* **32**, 7082–7090.

[29] Cohen, D, van Swinderen, B, & Tsuchiya, N. (2018) Isoflurane impairs Low-Frequency feedback but leaves High-Frequency feedforward connectivity intact in the fly brain. *eNeuro* **5**.

[30] Cauller, L. J & Kulics, A. T. (1988) A comparison of awake and sleeping cortical states by analysis of the somatosensory-evoked response of postcentral area 1 in rhesus monkey. *Exp. Brain Res.* **72**, 584–592.

[31] Hayat, H, Marmelshtein, A, Krom, A. J, Sela, Y, Tankus, A, Strauss, I, Fahoum, F, Fried, I, & Nir, Y. (2022) Reduced neural feedback signaling despite robust neuron and gamma auditory responses during human sleep. *Nat. Neurosci.* **25**, 935–943.

[32] Boly, M, Garrido, M. I, Gosseries, O, Bruno, M.-A, Boveroux, P, Schnakers, C, Massimini, M, Litvak, V, Laureys, S, & Friston, K. (2011) Preserved feedforward but impaired top-down processes in the vegetative state. *Science* **332**, 858–862.

[33] Dembski, C, Koch, C, & Pitts, M. (2021) Perceptual awareness negativity: a physiological correlate of sensory consciousness. *Trends Cogn. Sci.* **25**, 660–670.

[34] Gutschalk, A, Micheyl, C, & Oxenham, A. J. (2008) Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol.* **6**, e138.

[35] Dykstra, A. R, Halgren, E, Gutschalk, A, Eskandar, E. N, & Cash, S. S. (2016) Neural correlates of auditory perceptual awareness and release from informational masking recorded directly from human cortex: A case study. *Front. Neurosci.* **10**, 472.

[36] Eklund, R & Wiens, S. (2019) Auditory awareness negativity is an electrophysiological correlate of awareness in an auditory threshold task. *Conscious. Cogn.* **71**, 70–78.

[37] Schlossmacher, I, Dellert, T, Bruchmann, M, & Straube, T. (2021) Dissociating neural correlates of consciousness and task relevance during auditory processing. *Neuroimage* **228**, 117712.

[38] Nieder, A, Wagener, L, & Rinnert, P. (2020) A neural correlate of sensory consciousness in a corvid bird. *Science* **369**, 1626–1629.

[39] Dehaene, S & Naccache, L. (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* **79**, 1–37.

[40] Tononi, G, Boly, M, Massimini, M, & Koch, C. (2016) Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* **17**, 450–461.

[41] Lamme, V. A. F. (2018) Challenges for theories of consciousness: seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373**.

[42] Sikkens, T, Bosman, C. A, & Olcese, U. (2019) The role of Top-Down modulation in shaping sensory processing across brain states: Implications for consciousness. *Front. Syst. Neurosci.* **13**, 31.

[43] Mashour, G. A, Roelfsema, P, Changeux, J.-P, & Dehaene, S. (2020) Conscious processing and the global neuronal workspace hypothesis. *Neuron* **105**, 776–798.

[44] Storm, J. F, Klink, P. C, Aru, J, Senn, W, Goebel, R, Pigorini, A, Avanzini, P, Vanduffel, W, Roelfsema, P. R, Massimini, M, Larkum, M. E, & Pennartz, C. M. (2024) An integrative, multiscale view on neural theories of consciousness. *Neuron.*

[45] Balduzzi, D & Tononi, G. (2008) Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS Comput. Biol.* **4**, e1000091.

[46] Albantakis, L, Barbosa, L, Findlay, G, Grasso, M, Haun, A. M, Marshall, W, Mayner, W. G. P, Zaeemzadeh, A, Boly, M, Juel, B. E, Sasai, S, Fujii, K, David, I, Hendren, J, Lang, J. P, & Tononi, G. (2023) Integrated information theory (iit) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLOS Computational Biology* **19**, 1–45.

[47] Chatterjee, N & Sinha, S. (2007) in *Progress in Brain Research*, eds. Banerjee, R & Chakrabarti, B. K. (Elsevier) Vol. 168, pp. 145–153.

[48] van den Heuvel, M. P & Sporns, O. (2011) Rich-club organization of the human connectome. *J. Neurosci.* **31**, 15775–15786.

[49] Harriger, L, van den Heuvel, M. P, & Sporns, O. (2012) Rich club organization of macaque cerebral cortex and its role in network communication. *PLoS One* **7**, e46497.

[50] Crobe, A, Demuru, M, Didaci, L, Marcialis, G. L, & Fraschini, M. (2016) Minimum spanning tree and k-core decomposition as measure of subject-specific EEG traits. *Biomed. Phys. Eng. Express* **2**, 017001.

[51] van den Heuvel, M. P & Sporns, O. (2013) Network hubs in the human brain. *Trends Cogn. Sci.* **17**, 683–696.

[52] Royer, J, Bernhardt, B. C, Larivière, S, Gleichgerrcht, E, Vorderwülbecke, B. J, Vulliémoz, S, & Bonilha, L. (2022) Epilepsy and brain network hubs. *Epilepsia* **63**, 537–550.

[53] Liang, X, Hsu, L.-M, Lu, H, Sumiyoshi, A, He, Y, & Yang, Y. (2018) The Rich-Club organization in rat functional brain network to balance between communication cost and efficiency. *Cereb. Cortex* **28**, 924–935.

[54] Wang, S, Li, Y, Qiu, S, Zhang, C, Wang, G, Xian, J, Li, T, & He, H. (2020) Reorganization of rich-clubs in functional brain networks during propofol-induced unconsciousness and natural sleep. *Neuroimage Clin* **25**, 102188.

