Controlling target brain regions by optimal selection of input nodes

Karan Kabbur Hanumanthappa Manjunatha[1,6], Giorgia Baron[3], Danilo Benozzo[3], Erica Silvestri[3], Maurizio Corbetta[2,4,5], Alessandro Chiuso[3], Alessandra Bertoldo[2,3], Samir Suweis[1,2], Michele Allegra[1,2*],

1 Physics and Astronomy Department "Galileo Galilei", University of Padova, via Marzolo 8, 35121 Padova, Italy
2 Padova Neuroscience Center, University of Padova, Via Orus 2, 35131 Padova, Italy
3 Information Engineering Department, University of Padova, Via Gradenigo 6/b, 35131 Padova, Italy
4 Neuroscience Department, University of Padova, Via Giustiniani 5, 35128 Padova, Italy
5 Venetian Institute of Molecular Medicine (VIMM), Via Giuseppe Orus 2, 35129 Padova, Italy
6 Modeling and Engineering Risk and Complexity, Scuola Superiore Meridionale, Via Mezzocannone 4, 80138 Napoli, Italy

* michele.allegra@unipd.it

Abstract

The network control theory framework holds great potential to inform neurostimulation experiments aimed at inducing desired activity states in the brain. However, the current applicability of the framework is limited by inaccurate modeling of brain dynamics, and an overly ambitious focus on whole-brain activity control. In this work, we leverage recent progress in linear modeling of brain dynamics (effective connectivity) and we exploit the concept of target controllability to focus on the control of a single region or a small subnetwork of nodes. We discuss when control may be possible with a reasonably low energy cost and few stimulation loci, and give general predictions on where to stimulate depending on the subset of regions one wishes to control. Importantly, using the robustly asymmetric effective connectome instead of the symmetric structural connectome (as in previous research), we highlight the fundamentally different roles in- and out-hubs have in the control problem, and the relevance of inhibitory connections. The large degree of inter-individual variation in the effective connectome implies that the control problem is best formulated at the individual level, but we discuss to what extent group results may still prove useful.

Introduction

Brain controllability refers to the possibility of manipulating brain activity in a controlled way through external perturbations [1, 2], such as those that can be delivered non-invasively through transcranial magnetic stimulation (TMS). For this goal, one can exploit control theory, a general mathematical framework to design perturbations of dynamical systems with a desired effect. In a first approximation, neural dynamics can be modeled as a linear and time-invariant [3], and one can try to control brain activity

using the simple framework of of linear network control theory [4]. The activity of the whole network can be controlled by acting on a subset of "driver nodes", and theory predicts which nodes should be selected and which input signal should be applied to obtain desired activity states. Since the first proposal by Gu et al. [5], this idea has been extensively explored [1, 2] and debated [6, 7].

So far, however, there has been limited success in directly applying this framework to predict the outcomes of neurostimulation experiments [8–10]. In fact, the framework has been mainly applied in a relatively indirect way, by enriching the analysis of structural connectomes with a whole new set of tools based on *controllability metrics* [11]. The latter are node-wise metrics assessing the difficulty (energy cost) to reach desired states when specific nodes are selected as driver nodes, and they have proven very effective in summarizing features of the structural connectomes linked with cognitive function [12–20].

Among the obstacles hindering the practical applicability, and hence the widespread adoption of network control theory in neuromodulation experiments, a major one is a nearly exclusive focus on a quite ambitious objective, namely, whole-brain activity control. While a sufficiently dense network is in principle controllable with a single driver node, in practice a non-negligible fraction of the total number of nodes should be used as driver nodes to control the activity of the whole network with a realistic energy cost [21]. For large networks this means that many driver nodes are required. This is indeed the case for the brain: even with a coarse parcellation, $N \geq 60$ nodes are required to model the whole brain). However, current neurostimulation techiques such as TMS allow stimulating at most one (or two [22]) sites at the same time. Thus, a fine-grained control of whole brain activity is way beyond current experimental capabilities. A second relevant obstacle is the usage of inaccurate computational models of brain dynamics. The original proposal [5] assumed that the brain macroscopically follows linear dynamics with inter-areal couplings given by structural connectivity (SC), i.e., the amount of anatomical connections between areas estimated from diffusion MRI. This approach, however, was criticized by Tu et al. [6], who argued that couplings defined by structural connectivity miss important features of the dynamics. Dynamical coupling between brain areas is not simply proportional to anatomical connectivity: it can be asymmetric and include negative connections [23], whereas SC matrices inferred from diffusion imaging are always symmetric and positive. In fact, many authors have striven to develop powerful ways to fit functional MRI data at rest with a linear dynamical model and find the underlying *effective connectivity* (EC) structure [23–26].

In the present work, we propose a controllability approach relying on a realistic control objective and a proper dynamical model. On one side, we will focus on a more affordable goal: *target control*, which consists in controlling only a selected group of regions [27] rather than the whole brain. On the other side, we will frame the control problem using EC matrices rather than SC matrices. EC at the individual level will be inferred from functional megnetic resonance imaging (fMRI) data through sparse dynamic causal modeling (spDCM) [26]. This model is a recent improvement over previous DCMs for resting state fMRI [3, 24], allowing for accurate parameter inference by combining linearization of the hemodynamic response, discretization of the dynamics, and then a sparsity-inducing prior. Our proposal is illustrated by applying it to fMRI recordings of $N = 76$ subjects from a large public database (Leipzig Study for Mind-Body-Emotion Interactions - LEMON dataset [28]). We will first confirm the main difficulties of whole-brain controllability already highlighted by previous literature [6, 7, 29], showing that the control cost (energy) scales exponentially with the number of target nodes. Then, we will consider the simplest case of target control, where one wishes to control *a single target region* by acting on a remote brain region. Finally, we will move to the case where one wishes to control interconnected groups of regions defined by canonical resting state networks (RSNs). In all cases, we will address the problem of selecting

good driver nodes (yielding a low energy cost) depending on the target, showing that centrality metrics can assist the choice of drivers, and discussing to what extent an individualized or a group selection is convenient.
The approach we propose has the potential to inform neurostimulation experiments (e.g., with TMS) where one wishes to control the activity of a (small) set of target regions. In this context, our approahc allows indentifying the optimal driver region (or set of driver regions) to control the target, and assess the difficulty of the control problem in terms of control energy (the amplitude of the control signal to be applied).

Materials and methods

Data collection

The resting-state fMRI dataset employed in this study consists of resting-state scans on a subset of 295 healthy subjects from the publicly available MPI-Leipzig Mind-Brain-Body dataset (LEMON) [28]. The data selection was performed on the original dataset (consisting of 318 individuals) by excluding participants with structural images heavily affected by artefacts or functional images with high head motion (less than 400 volumes with a mean framewise displacement ¡ 0.4 mm) or affected by pre-processing failures and/or unavailability of rs-fMRI data [30]. While the first half of the dataset (147 subjects) was employed for clustering purposes (see details in the following sections), a final age- and gender-balanced sample of 76 individuals (younger: 20-39 M=19, F=19, older: 40-80 M=19, F=19) was extracted from the second half and then included in the controllability analysis of EC.
Data acquisition was performed with a 3T Siemens Magnetom Verio scanner, equipped with a 32-channel head coil. The protocol included a T1-weighted 3D magnetization-prepared 2 rapid acquisition gradient echoes (MP2RAGE; TR = 5,000 ms, TE = 2.92 ms, TI1 = 700 ms, TI2 = 2,500 ms, first flip angle = 4°, second flip angle = 5°, FOV = $256 \times 240 \times 176$ mm, voxel size = $1 \times 1 \times 1$ mm, multiband acceleration factor [MBAccFactor] = 3), rs-fMRI scans (TR = 1,400 ms, TE = 39.4 ms, flip angle = 69°, FOV = 202×202 mm, voxel size = $2.3 \times 2.3 \times 2.3$ mm, volumes = 657, MBAccFactor = 4) and two spin echo acquisitions (TR = 2,200 ms, TE = 52 ms, flip angle = 90°, FOV = 202×202 mm, voxel size = $2.3 \times 2.3 \times 2.3$ mm). During rs-fMRI scans, the subjects were asked to keep their eyes opened and to lie down as still as possible.

Data preprocessing

For each control an individual pseudo-T1w image was obtained by multiplying the T1w 3D-MP2RAGE image with its second inversion time image and the structural preprocessing performed on this pseudo-T1w image included bias field correction (N4BiasFieldCorrection [31], skull-stripping (MASS [32]) and nonlinear diffeomorphic registration [33] to the symmetric MNI152 2009c atlas [34]. Pre-processing of rs-fMRI data consisted of slice timing (Smith et al. 2004), distortion (TOPUP [35]) and motion correction (MCFLIRT [36]) and nonlinear normalization to the symmetric MNI atlas [34] passing through the pseudo-T1w image via a boundary-based registration [37]. As a second step an ICA-based denoising was performed. The GIFT toolbox (http://trendscenter.org/software/gift/) was used to decompose the functional pre-processed data into independent components (ICs) by performing a group spatial-ICA. As detailed in [38]. The ICs were classified into artefactual or resting-state network related in accordance with Refs. [39, 40]. As a result, ICs that were related to banding artifacts, vascular or noise components were discarded. Then, 10 principal components related to CSF and white matter signal (5 from WM, 5 from CSF) were regressed out

from rsfMRI timeseries as well as the 6 standard head motion parameters and their temporal derivatives. Then the denoised signal was high-passed with a filtering cut-off equal to 1/128 Hz.

Parcellation and networks

We then extracted subject-level time series data from a 100-area parcellation scheme of the cortex provided by the Schaefer atlas [41], which maps to 7 resting-state functional networks: Control network (CON, 10 parcels), Default mode network (DMN, 16 parcels), Dorsal attention network (DAN, 9 parcels), Limbic network (LIM, 5 parcels), Saliency/Ventral attention network (VAN, 11 parcels), Somatomotor network (SMN, 6 parcels), Visual network (VIS, 5 parcels). We also defined a set of 12 subcortical and cerebellar regions based on the AAL3 segmentation [42]: for each hemisphere, 6 regions consisting of thalamus, caudate, putamen, pallidum, hippocampus and cerebellum (SUB, 12 parcels).

In addition, we assigned to each subject a binary temporal mask accounting for brain volumes corrupted by head motion ($FD > 0.4mm$) and we applied volume despiking to the time series by means of the icatb_despike_tc function of the GIFT toolbox. Moreover, the temporal traces were band-pass filtered (0.008 to 0.1 Hz).

Given the need to keep the computational load of sparse DCM at a reasonable level, a Consensus Clustering Evidence Accumulation (CCEA) procedure [43] was applied to reduce the number of cortical parcels derived from the Schaefer atlas. In order to account for hemodynamic differences across spatially distant parcels, this procedure was performed selectively for subsets of adjacent cortical regions referring to the same functional network. This additional constraint implied that only functionally homogeneous and spatially contiguous parcels could be grouped together, ensuring the consistency of hemodynamic properties of each cluster. The resulting clustering procedure provided 62 cortical clusters, from which demeaned fMRI time courses (i.e., within-cluster mean BOLD signal) were extracted and supplied as inputs to sparse DCM together with the BOLD signals from subcortical sources.

Sparse DCM

Dynamical Causal Modelling (DCM) was first introduced by *Friston et al.* [44]. It is a generative model of measured brain responses, where the output haemodynamic responses are evoked either by an underlying (unobserved) brain activity arising from experimental stimuli (during tasks) or spontaneous neural fluctuations (at rest). Here, we use the *sparse DCM* approach by Prando et al. [26]. This DCM variant implements a sparsity inducing mechanism that automatically prunes irrelevant connections, thereby avoiding the need to perform a selection between competing network structures. The algorithm has been further adjusted to account for the signal reliability of the temporal frames, by introducing the binary temporal mask as a weighting measure during the estimation procedure. The model includes two layers: i) a coupled ODE system modeling neuronal activation $x(t)$, and ii) a mapping from neuronal activity $x(t)$ to the BOLD fMRI signal $y(t)$ (hemodynamic response). In formulas:

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \boldsymbol{\nu}(t) \tag{1a}$$

$$\mathbf{y}(t) = h(\mathbf{x}(t); \theta_h) + \mathbf{e}(t) \tag{1b}$$

where $\mathbf{x}(t) = [x_1(t)...x_n(t)]^T$ is the hidden neural activity of n brain regions at time t, A is the effective connectivity matrix, $\boldsymbol{\nu}(t)$ is a stochastic term driving intrinsic brain fluctuations, $\mathbf{y}(t)$ is the BOLD fMRI response at time t, θ_h denotes collectively a set of biophysical parameters regulating the haemodynamic response (which is modelled with

the Balloon-Windkessel model [44]), and $\mathbf{e}(t) \sim \mathcal{N}(0, R)$ is a Gaussian observation noise with covariance matrix R.

All model parameters, including the effective connectivity matrix A, need to be estimated by inverting the model on the measured fMRI data. To simplify the estimation procedure, Prando et al. [26] took two steps. First, in a *discretization* step, justified by the low temporal resolution of fMRI scanners with sampling time $T_R \sim 0.5s$ to $3s$, the equation is integrated in steps of T_R. If we measure time in units of T_R, leading to the finite difference equation

$$\mathbf{x}(k+1) = e^A \mathbf{x}(k) + \mathbf{w}(k) \tag{2}$$

If we assume that the stochastic term $\boldsymbol{\nu}(t)$ is a white Gaussian noise with diagonal covariance matrix $\sigma^2 I_n$, then $\mathbf{w}(k)$ is also white Gaussian and its corresponding variance is given by $Q = \sigma^2 \int_0^1 e^{A\tau} e^{A^T \tau} d\tau$. Second, in a *linearization* step the non-linear haemodynamic response is linearized by assuming a finite impulse response (FIR) for brain region i

$$y_i(k) = \sum_{l=0}^{s-1} h_{i,l} x_i(k-l) \tag{3}$$

where $h_i = [h_{i,0}, ..., h_{i,s-1}]^T$ are the FIR parameters for region i, with s large enough to maintain temporal dependencies. The combination of these two simplifying moves reduces the model to a linear stochastic model

$$\mathbf{x}^{(s)}(k+1) = A^{(s)} \mathbf{x}^{(s)}(k) + \mathbf{w}^{(s)}(k) \tag{4}$$

$$\mathbf{y}(k) = H^{(s)} \mathbf{x}^s(k) + \mathbf{e}(k) \tag{5}$$

where $\mathbf{x}^{(s)}(k) = [\mathbf{x}^T(k)\ \mathbf{x}^T(k-1)\ ...\ \mathbf{x}^T(k-s+1)]^T \in R^{n \times s}$ is the time-delayed activity signal, $A^{(s)}$ and $H^{(s)}$ are matrices containing the EC parameters $(A)$and the FIR parameters (H), respectively, $\mathbf{w}^{(s)}$ is a Gaussian noise terms with covariance matrix $Q^{(s)}$ (with blocks equal to Q), and $\mathbf{e}$ is a Gaussian noise with covariance matrix R.

