

SCIENTIFIC REPORTS



OPEN

Optimally controlling the human connectome: the role of network topology

Richard F. Betzel¹, Shi Gu¹, John D. Medaglia^{1,2}, Fabio Pasqualetti³ & Danielle S. Bassett^{1,4}

Received: 20 April 2016

Accepted: 07 July 2016

Published: 29 July 2016

To meet ongoing cognitive demands, the human brain must seamlessly transition from one brain state to another, in the process drawing on different cognitive systems. How does the brain's network of anatomical connections help facilitate such transitions? Which features of this network contribute to making one transition easy and another transition difficult? Here, we address these questions using network control theory. We calculate the optimal input signals to drive the brain to and from states dominated by different cognitive systems. The input signals allow us to assess the contributions made by different brain regions. We show that such contributions, which we measure as energy, are correlated with regions' weighted degrees. We also show that the network communicability, a measure of direct and indirect connectedness between brain regions, predicts the extent to which brain regions compensate when input to another region is suppressed. Finally, we identify optimal states in which the brain should start (and finish) in order to minimize transition energy. We show that the optimal target states display high activity in hub regions, implicating the brain's rich club. Furthermore, when rich club organization is destroyed, the energy cost associated with state transitions increases significantly, demonstrating that it is the richness of brain regions that makes them ideal targets.

One of the goals of modern biology is to understand how a system's form influences its function. Human brain networks manifest structure-function relationships, with converging evidence suggesting that the brain's network of white-matter fiber pathways (*structural connectivity*; SC) constrains the intrinsic functional interactions among brain regions at rest (*functional connectivity*; FC), thereby shaping the emergence of coherent spatiotemporal patterns of neural activity^{1–5}. The effect of these constraints is that over long periods of time (hours, days) resting FC largely recapitulates the underlying SC⁶ so that the strongest functional interactions are often mediated by direct anatomical projections⁷. Over shorter timescales, however, FC is more variable, decoupling from the underlying anatomy to engage specific cognitive systems both at rest^{8–10} and in order to meet ongoing cognitive demands^{11–13}.

How does the brain smoothly transition from the activation of one cognitive system to the activation of another? What are the anatomical and topological substrates that facilitate such transitions? One approach for addressing these and similar questions is to model the human brain as a dynamical system, treating brain regions as dynamic elements with time-dependent internal states. As the system evolves, each brain region's state is updated according to its own history and the states of its connected neighbors. Such models vary in their complexity and neurophysiological basis, ranging from biophysically plausible descriptions of interacting neuronal populations^{14–18} to abstract models based on oscillations, diffusion, and epidemic spreading^{19–22}.

In most applications, the question of how to control distributed brain dynamics is not explicitly considered. Here, control refers to the possibility of manipulating a dynamical system so that it evolves to follow a particular trajectory through its state space. We posit that the nature of brain state transitions can be meaningfully addressed with network control theory, which offers a mathematical framework for studying and, ultimately, controlling the evolution of dynamical systems on networks^{23,24}.

Most dynamical systems can be framed in a control perspective by introducing exogenous input to the system through a set of control sites (network nodes) in the form of time-varying signals. The effect of such inputs is to drive the system along a trajectory through its state space; different inputs, then, result in different trajectories^{24,25}.

¹Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, 19104, USA. ²Department of Psychology, University of Pennsylvania, Philadelphia, PA, 19104, USA. ³Department of Mechanical Engineering, University of California, Riverside, Riverside, CA, 92521, USA. ⁴Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, 19104, USA. Correspondence and requests for materials should be addressed to D.S.B. (email: dsb@seas.upenn.edu)

The effect of input on a system's trajectory depends upon (i) the system's dynamics, (ii) the composition of the control sites, and (iii) the configuration of the system's nodes and edges into a network (its topology)²⁴. Understanding how control occurs in the brain and how these factors contribute to enacting control on the brain is of critical importance, with clear clinical and engineering implications. For example, the efficacy of implantable neuromodulatory devices for suppressing Parkinsonian symptoms²⁶, seizure abatement in epilepsy²⁷, and other methods for manipulating brain activity, such as transcranial magnetic stimulation²⁸, depend on our ability to modify a network's function by introducing external electromagnetic signals.

In the current work we use network control theory to identify minimum energy input signals that cause the system to transition to and from specific brain states. The input signals are introduced through control sites—brain regions—and can be thought of in two complementary ways. One view is that control signals are issued directly from brain regions, themselves acting as local computational elements administering control over the network. Alternatively, control signals can be viewed as having extracranial provenance, originating from implanted electrodes or other neuromodulatory devices and thereby acting on specific brain regions. In either case, a region's local contribution can be modeled as an input of energy over time and interpreted as a measure of the amount of effort it puts forth during a control task^{25,29}.

We seek to better understand the role of brain network topology in determining a region's energy—what topological factors contribute to making transitions between different brain states more or less effortful? Previous investigations addressing similar questions have focused on how the underlying networks' statistical properties (e.g. the shape of the degree distribution) and global metrics (e.g. modularity, clustering, etc.) influence the minimum number of control sites necessary to render the network controllable^{24,30–33}. Such approaches, while illuminating, are limited. First, the focus on global network statistics makes it difficult to assess the contributions of individual nodes or edges. Second, the classification of networks as either “controllable” or “uncontrollable” overlooks finer gradations in the amount of energy required for control. Though some node-level metrics have been proposed³⁴, the precise roles of individual nodes and other topological features in facilitating control is not well understood. Previous applications of network control theory to brain networks investigated related questions, by studying *all possible state transitions* and assuming an infinite time horizon^{18,35}. Here, we focus on finite-time transitions between a limited set of accessible states, which we choose to correspond to previously-defined brain systems³⁶.

The remainder of this report is divided into two sections. In the first, we explore a set of progressively more difficult control tasks, demonstrating that a brain region's weighted degree (strength) is highly correlated with its control energy, suggesting that regions with many, strong connections contribute disproportionately more energy than regions with few or weak connections. We also study the effect of suppressing input to select sets of brain regions, which reveals a network of compensatory interactions among brain regions. We show that the degree to which one brain region compensates for the removal of another can be predicted with the network measure “communicability”, which measures the strength of direct and indirect pathways between network nodes.

The second section builds on the results from the first. Rather than use a pre-defined set of initial and target states, we objectively assign nodes to initial and target states so as to minimize the cost of the transitions among these states. We find that the optimal assignments implicate highly connected brain regions as ideal targets and weakly connected regions as ideal observers (i.e. they play no role in control). We also show that the energy associated with these optimal assignments is less than what would be expected given a degree-preserving random network model. Finally, we show that when the network's rich club³⁷ is disrupted the energy increases, further suggesting that the configuration of connections among hub regions supports low-energy transitions.

Mathematical model

We studied a dynamical system in which the brain's network of white matter fiber tracts among brain regions constrained the following linear time-invariant nodal dynamics^{35,38}:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}_{\mathcal{K}}\mathbf{u}_{\mathcal{K}}(t). \quad (1)$$

Here, $\mathbf{x}(t) \in \mathbb{R}^{n \times 1}$, is the state vector whose element, $x_i(t)$, represents the state (activity level) of brain region i at time t . The matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the symmetric and weighted adjacency matrix, whose element A_{ij} is the number of detected tracts between regions i and j normalized by the sum of their volumes. The input matrix, $\mathbf{B}_{\mathcal{K}}$, specifies the set of control points, $\mathcal{K} = \{k_1 \dots k_m\}$, such that:

$$\mathbf{B}_{\mathcal{K}} = [\mathbf{e}_{k_1}, \dots, \mathbf{e}_{k_m}], \quad (2)$$

where \mathbf{e}_{k_i} is the i th canonical column vector of dimension n . The time-varying input signals are denoted as $\mathbf{u}_{\mathcal{K}}(t) \in \mathbb{R}^{m \times 1}$ where $u_{k_i}(t)$ gives the input at control point k_i at time t (Fig. 1).

We were interested in the control task where the system transitions from some initial state, $\mathbf{x}_0 = \mathbf{x}(t=0)$, to some target state, $\mathbf{x}_T = \mathbf{x}(t=T)$. We solved this task using an optimal control framework, deriving the set of minimum-energy inputs, $\mathbf{u}_{\mathcal{K}}^*(t)$, for accomplishing this task (Supporting Information). Each control point's energy was defined as $E_{k_i} = \int_{t=0}^T \|u_{k_i}^*(t)\|^2 dt$ and the total energy was given by $E = \sum_{i=1}^m E_{k_i}$.

Rather than investigate all possible transitions, we considered a limited repertoire, focusing on a set of eight states based on systems previously identified in intrinsic functional connectivity studies³⁶ (Materials and Methods, Cognitive systems). We analyzed all system-to-system transitions (excluding self-transitions), resulting in 56 possible control tasks. Importantly, because our dynamical model is linear, any possible transition can be written as a linear combination of these transitions (though the resulting transition may not be optimal, in terms of minimum energy). Thus, our results are generally relevant to all transitions.