[55] Bertolero, M. A, Yeo, B. T. T, & D'Esposito, M. (2015) The modular and integrative functional architecture of the human brain. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6798–807.

[56] Chen, X, Necus, J, Peraza, L. R, Mehraram, R, Wang, Y, O'Brien, J. T, Blamire, A, Kaiser, M, & Taylor, J.-P. (2021) The functional brain favours segregated modular connectivity at old age unless affected by neurodegeneration. *Commun Biol* **4**, 973.

[57] Van Essen, D. C, Smith, S. M, Barch, D. M, Behrens, T. E. J, Yacoub, E, Ugurbil, K, & WU-Minn HCP Consortium. (2013) The WU-Minn human connectome project: an overview. *Neuroimage* **80**, 62–79.

[58] Fox, K. C. R, Shi, L, Baek, S, Raccah, O, Foster, B. L, Saha, S, Margulies, D. S, Kucyi, A, & Parvizi, J. (2020) Intrinsic network architecture predicts the effects elicited by intracranial electrical stimulation of the human brain. *Nat Hum Behav* **4**, 1039–1052.

[59] Borchers, S, Himmelbach, M, Logothetis, N, & Karnath, H.-O. (2011) Direct electrical stimulation of human cortex - the gold standard for mapping brain functions? *Nat. Rev. Neurosci.* **13**, 63–70.

[60] Yarkoni, T, Poldrack, R. A, Nichols, T. E, Van Essen, D. C, & Wager, T. D. (2011) Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670.

[61] Margulies, D. S, Ghosh, S. S, Goulas, A, Falkiewicz, M, Huntenburg, J. M, Langs, G, Bezgin, G, Eickhoff, S. B, Castellanos, F. X, Petrides, M, Jefferies, E, & Smallwood, J. (2016) Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12574–12579.

[62] Ito, T, Brincat, S. L, Siegel, M, Mill, R. D, He, B. J, Miller, E. K, Rotstein, H. G, & Cole, M. W. (2020) Task-evoked activity quenches neural correlations and variability across cortical areas. *PLoS Comput. Biol.* **16**, e1007983.

[63] Satterthwaite, T. D, Elliott, M. A, Gerraty, R. T, Ruparel, K, Loughead, J, Calkins, M. E, Eickhoff, S. B, Hakonarson, H, Gur, R. C, Gur, R. E, & Wolf, D. H. (2013) An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256.

[64] Luppi, A. I & Stamatakis, E. A. (2021) Combining network topology and information theory to construct representative brain networks. *Netw Neurosci* **5**, 96–124.

[65] Schaefer, A, Kong, R, Gordon, E. M, Laumann, T. O, Zuo, X.-N, Holmes, A. J, Eickhoff, S. B, & Yeo, B. T. T. (2018) Local-Global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114.

[66] Tian, Y, Margulies, D. S, Breakspear, M, & Zalesky, A. (2020) Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nat. Neurosci.* **23**, 1421–1432.

[67] Deco, G, Vidaurre, D, & Kringelbach, M. L. (2021) Revisiting the global workspace orchestrating the hierarchical organization of the human brain. *Nat Hum Behav.*

[68] Barnett, L & Seth, A. K. (2014) The MVGC multivariate granger causality toolbox: a new approach to granger-causal inference. *J. Neurosci. Methods* **223**, 50–68.

[69] Rubinov, M & Sporns, O. (2010) Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059–1069.

[70] van den Heuvel, M. P, de Lange, S. C, Zalesky, A, Seguin, C, Yeo, B. T. T, & Schmidt, R. (2017) Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. *Neuroimage* **152**, 437–449.

[71] Kitazono, J, Kanai, R, & Oizumi, M. (2020) Efficient search for informational cores in complex systems: Application to brain networks. *Neural Netw.* **132**, 232–244.

[72] Kitazono, J, Kanai, R, & Oizumi, M. (2018) Efficient algorithms for searching the minimum information partition in integrated information theory. *Entropy* **20**.

[73] Marcus, D. S, Harwell, J, Olsen, T, Hodge, M, Glasser, M. F, Prior, F, Jenkinson, M, Laumann, T, Curtiss, S. W, & Van Essen, D. C. (2011) Informatics and data mining tools and strategies for the human connectome project. *Front. Neuroinform.* **5**, 4.

[74] Yeo, B. T. T, Krienen, F. M, Sepulcre, J, Sabuncu, M. R, Lashkari, D, Hollinshead, M, Roffman, J. L, Smoller, J. W, Zöllei, L, Polimeni, J. R, Fischl, B, Liu, H, & Buckner, R. L. (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165.

[75] Müller, E. J, Munn, B, Hearne, L. J, Smith, J. B, Fulcher, B, Arnatkevičiūtė, A, Lurie, D. J, Cocchi, L, & Shine, J. M. (2020) Core and matrix thalamic sub-populations relate to spatio-temporal cortical connectivity gradients. *Neuroimage* **222**, 117224.

[76] Shafiei, G, Markello, R. D, Vos de Wael, R, Bernhardt, B. C, Fulcher, B. D, & Misic, B. (2020) Topographic gradients of intrinsic dynamics across neocortex. *Elife* **9**.

[77] Fornito, A, Arnatkevičiūtė, A, & Fulcher, B. D. (2019) Bridging the gap between connectome and transcriptome. *Trends Cogn. Sci.* **23**, 34–50.

[78] Coifman, R. R, Lafon, S, Lee, A. B, Maggioni, M, Nadler, B, Warner, F, & Zucker, S. W. (2005) Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7426–7431.