The parameters $\theta = \{A, H, Q, R\}$, are estimated within a Bayesian framework by taking into account the observed values of the BOLD signal as well as the prior distribution of the parameters, chosen to be in this factorized form:

$$p(\boldsymbol{\theta}) \propto p_\gamma(A) p(Q) p(H) p(R) \tag{6}$$

Here, $p(Q)$ and $p(R)$ are uninformative priors, $p(H)$ is Gaussian (with means and variances fixed from knowledge of the typical haemodynamic responses [44]), and $p_\gamma(A)$ is a *sparsity inducing prior*,

$$p_\gamma(A) \sim \mathcal{N}(0, diag(\gamma_1, ..., \gamma_{n^2})) \tag{7}$$

Parameters are estimated by maximum a-posteriori estimates, using the expectation-maximization algorithm. The hyper-parameters γ_i are estimated through marginal likelihood maximization, ensuring that a controlled fraction of the γ_i are small and thus effectively inducing sparsity in A.

Controllability

In our control framework, we neglect noise and assume that input is provided to a set of *driver nodes*. The system's dynamics become

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t) \tag{8}$$

where $\mathbf{u}(t)$ is a time-dependent $r \times 1$ vector representing r external inputs $(r \leq n)$, $\mathbf{u}(t) = (u_1(t), ..., u_r(t))^T$ and B is an $n \times r$ input matrix with which identifies the driver nodes, with $B_{ij} = 1$ if control input $u_j(t)$ is imposed on node i.

The *Kalman's controllability condition* [45] states that the system is controllable if and only if the *controllability Gramian* W

$$W = \int_0^\infty e^{At} B B^T e^{A^T t} \tag{9}$$

W is positive definite, $W > 0$ or $\lambda_{min}(W) > 0$ where λ_{min} is the minimum eigenvalue. Due to numerical inaccuracies, it is impossible to assess whether an eigenvalue value is exactly 0. Following common practice [46], we consider an eigenvalue to be 0 whenever is it below a very low numerical threshold $\epsilon = 10^{-12}$.

The *control energy* is defined as the (integrated) amplitude of the control signal used to steer the system from a given initial state $\mathbf{x}_0$ to a given final state $\mathbf{x}_f$,

$$E(u) = \int_0^{t_f} dt ||\mathbf{u}(t)||^2 \tag{10}$$

Note that, if t is measured in units of T_R, $E(u)$ is adimensional. The magnitude of $E(u)$ is related to the magnitude of the control signal, as $||\mathbf{u}|| \approx \sqrt{E/\tau}$ where τ is the time for which $||u||$ is significantly different from 0. As matrix elements of B are of value 1, the magnitude of the term $B\mathbf{u}$ in Eq.(8) is of order $\sqrt{E/\tau}$. This is to be compared with the magnitude of the initial and final states, $||\mathbf{x}_0||, ||\mathbf{x}_f|| = 1$. If $E(u) = 10^{12}$ and $\tau = 10^2$, this means that $||u|| \approx 10^5$, which means that the external driving must force the system through trajectories that pass through activity vectors of magnitude 10^5 times larger than the initial and final activity vectors.

Let $\mathbf{u}^*$ be the optimal control input minimizing the control energy for a given pair $(\mathbf{x}_0, \mathbf{x}_f)$. In the limit $t_f \to \infty$, for normalized $(\mathbf{x}_0, \mathbf{x}_f)$, one has

$$E(\mathbf{u}^*) \leq \mathcal{E} \equiv \frac{1}{\lambda_{min}(W)} \tag{11}$$

where the λs are simply the eigenvalues of W. A common metric to assess the difficulty of steering the system is the upper bound $\mathcal{E}$, which gives the control energy required to steer the system along the worst possible eigendirection of the Gramian W.

In *target control* [27], one aims to control only a selected subset of target nodes. Let $\mathcal{T} = \{c_1, c_2, c_3 ... c_S\}$ be the target node set (of cardinality S) and let

$$\mathbf{y}(t) = C\mathbf{x}(t) \tag{12}$$

be the output vector describing the activity of the the target nodes we want to control ($\mathbf{y}(t) \in \mathbb{R}^S$), with $C_{ij} = 1$ if and only if $i = j$ and $j \in \mathcal{C}$.

The definition of target controllability follows from that of standard (Kalman) controllability, where the system is now defined by the triple (A, B, C) instead of the pair (A, B) [27]. The system (A, B, C) is *target controllable* with respect to target node set $\mathcal{C}$ if the target controllability Gramian

$$W_C = CWC^T \tag{13}$$

is positive definite. Similar to the one we have seen in the case of full controllability, we have for the control energy

$$E(\mathbf{u}^*) \leq \mathcal{E}^{target} = \frac{1}{\lambda_{min}(W_C)} \tag{14}$$

If a single driver node i is used, and the target is in turn a single node j, the expression of the control energy significantly simplifies. We have $B = \mathbf{b}_i^T$ and $C = \mathbf{b}_j$, where $\mathbf{b}_i$ is the i-th canonical basis vector. Thus

$$\mathcal{E}_{i \to j} \equiv E_{min}^{target} = (W_{jj}^{(i)})^{-1} = \frac{1}{\int_0^\infty dt [e^{At}]_{ij}^2} \tag{15}$$

To better highlight the controllability properties of each node within a network, we define two quantities, the *driver centrality* and the *target centrality* by summing the pairwise control energy over all possible targets and all drivers, respectively:

$$\mathcal{E}_i^d = \sum_j \mathcal{E}_{i\to j} \tag{16a}$$

$$\mathcal{E}_i^t = \sum_j \mathcal{E}_{j\to i} \tag{16b}$$

Centrality measures and shortest paths

EC Centralities. A possible way to select driver nodes is based on centrality measures computed on the graph defined by the EC matrix A_{ij}. Nodes of the network are ranked according to a selected centrality measure, and high-ranking nodes are selected as driver nodes. A viable set of centrality measures appropriate for this approach is the following, which includes both on classical and controllability-tailored measures i) Out-strength (sum of absolute strength of outgoing connections) $\mathcal{A}_i^{out} = \sum_j |A_{ij}|$ and in-strength (sum of absolute strength of incoming connections) $\mathcal{A}_i^{in} = \sum_j |A_{ji}|$; note that $\mathcal{A}_i^{out} \neq \mathcal{A}_i^{in}$ in general because EC matrices are non-symmetric, and we take absolute values since entries of A can have both positive and negative sign ii) Page rank [47], which determines the nodes centrality based on the number and quality of its incoming connections (to avoid issues related to the presence of negative weights in the computation of PageRank, we considered an unweighted version of the network replacing all nonzero links with ones) iii) The ratio of absolute out-strength and in-strength $r_w(i) = \sum_{j=1}^N |A_{ji}| / \sum_{j=1}^N |A_{ij}|$, which was argued to be a good centrality measure to select driver nodes in the context of controllability [48], iv) the control centrality proposed by Lindmark et al [49], $r_{quot} = p_i/q_i$. In the last centrality measure, $p_i = Tr(W^{(i)})$, where $W^{(i)} = \int_0^\infty e^{At} b_i b_i^T e^{A^T t}$ is the controllability Gramian corresponding to using node i as a driver; $q_i = Tr(M^{(i)})$, where $M^{(i)}$ is the observability Gramian $M^{(i)} = \int_0^\infty e^{A^T t} b_i b_i^T e^{At}$. Respectively, p_i and q_i measure the ability to control other nodes from node i, and the ability to control node i indirectly from other nodes.

Energy Centralities. Based on the single-driver-single-target energy (15) we define two quantities, the *driver centrality* $\mathcal{E}_i^d = \sum_j \mathcal{E}_{i\to j}$ and the *target centrality* $\mathcal{E}_i^t = \sum_j \mathcal{E}_{j\to i}$, by summing the pairwise control energy over all possible targets and all drivers, respectively. The driver energy represents the average energy with which we can control another node, using node i as a driver. The target energy represents the average energy with which we can control node i using another node as a driver.

Shortest paths. In the graph defined by A_{ij}, we defined the length of a path by summing the length of each edge, assigning to the edge between nodes k and l a length $1/|A_{kl}|$, i.e., inversely proportional to the effective connection between k and l. We can thus compute shortest paths in the graph through Dijkstra's algorithm [50]. We denote by ℓ_{ij} the length of the shortest path between nodes i and j.

Optimal node placement and rank aggregation

For a given subset of target nodes $\mathcal{T}$, we rank nodes according to different centralities, and select as driver nodes the n_d nodes with lowest rank. Then, we can identify which centrality allows achieving the lowest value of control energy ('optimal centrality'), rank nodes according to the optimal centrality, and select as drivers the n_d driver nodes with lowest rank ('optimal drivers'). In a fist step, we can identify optimal drivers for each

subject independently, yielding a subject-dependent set of optimal drivers $\mathcal{O}(s)$ where s is a subject index.

Then, given a certain set of nodes $\mathcal{D}$, non-overlapping with $\mathcal{T}$ ($\mathcal{D} \cap \mathcal{T} = \emptyset$), we can test whether optimal drivers are preferentially selected from $\mathcal{D}$. This problem is analogous to the problem where one has balls of two colors (blue/red) divided in two urns (A/B), and wants to test whether urn A contains an anomalous proportion of blue balls (i.e., statistically unlikely if balls are randomly placed in the two urns regardless of color). This problem can be solved performing a Fisher exact test [51]. Here, we have optimal/non-optimal nodes (belonging respectively to $\mathcal{O}$ and its complement $\bar{\mathcal{O}}$), divided in two sets . Two know whether optimal nodes are preferentially selected from $\mathcal{D}$, we perform a Fisher exact test on the quantities

$$n_1 = \sum_s |\mathcal{O}(s) \cap \mathcal{D}|, \quad n_2 = \sum_s |\mathcal{O}(s) \cap \bar{\mathcal{D}}|, \quad n_3 = \sum_s |\bar{\mathcal{O}(s)} \cap \mathcal{D}|, \quad n_4 = \sum_s |\bar{\mathcal{O}(s)} \cap \bar{\mathcal{D}}|,$$

which correspond to the number of optimal/non-optimal that belong/not belong to $\mathcal{D}$. One can use rank aggregation to obtain a group-wise set of optimal nodes. Among the possible approaches to rank aggregation [52], we used the most basic approach, namely, computing the average rank (other common criteria such as Borda and Dowdall [52] give very similar results). For each subject, we rank nodes according to the optimal centrality, produce a group ranking using rank aggregation, and and select as drivers the n_d driver nodes with lowest rank.

Results

Effective connectivity matrices

We considered resting state fMRI data of $N = 76$ participants, parcellated into $n = 74$ regions (58 cortical regions + 16 subcortical regions). Applying sparse DCM [26] to the regional time series, we obtained individual effective connectivity matrices A_{ij}. The linear model given by DCM obtained a very good fit of the data, with a correlation between the functional connectivity (FC, standard Pearson correlation matrix between the BOLD signals of all areas) of the model and the actual FC of 0.78 (on average over subjects). Example matrices are shown in Supplementary Fig. 7. The EC matrices have nonzero, negative diagonal entries, $A_{ii} < 0$, as required for dynamic stability. Effective connections are sparse: on average over all subjects, the link density was 0.39 (61% of matrix entries are zero). On average 59.9% of links were positive ("excitatory"), and 40.1% negative ("inhibitory"). The EC matrices exhibit a large inter-subject variability. To assess the degree of inter-subject consistency, we evaluated the Pearson correlation between the EC matrices of all pairs of subjects: on average over all pairs, the correlation was 0.49 (s.d. 0.03). This figure is comparable with the inter-subject consistency of FC matrices (average 0.49, s.d. 0.08).

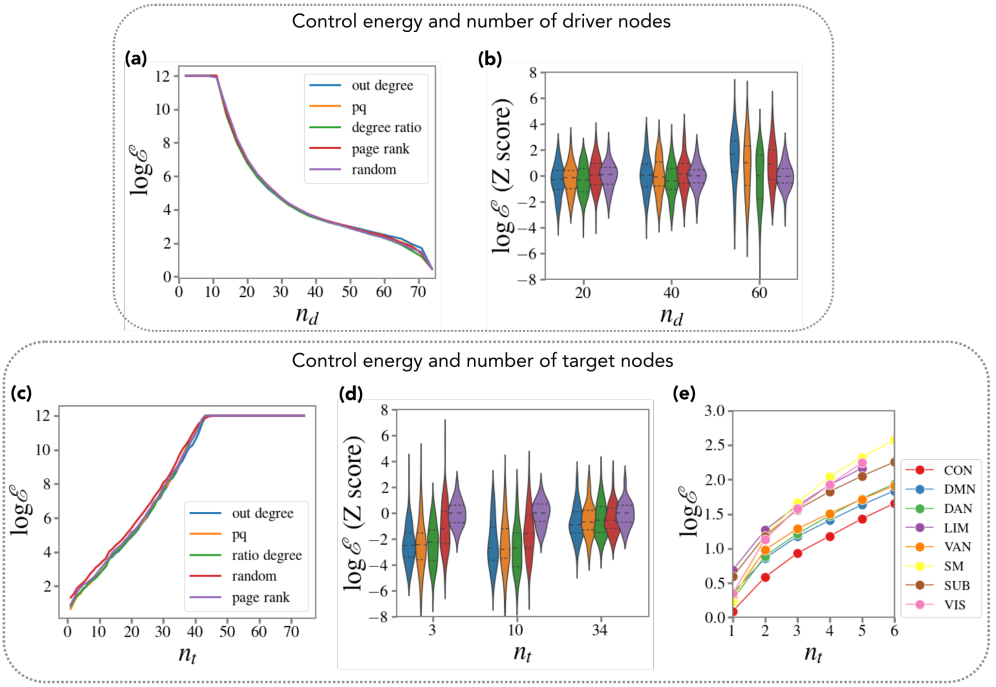


Fig 1. *Dependence of the control energy on the number of driver and target nodes.* **(a)** Energy to control the whole brain network (median over subjects) as a function of the number of driver nodes n_d. For each n_d, nodes were selected based on a ranking of centrality measures. **(b)** Energy to control the whole brain network (distribution over subjects), for three values of n_d. For each subject, energy values were z-scored with respect to the mean of the random node selection. **(c)** Energy to control a varying number of target nodes, using $n_d = 5$ driver nodes selected according to different centrality measures as well as randomly. Lines represent the average control energy over subjects (over both subjects and realizations for the random curve) **(d)** Energy to control target nodes, using $n_d = 10$ driver nodes (distribution over subjects). For each subject, energy values were z-scored with respect to the mean of the random node selection. **(e)** Energy to control a varying number of target nodes within each of 8 RSNs, using $n_d = 10$ driver nodes selected according to a ranking based on the out-strength. In (a-d), all random curves were obtained by averaging over $M = 100$ random selections of driver nodes.