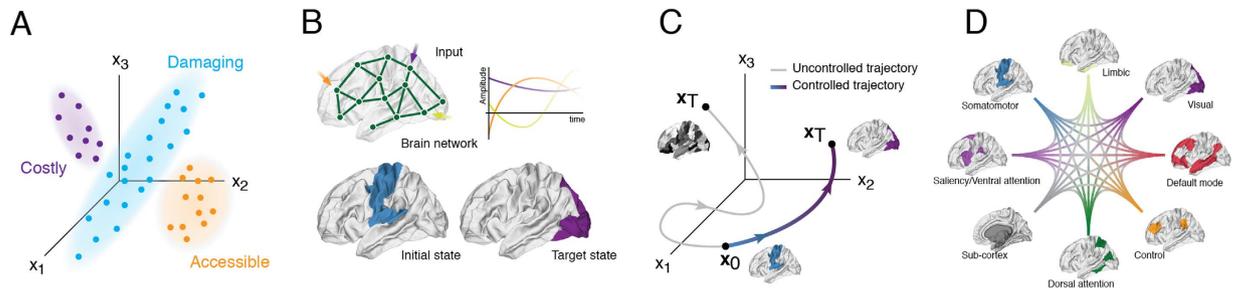


Figure 1. Schematic illustrating structural controllability framework. (A) The space of all possible states includes regions that require prohibitively large energy to access as well as damaging configurations. Our analyses are restricted to the accessible region of state space populated by viable configurations. (B) A set of time-varying inputs are injected into the system at different control points (nodes; brain regions). The aim is to drive the system from some particular initial state to a target state (e.g. from somatosensory to visual system). (C) Example trajectory through state space. Without external input (control signals) the system's passive dynamics leads to a state where random brain regions are more active than others; with input the system is driven into the desired target state. (D) Schematic illustration of all possible system-to-system transitions.

For a given control task we classified brain regions based on their states at times $t = 0$ and $t = T$:

- The *initial* class includes all regions active at $t = 0$ and inactive at $t = T$ (denoted as \mathbf{x}_0).
- The *target* class included all regions inactive at $t = 0$ but active at $t = T$ (denoted as \mathbf{x}_T).
- The *bulk* class included all regions inactive at both $t = 0$ and $t = T$.

In principal, there is a fourth class of regions active at both $t = 0$ and $t = T$. However, given the control tasks we consider here, this case never arises. In subsequent sections we will present our results within this classification scheme.

Results

Predicting energy from topology. One of the principal aims of this report was to determine what features of the brain's topology influence control energy. In this section we assessed the role of topology in determining the energy associated with each of the 56 control tasks. We began this investigation with a simple scenario in which all brain regions served as control points. We observed that $E_{target} > E_{initial} > E_{bulk}$ (one-way ANOVA comparing log-transformed energies across all control tasks and participants yielded minimum omnibus test-statistic of $F(2, 126) = 83.0$, and a maximum p -value of $p = 1.0 \times 10^{-23}$; pairwise t -tests for $E_{target} > E_{initial}$ and $E_{initial} > E_{bulk}$ yielded minimum t -statistic of $t = 8.31$ and maximum p -value of $p = 1.6 \times 10^{-9}$, Bonferroni adjusted) (Fig. 2A).

To explain this ranking of energies, we need to consider the system's dynamics under free evolution—the absence of input. In such a case, the system evolves as $\mathbf{x}(t) = e^{At}\mathbf{x}_0$, which we obtained by solving Eq. 1. The vector $\mathbf{v}(t) = \mathbf{x}_T - \mathbf{x}(t)$, then, specified the distance, $v_i(t)$, of each region from its target state at time t . Intuitively, as distance increased, more energy was required to drive the system towards its desired configuration (Figure S1). Moreover, distance followed a class-specific trajectory (Fig. 2B). At the control horizon, $t = 1$, bulk regions were nearest their target state, followed by initially active regions, followed by target regions, explaining why specific classes required more or less energy.

The previous analysis demonstrated that, foremost, regional control energy depended on class. Within each class, however, we found that much of the remaining variance could be accounted for by regions' weighted degrees (*strength*; $s_i = \sum_j A_{ij}$). Across participants and control tasks, the logarithm of control energy for regions assigned to initial and bulk classes was both positively correlated with the logarithm of strength (median(interquartile range) correlations of $r = 0.85(15)$ and $r = 0.79(15)$, respectively) while the opposite was true for target regions $r = -0.84(13)$) (Fig. 2C–E; Figure S2 for summary across all participants). Collectively, these results imply that predicting a brain region's energy contribution requires both topology and contextualizing a region's role – i.e. initial, target, bulk—in a given control task.

Simulated suppression and compensatory effects. In the previous section we investigated a scenario in which all brain regions served as control points, making it possible to control any region's state directly. In the current section we imagined a more difficult scenario in which the system performed the same control tasks but with specific subsets of brain regions excluded from the control set. These excluded regions, then, could only be manipulated via indirect input from their neighbors—they required help from the rest of the network to achieve their desired state. This framework, which we refer to as “simulated suppression”, is analogous to inhibitory neuromodulation, which can be achieved externally using transcranial magnetic stimulation (TMS), in which a brain region's local activity is suppressed but still receives inputs from surrounding areas^{39,40}. This notion of simulated suppression could also be achieved internally via competitive or inhibitory dynamics among neuronal populations.

Using simulated suppression we investigated how the suppression of input to specific regions changed the energies of the remaining, unsuppressed, regions. We interpreted such changes as a measure of compensation. In

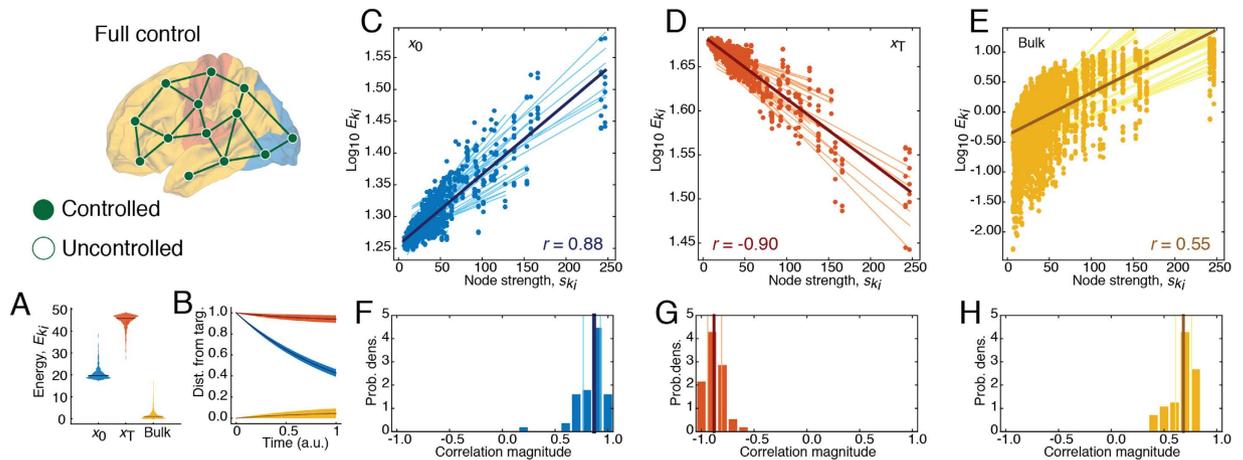


Figure 2. Energy depends on class and strength when all regions are controlled. (A) Violin plot of node-level control energies aggregated across all control tasks and divided into initial (x_0), target (x_T), and bulk classes. (B) Average distance of $x(t)$ from target state, x_T , under free evolution (i.e. no input signals) as a function of time and averaged within each class. (C–E) Scatterplots of node-level control energies, $\log_{10} E_{k_j}$, versus strengths, s_{k_j} , (total normalized streamline counts) across all control tasks. Light lines represent linear regression lines fit to specific control tasks; dark lines represent average regression line. (F–H) Distributions of Pearson's correlation coefficients for best-fit lines shown in panels (C–E).

this section, we explored a series of progressively more difficult control scenarios in which we suppressed both individual brain regions and entire classes of regions.

Suppression of individual brain regions. In this section we suppressed individual brain regions, one at a time, and repeated the same control tasks as the previous section. We expected that with fewer control points the total control energy, E , would increase. We found that this was largely the case, with the greatest percent changes in energy occurring among bulk regions (Fig. 3A).

The simulated suppression framework allowed us to calculate the extent to which brain regions engaged in compensatory relationships with one another, wherein the suppression of one region consistently resulted in increased energy of another region. We were also interested in determining to what extent these compensatory relationships could be predicted based on topological properties of the network. We hypothesized that the strength of compensatory relationships should depend upon the extent to which they were interconnected. We further reasoned that even indirectly-connected regions should be able to compensate for one another provided that they were linked by many multi-step paths. This intuition can be formalized as the *communicability* between two regions, a statistic that quantifies the strength of both direct and indirect pathways between node pairs⁴¹. Communicability can be thought of as the capacity for two regions to communicate with one another by pathways of all topological lengths (Fig. 3B) (Materials and Methods, Network communicability).

For each control task we calculated the percent change in energy of region k_j after suppressing region k_i , which we denote as $\% \Delta E_{k_i k_j}$ (Fig. 3C). When averaged across all control tasks we found excellent correspondence between this measure and weighted communicability, with a correlation of $r = 0.95(0.01)$ across participants (Fig. 3D). The same relationship persisted (albeit attenuated) when we examined specific control tasks and sub-divided regions according to their class: for initial $r = 0.74(0.08)$, for target $r = 0.80(0.06)$, and for bulk $r = 0.68(0.11)$ (Fig. 3E–G). These results suggest that both direct and indirect connections play important roles in compensating for suppressed brain regions (Figure S3 for summary across all participants).