[79] Koch, C, Massimini, M, Boly, M, & Tononi, G. (2016) Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* **17**, 307–321.

[80] Marshel, J. H, Kim, Y. S, Machado, T. A, Quirin, S, Benson, B, Kadmon, J, Raja, C, Chibukhchyan, A, Ramakrishnan, C, Inoue, M, Shane, J. C, McKnight, D. J, Yoshizawa, S, Kato, H. E, Ganguli, S, & Deisseroth, K. (2019) Cortical layer-specific critical dynamics triggering perception. *Science* **365**.

[81] Filipchuk, A, Schwenkgrub, J, Destexhe, A, & Bathellier, B. (2022) Awake perception is associated with dedicated neuronal assemblies in the cerebral cortex. *Nat. Neurosci.* **25**, 1327–1338.

[82] Aru, J, Suzuki, M, Rutiku, R, Larkum, M. E, & Bachmann, T. (2019) Coupling the state and contents of consciousness. *Front. Syst. Neurosci.* **13**, 43.

[83] Ward, L. M. (2011) The thalamic dynamic core theory of conscious experience. *Conscious. Cogn.* **20**, 464–486.

[84] Slagter, H. A, Mazaheri, A, Reteig, L. C, Smolders, R, Figee, M, Mantione, M, Schuurman, P. R, & Denys, D. (2017) Contributions of the ventral striatum to conscious perception: An intracranial EEG study of the attentional blink. *J. Neurosci.* **37**, 1081–1089.

[85] Afrasiabi, M, Redinbaugh, M. J, Phillips, J. M, Kambi, N. A, Mohanta, S, Raz, A, Haun, A. M, & Saalmann, Y. B. (2021) Consciousness depends on integration between parietal cortex, striatum, and thalamus. *Cell Syst* **12**, 363–373.e11.

[86] Raccah, O, Block, N, & Fox, K. C. R. (2021) Does the prefrontal cortex play an essential role in consciousness? insights from intracranial electrical stimulation of the human brain. *J. Neurosci.* **41**, 2076–2087.

[87] Tononi, G. (2004) An information integration theory of consciousness. *BMC Neurosci.* **5**, 42.

[88] Boly, M, Massimini, M, Tsuchiya, N, Postle, B. R, Koch, C, & Tononi, G. (2017) Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? clinical and neuroimaging evidence. *J. Neurosci.* **37**, 9603–9613.

[89] Lamme, V. A & Roelfsema, P. R. (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* **23**, 571–579.

[90] Dehaene, S, Kerszberg, M, & Changeux, J. P. (1998) A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14529–14534.

[91] Glover, G. H. (2011) Overview of functional magnetic resonance imaging. *Neurosurg. Clin. N. Am.* **22**, 133–9, vii.

[92] Gregoriou, G. G, Gotts, S. J, Zhou, H, & Desimone, R. (2009) High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* **324**, 1207–1210.

[93] Baldauf, D & Desimone, R. (2014) Neural mechanisms of object-based attention. *Science* **344**, 424–427.

[94] van Vugt, B, Dagnino, B, Vartak, D, Safaai, H, Panzeri, S, Dehaene, S, & Roelfsema, P. R. (2018) The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science* **360**, 537–542.

[95] Barch, D. M, Burgess, G. C, Harms, M. P, Petersen, S. E, Schlaggar, B. L, Corbetta, M, Glasser, M. F, Curtiss, S, Dixit, S, Feldt, C, Nolan, D, Bryant, E, Hartley, T, Footer, O, Bjork, J. M, Poldrack, R, Smith, S, Johansen-Berg, H, Snyder, A. Z, Van Essen, D. C, & WU-Minn HCP Consortium. (2013) Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189.

[96] Achard, S, Salvador, R, Whitcher, B, Suckling, J, & Bullmore, E. (2006) A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* **26**, 63–72.

[97] Achard, S & Bullmore, E. (2007) Efficiency and cost of economical brain functional networks. *PLoS Comput. Biol.* **3**, e17.

[98] Buckner, R. L, Sepulcre, J, Talukdar, T, Krienen, F. M, Liu, H, Hedden, T, Andrews-Hanna, J. R, Sperling, R. A, & Johnson, K. A. (2009) Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to alzheimer's disease. *J. Neurosci.* **29**, 1860–1873.

[99] Xu, T, Nenning, K.-H, Schwartz, E, Hong, S.-J, Vogelstein, J. T, Goulas, A, Fair, D. A, Schroeder, C. E, Margulies, D. S, Smallwood, J, Milham, M. P, & Langs, G. (2020) Cross-species functional alignment reveals evolutionary hierarchy within the connectome. *Neuroimage* **223**, 117346.

[100] Goulas, A, Bastiani, M, Bezgin, G, Uylings, H. B. M, Roebroeck, A, & Stiers, P. (2014) Comparative analysis of the macroscale structural connectivity in the macaque and human brain. *PLoS Comput. Biol.* **10**, e1003529.

[101] Eichert, N, Robinson, E. C, Bryant, K. L, Jbabdi, S, Jenkinson, M, Li, L, Krug, K, Watkins, K. E, & Mars, R. B. (2020) Cross-species cortical alignment identifies different types of anatomical reorganization in the primate temporal lobe. *Elife* **9**.