Scaling of control energy with the number of driver and target nodes

Two key parameters determining the control energy (Eq.11) are the number of driver nodes, n_d, and the number of target nodes, n_t. In the literature, the case $n_t = n$ is usually considered, where one tries to control the whole network. We thus first fix $n_t = n$ and analyze the control energy $\mathcal{E}$ as a function of the number of driver nodes n_d (Fig. 1a). For each value of n_d, we selected driver nodes as high-ranking nodes according to different centrality measures (Methods). Results for a random selection of nodes are also presented. The control energy is exceedingly high ($> 10^{12}$) unless a significant ($\gtrsim 15\%$) fraction of the nodes are used as driver nodes. This result is fully in agreement with the previous results of Tu et al. [6] and resonates with previous theoretical controllability studies. On one side, since the matrix A has nonzero diagonal entries, the maximum matching theorem ensures that the system is controllable by applying a single external input jointly to all nodes, i.e. $B = [1, 1, 1, ..., 1]^T$ [21,53]. However, when computing the minimum eigenvalue of the corresponding Gramian matrix, we systematically obtain very small values (of the order 10^{-13}). Therefore, this simple control solution is not applicable in practice. In fact, as highlighted in Ref. [21], unless a considerable fraction of the nodes are controlled, control energy is astronomically large, and control trajectories extremely long and numerically unstable. Fig. 1b shows the distribution (over subjects) of control energies obtained with different (centrality-based) choices of driver nodes, with energies z-scored to the mean of the distribution obtained with a random choice of driver nodes. The control energy depends quite weakly on the choice of driver nodes, with centrality measures not affording any significant advantage over a random choice of nodes.

Given the difficulties with whole-brain network controllability, we next consider the dependence of the control energy on the total number of target nodes n_t. In Fig.1c we plot the energy required to control a varying number of target nodes. Here, target nodes were chosen randomly (nodes were randomly sorted and an increasing number of nodes was included in the set of target nodes following the ranking). We used $n_d = 5$ driver nodes selected according to 4 different centrality measures, as well as randomly. The control energy scales *exponentially* with n_t. Since current techniques allow perturbing only one or a few nodes simultaneously, this implies that the control problem is feasibile only for a low number of target nodes. The strictly exponential scaling depends on the fact that target nodes were chosen randomly: therefore, target nodes were on average not strongly connected to driver nodes. If one restricts attention to groups of strongly connected nodes, such as those belonging to the same resting state network (RSN), we observe a deviation from the exponential scaling (Fig. 1e). In particular, the scaling of $\log \mathcal{E}$ with n_d is weakly sublinear, showing a weak "saturation effect" whereby adding new nodes to the set of target nodes is progressively less costly. Fig. 1c also shows that selection of driver nodes has an effect on the control energy. In particular, the random choice systematically yields larger energies than centrality-based choices. In Fig.1d we show the distribution (over subjects) of the (log-)control energies, z-scored to the mean of the distribution obtained with a random choice of driver nodes. For all values of n_t, a centrality-based choice of driver nodes affords a significant advantage over a random choice of driver nodes (T-test, $T(75) < -8$, $p < 10^{-10}$ corrected for 5 multiple comparisons). The effect is more pronounced for low n_t. These results do not depend either on the specific n_d used (analogous results are obtained with $n_d = 5, n_d = 20$). *In summary, the control energy scales exponentially with the number of target nodes. When the number of target nodes is large, the energy is exceedingly large unless a significant fraction of the nodes is used as driver nodes. These results imply that whole-brain controllability is unfeasible with current techniques. The dependence of the control energy on the choice of driver nodes is appreciable for a low number of driver and target nodes.*

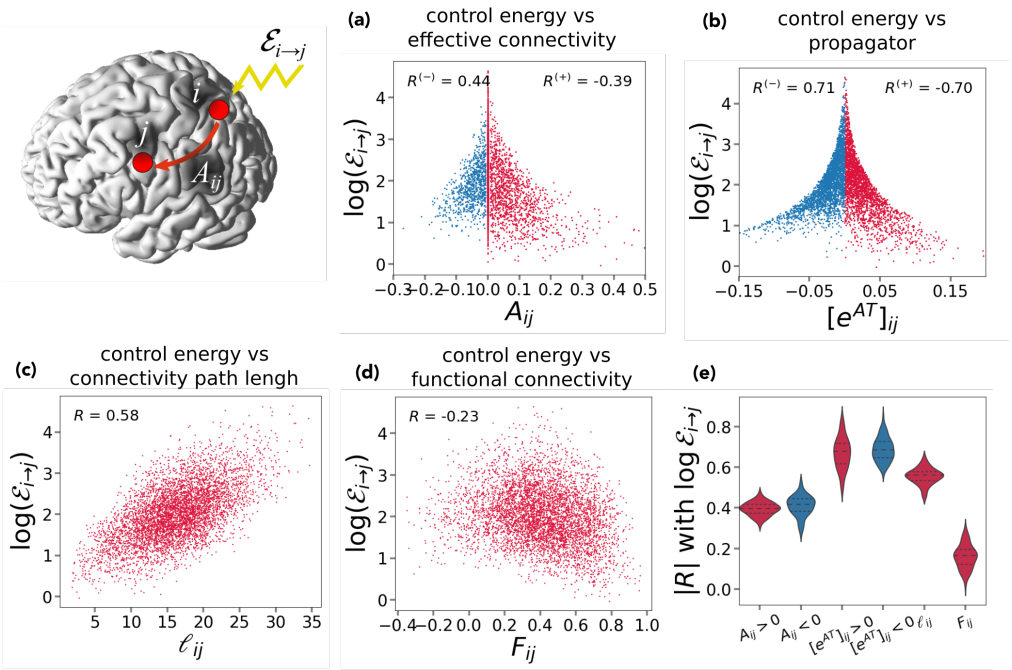


Fig 2. *Relation between single-driver single-target centrality and effective connectivity.* **(a)** Control energy $\mathcal{E}_{i\to j}$ (energy required to control a single node j using a single node i as driver) vs. the effective connectivity between i and j, A_{ij} for a single representative subject. Positive ($A_{ij} > 0$) and negative ($A_{ij} < 0$) effective connections are highlighted in blue and red respectively. The value of Spearman correlation between A_{ij} and $\log \mathcal{E}_{i\to j}$ is shown for positive ($R^{(+)}$) and negative ($R^{(-)}$) connections respectively. **(b)** Control energy $\mathcal{E}_{i\to j}$ vs. the absolute value of the i, j matrix element of the propagator e^{AT} for $T = 10$ for a single representative subject. Positive ($[e^{AT}]_{ij} > 0$) and negative ($e^{AT}_{ij} < 0$) effective connections are highlighted in blue and red respectively, along with the corresponding values of Spearman correlation with $\log \mathcal{E}_{i\to j}$. **(c)** Control energy $\mathcal{E}_{i\to j}$ vs. the length of the shortest path ℓ_{ij} connecting i and j using effective connections for a single representative subject. **(d)** Control energy $\mathcal{E}_{i\to j}$ vs. the i, j matrix element of the functional connectivity F for a single representative subject. **(e)** Distribution (over subjects) of the absolute value of Spearman correlation $|R|$ between $\mathcal{E}_{i\to j}$ and A_{ij} (positive and negative connections), $[e^{AT}]_{ij}$ (positive and negative connections), ℓ_{ij}
, and F_{ij}

Single-node target Controllability

Given the unfeasibility of whole-brain controllability, in the remainder we concentrate on target controllabilty of selected brain regions or gorups of regions. We first consider the simplest, and experimentally most accessible target controllability problem: controlling *a single target node* by using *a single driver node*. This case corresponds to the typical experimental setting where one wishes to activate/deactivate a specific brain region by stimulating a (single) remote region. Furthermore, it allows clarifying general relations between effective connections matrices and controllability.

Which connections contribute to control. The control energy for the single-driver-single-target case is given by Eq. (15), which determines the energy $\mathcal{E}_{i\to j}$ required to control node j through node i. In Fig.2a we plot $\mathcal{E}_{i\to j}$ against A_{ij} for a single subject.

Unsurprisingly, for positive links ($A_{ij} > 0$, blue) $\mathcal{E}_{i\to j}$ is negatively correlated with A_{ij} (Spearman $R \approx -0.39$ for this subject). This means is there is a large effective connection between i and j, it is less costly to control j through i (control energy decreases). However, for negative connections ($A_{ij} < 0$) $\mathcal{E}_{i\to j}$ is *positively* correlated with A_{ij} (Spearman $R \approx 0.44$ for this subject). Thus, negative connections have a positive, not a detrimental effect for controllability: if there is a large, negative effective connection between i and j, it is less costly to control j through i. Stated otherwise, effective connections reduce the required control energy with a contribution dependent on their strength, but independent of their sign. Group results confirm this finding (Fig. 2e). The average Spearman correlation between EC and control energy is $R = -0.40 \pm 0.03$ (mean $\pm$ s.d.) for positive connections and $R = 0.41 \pm 0.05$ for negative connections. From Fig.2a we also see that large effective connections are a sufficient, yet not necessary condition to have low control energy. We hypothesized that this is due to indirect connections. Indeed, mathematically, the influence of node i onto j over a time scale t is exerted though the propagator e^{At}, rather than A. The matrix element $[e^{At}]_{ij}$ effectively integrates the effect of direct and indirect paths between i and j. In Fig. 2b we plot $\mathcal{E}_{i\to j}$ against $[e^{AT}]_{ij}$ for a single subject, where $T = 10$ (corresponding to a typical time scale of duration of a control signal), obtaining stronger correlations ($R \approx -0.70$,$R \approx 0.71$ for positive and negative connections respectively). Over all subjects (Fig. 2e), we obtain $R = 0.67 \pm 0.08$ for positive connections and $R = 0.69 \pm 0.06$ for negative connections. To strengthen the conjecture that the value of the control energy $\mathcal{E}_{i\to j}$ is related to the presence of direct and indirect connections between i and j, in Fig.2c we plot $\mathcal{E}_{i\to j}$ against ℓ_{ij}, the length of the shortest path between i and j in the graph defined by A_{ij}. We observe a strong positive correlation ($R \approx 0.58$): if nodes i and j are "near" (i.e., linked by strong direct or indirect connections), the control energy is lower. Over all subjects (Fig.2e), the average correlation coefficient is 0.56 ± 0.04. Finally, we note that the value of the control energy $\mathcal{E}_{i\to j}$ is poorly predicted by the standard functional connectivity F_{ij} between between nodes i and j (Fig.2d), as we observe only a weak negative correlation ($R \approx -0.23$). Over all subjects (Fig.2e), the average correlation coefficient is -0.16 ± 0.06. *In summary, the presence of large (direct and indirect, positive or negative) connections between i and j determines a low control energy $\mathcal{E}_{ij}$.*

Optimal driver and target nodes. Based on the results of the previous section, we assumed that nodes with strong incoming connections would require a low energy to be controlled, and nodes with strrong outgoing connections would require low energy to control onther nodes. We verified this hypothesis by computing the link between the driver/target centrality $\mathcal{E}_i^d, \mathcal{E}_i^t$ of a node (Methods), representing the average energy when using a node as a driver or target, and the in- and out-strength of that node $\mathcal{A}_i^{in}, \mathcal{A}_i^{out}$. On average over subjects (Supplementary Fig. 8a), $\mathcal{E}_i^d$ is strongly negatively correlated with $\mathcal{A}_i^{out}$ ($R = -0.71 \pm 0.05$, mean $\pm$ s.d.) but uncorrelated with $\mathcal{A}_i^{in}$ ($R = 0.00 \pm 0.15$). Conversely, $\mathcal{E}_i^t$ is strongly negatively correlated with $\mathcal{A}_i^{in}$ ($R = -0.72 \pm 0.07$) but weakly correlated with $\mathcal{A}_i^{out}$ ($R = -0.10 \pm 0.17$). This implies that asymmetries between incoming and outgoing connections have large significance for control. These asymmetries can be appreciated when considering EC (which is non- symmetric), but not standard functional connectivity, FC (which is by definition symmetric). In fact, when considering the FC strength ($\mathcal{F}_i = \sum_j F_{ij}$ where F is the functional connectivity matrix), we we did not find any relation with either the driver centrality ($R = -0.02 \pm 0.12$), or the target centrality ($R = 0.17 \pm 0.13$).