Suppressing initial or target classes. We extended the simulated suppression framework by simultaneously suppressing entire classes of brain regions, resulting in considerably more difficult control tasks. In particular, we suppressed all regions assigned to either initial or target classes. As in the previous section, we found that bulk regions exhibited the greatest percent change in their energies (Fig. 4A,F). Also in agreement with the previous section, we found that the percent change in a region's energy was predicted by its total communicability to the suppressed class ($\sum_{k_j \in x_0} G_{k_i k_j}$ and $\sum_{k_j \in x_T} G_{k_i k_j}$) (Fig. 4B,C,G,H). When the target class was suppressed, the correlation of communicability and change in energy was $r = 0.76(0.24)$ and $r = 0.38(0.36)$ for the remaining initial and bulk regions, respectively. When the initial class was suppressed, the correlation of communicability and change in energy was $r = 0.72(0.30)$ and $r = 0.32(0.41)$ for the remaining target and bulk regions (Figure S4 for summary across all participants).

Suppressing initial and target classes. Finally, we explored the consequences of simultaneously suppressing both initial and target regions, leaving only the bulk as control points. In the previous sections if suppression was applied to initial or target nodes, the remaining bulk could perform the same control task as the missing nodes. Here, however the bulk was tasked with performing two duties: simultaneously turning on and off target and initial regions, respectively. The most energetic members of the bulk were those with high levels of

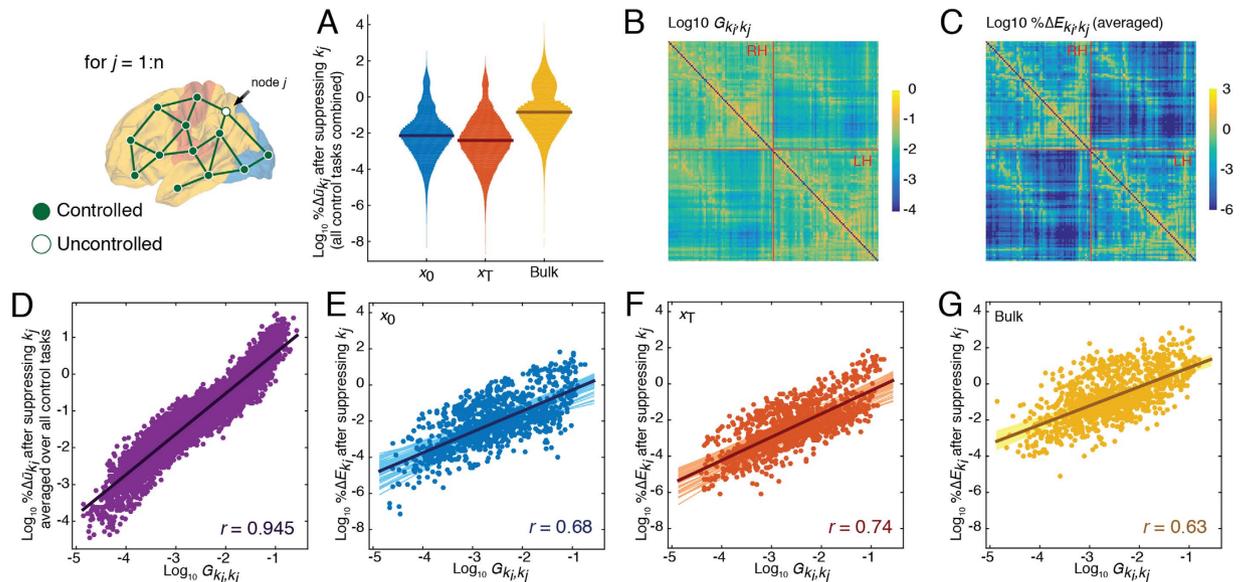


Figure 3. Communicability predicts compensation (or change in energy) when individual regions are suppressed. (A) Violin plot of percent change in regions' energies ($\% \Delta E_{k_j}$) after suppressing region k_j aggregated by class. (B) Compensation network in matrix form; the weights of the compensation network are defined as the percent change in energy of region k_i ($\% \Delta E_{k_i}$) after suppressing region k_j averaged across all control tasks. (C) Weighted communicability matrix, which measures the total strength of direct and indirect connections between two regions. (D) Scatterplot showing the correlation of communicability and the compensation network weights. (E–G) Percent change in the energy of region k_i after suppressing region k_j . Each class is shown separately – x_0 are active at time $t = 0$, x_T are active at time $t = T$, and bulk nodes are inactive at both $t = 0$ and $t = T$. Matrices in panels (B,C) are shown with logarithmic scaling.

communicability to either initial or target regions, thereby affording them the possibility of directly and indirectly controlling the states of those classes (Fig. 5A) (across participants, the correlation of E_{k_i} and $\sum_{k_j \in \{x_0 \cup x_T\}} G_{k_p k_j}$, the communicability of region k_i to both initial and target regions, was $r = 0.57(0.25)$). These results further suggest that bulk regions—those not actively changing their own state during a control task—nonetheless acted as compensators when other regions become compromised.

We also were able to predict bulk regions whose energy contributions increased by the greatest amount. As noted earlier, the vector $\mathbf{v}(t) = \mathbf{x}_T - e^{At} \mathbf{x}_0$ gives the Euclidean distance in nodes' states at time t with their respective target states. We hypothesized that the greater the distance a bulk region's neighbors were from their target states would be related to how much additional energy that region would have to contribute. To this end, we calculated $\theta \in \mathbb{R}^{m \times 1}$ whose element $\theta_{k_i} = \sum_j A_{k_i j} \|v_j\|$ gives the total distance of k_i 's neighbors from their target states weighted by the strength of k_i 's connection to those nodes. We found that the logarithm of θ_{k_i} and E_{k_i} were robustly negatively correlated across both participants and control tasks ($r = -0.33(0.18)$) (Fig. 5B) (Figure S5 for summary across all participants).

Collectively, the results reported in this section make two important points. First, bulk regions—those not changing their state from active to inactive (or *vice versa*)—exhibited the greatest increase in energy following the suppression of other regions, suggesting that these “bystanders” may play important compensatory roles in the control of brain dynamics. Secondly, in demonstrating that brain region's compensatory relationships are correlated with their communicability, we implicate indirect connections as important pathways through which compensatory relationships emerge.

Optimal class assignments. The human brain has evolved to exhibit many near-optimal features. For example, drives to reduce total wiring length^{42,43} and metabolic expenditure^{44,45} are thought to impose powerful constraints on the brain's organization, development, and growth. In addition to costs associated with architecture and energy, it has been suggested that the brain is capable of implementing computational algorithms for minimizing arbitrary cost functions⁴⁶. In this section, we ask the question: *if* the brain were capable of implementing an optimal control algorithm and wished to minimize total energy expenditure, which brain regions would it choose as its initial and target states? One approach for answering this question would be to enumerate all possible divisions of brain regions into the three aforementioned classes and then to apply optimal control to determine the total transition energy. The optimal class assignments, then, would be the division associated with the smallest transition energy. This approach, however, is computationally intractable due to the prohibitively large number of possible divisions. An attractive alternative is to re-state this question in terms of an optimization problem and to use some heuristic for estimating the optimal division. In this section we investigated such an approach. Specifically, we identified class assignments that minimized the objective function, $\mathcal{E} = \mathbf{v}^T \mathbf{v}$, where \mathbf{v} is