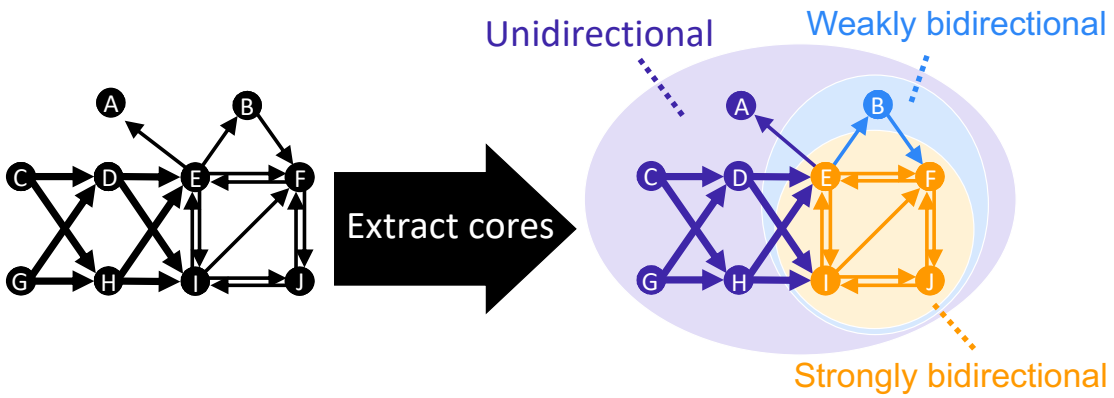
[102] Fulcher, B. D, Murray, J. D, Zerbi, V, & Wang, X.-J. (2019) Multimodal gradients across mouse cortex. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4689–4695.

[103] Luppi, A. I, Craig, M. M, Pappas, I, Finoia, P, Williams, G. B, Allanson, J, Pickard, J. D, Owen, A. M, Naci, L, Menon, D. K, & Stamatakis, E. A. (2019) Consciousness-specific dynamic interactions of brain integration and functional diversity. *Nat. Commun.* **10**, 4616.

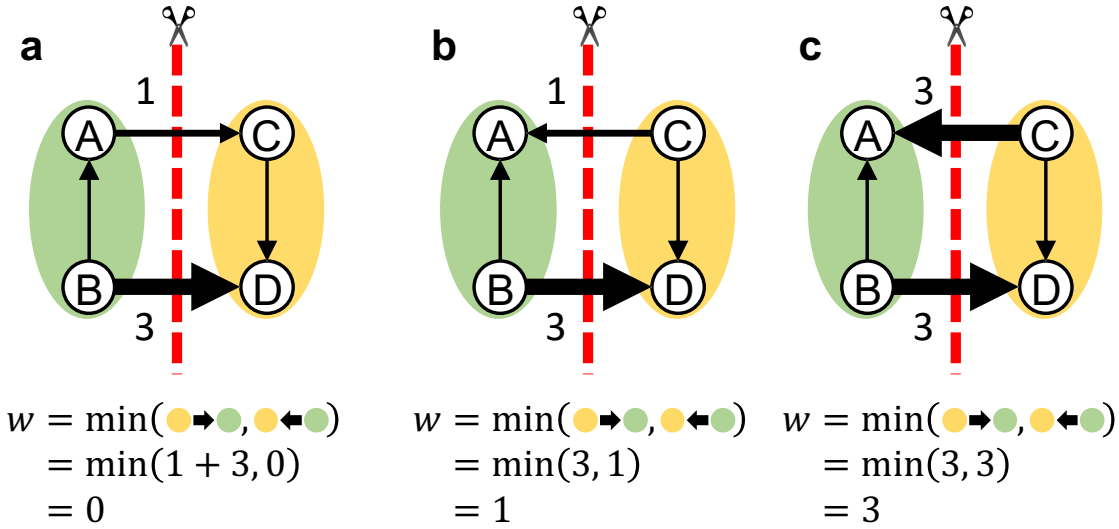
[104] Luppi, A. I, Golkowski, D, Ranft, A, Ilg, R, Jordan, D, Menon, D. K, & Stamatakis, E. A. (2021) Brain network integration dynamics are associated with loss and recovery of consciousness induced by sevoflurane. *Hum. Brain Mapp.* **42**, 2802–2822.

[105] Onoda, K & Akama, H. (2023) Complex of global functional network as the core of consciousness. *Neurosci. Res.* **190**, 67–77.