The above result suggest that "in-hubs" of effective connectivity are the easiest nodes to control, while "out-hubs" of effective connectivity are the best nodes to use to control other nodes, and should possibly chosen as driver nodes. However, an important *caveat* to this result is that out-hubs are poorly consistent across subjects. In fact, we assessed

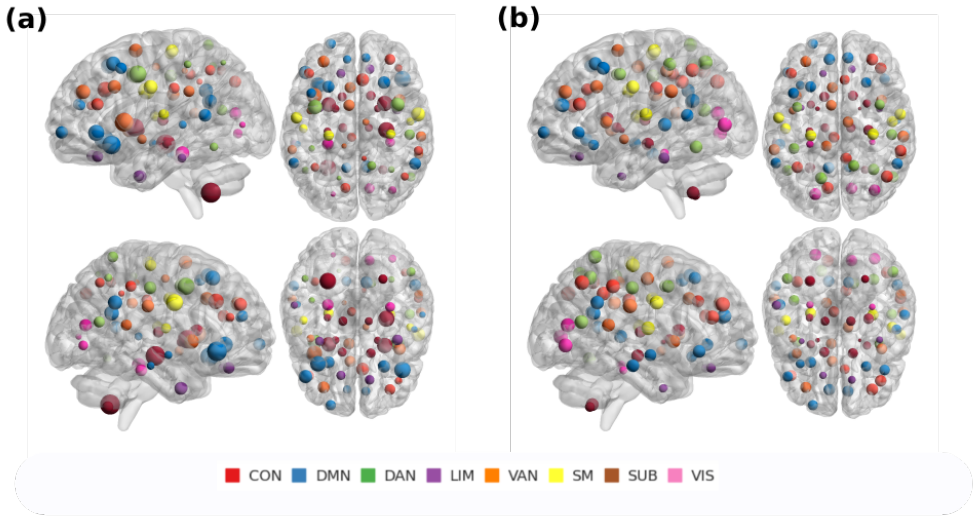


Fig 3. *Driver nodes and target nodes* **(a)** We show a rendering of the brain, with each dot representing the center of one of the 74 regions. Node size is inversely proportional to the node driver centrality $\mathcal{E}_i^d$, while nodes color corresponds to resting state network affiliation. **(b)** as in (a), but node size is inversely proportional to the node target centrality $\mathcal{E}_i^t$.

the consistency of in- and out-strength over subject by computing the coefficient of variation c_V (standard deviation/mean; small values imply high consistency and vice versa). We obtained and for out-strength $c_V = 0.31 \pm 0.07$ (mean $\pm$ st. dev. over nodes) and for in-strength($c_V = 0.17 \pm 0.03$ (Supplementary Fig. 8b). Thus, the in-strength is much more consistent than the out-strength (T test, $T(73) = -19.2$, $p < 10^{-30}$). Correspondingly, the target centrality ($c_V = 0.10 \pm 0.02$) is much more consistent than the driver centrality ($c_V = 0.19 \pm 0.03$; T test, $T(73) = -19.5$, $p < 10^{-30}$). Thus, the easy-to-control nodes are more consistent across subjects, than the good "input nodes" by which one can control other nodes.

To identify good driver nodes, we ranked nodes based on $\mathcal{E}_i^d$. Fig. 3a shows a rendering of the brain, with node size inversely proportional to the average node rank (average over subjects) based on $\mathcal{E}_i^d$. The average $\mathcal{E}_i^t$ tends to decrease along the posterior/anterior axis, with posterior nodes generally corresponding to larger target energies (the correlation between $\mathcal{E}_i^t$ and the sagittal coordinate y of the nodes is significant, $R = -0.34, p = 0.003$). Nodes with low rank (low $\mathcal{E}_i^d$), on average over subjects, include portions of the anterior DMN (ventrolateral prefrontal cortex/ba47 and dorsomedial prefrontal cortex/ba8), the anterior portion of the VAN (dorsolateral prefrontal cortex), the anterior portion of DAN (frontal eye field), primary motor cortex, putamen, left cerebellum, right hippocampus. Nodes with low rank (high $\mathcal{E}_i^d$) include thalamus, caudate, the temporal portion of DMN, primary visual cortex, the posterior portion of the DAN. However, in agreement with the above states caveat, the distribution (over subjects) of node ranks is quite broad for all nodes, with a st. dev. of ≈ 20 for all nodes, implying that ranks are not consistent across subject (Supplementary Fig. 9a). Therefore, while we can identify nodes that tend to be better/worse as driver nodes across subjects, no nodes are consistently good/bad for all subjects. Supplementary Fig. 9a also shows node affiliation to one of eight resting state networks (RSN). The node ranking does not clearly correlate with RSNs affiliation: no networks are consistently associated with low/high ranks.

To identify nodes that are easy to control, we ranked nodes based on $\mathcal{E}_i^t$. Fig. 3b shows a rendering of the brain, with node size size inversely proportional to the average node

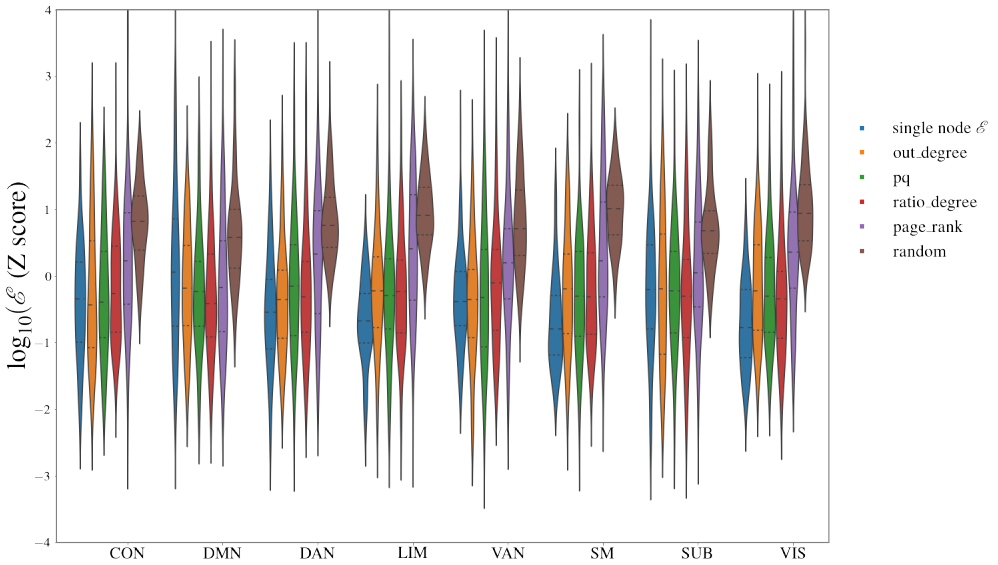


Fig 4. *Target RSN controllability.* For each of the eight RSNs, we computed the control energy required to control all nodes belonging to that RSN, when $n_d = 5$ driver nodes are selected among all remote nodes (*not* belonging to that RSN) according to rankings based on centrality measures. For all RSNs and all choices of driver nodes, we plot the distribution over subjects of the average (log) energy per node.

rank. The average $\mathcal{E}_i^t$ tends to decrease along the ventral/dorsal axis, and to increase along the posterior/anterior axis, with ventral and anterior nodes generally corresponding to larger target energies (we found a significant correlation between $\mathcal{E}_i^t$ and the axial coordinate z of the nodes, $R = -0.63$, $p = 4 \cdot 10^{-9}$, and a significant correlation between $\mathcal{E}_i^t$ and the sagittal coordinate y of the nodes, $R = -0.32$, $p = 0.005$). Nodes with low $\mathcal{E}_i^t$ include primary visual cortex, posterior nodes of the DAN, posterior nodes of the CON, right anterior nodes of the CON, and the medial prefrontal cortex portion of the DMN. In terms of RSN affiliartion, nodes of the Limbic network and subcortical nodes are generally associated with very high ranks. On the contrary, nodes of the control and the sensorimotor network are generally associated to low ranks. The distribution (across subjects) of node ranks, for each node, is shown in Supplementary Fig. 9b. In agremeement with the above discussion of consistency, the rank distribution is much sharper than that obtained with $\mathcal{E}_i^d$, with a st. dev. of < 10 for many nodes, Therefore, not only can we identify nodes that tend to be better/worse as target nodes (in terms of control energy) across subjects, but we find nodes that are consistently good/bad for all subjects.

In summary, in-hubs of EC are easy to control, out-hubs of EC are the best nodes to use as driver nodes; In-hubs are consistent over subjects, and generally located dorsally; out-hubs are poorly consistent over subjects, and mostly locate frontally.

RSN target Controllability

We have showed that the control energy needed to control a target node depends on the choice of the driver node, and we linked this variability to the structure of effective couplings. Here, we address the general problem of selecting driver nodes when wishing to control more than one target nodes. Due to the general findings in the "scaling" subsection, we consider only small sets of target nodes. A natural choice is to consider as targets groups of nodes belonging to the same resting state network (RSNs). RSNs

correspond to integrated neurocognitive systems [54–57] and are jointly affected in major brain disorders [58, 59].

Driver node selection. We computed the control energy required to control each of eight RSNs, using a fixed number of driver nodes n_d. We systematically analyzed the effect of driver node selection, by comparing results obtained selecting driver nodes: i) based on a driver energy rank ii) based on EC centrality iii) randomly. Nodes belonging to the target RSN where excluded from the set of possible driver nodes. Results for $n_d = 5$ are shown in Fig. 4, where we show the average (log)energy to control each RSN with different driver node selection. For each RSN, energies were z-scored to the mean energy (over subjects and driver node selection). Performing a two-way repeated measures ANOVA on the z-scored energy values, with RSN and driver nodes selection criterion as factors, we obtained a significant effect of selection criterion ($F(5, 375) = 195, p < 10^{-10}$), and a significant criterion $\times$ RSN interaction ($F(35, 2625) = 5.2, p < 10^{-10}$). Post-hoc T-tests comparing different criteria to select control nodes show that selecting nodes based on driver centrality or EC centralities except page-rank (pq, ratio degree and out degree centralities) systematically yields lower energies than random ($T(75) < -21.6$, $p < 10^{-10}$). The strongest effects of node selection are felt in the small networks (LIM, VIS and SMN networks). For each target RSN, we identified the centrality yielding the lowest energy ("optimal centrality"), shown in Tab. 1.

Energy to control a target RSN We computed the control energy required to control each of eight RSNs, using a fixed number of driver nodes n_d. Nodes were selected according to the optimal centrality. In supplementary Fig. 10 we show how the energy scales as a function of n_d. Note that each RSN includes a different number of nodes, hence we cannot directly compare $\mathcal{E}$ for different RSNs (the DMN, which comprises 16 nodes, is generally more difficult to control than the VIS and LIM, that comprise 5 nodes). For $n_d = 2$, energies are $> 10^3$ for all RSNs ($10^3 - 10^8$). For $n_d = 5$, energies are in the range $10^{2.5} - 10^{4.5}$. With $n_d = 10$, energies are in the range $10^2 - 10^{3.5}$. Thus, 5 to 10 driver nodes would generally be required to control a target RSN with fairly low energy.

Controlling a RSN from another RSN. For each target RSN, we ranked nodes based on the optimal centrality and identified a set of 'optimal driver nodes' $n_d = 10$ nodes with lowest rank (at the individual level). We asked whether the optimal nodes to control a given target RSN preferentially belong to specific driver RSNs. In Fig.5a we show, for each target RSN, the average percentage of optimal nodes belonging to each driver RSN. For each pair driver RSN/target RSN, we we tested whether this fraction was higher or lower than expected randomly. Intuitively, if optimal nodes were selected randomly from any driver RSN, the fraction of optimal nodes from a given RSN should approximately match the fraction of nodes belonging to that RSN. More formally, a Fisher exact test (Methods) can be performed to identify when the fraction of nodes from a given driver RSN is lower/higher than chance (marked with $^{*+}$, $^{*-}$ in Fig.5a). Notably, the DMN is overrepresented among good drivers of nearly all networks. The VAN and DAN are overrepresented in the control of each other and the DMN and CON. Conversely, the LIM is systematically underrepresented.

Individual vs. group selection of driver nodes. For each target RSN, we used rank aggregation to combine individual rankings in a unique ranking (Methods), obtaining a group-wise set of optimal driver nodes. In Fig. 5b we show, for each target RSN, the average percentage of group-optimal nodes belonging to each driver RSN. Results are very similar to Fig. 5a, but rank aggregation tends to sparsify the matrix. We

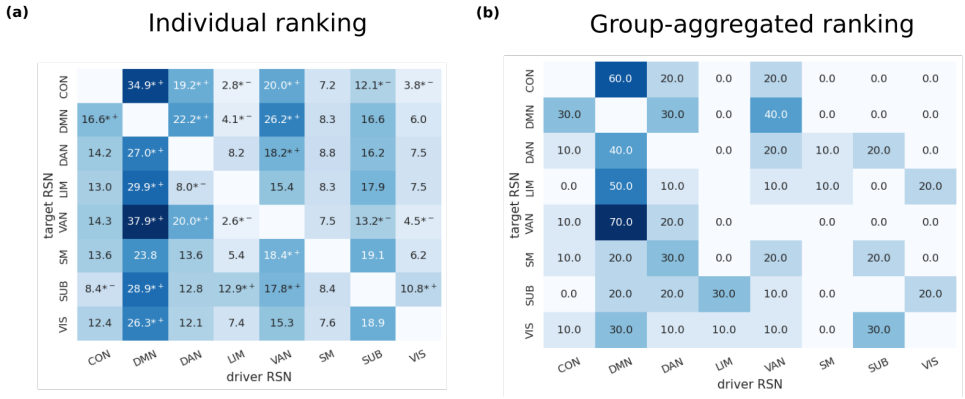


Fig 5. *Relation between driver and target RSNs.* For each target RSN and each subject, we obtained a ranking of driver nodes based on the centrality measure yielding the least average energy to control the target network. **a** For each RSN and subject, we considered the top 10 driver nodes according to the ranking . We plot the number of the top 10 driver nodes belonging to each RSN, on average over subjects. We mark with $*+$, $*-$ nodes that significantly more/less represented among the top 10 driver nodes than expected by chance (Fisher exact test, $P < 0.05$, false-discovery-rate corrected for multiple comparisons (as we have 8 RSNs, we have $8 \times 8 - 8 = 56$ comparisons, as we consider all possible pairs of driver RSN - target RSN excluding identical pairs). **b** Using rank aggregation, for each target RSN we obtained a single ranking for all subjects and considered the top 10 driver nodes. We plot the number of the top 10 driver nodes belonging to each RSN

observe that the SMN is very underrepresented among top-ranking driver nodes. This is probably a consequence of the fact that effective connections of the SMN are highly variable among subjects, so that no nodes of the SMN consistently appear among the top-ranking for many subjects. In Fig. 6 we show the 10 top-ranking nodes according to the aggregated ranking, for two example target RSNs (CON and SUB). In supplementary Fig. 11 we show results for all target RSNs. Among the nodes frequently represented we find: the ventrolateral prefrontal cortex nodes of the DMN and VAN (which among the top-10 ranking nodes for nearly all target networks), the frontal nodes of the DAN, the dorsomedial nodes of the DMN, the precuneus, the striatum and the left cerebellum. We asked to what extent group results, i.e., the aggregated ranking, can be used to select driver nodes. Therefore, we compared the energy to control each target RSN, averaged over subjects, when nodes were selected based on an individual node ranking ($\langle \log_{10} \mathcal{E} \rangle$) or the aggregated ranking $\langle \log_{10} \mathcal{E} \rangle_{agg}$. Results are shown in Table 1. Obviously, the individual ranking is more efficient ($\Delta\mathcal{E} = \langle \log_{10} \mathcal{E} \rangle - \langle \log_{10} \mathcal{E} \rangle_{agg} > 0$). However, the difference is small, ranking from a $\Delta\mathcal{E} = 0.075$ for DAN (corresponding to a factor 1.3 in energy) to $\Delta\mathcal{E} = 0.205$ for VAN (corresponding to a factor 1.6 in energy). Therefore, the group results can be used to inform the node selection.