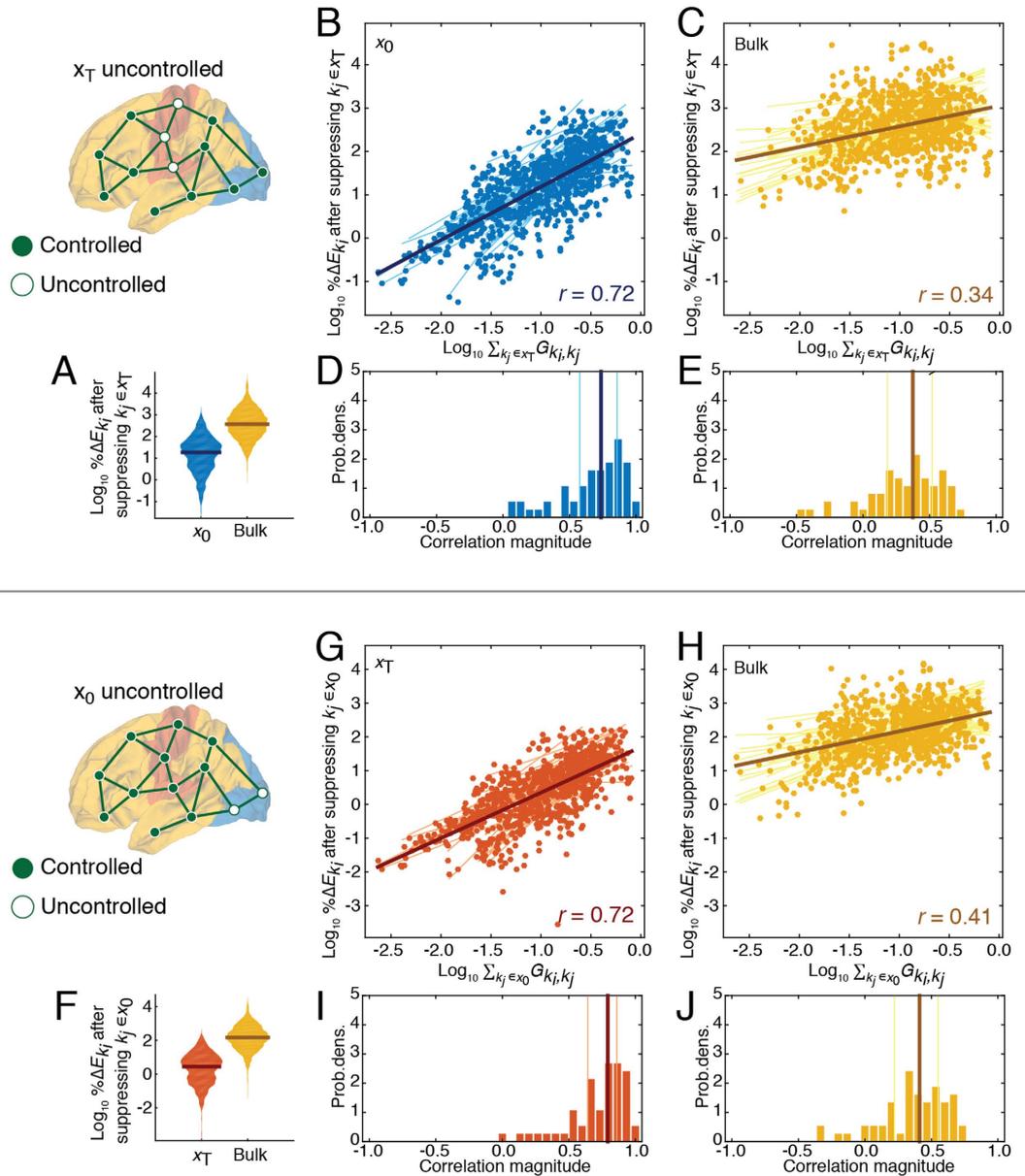


Figure 4. Summary of initial or target class suppression. (A) Violin plot showing percent changes in energy $\% \Delta E_{k_p, k_j}$ for the remaining nodes after excluding all target nodes from the control set. (B,C) Scatterplot of the percent change of region k_i 's energy after suppressing all target nodes against k_i 's communicability to all target nodes. (D,E) Distributions of Pearson's correlation coefficients between change in energy and communicability, \hat{r}_{G, u^*} , for initial and bulk nodes. Panels (F–J) recapitulate (A–E) but with initial nodes excluded from the control set rather than target nodes.

defined in the same way as before. We chose this particular objective function because it is highly correlated with control energy (Figure S1) and also for computational ease, as it can be calculated in a more straightforward manner than energy, which requires first deriving the optimal control signals.

We used a simulated annealing algorithm to minimize \mathcal{E} and generate estimates of the probability with which each brain region was assigned to initial, target, and bulk classes. Across participants we found that the probability of a region being assigned to either the target and bulk classes was highly correlated with its weighted degree ($r_{\text{target}} = 0.88(0.03)$ and $r_{\text{bulk}} = -0.73(0.07)$; maximum p -value of $p = 2.5 \times 10^{-16}$), while the nodes assigned to the initial class were not closely associated with weighted degree ($r = -0.02(0.09)$; minimum p -value of $p = 0.27$). These results suggest that highly and weakly connected regions are ideal targets and bystanders, respectively.

What aspects of the network's organization determines class assignments? One possibility is that the low-energy transitions facilitated by these assignments are merely a consequence of the number and weight of connections a brain region makes—i.e. they do not depend on the actual configuration of a network's connections. To test this hypothesis, we randomized connection placement while preserving the number of connections that each brain region makes (generating 100 random networks for each participant) and evaluated the objective

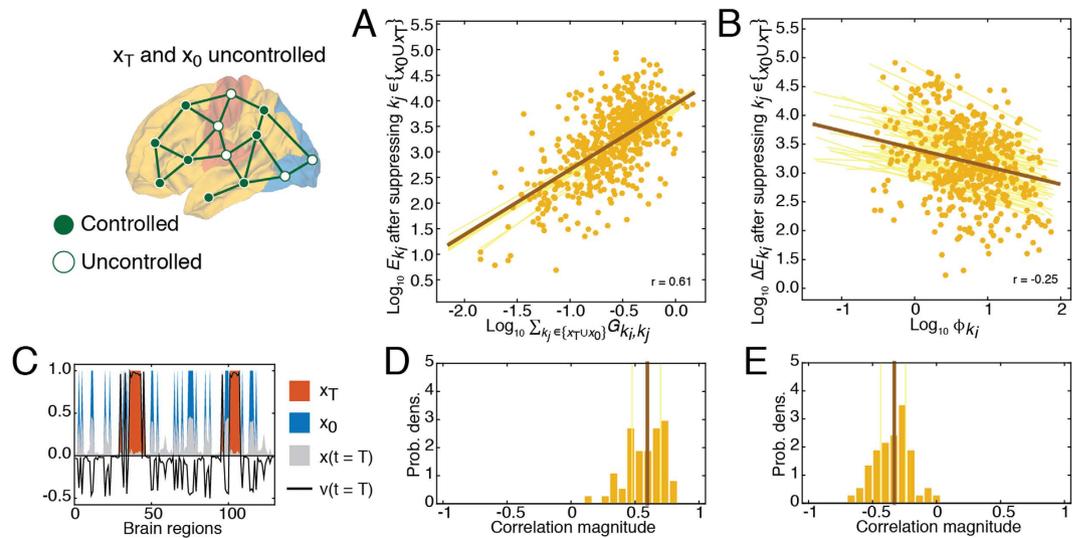


Figure 5. Summary of initial and target class suppression. (A) Only bulk nodes are directly controlled. Their energies, E_{k_j} , are strongly correlated with their communicability to the initial and target nodes, $\sum_{k_j \in \{x_0 \cup x_T\}} G_{k_p, k_j}$. (B) The percent change in total energy going from the *full control* experiment to the *no control* experiment was negatively correlated with $\theta_{k_i} = \sum_j A_{k_i, j} \|v_j\|$, or the weighted distance of controller k_i 's neighbors from their respective target states relative to their states under free evolution. (C) An illustrative example of the components that go into calculating θ_{k_i} . The blue and orange bars represent the state of the system at times $t=0$ and $t=T$. The blue bars indicate the activation of DMN nodes, in this case, while the orange bars indicate activation of VIS nodes. The grey bars represent the state of the system at time $t=T$ under free evolution after starting in $x_0 = \text{DMN}$. Finally, the black line is equal to $v(t) = x_T - x_0$, or the element-wise difference between the orange and grey bars. (D,E) Distributions of Pearson's correlation coefficients from the best-fit lines shown in panels (A,B), respectively.

function for the optimal class assignments given these networks. We observed that across all class compositions the randomized networks were associated with significantly greater energies than the observed networks (non-parametric test, $\max p < 1 \times 10^{-15}$).

Rich club promotes low-energy transitions. The observation that highly-connected brain regions make good targets and poor bystanders suggests that hub regions and, perhaps, rich clubs play an important role in facilitating low-energy control^{37,47}. Intuitively, a rich club is a collection of hubs—highly connected, highly central regions—that are more densely interconnected to one another than expected. This type of organization is thought to promote rapid transmission and integration of information among brain regions⁴⁸. Indeed, rich club regions were more likely to be assigned to the optimal target state (Figure S6). The previous null model, wherein all connections were randomized, tested the null hypothesis that structureless networks could produce comparable levels of energy. To test whether a network's rich club influences energy requires a more subtle and specific null model. Moreover, a network's rich club is a pseudo-continuous structure and can be defined at multiple resolutions. Our focus was on the neighborhood of binary rich clubs identified at $k=84$ —i.e. all brain regions included in the rich club must have degree of at least 84—which corresponds to the maximum normalized rich club coefficient obtained across all participants (Fig. 6A). At this level, the most consistent rich club regions across participants included subcortical regions thalamus, caudate, putamen and hippocampus as well as precuneus, isthmus cingulate, posterior cingulate, lateral orbito-frontal, and insular cortex. These regions are in close agreement with previously-described rich clubs and hubs^{1,37,47,49,50} (This particular rich-club composition is consistent within a neighborhood of $k=84$; Figure S7).

We implemented a null model where we rewired only connections among rich club members (while preserving degree) so that the density of connections among rich club members was as low as possible. We observed that as long as eight brain regions (the smallest increment that we considered) were assigned as initially-active regions, then rewiring connections to dissolve the rich club always yields increased energy (non-parametric test, $\max p < 1 \times 10^{-15}$), suggesting that the brain's rich club, specifically, supports low-energy transitions from a diverse set of initial states to a target state of high-strength hub regions (we verify that this result holds for rich clubs defined at $k=80$ to $k=88$; Figure S8).