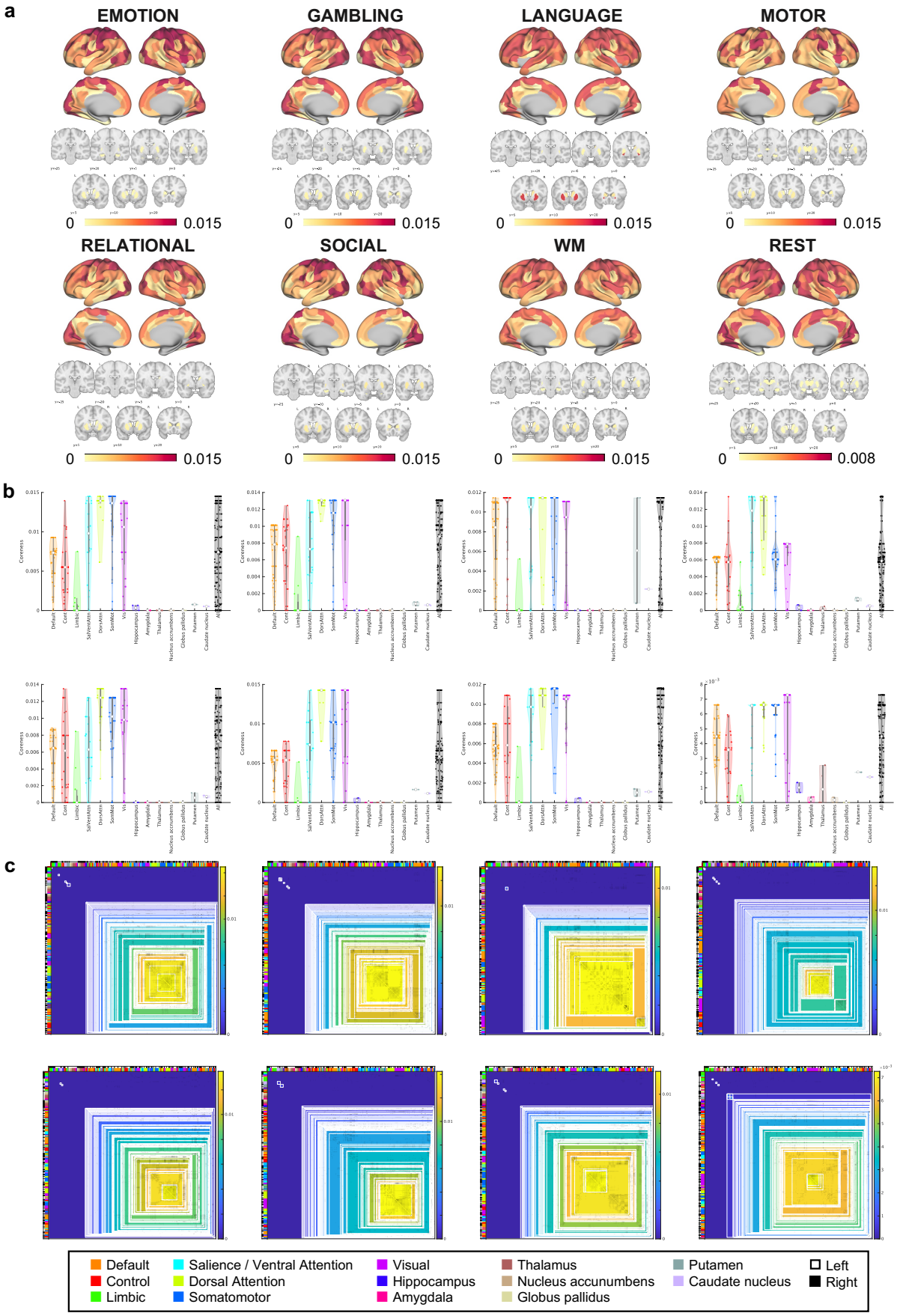
Supplementary information



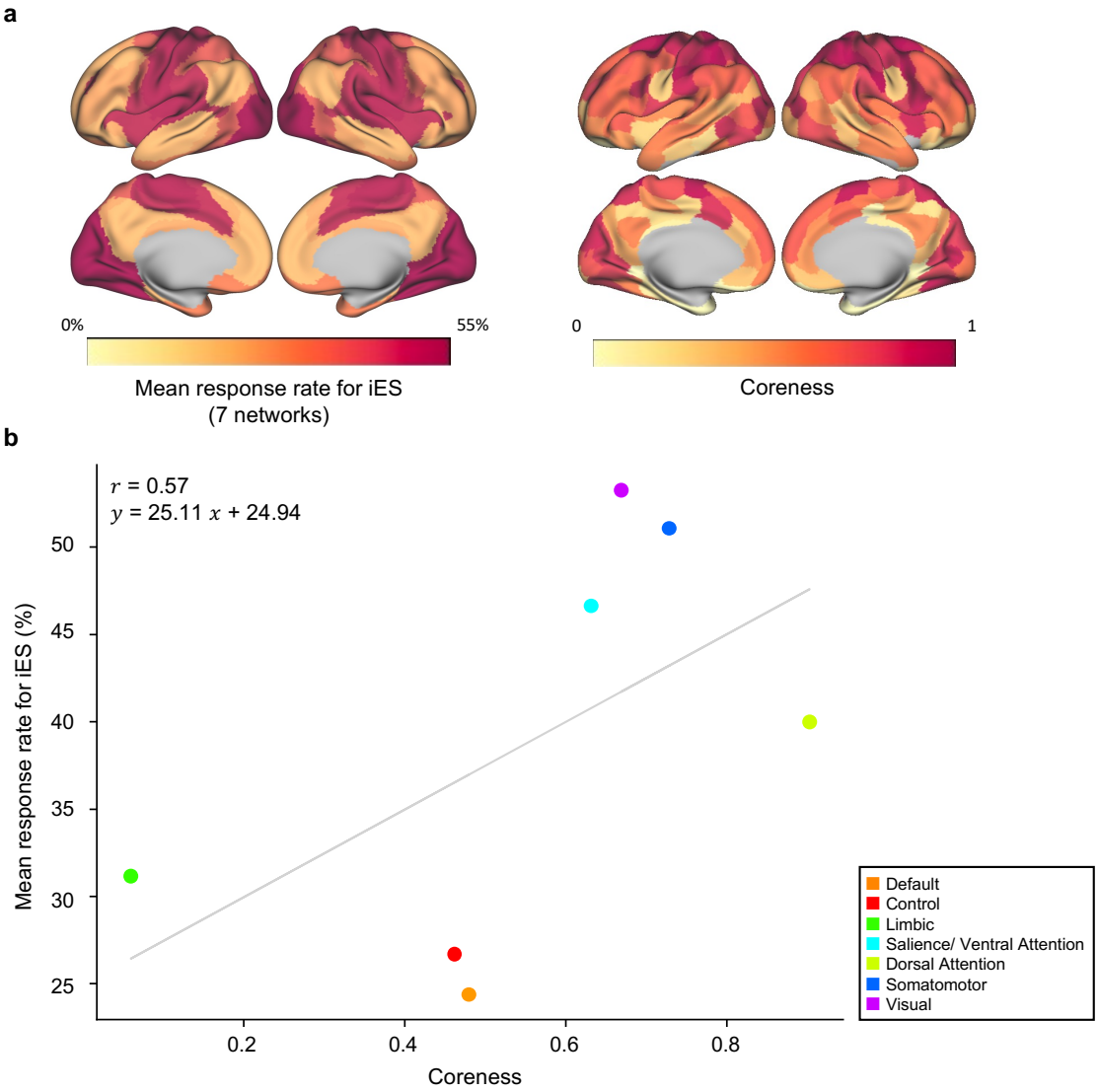
Supplementary Figure 1: **Overview of core extraction in a toy network.** Nodes BEFIJ are connected bidirectionally whereas nodes ACDGH are connected in a feedforward manner. Strongly bidirectional cores in this network are indicated by a colored background, wherein subnet EFIJ, in orange, has the strongest bidirectional connections, followed by BEFIJ, in blue. Generally, complexes with stronger bidirectional connections are included in those with weaker bidirectional connections to form a unimodal or multimodal hierarchical structure.