Discussion

Effective-connectivity-based controllability. We proposed an approach to brain controllability based on effective connectivity (EC) inferred from fMRI, instead of structural connectivity (SC) as in the standard approach. To what extent EC depends on the underlying SC is an open question [60]. The EC model is in principle better suited to represent activity propagation, but we are not aware previous studies presenting a

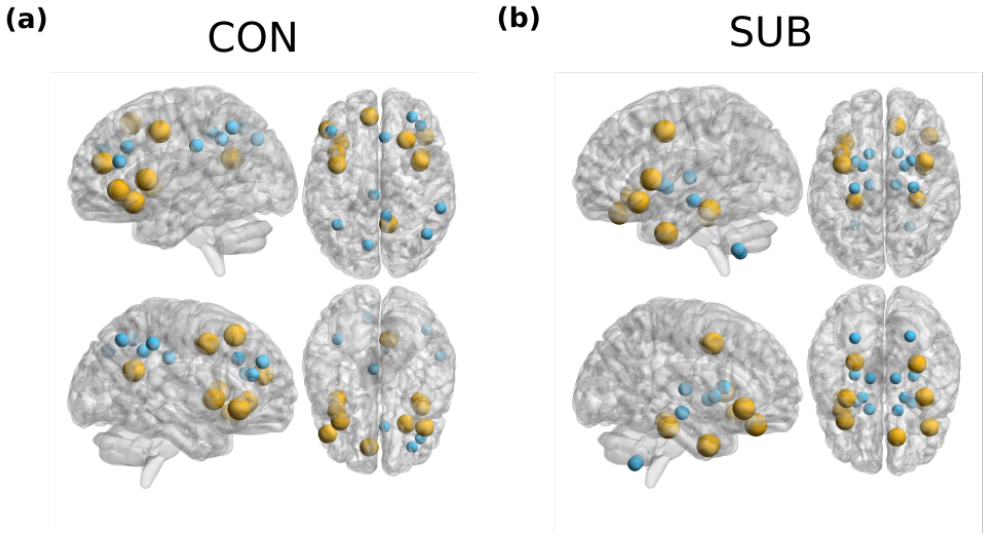


Fig 6. *Optimal driver nodes for RSN*). Two example RSNs with the corresponding target nodes(small blue markers) and top 10 aggregate driver nodes(orange markers) are shown

Table 1

region	centrality	n_t	n_d	$\langle \log_{10}(\mathcal{E}) \rangle$	$\langle \log_{10}(\mathcal{E}) \rangle_{agg}$	$\Delta\mathcal{E}$
CON	out_degree	10	10	2.446	2.536	0.090
DMN	out_degree	16	10	3.521	3.698	0.177
DAN	singlenode	9	10	2.602	2.677	0.075
LIM	singlenode	5	10	2.145	2.350	0.205
VAN	out_degree	11	10	2.797	2.992	0.195
SMN	singlenode	6	10	2.608	2.760	0.152
SUB	pq	12	10	3.298	3.472	0.174
VIS	singlenode	5	10	2.252	2.401	0.151

thorough analysis of controllability properties of whole-brain EC networks. Ref. [17] computed controllability metrics on EC networks, but the entire analysis focused on a small subset regions involved in cognitive control. Recent work [61, 62] assessed controllability properties of functional rather than structural networks, but used standard functional connectivity (FC) networks, rather than effective connectivity (EC) networks. We stress that standard FC just measures correlations between the signals of two areas, and it is not in principle suited to model activity propagation. A key trait of EC networks is that they are asymmetric, contrary to FC networks (symmetric by definition) and SC networks (symmetric due to limitations of dMRI). Consistently with other works [23], we observed substantial asymmetries in EC connections. We showed that asymmetries are important for control, as incoming and outgoing connections, and correspondingly in- and out-hubs, play a different role (Fig. 8a). This differences cannot be detected with an SC- or FC-based analysis. In future work, it is certainly of interest to compare SC- and EC-based predictions of activity propagation following neuromodulation in a TMS-EEG experiment (a recent study [63] compared SC- and FC- based prediction, but FC cannot adequately predict propagation patterns).

Impracticability of whole-brain control. Based on Kalman's criterion, Gu et al. [5] originally argued that the whole brain could be controlled by acting on a single

driver node. However, Kalman's criterion does not ensure practical controllability. The latter requires a reasonably low power of the control signal, i.e., a reasonably low control energy. Very large values of the control energy are problematic for the following reasons: i) experimentally, control signals are subjected to bandwidth and power constraints ii) control trajectories become highly non-local and extremely long, and, consequently, numerically unstable; small numerical errors in the Gramian W (of magnitude comparable to machine precision) imply very large deviations in the final state, and one is not practically able to reach the desired target [21] iii) numerical instabilities can be exacerbated by noise and model inaccuracies. Focusing on SC controllability, Tu et al. [6] showed that control energies are astronomically large unless an important fraction of the network nodes ($\gtrsim 20\%$) are used as driver nodes. Here, we replicate Tu et al.'s findings with EC controllability, showing that whole-brain control implies exceedingly high control energies unless $15-20\%$ of the nodes are used as driver nodes (Fig. 1a). In fact, huge control energies should be expected whenever the ratio between number of driver nodes and the number of target nodes is small, irrespective of details of the system or model at hand. Indeed, the control energy scales (roughly) exponentially with the number of target nodes to be controlled (Fig. 1c).

Target controllability of brain networks. Currently available techniques for non-invasive brain stimulation, such as transcranial magnetic stimulation (TMS), do not allow for stimulation of multiple sites. Standard TMS allows stimulation of one site, and recent experiments with dual coil TMS stimulation (also known as cortico-cortical paired associative stimulation [22, 64]) allow stimulating two cortical sites. Proposals to implement multi-site stimulation are currently under development [65], but formidable technical difficulties must be surpassed before stimulation of more than a few sites simultaneously becomes possible. Therefore, we decided to focus our analysis on control of one or a few target nodes, adopting the framework of target controllability [27, 66]. While recent work considered target controllability of whole-brain brain networks [67,68], we are not aware of a systematic study of target controllability on EC networks. We first analysed the case where a single driver node is used to control a single target nodes (corresponding to experiments where one tries to activate a chosen brain region by acting on a remote region). Controlling a single target region is feasible with limited control energy ($\lesssim 10^1$). We then considered the case of controlling a subset of nodes belonging to the same resting state network (RSN). Control of target RSNs demands large control energies unless at least $n_d \approx 5$ driver nodes are used. Therefore, the control objective is not easily accomplished with current technology. We stress, however, that controlling a subnetwork (in the sense of target controllability) is equivalent to being able to generate an arbitrary activity pattern in the subnetwork. This is considerably more difficult than producing a generic overall activation of the subnetwork (as in [69]).

Criteria for selection of driver nodes. When considering targets of small size, it becomes relevant to properly select driver nodes, ensuring a low control energy. A random selection of nodes is generally inefficient (Fig. 1). In the single-driver-single-target case, control energies span 3 orders of magnitude depending on which node is selected (Fig. 2a). Good driver nodes should be connected to the target: owing to network effects, both direct and indirect connections count (Figs. 2a-b), such that there is a general relation between control energy and network distance between driver and target (Fig. 2c). Negative (inhibitory) connections appear to have the same effectiveness for control as positive (excitatory) connections (Fig. 2e). To identified node that are generally "good drivers" or "good targets", we defined two control metrics, a driver centrality and target centrality", averaging, respectively, the energy required to control other nodes of the network from a given driver node and the energy required to control

a given target node from other nodes. The driver and target centralities are strongly related to EC graph centralities. In particular, out-hubs of EC (nodes with a large weight of outgoing connections) serve as good control drivers, while in-hubs of EC (nodes with a large weight of incoming connections) are easy target nodes. We thus replicate qualitatively a major finding of previous studies on brain controllability [1, 5], namely, that hubs correspond to optimal driver nodes. However, within our picture there is a clear distinction between in-hubs and out-hubs - a difference that can be appreciated only when considering asymmetric effective connections. When using several driver nodes (e.g., to control a target RSN), it is impossible to perform an exhaustive search of the optimal subset driver nodes. However, node centralities can assist the choice of drivers. Selecting driver nodes from a ranking based on driver centrality or EC graph centralities (pq, ratio-degree or out-strength centrality) yields a significant energy advantage over a random selection of driver nodes (Fig. 4). All these metrics largely correlate with the out-strength centrality.

Individual vs. group selection of drivers. While in-hubs of EC were consistent over subjects, out-hubs exhibited a much larger inter-subject variability. Therefore, the location of a single "optimal driver node" may strongly vary among different individuals. This finding strengthens the case for an individual optimization of the stimulation targets, as advocated by recent contributions [69,70]. However, when using multiple driver nodes, selecting nodes based on a group-averaged node ranking does not entail a significant additional energy cost with respect to using a subject-dependent ranking. Possibly, the usage of multiple nodes may partially offset fine-grained individual differences in the connectome profiles. This result is interesting in the light of the development of multiple-site neurostimulation paradigms: it implies that using a standardized protocol over different subjects, that is certainly convenient especially in a development stage, may not determine severe efficiency trade-offs.

Optimal driver regions. Nodes with low driver centrality (good drivers) are predominantly located frontally (Fig-3a). This finding is in agreement with a previously reported anterior-to-posterior information flow in the slow frequency range [71]. Good drivers include dorsolateral and ventrolateral prefrontal nodes of the CON, DAN and DMN, primary motor areas, and left cerebellum. The driver centrality, being an average measure, is not sensistive to different targets. By using rank aggregation, we found nodes that are frequently ranked among good driver nodes for several target RSNs. The most recurring nodes are DAN and DMN nodes located in the ventrolateral prefrontal cortex, the frontal eye field (DAN), and hubs of the DMN (dorsomedial nodes of the DMN and precuneus), the cerebellum and the striatum. Dorsolateral and ventrolateral prefrontal regions are among the key regions involved in cognitive control [72–77]. The frontal eye field is a key region mediating attentional control [78]. Primary motor regions, that are involved not only in motor control but also a wide array of top-down processes [79,80], have been previously associated to a high control centrality [5,68]. The left cerebellum is strongly involved in motor control of the dominant hemisphere [81]). The striatum is widely implicated in learning and reward [82]. The topography of good driver nodes generally aligns with the cortical hierarchy [83]. Areas from attentional and association networks high in the cortical hierarchy (DAN, DMN, VAN) are consistently identified as good driver ndoes, contrary to areas low in the cortical hierarchy (including somatomotor, visual and limbic areas). This trend culminates with the DMN, which sits on top of the cortical hierarchy and is overrepresented among driver nodes of all target networks, coherently with a hypothesized central integrative role of the DMN in the brain [84]. We note that these findings do not fully align with those of Ref. [68], which indicated only motor regions as optimal driver regions. However, Ref. [68] based the

controllability analysis on a "directed structural connectome" obtained by normalizing outgoing connections by node degree, assuming a diffusive process on the network ([85]). The ensuing network considerably differs from EC network empirically observed in fMRI, and implies robust outgoing connections for nodes with low degree, which are mainly located in the somatomotor cortex.

Optimal targets. Target centrality is organized along the axial direction, with ventral nodes being generally associated with higher target centrality (Fig. 3b). In particular, subcortical nodes and nodes of the limbic network correspond to particularly large control energies, implying that they are much less easy to activate and control from remote regions, in agreement with previous findings [70]. Among the nodes with lowest target centrality (hence, among the nodes that are most easy to perturb remotely) are frontoparietal nodes belonging to the control network, which integrate input from several regions and plays a central role in cognitive control [73].