Discussion

In this report we used network control theory to investigate the role of the brain's anatomical network in supporting transitions among different brain systems. In the first section we focused on transitions among a pre-defined set of brain systems and demonstrated, in agreement with earlier work, that brain regions with many strong connections were associated with increased energies. We further demonstrated that when the control set is perturbed,

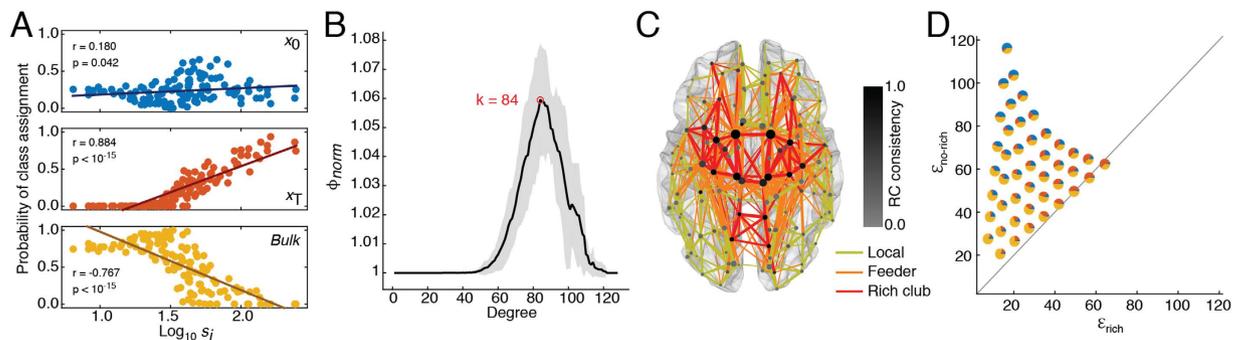


Figure 6. Rich club detection and optimal class assignments. (A) Class assignment probabilities as a function of the logarithm of each brain region's strength. (B) Normalized rich club coefficient as a function of degree. Gray envelope represents \pm one standard deviation. (C) Topographic distribution of rich club nodes and classification of edges (top 5% for visualization) as either "local" (links two non-rich club regions), "feeder" (links a rich club region to a non-rich club region), or "rich club" (links a rich club region to another rich club region). The size of each node is proportional to its degree and the darkness of each point indicates the likelihood, across participants, that that node was assigned to the $k = 84$ rich club. (D) Comparison of energy with rich club intact (\mathcal{E}_{rich}) versus rewired rich club ($\mathcal{E}_{no-rich}$). Each point (pie chart) represents a particular composition of class assignments. The gray line indicates the "break-even" line—points along the line correspond to optimal class assignments that have approximately equal average energy both with and without a rich club. Points above this line indicate that rewiring the rich club leads to increased energy associated with the optimal class assignments.

by suppressing control of both individual brain regions and entire classes, the remaining regions compensate for the loss by increasing their own energies. Moreover, we showed that the percent change in a controller's energy could be predicted by its communicability to the suppressed regions, highlighting the role of indirect communication paths. In the second section, we sought to objectively identify initial and target states that could be transitioned to and from for little energy. We found that the optimal initial states were diverse while the probability that a region was among the target state was highly correlated with its weighted degree. We showed that transitions among these optimal class assignments were, in part, facilitated by the brain's rich club; when connections among rich club regions were rewired, the energy associated with such transitions consistently increased.

From descriptive to predictive network models. The study of networked neural systems has advanced rapidly in the past decade. While early analyses focused on the topological properties of SC networks such as their small-worldness^{51,52} or the presence of hubs, rich clubs, and modules^{1,37}, the focus of recent work has shifted from static descriptions to dynamical systems models, making it possible to investigate how network topology shapes passive dynamics (i.e. no inputs)^{19–22}. A natural extension of these and other recent studies is to incorporate exogenous input into the dynamical model. Such an extension makes it possible to begin addressing questions related to the control of the brain. At the level of large-scale human brain networks, these theoretical questions are only now beginning to be addressed^{18,35}, though the utility of this approach is obvious, showing promise in stimulation-based treatment of epilepsy²⁷. These first studies offered statistics for characterizing the extent to which brain regions contribute to making the entire state space accessible for a network—i.e. rendering it controllable. The entire state space, however, likely contains states that, for one reason or another, should actively be avoided by the system. In this present study we sought to characterize the energy contributions of brain regions based on transitions among a limited state space populated by seemingly neurophysiologically plausible states.

Structural predictors of the ease or difficulty of control. The aim of this report was to shed light on the features of a network's topology that contribute to making a control task easier or more difficult. Toward this end we made a number of contributions and novel observations. First, we presented a classification system of nodes for studying specific control tasks. We showed that for the control tasks we investigated, a node's control energy is highly correlated with its strength. The nature of this relationship, however, depends critically on whether a node is classified as part of the initial, target, or bulk set—the energies of nodes that are initially active and later inactive (initial) as well as nodes that, at the boundary conditions of $t = 0$ and $t = T$, are inactive (bulk) maintain strong positive relationships with node strength; the opposite is true for target nodes, which are negatively correlated with node strength. In other words, to predict a brain region's control energy we require both its context – its role within a given control task – and its embedding within the network's topology.

Under this framework, we explored a series of progressively more difficult control tasks in which we suppressed specific subsets of nodes. This set of experiments highlighted the brain's compensation network—a network whose edge weights represent the percent change in a node's energy when other nodes are excluded from the control set. We went on to show that *communicability* between two nodes was highly correlated with their compensation weight, which is interesting for several reasons. Communicability measures the strength of direct and indirect connections between two nodes⁵³, which suggests a possible functional role for multi-step pathways in the human brain⁵. Namely, that when a brain region's capacity for control is compromised—e.g. acute ischemic stroke or electro-magnetic simulation—the regions that "pick up the slack" and take on expanded control

roles include, as one might anticipate, those with direct connections to the compromised region, but also those with many indirect and potentially long-distance connections. This observation serves as a potential mechanistic account of diaschastic phenomena, where the effect of a focal lesion on brain function is observed at some non-trivial distance from the lesion site^{54–56}. Additionally, this observation is in line with recent work showing that including multi-step pathways in predictive models of FC leads to decreased error rates and improved predictions of resting state functional connectivity^{5,7}.

Interestingly, we observed that bulk regions exhibited the greatest percent increase in their energy. This suggests that regions not directly involved in a particular control task actually play a disproportionately greater compensatory role than those directly involved—i.e. initial and target regions. Intuitively, these results may provide a context to understand cognitive dysfunction observed in neurological conditions that involve region damage or loss, such as traumatic brain injury or Alzheimer's disease. From a control theoretic perspective, as brain regions suffer damage, the increased burden to bulk nodes for performing multiple control tasks may interfere with one another. Bulk regions may become sites of processing interference due to competition between new compensatory roles and older noncompensatory roles. This increased competition between control tasks could have implications for the consequences of brain injury and neurodegeneration over longer timescales⁵⁷. In particular, increased processing burdens may result in later deleterious effects on regions that assume a disproportionate share of compensatory burdens⁵⁸, resulting in a cascade of later failures across the brain.

Regional suppression is also relevant to models of conflict processing and the cognitive effort expended in doing so⁵⁹. In such models, computational executive control mechanisms can be deployed for a limited number of simultaneous tasks, resulting in opportunity costs. With increased opportunity costs comes the phenomenology of effort (motivation, fatigue, boredom), which, in turn, alters the control mechanisms. In the context of our control theoretic view of task control, as node suppression introduces potential conflicts between control tasks among bulk regions, we might expect increases in perceived effort, difficulty managing sustained tasks or task switches, and negative consequences for task performance in a variety of disorders and tasks with control demands. Future studies could test this prediction by mapping the cognitive difficulties and perceived effort that arise from conflicting control demands following real region damage or following non-invasive temporary suppression. Specifically, we would expect that cognitive performance would be low and perceived effort high in general in individuals where a relatively small set of bulk regions assume a relatively high set of compensatory roles across control tasks. Thus, our control theoretic approach may provide a practical means to quantify effort and processing conflicts in health and disease.

The rich club as a control backbone. Whereas we have approached control from an engineering perspective, in cognitive neuroscience the term control refers to a set of processes, which include memory and attentional systems, for guiding behavior towards a particular goal^{60,61}. Control is thought to be instantiated via an anatomical substrate consisting of cortical⁶² and sub-cortical areas⁶² acting largely through inhibitory mechanisms. An important question concerns how this neuro-mechanistic account of control interfaces with current theories on the role of brain connectivity in normative brain function and to what extent it is related to the engineering-focused approach adopted here. In the second section we identified brain regions that act, in an objective sense as optimal initial, target, and bulk classes for control. We showed that the probability of a region being classified as target or bulk was closely related to its strength, with high and low strength regions more likely to be classified as targets and bulk, respectively. The initial class, on the other hand, had a more diverse constituency and was not obviously related to strength. We went on to show that the energies associated with these class assignments were much lower than those obtained from degree-preserved random networks, indicating that the class assignments were driven by some non-trivial aspect of the network's topology. We further showed that when the brain's rich club was dissolved the energy associated with these optimal assignments increased, suggesting that the rich club contributes in facilitating low-energy transitions from a diverse set of initial states to target states composed of high-strength, high-degree rich club regions.

The rich club is typically thought of as an integrative backbone that allows hub regions to communicate with one another, facilitating rapid communication and transmission of signals across the entire brain^{48,63,64}. Despite this supposition, there is some controversy as to the rich club's precise role in network communication, with some indication that rich club hubs are primarily drivers of network dynamics—acting in a top-down fashion to influence different sub-systems^{65–67}. On the other hand, rich club regions are highly connected to the rest of the network and, by this virtue, are not only in a position to influence their neighbors, but also to be influenced by their neighbors. Indeed, computational models suggest that rich clubs adopt stable, regular behavior as a consequence of receiving input from a multitude of sources⁶⁸. Our results agree with this latter account in some respects. We observed that high-degree, rich club regions are best suited for roles as targets that can be transitioned into from a diverse set of initial states. Part of why rich club regions are so successful in this respect is by virtue of their many connections—no matter where input is injected into the system, there is a high probability that the signal will propagate from its source to activate regions in the rich club, suggesting a possible explanation for why rich club regions, which overlap considerably with the brain's default mode network, tend to be active in the default mode of resting state function⁶⁷. More relevant to the present discussion, this also suggests that under certain conditions the rich club (or at least highly connected regions) may be in a better position to be indirectly controlled than weakly connected regions on the network periphery.