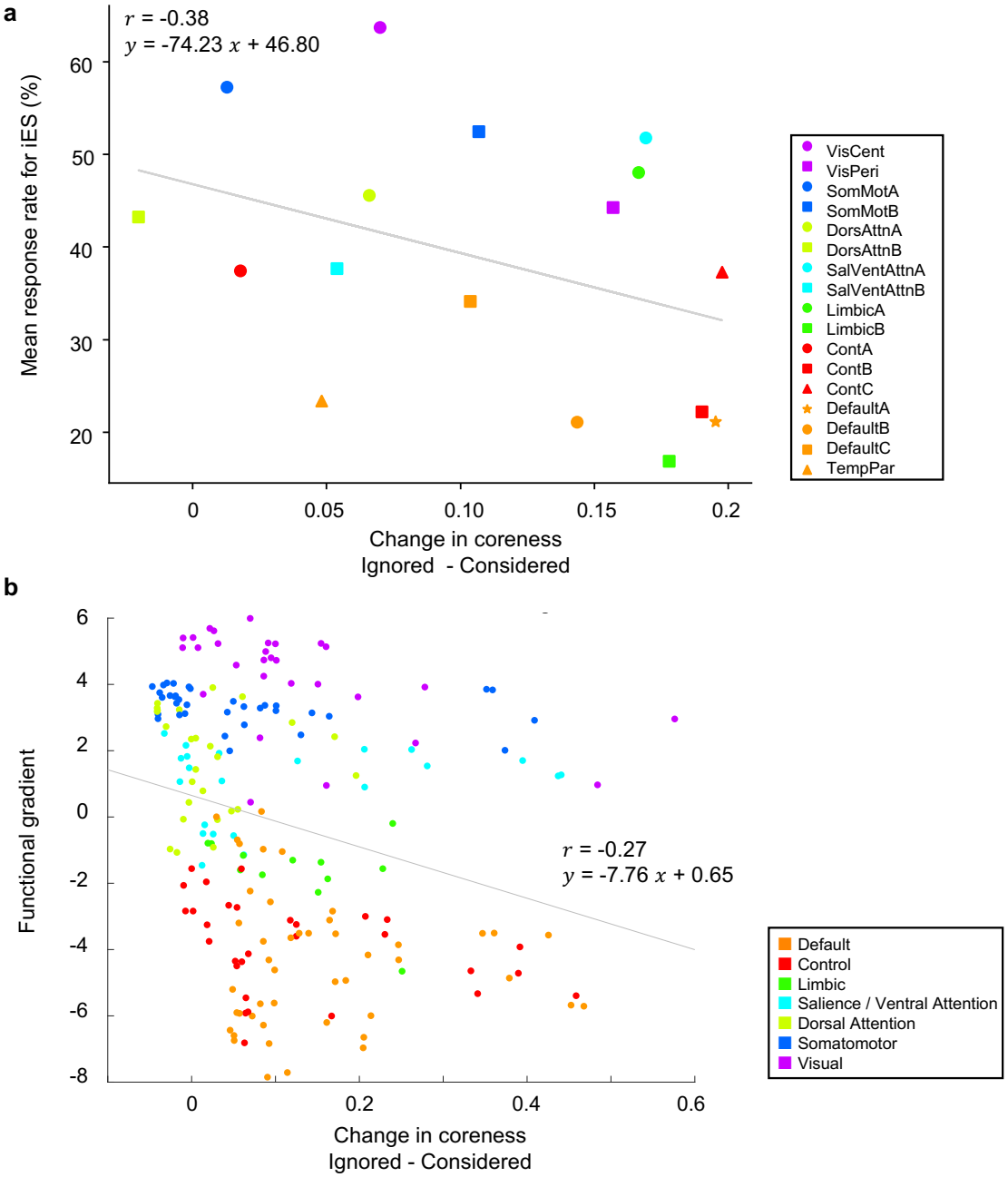
Supplementary Figure 2: **Overview of graph cut.** **a**, When two parts of a network are linked unidirectionally, the strength of bidirectional connections $w(V_{\mathrm{L}}; V_{\mathrm{R}})$ equals 0. **b**, When a connection in one direction is strong ($w(V_{\mathrm{L}} \to V_{\mathrm{R}}) = 3$), while the connection in the other direction is weak ($w(V_{\mathrm{R}} \to V_{\mathrm{L}}) = 1$), the strength of bidirectional connections is small ($w(V_{\mathrm{L}}; V_{\mathrm{R}}) = 1$). **c**, When connections in both directions are strong ($w(V_{\mathrm{L}} \to V_{\mathrm{R}}) = w(V_{\mathrm{R}} \to V_{\mathrm{L}}) = 3$), the strength of bidirectional connections is large ($w(V_{\mathrm{L}}; V_{\mathrm{R}}) = 3$).