Limitations. We finally discuss possible limitations of the present work. *Cohort and recordings.* The sample used ($N = 76$) was not large enough to split our analysis into a training and a validation cohort. Furthermore, the relatively short fMRI time-series (here, 657 time points) are not optimal for the stability of individual-level EC matrices. Thus, a replication in a different data set involving a larger cohort and longer recordings may considerably strengthen our analysis. A larger cohort would possibly allow linking control properties (such as driver and target energies) with individual traits, e.g., demographic data, allowing to look for age-, sex- and parenthood-related effects ([68,86]). *Parcellation.* The general findings of our manuscript do not depend on the specific parcellation used, but some of the more specific findings may not be parcellation-invariant. For instance, while the relation between out-degree and driver centrality is expected to be general, the specific identity of the optimal driver nodes may depend on the parcellation, as the in- and out-degree of different nodes may slightly vary in different parcellations. We also warn the reader that the specific values of control energy obtained depend on the parcellation used: more fine-grained parcellations imply a larger number of nodes, and hence an increased difficulty of the control problem. *Control cost.* In the present work we used the control energy as general measure of control cost. Control energy is a worst-case-scenario metric, as it measures the maximal square amplitude of the control signal required to generate a desired activity pattern in the target region (maximum over all possible patterns). Therefore, our estimates of controllability are generally quite conservative. *Controllability framework.* Here, we used the standard framework of linear controllability. Obviously, a limitation of this framework is the assumption that the dynamics is linear. However, we acknowledge two additional, potentially more relevant limitations. First, the framework is not fault-tolerant, as it neglects noise and aims at inducing an exact target pattern. In a recent publication, Kamiya et al. [87] frameed the control problem probabilistically: the system's state is not a specific activity pattern, but a distribution, and the control objective is to reach a target distribution centered around a specific pattern. In principle, one could use this approach in combination with spDCM to achieve a fault-tolerant control approach, but this would require some advances: in its current formulation, Kamiya et al's approach assumes that one can control all network nodes, and we should therefore adapt it to embed stronger constraints on the driver nodes that can be used. A second limitation is the framework allows for arbitrary control signals u: in practice, there are often constraints in the control signals that can be generated, e.g., in terms of bandwidth. Finally, in terms of applications, the control approach we are using is suitable for a "single shot" application where one temporarily induces a desired activity state. For clinical applications, it would be relevant to understand how repeated stimulation can leverage plasticity mechanisms

inducing long-term changes [88]. *Rapid effects of brain stimulation.* EC inferred from fMRI is represents infraslow activity (which mostly represents slow modulations of gamma activity [89]), and hence it can captures a slow activity propagation following local activity increases induced by stimulation. It cannot, however, account fora rapid non-local effect of neurostimulation, which induce local spiking and hence also a fast activity spread along anatomical pathways [90]. How to model this effect remains an open challenge.

Acknowledgments

Giorgia Baron and Danilo Benozzo were supported by the DEI Proactive grant "Personalized whole brain models for neuroscience: inference and validation" from the Department of Information Engineering of the University of Padova (Italy).
Maurizio Corbetta was supported by Fondazione Cassa di Risparmio di Padova e Rovigo (CARIPARO) Grant Agreement number 55403; Ministry of Health, Italy (RF-2008 -12366899) Brain connectivity measured with high-density electroencephalography: a novel neurodiagnostic tool for stroke- NEUROCONN; BIAL foundation grant (Grant Agreement number 361/18); H2020 European School of Network Neuroscience (euSNN); H2020 Visionary Nature Based Actions For Heath, Wellbeing & Resilience in Cities (VARCITIES); Ministry of Health Italy (RF-2019-12369300): Eye-movement dynamics during free viewing as biomarker for assessment of visuospatial functions and for closed-loop rehabilitation in stroke (EYEMOVINSTROKE); European Union ("ERC-2022-SYG NEMESIS Grant number 101071900). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Code and Data Availability Statement

The EC matrices for individual subjects, as well as the code used to produce all figures in the manuscript, is made available at https://github.com/karankabbur/Effective-connectivity-and-Controllability-in-brain-dynamics.

Bibliography

References

1. Tang E, Bassett DS. Colloquium: Control of dynamics in brain networks. Reviews of modern physics. 2018;90(3):031003.

2. Lynn CW, Bassett DS. The physics of brain network structure, function and control. Nature Reviews Physics. 2019;1(5):318–332.

3. Friston KJ, Kahan J, Biswal B, Razi A. A DCM for resting state fMRI. NeuroImage. 2014;94:396–407. doi:10.1016/j.neuroimage.2013.12.009.

4. Liu YY, Slotine JJ, Barabási AL. Controllability of complex networks. nature. 2011;473(7346):167–173.

5. Gu S, Pasqualetti F, Cieslak M, Telesford QK, Yu AB, Kahn AE, et al. Controllability of structural brain networks. Nature communications. 2015;6(1):1–10.

6. Tu C, Rocha RP, Corbetta M, Zampieri S, Zorzi M, Suweis S. Warnings and caveats in brain controllability. NeuroImage. 2018;176:83–91.

7. Pasqualetti F, Gu S, Bassett DS. RE: Warnings and caveats in brain controllability. NeuroImage. 2019;197:586–588.

8. Beynel L, Deng L, Crowell CA, Dannhauer M, Palmer H, Hilbig S, et al. Structural controllability predicts functional patterns and brain stimulation benefits associated with working memory. Journal of Neuroscience. 2020;40(35):6770–6778.

9. Muldoon SF, Pasqualetti F, Gu S, Cieslak M, Grafton ST, Vettel JM, et al. Stimulation-based control of dynamic brain networks. PLoS computational biology. 2016;12(9):e1005076.

10. Stiso J, Khambhati AN, Menara T, Kahn AE, Stein JM, Das SR, et al. White matter network architecture guides direct electrical stimulation through optimal state transitions. Cell reports. 2019;28(10):2554–2566.

11. Pasqualetti F, Zampieri S, Bullo F. Controllability metrics, limitations and algorithms for complex networks. IEEE Transactions on Control of Network Systems. 2014;1(1):40–52.

12. Kenett YN, Medaglia JD, Beaty RE, Chen Q, Betzel RF, Thompson-Schill SL, et al. Driving the brain towards creativity and intelligence: A network control theory analysis. Neuropsychologia. 2018;118:79–90.

13. Tang E, Giusti C, Baum GL, Gu S, Pollock E, Kahn AE, et al. Developmental increases in white matter network controllability support a growing diversity of brain dynamics. Nature communications. 2017;8(1):1–16.

14. Cornblath EJ, Tang E, Baum GL, Moore TM, Adebimpe A, Roalf DR, et al. Sex differences in network controllability as a predictor of executive function in youth. NeuroImage. 2019;188:122–134.

15. Lee WH, Rodrigue A, Glahn DC, Bassett DS, Frangou S. Heritability and cognitive relevance of structural brain controllability. Cerebral Cortex. 2020;30(5):3044–3054.

16. Hahn T, Jamalabadi H, Emden D, Goltermann J, Ernsting J, Winter NR, et al. A Network Control Theory Approach to Longitudinal Symptom Dynamics in Major Depressive Disorder. arXiv preprint arXiv:210710178. 2021;.

17. Cai W, Ryali S, Pasumarthy R, Talasila V, Menon V. Dynamic causal brain circuits during working memory and their functional controllability. Nature Communications. 2021;12(1):1–16.

18. Fang F, Gao Y, Schulz PE, Selvaraj S, Zhang Y. Brain controllability distinctiveness between depression and cognitive impairment. Journal of Affective Disorders. 2021;294:847–856.

19. Parkes L, Moore TM, Calkins ME, Cieslak M, Roalf DR, Wolf DH, et al. Network controllability in transmodal cortex predicts psychosis spectrum symptoms. Biological Psychiatry. 2021;89(9):S370–S371.

20. Zöller D, Sandini C, Schaer M, Eliez S, Bassett DS, Van De Ville D. Structural control energy of resting-state functional brain states reveals less cost-effective brain dynamics in psychosis vulnerability. Human brain mapping. 2021;42(7):2181–2200.

21. Sun J, Motter AE. Controllability Transition and Nonlocality in Network Control. Physical Review Letters. 2013;110(20). doi:10.1103/physrevlett.110.208701.

22. Momi D, Neri F, Coiro G, Smeralda C, Veniero D, Sprugnoli G, et al. Cognitive enhancement via network-targeted cortico-cortical associative brain stimulation. Cerebral Cortex. 2020;30(3):1516–1527.

23. Gilson M, Moreno-Bote R, Ponce-Alvarez A, Ritter P, Deco G. Estimation of directed effective connectivity from fMRI functional connectivity hints at asymmetries of cortical connectome. PLoS computational biology. 2016;12(3):e1004762.

24. Razi A, Kahan J, Rees G, Friston KJ. Construct validation of a DCM for resting state fMRI. Neuroimage. 2015;106:1–14.

25. Frässle S, Lomakina EI, Kasper L, Manjaly ZM, Leff A, Pruessmann KP, et al. A generative model of whole-brain effective connectivity. Neuroimage. 2018;179:505–529.

26. Prando G, Zorzi M, Bertoldo A, Corbetta M, Zorzi M, Chiuso A. Sparse DCM for whole-brain effective connectivity from resting-state fMRI data. NeuroImage. 2020;208:116367. doi:10.1016/j.neuroimage.2019.116367.

27. Gao J, Liu YY, D'Souza RM, Barabási AL. Target control of complex networks. Nature Communications. 2014;5(1). doi:10.1038/ncomms6415.

28. Babayan A, Erbey M, Kumral D, Reinelt JD, Reiter AM, Röbbig J, et al. A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults. Scientific data. 2019;6(1):1–21.

29. Suweis S, Tu C, Rocha RP, Zampieri S, Zorzi M, Corbetta M. Brain controllability: not a slam dunk yet. Neuroimage. 2019;200:552–555.

30. Moretto M, Silvestri E, Zangrossi A, Corbetta M, Bertoldo A. Unveiling whole-brain dynamics in normal aging through Hidden Markov Models. Human Brain Mapping. 2022;43(3):1129–1144.

31. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. IEEE transactions on medical imaging. 2010;29(6):1310–1320.

32. Doshi J, Erus G, Ou Y, Gaonkar B, Davatzikos C. Multi-atlas skull-stripping. Academic radiology. 2013;20(12):1566–1576.

33. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage. 2011;54(3):2033–2044.

34. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL, et al. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage. 2011;54(1):313–327.

35. Andersson JL, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage. 2003;20(2):870–888.

36. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage. 2002;17(2):825–841.

37. Greve DN, Fischl B. Accurate and robust brain image alignment using boundary-based registration. Neuroimage. 2009;48(1):63–72.

38. Silvestri E, Moretto M, Facchini S, Castellaro M, Anglani M, Monai E, et al. Widespread cortical functional disconnection in gliomas: an individual network mapping approach. Brain communications. 2022;4(2):fcac082.

39. Damaraju E, Allen EA, Belger A, Ford JM, McEwen S, Mathalon D, et al. Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. NeuroImage: Clinical. 2014;5:298–308.

40. Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud G, Sexton CE, et al. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. Neuroimage. 2014;95:232–247.

41. Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo XN, Holmes AJ, et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cerebral cortex. 2018;28(9):3095–3114.

42. Rolls ET, Huang CC, Lin CP, Feng J, Joliot M. Automated anatomical labelling atlas 3. Neuroimage. 2020;206:116189.

43. Ryali S, Chen T, Padmanabhan A, Cai W, Menon V. Development and validation of consensus clustering-based framework for brain segmentation using resting fMRI. Journal of neuroscience methods. 2015;240:128–140.

44. Friston KJ, Harrison L, Penny W. Dynamic causal modelling. Neuroimage. 2003;19(4):1273–1302.

45. Sontag ED. Mathematical control theory: deterministic finite dimensional systems. vol. 6. Springer Science & Business Media; 2013.

46. Yuan Z, Zhao C, Di Z, Wang WX, Lai YC. Exact controllability of complex networks. Nature Communications. 2013;4(1). doi:10.1038/ncomms3447.

47. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems. 1998;30(1-7):107–117.

48. Lindmark G, Altafini C. Minimum energy control for complex networks. Scientific Reports. 2018;8(1). doi:10.1038/s41598-018-21398-7.

49. Lindmark G, Altafini C. Combining centrality measures for control energy reduction in network controllability problems. In: 2019 18th European Control Conference (ECC); 2019. p. 1518–1523.

50. Van Steen M. Graph theory and complex networks. An introduction. 2010;144.

51. Corder GW, Foreman DI. Nonparametric statistics: A step-by-step approach. John Wiley & Sons; 2014.

52. Lin S. Rank aggregation methods. Wiley Interdisciplinary Reviews: Computational Statistics. 2010;2(5):555–570.

53. Cowan NJ, Chastain EJ, Vilhena DA, Freudenberg JS, Bergstrom CT. Nodal dynamics, not degree distributions, determine the structural controllability of complex networks. PloS one. 2012;7(6):e38398.

54. Mesulam MM. Large-scale neurocognitive networks and distributed processing for attention, language, and memory. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society. 1990;28(5):597–613.

55. Damoiseaux JS, Rombouts S, Barkhof F, Scheltens P, Stam CJ, Smith SM, et al. Consistent resting-state networks across healthy subjects. Proceedings of the national academy of sciences. 2006;103(37):13848–13853.

56. Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proceedings of the National Academy of Sciences. 2005;102(27):9673–9678.

57. Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, et al. Functional network organization of the human brain. Neuron. 2011;72(4):665–678.

58. Greicius M. Resting-state functional connectivity in neuropsychiatric disorders. Current opinion in neurology. 2008;21(4):424–430.

59. Fox MD, Greicius M. Clinical applications of resting state functional connectivity. Frontiers in systems neuroscience. 2010; p. 19.

60. Benozzo D, Baron G, Coletta L, Chiuso A, Gozzi A, Bertoldo A. Macroscale coupling between structural and effective connectivity in the mouse brain. bioRxiv. 2023;doi:10.1101/2023.02.22.529400.

61. Li Q, Yao L, You W, Liu J, Deng S, Li B, et al. Controllability of Functional Brain Networks and Its Clinical Significance in First-Episode Schizophrenia. Schizophrenia Bulletin. 2022;.

62. Fang F, Godlewska B, Cho RY, Savitz SI, Selvaraj S, Zhang Y. Effects of escitalopram therapy on functional brain controllability in major depressive disorder. Journal of Affective Disorders. 2022;310:68–74.

63. Momi D, Ozdemir RA, Tadayon E, Boucher P, Di Domenico A, Fasolo M, et al. Perturbation of resting-state network nodes preferentially propagates to structurally rather than functionally connected regions. Scientific reports. 2021;11(1):12458.

64. Chiappini E, Silvanto J, Hibbard PB, Avenanti A, Romei V. Strengthening functionally specific neural pathways with transcranial brain stimulation. Current Biology. 2018;28(13):R735–R736.

65. Koponen LM, Nieminen JO, Ilmoniemi RJ. Multi-locus transcranial magnetic stimulation—theory and implementation. Brain Stimulation. 2018;11(4):849–855.

66. Yan G, Vértes PE, Towlson EK, Chew YL, Walker DS, Schafer WR, et al. Network control principles predict neuron function in the Caenorhabditis elegans connectome. Nature. 2017;550(7677):519–523.