Methodological considerations. Several important methodological considerations are pertinent to this work. First, we relied on diffusion spectrum imaging and tractography to infer the presence of large-scale fiber tracts in participants' brains. These methodologies are imperfect and can detect spurious tracts or fail to detect existing tracts^{69,70}. However, at present there exist no other non-invasive methods for reconstructing human

structural connectivity networks. Future work will likely address these shortcomings by introducing improved tractography algorithms and imaging techniques⁷¹.

Another important consideration is our use of a linear dynamical model despite the fact that, by most accounts, brain activity is fundamentally non-linear⁶. Our justification for using such dynamics is twofold. First, the emphasis of this paper is on the role that the brain's structural connectivity network plays in control. While the form of dynamics certainly contributes to making a system controllable or not, we focus primarily on the contribution of the network's topology. Secondly, a linear dynamics is in line with other papers that investigate control theory with an emphasis on topology^{24,29,72}. Moreover, describing non-linear systems in terms of a linear approximation in the neighborhood of its equilibrium points is common⁷³, and makes it possible to apply linear control to otherwise intractable systems³⁵. It should also be noted that even if we wish to proceed using a linear dynamical model, the coupling matrix—here selected to be the brain's structural connectivity matrix—can be further modified. One possibility is to use a coupling matrix based on the observed pattern of co-fluctuations (correlations) between brain regions' activity time series⁷⁴. Should such an approach be used in the context of control, the assumptions of the dynamical model would need to be justified carefully, the predictions validated, and the associated statistical methods potentially extended to account for the differences in the mathematical properties of correlation matrices (positive semi-definite) in comparison to structural matrices built on streamline counts. Here we restrain ourselves to building on the justifications, validations, and statistical extensions for structural matrices described in our prior work³⁵. We expand on this discussion in the Supplementary Material under the heading *Justification for studying linear dynamics*.

An additional concern is the decision to represent functional systems as non-overlapping—i.e. brain regions can be assigned to one system and one system only. In reality, of course, there is no clear one-to-one mapping of brain regions to cognition and behavior; many brain areas are poly-functional. There is, however, a trend in cognitive neuroscience, and especially network neuroscience, to treat brain regions as non-overlapping (see^{36,75,76}). Following this tradition was practically useful for our analyses, as we took advantage of the non-overlapping nature of systems to assign brain regions uniquely to initial, target, and bulk classes—had we used overlapping systems it would have been possible for regions to appear in both the initial and target classes, for example. Future work should explore techniques for obtaining overlapping clusters (and therefore overlapping network states) and investigating optimal control in this context.

A final limitation is the form of the communicability measure, which (when applied to a binary networks) weighs longer paths exponentially less than shorter paths. While this standard form appears sufficient for our purposes here, alternative weighting schemes may provide additional insight in the role of multi-step paths on network control.

Conclusion

Understanding how control occurs in the brain, and how we can use external interventions to affect that control, has broad implications across the cognitive and clinical neurosciences. By examining control strategies implemented in finite time and with limited energy, we were able to uncover fundamental principles of brain structure that impact the ease or difficulty of control tasks informed by the systems known to perform diverse cognitive functions. It is intuitively plausible that these principles may be altered by psychiatric disease and neurological disorders, via a change in underlying structural connectivity. In future, it will be interesting to understand how individual differences in brain structure affect individual differences in the natural implementation of control (e.g., cognitive control) and resulting behavior in executive domains. Moreover, it will be interesting to understand how these individual differences might also directly affect susceptibility and response to external interventions via invasive or non-invasive neuromodulation. Such an understanding would provide critical groundwork for personalized medicine.

Materials and Methods

Data acquisition and processing. Diffusion spectrum images (DSI) were acquired for a total of 30 subjects along with a T1-weighted anatomical scans. We followed a parallel strategy for data acquisition and construction of streamline adjacency matrices as in previous work³⁵. DSI scans sampled 257 directions using a Q5 half-shell acquisition scheme with a maximum b -value of 5,000 and an isotropic voxel size of 2.4 mm and an axial acquisition with the following parameters: repetition time (TR) = 5 s, echo time (TE) = 138 ms, 52 slices, field of view (FoV) (231, 231, 125 mm). All procedures were approved in a convened review by the University of Pennsylvania's Institutional Review Board and were carried out in accordance with the guidelines of the Institutional Review Board/Human Subjects Committee, University of Pennsylvania. All participants volunteered with informed consent in writing prior to data collection.

DSI data were reconstructed in DSI Studio (www.dsi-studio.labsolver.org) using q -space diffeomorphic reconstruction (QSDFR)⁷⁷. QSDFR first reconstructs diffusion-weighted images in native space and computes the quantitative anisotropy (QA) in each voxel and warped to a template QA volume in Montreal Neurological Institute (MNI) space using the statistical parametric mapping (SPM) nonlinear registration algorithm. Once in MNI space, spin density functions were reconstructed with a mean diffusion distance of 1.25 mm using three fiber orientations per voxel. Fiber tracking was performed in DSI studio with an angular cutoff of 55°, step size of 1.0 mm, minimum length of 10 mm, spin density function smoothing of 0.0, maximum length of 400 mm and a QA threshold determined by DWI signal in the colony-stimulating factor. Deterministic fiber tracking using a modified FACT algorithm was performed until 1,000,000 streamlines were reconstructed for each individual.

Anatomical scans were segmented using FreeSurfer59 and parcellated using the connectome mapping toolkit⁷⁸. A parcellation scheme including $n = 129$ regions was registered to the B0 volume from each subject's DSI data. The B0 to MNI voxel mapping was used to map region labels from native space to MNI coordinates. To extend region labels through the grey-white matter interface, the atlas was dilated by 4 mm⁷⁹. Dilation was

accomplished by filling non-labelled voxels with the statistical mode of their neighbors' labels. In the event of a tie, one of the modes was arbitrarily selected. From these data, we constructed a structural connectivity matrix, \mathbf{A} whose element A_{ij} represented the number of streamlines connecting different regions, divided by the sum of volumes for regions i and j .

Cognitive systems. The human brain can also be studied as a network of functional connections. Functional connectivity networks are modular, which means that they can be partitioned into non-overlapping sub-systems^{36,75,80}. These sub-systems are referred to as *intrinsic connectivity networks* (ICNs) and have distinct cognitive and behavioral fingerprints^{81,82}. The ICN definition used here was based on a previously-delineated set of system boundaries³⁶, and included default mode (DMN), control (CONT), dorsal attention (DAN), saliency/ventral attention (SAL/VAN), somatomotor (SM), visual (VIS), limbic (LIM), and sub-cortical (SUB) systems. In order to assign each region of interest to a single system, we mapped both atlases to a common surface template (*fsaverage*) and calculated the overlap (number of common vertices) of each region of interest with each of the seven ICNs. A region's ICN assignment was defined as the system with which it overlapped to the greatest extent (Figure S9).

Control tasks. We considered transitions from an initial state, \mathbf{x}_0 , to a target state, \mathbf{x}_T , where $T=1$ is the control horizon. Initial and target states were selected to correspond to specific cognitive systems. For example, one possible control task was to start in a state where the default mode network (DMN) was maximally active, and to transition to a state where the visual system (VIS) becomes maximally active. Intuitively, such a transition might correspond to the presentation of a visual stimulus at rest, eliciting activation of visual cortex while suppressing activation of the default mode system. In this context, the control question one asks is which nodes play a role in the minimum energy trajectory between these states. We modeled this control task by starting DMN regions in an active state at $t=0$:

$$x_i(t=0) = \begin{cases} 0, & \text{if } i \notin \text{DMN} \\ 1, & \text{if } i \in \text{DMN}. \end{cases}$$

Similarly, when $t=T$, only visual regions were in an active state:

$$x_i(t=T) = \begin{cases} 0, & \text{if } i \notin \text{VIS} \\ 1, & \text{if } i \in \text{VIS}. \end{cases}$$

Given these boundary conditions, we calculated the optimal inputs, $\mathbf{u}_k^*(t)$ to effect the transition from specified initial state to specified target state. In general, state vectors at other times can take on any real value. Though we considered this limited set of states, the linear model of brain dynamics means that any transition can be written as a linear combination of the transitions we studied here.

Network communicability. The adjacency matrix, \mathbf{A} , encodes a network's direct connections. In addition, communication between pairs of nodes can also take advantage of indirect connections. To quantify the extent to which nodes are connected indirectly, one can calculate the communicability matrix⁴¹, $\mathbf{G} \in \mathbb{R}^{n \times n}$, whose element $G_{ij} = \sum_{k=0}^{\infty} \binom{A^k}{k!}_{ij} = (e^{\mathbf{A}})_{ij}$. The element G_{ij} , then, represents the weighted sum of walks of all lengths. The $k!$ in the denominator means that longer walks contribute disproportionately less compared to shorter walks. Communicability has been generalized to weighted networks, such as those considered here⁸³. Specifically, $G_{ij}^w = e^{\mathbf{A}'}$, where $\mathbf{A}' = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ and $\mathbf{D} \in \mathbb{R}^{n \times n}$ is the square matrix whose diagonal elements $D_{ii} = \sum_j A_{ij}$. Communicability is also related to the linear dynamics we study here. Solving (1) for the special case where there is no input, the evolution of the system is described by the matrix exponential of the connectivity matrix.