a
EMOTION
GAMBLING
LANGUAGE
MOTOR
RELATIONAL
SOCIAL
WM
REST
0
0.015
0.008
b
Coreness
c
Default
Control
Limbic
Salience / Ventral Attention
Dorsal Attention
Somatomotor
Visual
Hippocampus
Amygdala
Thalamus
Nucleus accumbens
Globus pallidus
Putamen
Caudate nucleus
Left
Right

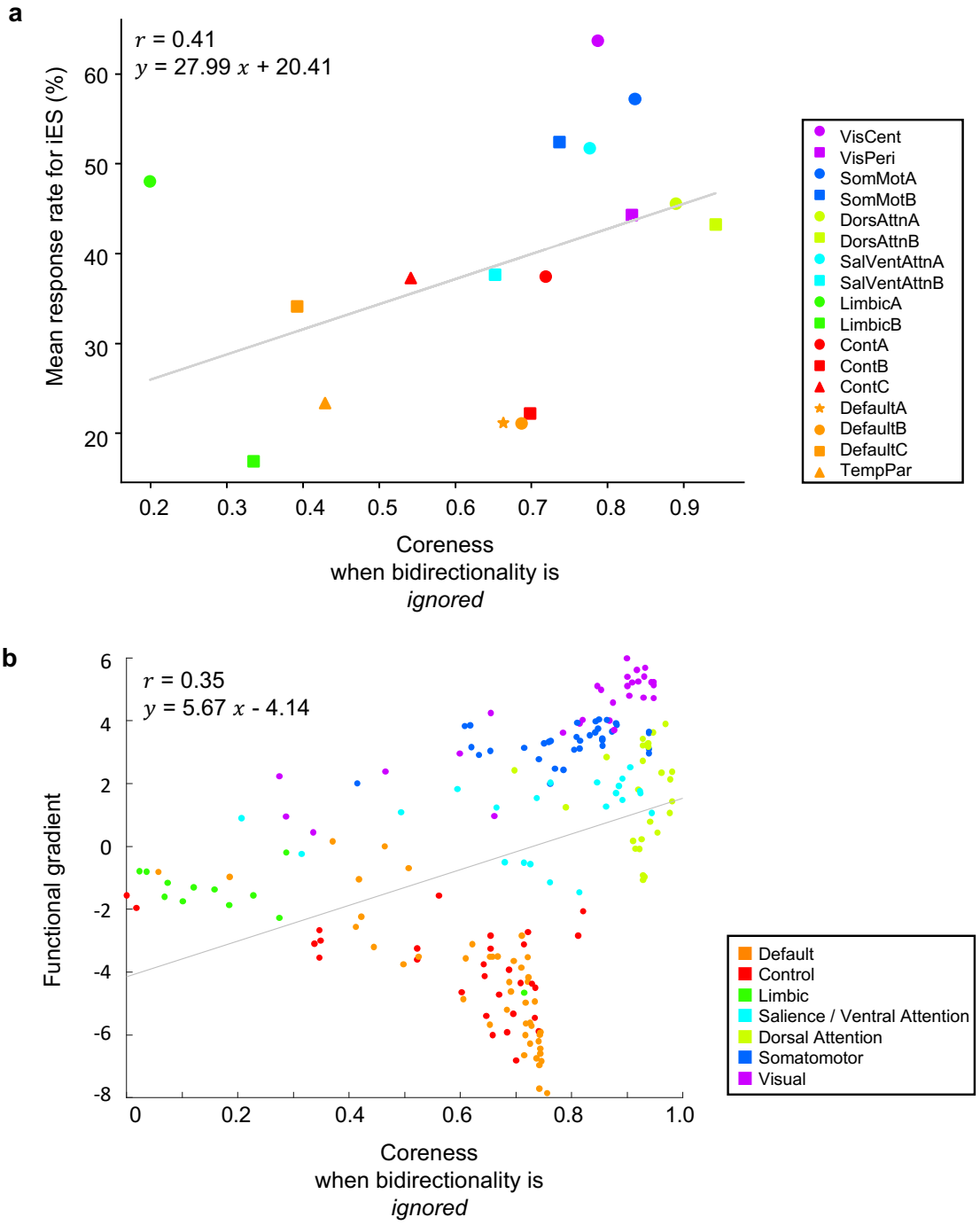
Supplementary Figure 3: **Results of the cores extracted for each task.** **a**, Coreness at rest and during seven tasks in the cerebral cortex (top) and the subcortex shown in seven coronal slices (bottom). Here, coreness is not normalized. **b**, Violin plots showing the distribution of coreness at rest and during seven different tasks, categorized by the Yeo-7 network atlas for cortical regions and major divisions for subcortical areas. **c**, Hierarchical structure of the cores with bidirectional connections at rest and during seven tasks. The color-shading of each element in the matrix represents the weights of edges in the estimated directed network. The white and black at the left and top of the matrix indicate the left and right hemispheres, respectively, whereas the color bar represents the divisions of the Yeo-7 network atlas and major divisions of subcortical areas. The directed network is arranged according to the hierarchical structure of the complexes, with white lines separating each hierarchical level. **a, b, c** show, from left to right, the results for emotion, gambling, language, motor, relational, social, working memory, and rest, respectively.