67. Tahmassebi A, Meyer-Baese U, Meyer-Baese A. Structural target controllability of brain networks in dementia. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2021. p. 3978–3981.

68. Bassignana G, Lacidogna G, Bartolomeo P, Colliot O, De Vico Fallani F. The impact of aging on human brain network target controllability. Brain Structure and Function. 2022; p. 1–15.

69. Menardi A, Ozdemir RA, Momi D, Tadayon E, Boucher P, Vallesi A, et al. Effect of group-based vs individualized stimulation site selection on reliability of network-targeted TMS. NeuroImage. 2022;264:119714.

70. Menardi A, Momi D, Vallesi A, Barabási AL, Towlson EK, Santarnecchi E. Maximizing brain networks engagement via individualized connectome-wide target search. Brain Stimulation. 2022;15(6):1418–1431.

71. Hillebrand A, Tewarie P, Van Dellen E, Yu M, Carbo EW, Douw L, et al. Direction of information flow in large-scale resting-state networks is frequency-dependent. Proceedings of the National Academy of Sciences. 2016;113(14):3867–3872.

72. Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, et al. Dissociable intrinsic connectivity networks for salience processing and executive control. Journal of Neuroscience. 2007;27(9):2349–2356.

73. Cole MW, Schneider W. The cognitive control network: Integrated cortical regions with dissociable functions. Neuroimage. 2007;37(1):343–360.

74. Brosnan MB, Wiegand I. The dorsolateral prefrontal cortex, a dynamic cortical area to enhance top-down attentional control. Journal of Neuroscience. 2017;37(13):3445–3446.

75. Menon V, D'Esposito M. The role of PFC networks in cognitive control and executive function. Neuropsychopharmacology. 2022;47(1):90–103.

76. Badre D, Wagner AD. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia. 2007;45(13):2883–2901.

77. Badre D. Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. Trends in cognitive sciences. 2008;12(5):193–200.

78. Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. Nature reviews neuroscience. 2002;3(3):201–215.

79. Sohn H, Meirhaeghe N, Rajalingham R, Jazayeri M. A network perspective on sensorimotor learning. Trends in Neurosciences. 2021;44(3):170–181.

80. Bassett DS, Yang M, Wymbs NF, Grafton ST. Learning-induced autonomy of sensorimotor systems. Nature neuroscience. 2015;18(5):744–751.

81. Manto M, Bower JM, Conforto AB, Delgado-García JM, Da Guarda SNF, Gerwig M, et al. Consensus paper: roles of the cerebellum in motor control—the diversity of ideas on cerebellar involvement in movement. The Cerebellum. 2012;11:457–487.

82. Cox J, Witten IB. Striatal circuits for reward learning and decision-making. Nature Reviews Neuroscience. 2019;20(8):482–494.

83. Huntenburg JM, Bazin PL, Margulies DS. Large-scale gradients in human cortical organization. Trends in cognitive sciences. 2018;22(1):21–31.

84. Smallwood J, Bernhardt BC, Leech R, Bzdok D, Jefferies E, Margulies DS. The default mode network in cognition: a topographical perspective. Nature reviews neuroscience. 2021;22(8):503–513.

85. Avena-Koenigsberger A, Misic B, Sporns O. Communication dynamics in complex brain networks. Nature reviews neuroscience. 2018;19(1):17–33.

86. Jamalabadi H, Hahn T, Winter NR, Nozari E, Ernsting J, Meinert S, et al. Interrelated effects of age and parenthood on whole-brain controllability: protective effects of parenthood in mothers. bioRxiv. 2022; p. 2022–07.

87. Kamiya S, Kawakita G, Sasai S, Kitazono J, Oizumi M. Optimal control costs of brain state transitions in linear stochastic systems. Journal of Neuroscience. 2023;43(2):270–281.

88. Mantovani A, Neri F, D'Urso G, Mencarelli L, Tatti E, Momi D, et al. Functional connectivity changes and symptoms improvement after personalized, double-daily dosing, repetitive transcranial magnetic stimulation in obsessive-compulsive disorder: a pilot study. Journal of Psychiatric Research. 2021;136:560–570.

89. Drew PJ, Mateo C, Turner KL, Yu X, Kleinfeld D. Ultra-slow oscillations in fMRI and resting-state connectivity: neuronal and vascular contributions and technical confounds. Neuron. 2020;107(5):782–804.

90. Momi D, Ozdemir RA, Tadayon E, Boucher P, Shafi MM, Pascual-Leone A, et al. Network-level macroscale structural connectivity predicts propagation of transcranial magnetic stimulation. Neuroimage. 2021;229:117698.

91. Yan G, Ren J, Lai YC, Lai CH, Li B. Controlling Complex Networks: How Much Energy Is Needed? Phys Rev Lett. 2012;108:218703. doi:10.1103/PhysRevLett.108.218703.

92. Yan G, Tsekenis G, Barzel B, Slotine JJ, Liu YY, Barabási AL. Spectrum of controlling and observing complex networks. Nature Physics. 2015;11(9):779–786. doi:10.1038/nphys3422.

93. Duan G, Li A, Meng T, Wang L. Energy cost for target control of complex networks. Advances in Complex Systems. 2019;22(07n08):1950022.

Supporting information

Linear controllability A continuous linear time-invariant(LTI) system has a state equation of the form:

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t) \tag{17}$$

Here, $\mathbf{x}(t)$ is a $n \times 1$ state vector of the form $(x_1(t), ..., x_n(t))^T$, n being the number of nodes in the system. A is an $n \times n$ matrix describing the interaction strength between the system components. The system is stable if and only if all the eigenvalues of A have negative real part. B is an $n \times r$ input matrix ($r \leq n$) which identifies the nodes controlled by an outside controller, or *driver nodes*, with $B_{ij} = 1$ if control input $u_j(t)$ is imposed on node i; $\mathbf{u}(t)$ is a time-dependent $r \times 1$ input vector of the form $\mathbf{u}(t) = (u_1(t), ..., u_r(t))^T$ with r external inputs.

The system defined by continuous LTI Eq. 8 is said to be *controllable* if for a suitable choice of input signal vector $\mathbf{u}(t), t \in [0, t_f]$, it can be driven from any initial state $\mathbf{x}_0 = x(0)$ to any final state $\mathbf{x}_f = \mathbf{x}(t_f)$, where t_f is a finite time. Kalman's controllability rank condition states that a system is controllable if and only if the *controllability matrix*

$$\mathcal{C} = [B, AB, A^2B, ..., A^{n-1}B] \tag{18}$$

which is of dimension $N \times (N \cdot r)$, has full rank [45]:

$$rank(\mathcal{C}) = n \tag{19}$$

An equivalent condition can be formulated in terms of the *controllability Gramian* W, implicitly defined by the (continuous) Lyapunov equation

$$AW + WA^T = -BB^T \tag{20}$$

Eq. (20) has unique solution

$$W = \int_0^\infty e^{At} BB^T e^{A^T t} \tag{21}$$

Then the system is controllable if and only if W is positive definite,

$$W > 0 \tag{22}$$

Here, following standard notation, $W > 0$ means $\min(eig(W)) > 0$), i.e., the minimum eigenvalue of W is strictly positive. The conditions (19) and (22) are algebraically equivalent. The two conditions can be numerically verified by computing the minimum singular values of $\mathcal{C}$ and W, respectively. Note that, due to numerical inaccuracies, it is impossible to assess whether a singular value is exactly 0. Therefore, following common practice [46], we consider an eigenvalue to be 0 whenever is it below a very low numerical threshold $\epsilon = 10^{-14}$ (note that fixing a threshold can lead to seeming inconsistencies between the two conditions whenever the minimum eigenvalues are small) [6,29].

The *control energy* is defined as the (integrated) amplitude of the control signal used to steer the system from a given initial state $\mathbf{x}_0$ to a given final state $\mathbf{x}_f$,

$$E(\mathbf{u}) = \int_0^{t_f} dt ||\mathbf{u}(t)||^2 \tag{23}$$

Among infinitely many solutions for $\mathbf{u}$ which can drive the system $\mathbf{x}_0$ to final state $\mathbf{x}_f$ in time t_f, the optimal control input

$$\mathbf{u}^*(\mathbf{x}_0, \mathbf{x}_f, t_f) = B^T e^{A^T(t_f - t)} W(t_f)^{-1} (\mathbf{x}_f - e^{At_f}\mathbf{x}_0)$$

minimizes the control energy [91], i.e.,

$$\begin{aligned} min_{\{\mathbf{u},\ x(t_f)=x_f\}} E(\mathbf{u}) = E(\mathbf{u}^*(\mathbf{x}_0, \mathbf{x}_f, t_f)) = \\ = (\mathbf{x}_f - e^{At_f}\mathbf{x}_0)^T W(t_f)^{-1} (\mathbf{x}_f - e^{At_f}\mathbf{x}_0) = (\mathbf{x}_f)^T W(t_f)^{-1} \mathbf{x}_f \end{aligned} \tag{24}$$

where $W(t_f) = \int_0^{t_f} dt e^{At} BB^T e^{A^T t}$ and we assumed that the final state is normalized, $||\mathbf{x}_f||_2^2 = 1$. The energy cost can be thus bounded as follows:

$$\frac{1}{\lambda_{max}(W(t_f))} \equiv E_{min}(t_f) \leq E(\mathbf{u}^*(\mathbf{x}_0, \mathbf{x}_f, t_f)) \leq E_{max}(t_f) \equiv \frac{1}{\lambda_{min}(W(t_f))} \tag{25}$$

Since for a stable system the real parts of the eigenvalues of A are negative, the optimal control energy $E(\mathbf{u}^*(\mathbf{x}_0, \mathbf{x}_f, t_f))$ quickly decays to a (nonzero) asymptotic value $E^*(\mathbf{x}_0, \mathbf{x}_f) \equiv \lim_{t_f \to \infty} E(\mathbf{u}^*(\mathbf{x}_0, \mathbf{x}_f, t_f))$. In this limit, $W(t_f)$ coincides with the controllability Gramian, $\lim_{t_f \to \infty} W(t_f) = W$ [92]. Hence, the bounds on the energy given by Eq. 25 can be expressed as

$$\frac{1}{\lambda_{max}(W)} \equiv E_{min} \leq E^*(\mathbf{x}_0, \mathbf{x}_f) \leq E_{max} \equiv \frac{1}{\lambda_{min}(W)} \tag{26}$$

The typical metric to assess the difficulty of steering the system from one state to another is given by the upper bound of Eq. 11. This upper bound gives us control energy

cost required to steer the brain system to worst possible eigen direction of controllability gramian W and henceforth we consider this quantity as control energy.

$$\mathcal{E} = 1/\lambda_{min}(W) \tag{27}$$

where the λs are simply the eigenvalues of W.

In *target control*, one aims to control only a selected subset of target nodes [27]. The framework is the same, with the only difference that we focus on the final state of a subset of nodes $\mathbf{y}$:

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t) \tag{28}$$

$$\mathbf{y} = C\mathbf{x} \tag{29}$$

where $\mathbf{y} \in \mathbb{R}^S$ is the output vector describing the activity of the the target nodes we want to control. Given a network with n nodes, we can define a target node set $\mathcal{T} = \{\tau_1, \tau_2, \tau_3 ... \tau_S\}$ of size $S = |\mathcal{T}|$, where τ_i can be any node from $\{1, 2, 3..., n\}$. The output matrix C is defined as: $C = [\mathbb{I}^T_{\tau_1}, \mathbb{I}^T_{\tau_2}, ..., \mathbb{I}^T_{\tau_S}]^T \in \mathbb{R}^{S\times N}$ where $\mathbb{I}_{\tau_i}$ is τ_i-th row of the identity matrix $\mathbb{I}$.
The definition of target controllability follows from that of standard (Kalman) controllability, where the system is now defined by the triple (A, B, C) instead of the pair (A, B) [27]. The system (A, B, C) is said to be *target controllable* with respect to target node set $\mathcal{C}$ if there exists a time-dependent input vector $\mathbf{u}$ which can drive the state of the target nodes to any desired final state in finite time. It can be shown that the system is target controllable if and only if

$$rank[\mathcal{C}] = S \tag{30}$$

where $\mathcal{C} \equiv [CB, CAB, CA^2B, ..., CA^{n-1}B]$ is the target controllability matrix. The Gramian for target controllability is given by [93]:

$$W_C = CWC^T \tag{31}$$

where W is the standard Gramian, Eq. (9) for system (A, B). Analogously to the case of full controllability, the optimal control input [27]

$$\mathbf{u}^*_t = B^T e^{A^T(t_f - t)} C^T \left(CW(t_f)C^T\right)^{-1} \left(\mathbf{y}_f - CA^{t_f}\mathbf{x}_0\right)$$

which minimizes the control energy and drives the subsystem from initial state $\mathbf{y}_0 = C\mathbf{x}_0$ to final output state $\mathbf{y}_f$ is substituted into Eq. (10) and we obtain the optimal target control energy:

$$E(\mathbf{u}^*(\mathbf{y}_f, \mathbf{y}_0, t_f) = \mathbf{y}^T_f (CW(t_f)C^T)^{-1}\mathbf{y}_f \tag{32}$$

where we have assumed that the initial state is $\mathbf{x}_0 = \mathbf{0}$. The energy can be bounded as follows:

$$\frac{1}{\lambda_{max}(CW(t_f)C^T)} \equiv E_{min}(t_f) \leq E(\mathbf{u}^*(\mathbf{y}_f, \mathbf{y}_0, t_f) \leq E_{max}(t_f) \equiv \frac{1}{\lambda_{min}(CW_{t_f}C^T)}$$

Asymptotically, as $t_f \to \infty$, we can rewrite the above bounds as:

$$\frac{1}{\lambda_{max}(W_C)} \leq E^*(\mathbf{y}_f, \mathbf{y}_0) \leq \frac{1}{\lambda_{min}(W_C)} \tag{33}$$