Optimal class assignments and simulated annealing. In the second section of this paper—*optimal class assignments*—we sought a division of nodes into initial, target, and bulk classes corresponding to the smallest possible control energy. To identify the optimal division, we needed to solve the minimization problem:

$$\min_{\mathbf{x}_0, \mathbf{x}_T} (\mathcal{E}(\mathbf{x}_0, \mathbf{x}_T)). \quad (3)$$

Note that we are not minimizing control energy directly; instead, we minimize $\mathcal{E} = \mathbf{v}^T \mathbf{v}$ where $\mathbf{v} = \mathbf{x}_T - e^{\mathbf{A}} \mathbf{x}_0$, a more computationally tractable measure that is highly correlated with the actual control energy (see Figure S1). This minimization problem cannot be solved analytically and therefore the optimal division of the network into initial, target, and bulk classes can only be approximated. To generate good approximations of the optimal solution, we use the optimization heuristic *simulated annealing* (SA)⁸⁴. SA is a heuristic that can be applied to a wide range of minimization problems. Briefly, SA begins with an initial estimate of the solution (which can be generated randomly) and a temperature parameter. Next, the quality of the initial estimate is calculated according to an objective function (usually supplied by the user), after which the initial estimate of the solution is modified slightly with the addition of a small amount of noise. If the modification results in a higher quality estimate of the solution, then this modification is retained and propagated to the next iteration. Even if the solution results in decreased quality, it is still retained with some probability that decays monotonically as a function of the temperature parameter. This makes it possible, at high temperatures, to sample a range of possible solutions from the parameter space. Gradually, the temperature is reduced, meaning that the only solutions that get retained are

those of higher quality. The algorithm continues until some termination criteria are reached—e.g. a pre-specified number of iterations have passed without uncovering a new estimate of the optimal solution.

For our specific minimization problem, we started with a random division of nodes into initial, target, and bulk classes. The proportion of nodes assigned to each of the three classes was fixed from the beginning and did not change over the course of minimization. At each iteration, we propose a new estimate of the optimal solution by randomly selecting two nodes assigned to different classes and swapping their class assignments. If the new estimate reduces \mathcal{E} (i.e. is of increased quality), we accept the swap. If the estimate results in a $\mathcal{E}' > \mathcal{E}$, we accept the swap with probability $Pr = \exp[-(\mathcal{E}' - \mathcal{E})/\tau]$. We initialized the algorithm with a temperature $\tau_0 = 1$. In total, the algorithm consisted of 100 stages, each comprised of 5000 iterations. The temperature was fixed throughout each stage and, following each stage, was decreased by a factor of 0.925, so that by the 100th stage, the temperature was $\tau_{100} = \tau_0^{100} \approx 4.11 \times 10^{-4}$. Because we had no prior knowledge of what proportion of nodes should be assigned to initial, target, and bulk classes, we tested 44 possible proportions. These proportions were selected by fixing the number of bulk nodes to be 48, 56, 64, 72, 80, 88, 96, or 104. The nodes were then divided among initial and target classes in multiples of eight. For example, suppose we fixed the number of bulk nodes to be 64. This leaves 64 nodes to be divided among initial/target classes. The possible divisions are: 8/56, 16/48, 24/40, 32/32, 40/24, 48/16, and 56/8. Once the proportion was determined, we ran the SA algorithm 50 times and retained the 100 highest-quality, unique estimates of the optimal division.

Rich club detection. A rich club refers to a subset of nodes, all of degree $\geq k$, that are densely connected to one another^{85,86}. The density of any such subset is summarized by its rich club coefficient: $\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k}-1)}$. We calculated $\phi(k)$ for all possible values of k across all participants and subsequently normalized rich club coefficients by comparing them against the distribution of coefficients obtained from an ensemble of randomized networks (preserved degree sequence)⁸⁷. It is sometimes the case that there is no single rich club—there may be many statistically significant rich clubs spanning a range of k ³⁷. For ease of description, however, we focus on the rich club detected at $k = 84$, which corresponds to the greatest value of ϕ_{norm} averaged across participants (Fig. 6A). We explore the consistency of this rich club in Figure S6.

References

- Hagmann, P. et al. Mapping the structural core of human cerebral cortex. *PLoS Biol* **6**, e159 (2008).
- Honey, C. et al. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences* **106**, 2035–2040 (2009).
- Hermundstad, A. M. et al. Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proceedings of the National Academy of Sciences* **110**, 6169–6174 (2013).
- Atasoy, S., Donnelly, I. & Pearson, J. Human brain networks function in connectome-specific harmonic waves. *Nature Communications* **7** (2016).
- Becker, C. O. et al. Accurately predicting functional connectivity from diffusion imaging. *arXiv preprint arXiv:1512.02602* (2015).
- Deco, G., Jirsa, V. K. & McIntosh, A. R. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience* **12**, 43–56 (2011).
- Góñi, J. et al. Resting-brain functional connectivity predicted by analytic measures of network communication. *Proceedings of the National Academy of Sciences* **111**, 833–838 (2014).
- Allen, E. A. et al. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex* bhs352 (2012).
- Karahanoglu, F. I. & Van De Ville, D. Transient brain activity disentangles fmri resting-state dynamics in terms of spatially and temporally overlapping networks. *Nature communications* **6** (2015).
- Betzler, R. F., Fukushima, M., He, Y., Zuo, X.-N. & Sporns, O. Dynamic fluctuations coincide with periods of high and low modularity in resting-state functional brain networks. *NeuroImage* **127**, 287–297 (2016).
- Bassett, D. S. et al. Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences* **108**, 7641–7646 (2011).
- Braun, U. et al. Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences* **112**, 11678–11683 (2015).
- Mattar, M. G., Cole, M. W., Thompson-Schill, S. L. & Bassett, D. S. A functional cartography of cognitive systems. *PLoS Comput Biol* **11**, e1004533 (2015).
- Wilson, H. R. & Cowan, J. D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal* **12**, 1 (1972).
- Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M. & Friston, K. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol* **4**, e1000092 (2008).
- Marreiros, A. C., Kiebel, S. J. & Friston, K. J. A dynamic causal model study of neuronal population dynamics. *Neuroimage* **51**, 91–101 (2010).
- Freyer, F. et al. Biophysical mechanisms of multistability in resting-state cortical rhythms. *The Journal of Neuroscience* **31**, 6353–6361 (2011).
- Muldoon, S. F. et al. Stimulation-based control of dynamic brain networks. *arXiv preprint arXiv:1601.00987* (2016).
- Breakspear, M., Heitmann, S. & Daffertshofer, A. Generative models of cortical oscillations: neurobiological implications of the kuramoto model. *Frontiers in human neuroscience* **4** (2010).
- Haimovici, A., Tagliazucchi, E., Balenzuela, P. & Chialvo, D. R. Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Physical review letters* **110**, 178101 (2013).
- Abdelnour, F., Voss, H. U. & Raj, A. Network diffusion accurately models the relationship between structural and functional brain connectivity networks. *Neuroimage* **90**, 335–347 (2014).
- Mišić, B. et al. Cooperative and competitive spreading dynamics on the human connectome. *Neuron* **86**, 1518–1529 (2015).
- Kalman, R. E. Mathematical description of linear dynamical systems. *Journal of the Society for Industrial & Applied Mathematics, Series A: Control* **1**, 152–192 (1963).
- Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
- Sun, J. & Motter, A. E. Controllability transition and nonlocality in network control. *Physical Review Letters* **110**, 208701 (2013).
- Rosin, B. et al. Closed-loop deep brain stimulation is superior in ameliorating parkinsonism. *Neuron* **72**, 370–384 (2011).
- Taylor, P. N. et al. Optimal control based seizure abatement using patient derived connectivity. *Frontiers in Neuroscience* **9**, 202 (2015).