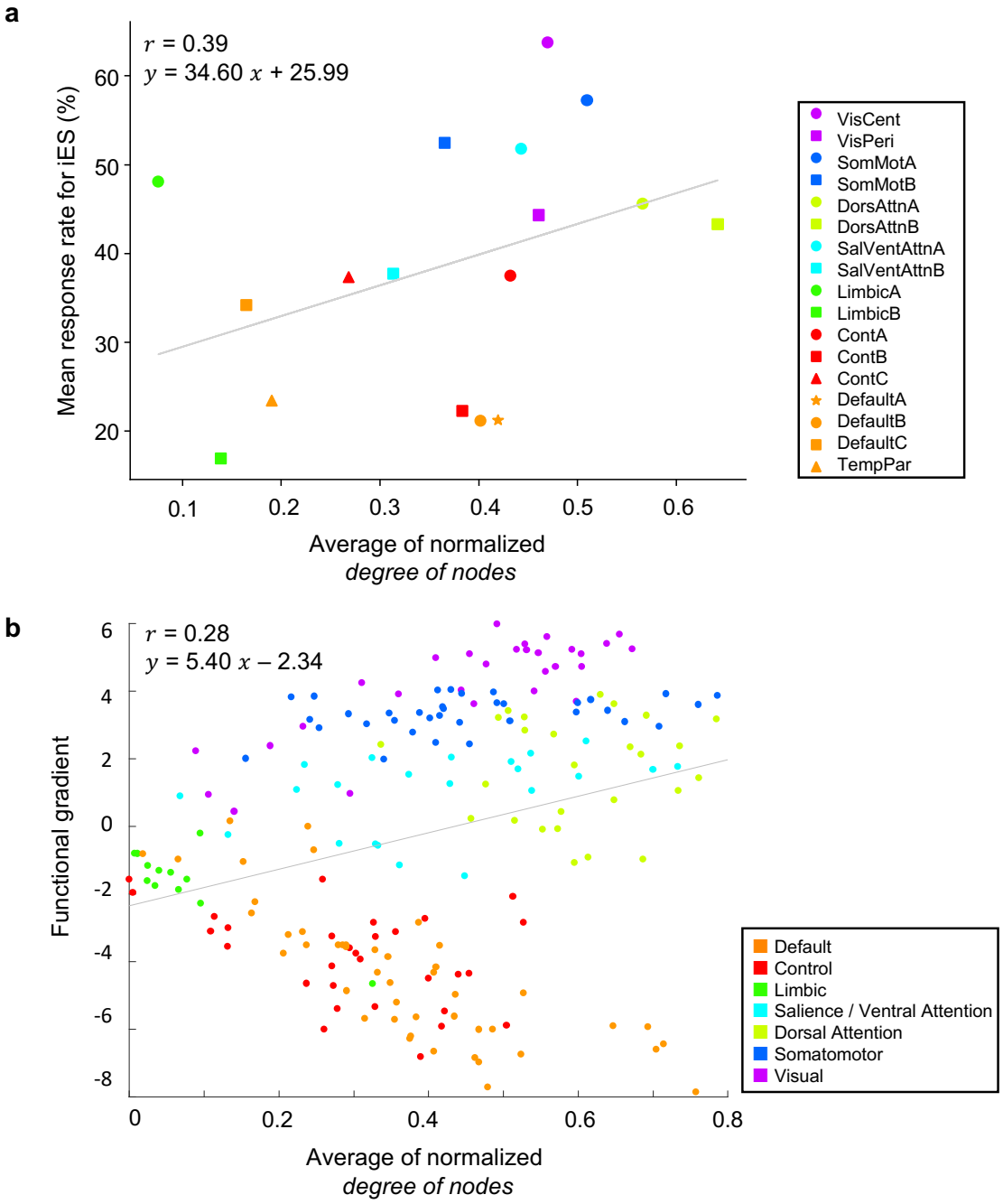
Supplementary Figure 4: **Comparison of the mean response rate (MRR) for iES and coreness in the Yeo-7 network atlas.** **a**, A cortical surface rendering of the MRR for iES (7 network) (left), and that of coreness (right). **b**, Comparison in the Yeo-7 network atlas revealed a positive correlation between the MRR for iES and coreness ($r = 0.57$). The solid line represents the regression line.



Supplementary Figure 5: **Comparison of changes in coreness when ignoring bidirectionality with the MRR for iES and the functional connectivity gradient. a**, There is a negative correlation between the change in coreness and the MRR for iES ($r = -0.38$). **b**, There is a negative correlation between the change in coreness and the functional connectivity gradient ($r = -0.27$). The solid lines represent the regression lines.



Supplementary Figure 6: **Comparison of coreness when ignoring bidirectionality with the MRR for iES and the functional connectivity gradient.** **a**, Compared to when bidirectionality is considered, the correlation between coreness and the MRR is weaker when bidirectionality is ignored ($r = 0.41$). **b**, Compared to when bidirectionality is considered, the correlation between coreness and the functional connectivity gradient is weaker when bidirectionality is ignored ($r = 0.35$). The solid lines represent the regression lines.



Supplementary Figure 7: **Comparison of node degree with the MRR for iES and the functional connectivity gradient.** **a**, The correlation between the degree and the MRR ($r = 0.39$) is weaker than that between the coreness when bidirectionality is considered and the MRR. **b**, The correlation between the degree and the functional connectivity gradient ($r = 0.28$) is weaker than that between coreness when considering bidirectionality and the functional connectivity gradient. The solid lines represent the regression lines.