The upper bound gives the energy required to steer the subsystem along the worst possible direction:

$$\mathcal{E}^{target} = 1/\lambda_{min}(W_C) \tag{34}$$

If a single driver node i is used, and the target is in turn a single node j, the expression of the control energy significantly simplifies. We have $B = b^T_i$ and $C = b_j$. Thus

$$\mathcal{E}_{i\to j} \equiv E^{target}_{min} = (W^{(i)}_{jj})^{-1} = \frac{1}{\int_0^\infty dt[e^{At}]^2_{ij}} \tag{35}$$

Optimal nodes to control RSNs according to rank aggregation For each target RSN, we have ranked nodes according to the centrality measure yielding the lowest average energy, when driver nodes are selected among the nodes with largest centrality. After rankings were estimated separately for each subject, we have used rank aggregation (average rank) to obtain a group ranking. Here, we focus on the 10 top-ranking nodes for each target RSN (Fig. 11). The top nodes to control the CON are located in ventrolateral prefrontal cortex (DMN and VAN), precuneus and right parietal (DMN), frontal eye fields (DAN), right dorsal frontal (DMN). The top nodes to control the DMN are located in left cerebellum, right parietal (DAN), left parietal (CON), frontal eye fields (DAN), ventrolateral PFC (VAN), dorsomedial prfrontal cortex (DMN/VAN). The top nodes to control the DAN are located in right parietal (DMN), right precentral (SMN), left striatum, ventrolateral PFC (VAN/DMN), dorsomedial frontal (VAN), right dorsal frontal (DMN). The top nodes to control the limbic nodes are located in right cerebellum, right temporal cortex (VIS), right precentral (SMN), right frontal eye field (DAN), right dorsal frontal (DMN), ventrolateral prefrontal cortex (DMN and VAN), left dorsolateral preforntal cortex (VAN). The top nodes to control the VAN are located in precuneus (DMN), right parietal (DMN), medial prefrontal cortex (DMN), ventrolateral prefrontal cortex (DMN), frontal eye fields (DAN), right dorsal frontal (DMN and CON). The top nodes to control the SMN are located in striatum, frontal eye fields (DAN), ventrolateral prefrontal cortex (VAN and DMN), dorsomedial frontal (CON), right precentral (DAN). The top nodes to control subcortical regions are located in ventral temporal cortex (limbic), temporal (VIS), ventrolateral prefrontal cortex (DMN and VAN), orbitofrontal cortex (limbic), frontal eye fields (DAN). The top nodes do control the VIS are located in left cerebellum, precuneus (DMN), striatum, right ventral temporal (limbic), ventrolateral prefrontal cortex (DMN/VAN), right frontal eye field (DAN), medial parietal (CON). In general, we observe a large prevalence of anterior nodes, in particular from ventrolateral prefrontal cortex.

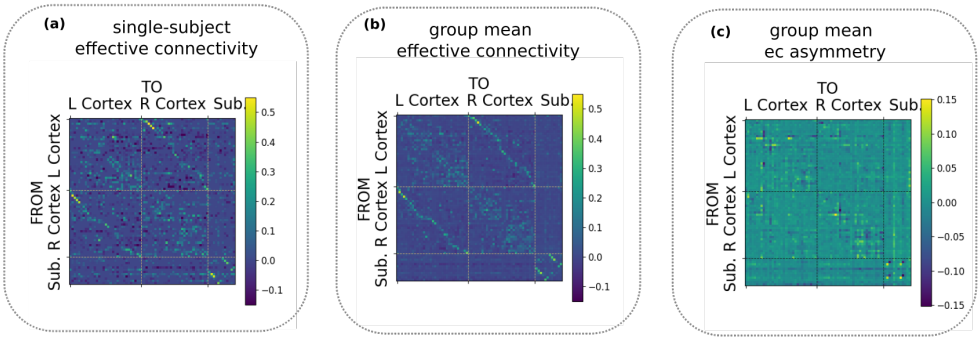


Fig 7. *Examples of effective connectivity matrices.*. The 74 brain areas are divided in left cortical, right cortical and subcortical areas. The figure in panel (c) represents the asymmetry in A, defined as $\delta A = A - A^T$.

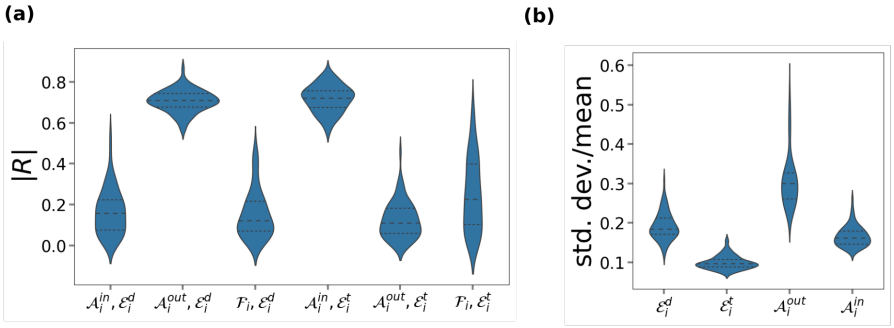
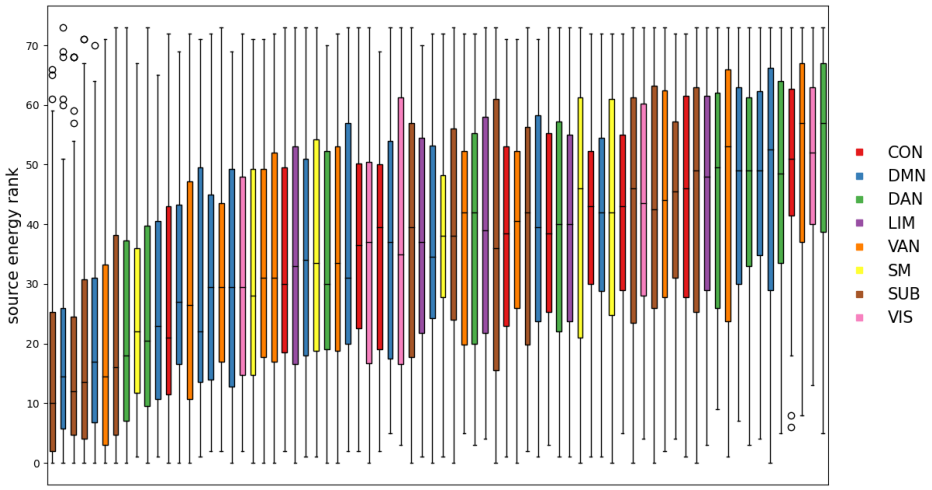
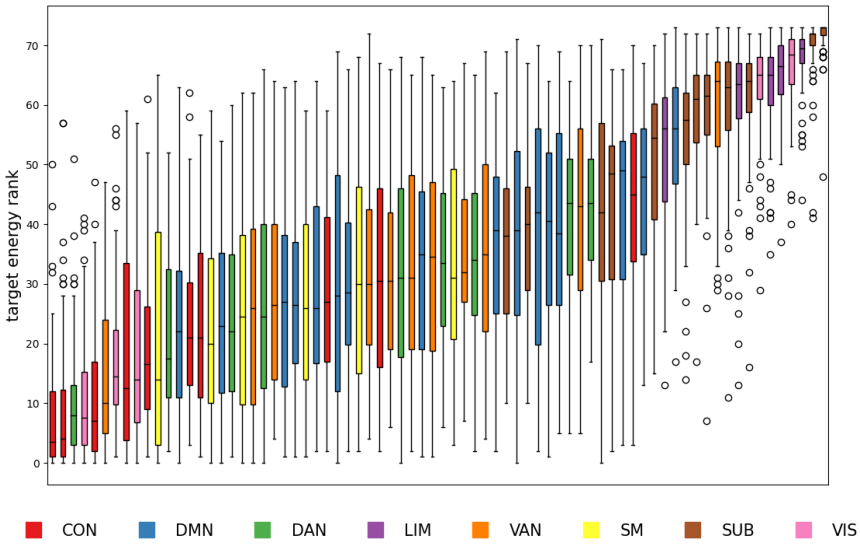


Fig 8. *Relation of driver and target control energy and effective connectivity.* **(a)** For each subject, we computed the Pearson correlation R between the node driver and target centrality $\mathcal{E}_i^d$, $\mathcal{E}_i^t$ (average energy to control other nodes from node i vs average energy to control node i from other nodes) and and the in-strength A_i^{in} and out-strength A_i^{out} of effective connections, as well as the strength of functional connections $\mathcal{F}_i$. We show the distribution of $|R|$ over subjects. **(b)** For each node, we computed the coefficient of variation (s.d./mean) over subjects of $\mathcal{E}_i^d$,$\mathcal{E}_i^t$, A_i^{out} A_i^{in}. We show the distribution of the coefficient of variation over nodes.



(a) node ranks based on $\mathcal{E}_i^d$



(b) node ranks based on $\mathcal{E}_i^t$

Fig 9. *Node ranks based on driver/target control energy.* **(a)** For each subject, we ranked all nodes based on the value of driver centrality $\mathcal{E}_i^d$. We show the rank distribution for all nodes, with nodes ordered according to the average rank, from lowest to highest. Nodes are colored according to the resting state network they belong to. **(b)** Same as (a), but ranks are based on target centrality $\mathcal{E}_i^t$.

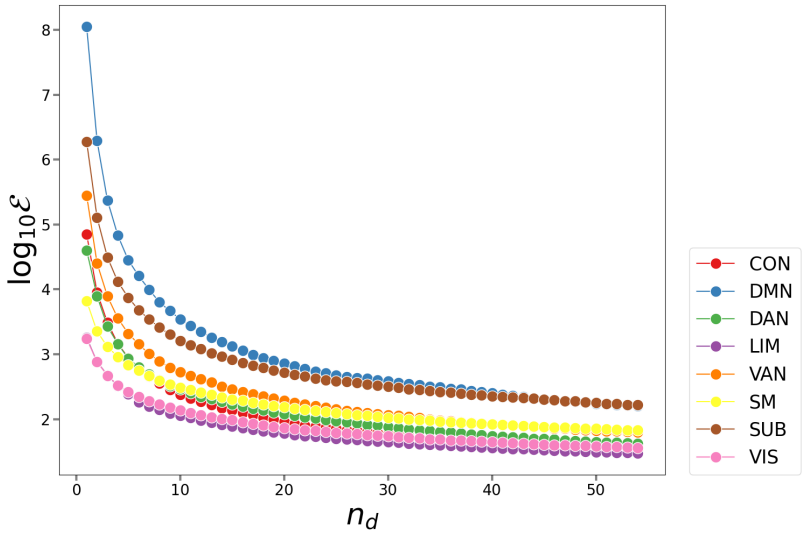


Fig 10. *Dependence of the RSN control energy on the number of driver nodes.* Energy to control RSN with a varying number of driver nodes, selecting driver nodes according to the driver centrality

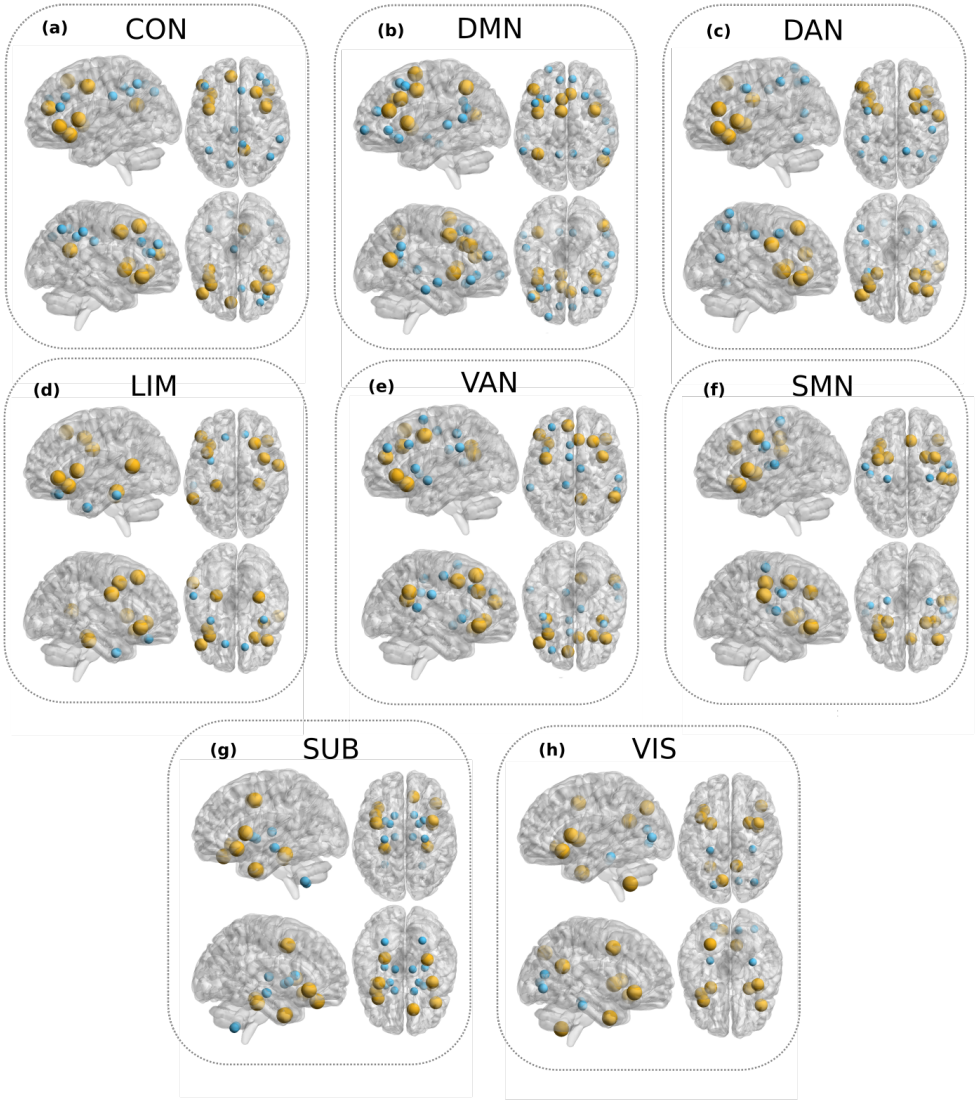


Fig 11. *Optimal driver nodes for RSN*). For each RSN we show the corresponding target nodes(small blue markers) and top 10 aggregate driver nodes(orange markers) are shown