28. Fox, M. D., Halko, M. A., Eldaief, M. C. & Pascual-Leone, A. Measuring and manipulating brain connectivity with resting state functional connectivity magnetic resonance imaging (fcmri) and transcranial magnetic stimulation (tms). *Neuroimage* **62**, 2232–2243 (2012).
29. Yan, G., Ren, J., Lai, Y.-C., Lai, C.-H. & Li, B. Controlling complex networks: How much energy is needed? *Physical review letters* **108**, 218703 (2012).
30. Wang, W.-X., Ni, X., Lai, Y.-C. & Grebogi, C. Optimizing controllability of complex networks by minimum structural perturbations. *Physical Review E* **85**, 026115 (2012).
31. Cowan, N. J., Chastain, E. J., Vilhena, D. A., Freudenberg, J. S. & Bergstrom, C. T. Nodal dynamics, not degree distributions, determine the structural controllability of complex networks. *PLoS one* **7**, e38398 (2012).
32. Pósfai, M., Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Effect of correlations on network controllability. *Scientific reports* **3** (2013).
33. Ruths, J. & Ruths, D. Control profiles of complex networks. *Science* **343**, 1373–1376 (2014).
34. Pasqualetti, F., Zampieri, S. & Bullo, F. Controllability metrics, limitations and algorithms for complex networks. *Control of Network Systems, IEEE Transactions on* **1**, 40–52 (2014).
35. Gu, S. *et al.* Controllability of structural brain networks. *Nature communications* **6** (2015).
36. Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology* **106**, 1125–1165 (2011).
37. van den Heuvel, M. P. & Sporns, O. Rich-club organization of the human connectome. *The Journal of neuroscience* **31**, 15775–15786 (2011).
38. Lin, C. T. Structural controllability. *Automatic Control, IEEE Transactions on* **19**, 201–208 (1974).
39. Pascual-Leone, A., Walsh, V. & Rothwell, J. Transcranial magnetic stimulation in cognitive neuroscience—virtual lesion, chronometry, and functional connectivity. *Current opinion in neurobiology* **10**, 232–237 (2000).
40. Silvano, J. & Muggleton, N. G. New light through old windows: Moving beyond the “virtual lesion” approach to transcranial magnetic stimulation. *Neuroimage* **39**, 549–552 (2008).
41. Estrada, E. & Hatano, N. Communicability in complex networks. *Physical Review E* **77**, 036111 (2008).
42. Roberts, J. A. *et al.* The contribution of geometry to the human connectome. *NeuroImage* **124**, 379–393 (2016).
43. Betzel, R. F. *et al.* Generative models of the human connectome. *NeuroImage* **124**, 1054–1064 (2016).
44. Laughlin, S. B. & Sejnowski, T. J. Communication in neuronal networks. *Science* **301**, 1870–1874 (2003).
45. Niven, J. E. & Laughlin, S. B. Energy limitation as a selective pressure on the evolution of sensory systems. *Journal of Experimental Biology* **211**, 1792–1804 (2008).
46. Marblestone, A., Wayne, G. & Kording, K. Towards an integration of deep learning and neuroscience. *arXiv preprint arXiv:1606.03813* (2016).
47. van den Heuvel, M. P. & Sporns, O. Network hubs in the human brain. *Trends in cognitive sciences* **17**, 683–696 (2013).
48. van den Heuvel, M. P., Kahn, R. S., Goñi, J. & Sporns, O. High-cost, high-capacity backbone for global brain communication. *Proceedings of the National Academy of Sciences* **109**, 11372–11377 (2012).
49. Betzel, R. F. *et al.* Changes in structural and functional connectivity among resting-state networks across the human lifespan. *NeuroImage* **102**, 345–357 (2014).
50. Crossley, N. A. *et al.* The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* **137**, 2382–2395 (2014).
51. Iturria-Medina, Y., Sotero, R. C., Canales-Rodríguez, E. J., Alemán-Gómez, Y. & Melie-García, L. Studying the human brain anatomical network via diffusion-weighted mri and graph theory. *NeuroImage* **40**, 1064–1076 (2008).
52. Gong, G. *et al.* Mapping anatomical connectivity patterns of human cerebral cortex using *in vivo* diffusion tensor imaging tractography. *Cerebral cortex* **19**, 524–536 (2009).
53. Estrada, E., Hatano, N. & Benzi, M. The physics of communicability in complex networks. *Physics reports* **514**, 89–119 (2012).
54. Grefkes, C. & Fink, G. R. Reorganization of cerebral networks after stroke: new insights from neuroimaging with connectivity approaches. *Brain* awr033 (2011).
55. Carrera, E. & Tononi, G. Diaschisis: past, present, future. *Brain* awu101 (2014).
56. Crofts, J. J. *et al.* Network analysis detects changes in the contralesional hemisphere following stroke. *NeuroImage* **54**, 161–169 (2011).
57. Sivanandam, T. M. & Thakur, M. K. Traumatic brain injury: a risk factor for alzheimer’s disease. *Neuroscience & Biobehavioral Reviews* **36**, 1376–1381 (2012).
58. Stam, C. J. Modern network science of neurological disorders. *Nature Reviews Neuroscience* **15**, 683–695 (2014).
59. Kurzban, R., Duckworth, A., Kable, J. W. & Myers, J. An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences* **36**, 661–679 (2013).
60. Schneider, W. & Shiffrin, R. M. Controlled and automatic human information processing: I. detection, search, and attention. *Psychological review* **84**, 1 (1977).
61. Norman, D. A. & Shallice, T. *Attention to action* (Springer, 1986).
62. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annual review of neuroscience* **24**, 167–202 (2001).
63. Senden, M., Deco, G., de Reus, M. A., Goebel, R. & van den Heuvel, M. P. Rich club organization supports a diverse set of functional network configurations. *NeuroImage* **96**, 174–182 (2014).
64. Collin, G., Sporns, O., Mandl, R. C. & van den Heuvel, M. P. Structural and functional aspects relating to cost and benefit of rich club organization in the human cerebral cortex. *Cerebral cortex* **24**, 2258–2267 (2014).
65. Harriger, L., Van Den Heuvel, M. P. & Sporns, O. Rich club organization of macaque cerebral cortex and its role in network communication. *PLoS one* **7**, e46497 (2012).
66. Towson, E. K., Vértés, P. E., Ahnert, S. E., Schafer, W. R. & Bullmore, E. T. The rich club of the *c. elegans* neuronal connectome. *The Journal of Neuroscience* **33**, 6380–6387 (2013).
67. van den Heuvel, M. P. & Sporns, O. An anatomical substrate for integration among functional networks in human cortex. *The Journal of neuroscience* **33**, 14489–14500 (2013).
68. Gollo, L. L., Zalesky, A., Hutchison, R. M., van den Heuvel, M. & Breakspear, M. Dwelling quietly in the rich club: brain network determinants of slow cortical fluctuations. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **370**, 20140165 (2015).
69. Thomas, C. *et al.* Anatomical accuracy of brain connections derived from diffusion mri tractography is inherently limited. *Proceedings of the National Academy of Sciences* **111**, 16574–16579 (2014).
70. Reveley, C. *et al.* Superficial white matter fiber systems impede detection of long-range cortical connections in diffusion mr tractography. *Proceedings of the National Academy of Sciences* **112**, E2820–E2828 (2015).
71. Pestilli, F., Yeatman, J. D., Rokem, A., Kay, K. N. & Wandell, B. A. Evaluation and statistical inference for human connectomes. *Nature methods* **11**, 1058–1063 (2014).
72. Müller, F.-J. & Schuppert, A. Few inputs can reprogram biological networks. *Nature* **478**, E4–E4 (2011).
73. Luenberger, D. Introduction to dynamic systems: theory, models, and applications (1979).
74. Vuksanović, V. & Hövel, P. Functional connectivity of distant cortical regions: Role of remote synchronization and symmetry in interactions. *NeuroImage* **97**, 1–8 (2014).
75. Power, J. D. *et al.* Functional network organization of the human brain. *Neuron* **72**, 665–678 (2011).

76. Bellec, P., Rosa-Neto, P., Lyttelton, O. C., Benali, H. & Evans, A. C. Multi-level bootstrap analysis of stable clusters in resting-state fmri. *Neuroimage* **51**, 1126–1139 (2010).
77. Yeh, F.-C., Wedeen, V. J. & Tseng, W.-Y. I. Estimation of fiber orientation and spin density distribution by diffusion deconvolution. *Neuroimage* **55**, 1054–1062 (2011).
78. Cammoun, L. *et al.* Mapping the human connectome at multiple scales with diffusion spectrum mri. *Journal of neuroscience methods* **203**, 386–397 (2012).
79. Cieslak, M. & Grafton, S. Local termination pattern analysis: a tool for comparing white matter morphology. *Brain imaging and behavior* **8**, 292–299 (2014).
80. Sporns, O. & Betzel, R. F. Modular brain networks. *Annual review of psychology* **67** (2015).
81. Smith, S. M. *et al.* Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences* **106**, 13040–13045 (2009).
82. Crossley, N. A. *et al.* Cognitive relevance of the community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences* **110**, 11583–11588 (2013).
83. Crofts, J. J. & Higham, D. J. A weighted communicability measure applied to complex brain networks. *Journal of the Royal Society Interface* rsif-2008 (2009).
84. Kirkpatrick, S., Vecchi, M. P. *et al.* Optimization by simulated annealing. *Science* **220**, 671–680 (1983).
85. Colizza, V., Flammini, A., Serrano, M. A. & Vespignani, A. Detecting rich-club ordering in complex networks. *Nature physics* **2**, 110–115 (2006).
86. McAuley, J. J., da Fontoura Costa, L. & Caetano, T. S. Rich-club phenomenon across complex network hierarchies. *Applied Physics Letters* **91**, 084103 (2007).
87. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).

Acknowledgements

This work was supported from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation, the Army Research Laboratory and the Army Research Office through contract numbers W911NF-10-2-0022 and W911NF-14-1-0679, the National Institute of Mental Health (2-R01-DC-009209-11), the National Institute of Child Health and Human Development (1R01HD086888-01), the Office of Naval Research, and the National Science Foundation (BCS-1441502 and BCS-1430087). FP acknowledges support from BCS-1430280. JDM acknowledges support from the Office of the Director at the National Institutes of Health (1DP5OD021352-01).

Author Contributions

R.F.B., S.G. and F.P. provided code. J.D.M. contributed data. R.F.B. and D.S.B. performed data analysis. All authors wrote manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Betzel, R. F. *et al.* Optimally controlling the human connectome: the role of network topology. *Sci. Rep.* **6**, 30770; doi: 10.1038/srep30770 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016