Control Theoretic Analysis of Human Brain Networks

Shi Gu
University of Pennsylvania, gus@sas.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/edissertations
Part of the Applied Mathematics Commons, Biomedical Commons, and the Neuroscience and Neurobiology Commons

Recommended Citation
http://repository.upenn.edu/edissertations/1751

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/1751
For more information, please contact libraryrepository@pobox.upenn.edu.
Control Theoretic Analysis of Human Brain Networks

Abstract
The brain is a complex system with complicated structures and entangled dynamics. Among the various approaches to investigating the brain's mechanics, the graphical method provides a successful framework for understanding the topology of both the structural and functional networks, and discovering efficient diagnostic biomarkers for cognitive behaviors, brain disorders and diseases. Yet it cannot explain how the structure affects the functionality and how the brain tunes its transition among multiple states to manipulate the cognitive control. In my dissertation, I propose a novel framework of modeling the mechanics of the cognitive control, which involves in applying control theory to analyzing the brain networks and conceptually connecting the cognitive control with the engineering control. First, I examine the energy distribution among different states via combining the energetic and structural constraints of the brain's state transition in a free energy model, where the interaction between regions is explicitly informed by structural connectivity. This work enables the possibility of achieving a whole view of the brain's energy landscape and preliminarily indicates the feasibility of control theory to model the dynamics of cognitive control. In the following work, I exploit the network control theory to address two questions about how the large-scale circuitry of the human brain constrains its dynamics. First, is the human brain theoretically controllable? Second, which areas of the brain are most influential in constraining or facilitating changes in brain state trajectories? Further, I seek to examine the structural effect on the control actions through solving the optimal control problem under different boundary conditions. I quantify the efficiency of regions in terms of the energy cost for the brain state transition from the default mode to task modes. This analysis is extended to the perturbation analysis of trajectories and is applied to the comparison between the group with mild traumatic brain injury (mTBI) and the healthy group. My research is the first to demonstrate how control theory can be used to analyze human brain networks.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Applied Mathematics

First Advisor
Danielle S. Bassett

Keywords
brain, cognitive control, connectivity, controllability, dynamics, energy landscape

Subject Categories
Applied Mathematics | Biomedical | Neuroscience and Neurobiology

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/1751
CONTROL THEORETIC ANALYSIS OF HUMAN BRAIN NETWORKS

Shi Gu

A DISSERTATION

in

Applied Mathematics and Computational Science

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

Danielle S. Bassett, Associate Professor of Bioengineering

Graduate Group Chairperson

Charles L. Epstein, Thomas A. Scott Professor of Mathematics

Dissertation Committee

Danielle S. Bassett, Associate Professor of Bioengineering

Fabio Pasqualetti, Assistant Professor of Mechanical Engineering

James C. Gee, Associate Professor of Radiologic Science in Radiology

Shane T. Jensen, Associate Professor of Statistics
CONTROL THEORETIC ANALYSIS OF HUMAN BRAIN NETWORKS

© COPYRIGHT

2016

Shi Gu

This work is licensed under the
Creative Commons Attribution
NonCommercial-ShareAlike 3.0
License

To view a copy of this license, visit

http://creativecommons.org/licenses/by-nc-sa/3.0/
To the moon, who consistently brightens my path in the darkness.
ACKNOWLEDGEMENT

I would like to first thank my parents who support me consistently and unreservedly from my birth to now. Their back-up and trust offer me the courage and determination along the journey of life.

I would like to give my special thanks to my advisor Dr. Danielle S. Bassett. Her insightful instruction directs my exploration in the filed of network neuroscience during my PhD period.

I would also like to give many thanks to all my committees, colleagues, collaborators, course professors, and friends for their kind helps in both academia and life.

I also appreciate it sincerely for the financial support from the program of Applied Mathematics and Computational Science, School of Arts and Science and National Science Foundation.
ABSTRACT

CONTROL THEORETIC ANALYSIS OF HUMAN BRAIN NETWORKS

Shi Gu
Danielle S. Bassett

The brain is a complex system with complicated structures and entangled dynamics. Among the various approaches to investigating the brain’s mechanics, the graphical method provides a successful framework for understanding the topology of both the structural and functional networks, and discovering efficient diagnostic biomarkers for cognitive behaviors, brain disorders and diseases. Yet it cannot explain how the structure affects the functionality and how the brain tunes its transition among multiple states to manipulate the cognitive control. In my dissertation, I propose a novel framework of modeling the mechanics of the cognitive control, which involves in applying control theory to analyzing the brain networks and conceptually connecting the cognitive control with the engineering control. First, I examine the energy distribution among different states via combining the energetic and structural constraints of the brain’s state transition in a free energy model, where the interaction between regions is explicitly informed by structural connectivity. This work enables the possibility of achieving a whole view of the brain’s energy landscape and preliminarily indicates the feasibility of control theory to model the dynamics of cognitive control. In the following work, I exploit the network control theory to address two questions about how the large-scale circuitry of the human brain constrains its dynamics. First, is the human brain theoretically controllable? Second, which areas of the brain are most influential in constraining or facilitating changes in brain state trajectories? Further, I seek to examine the structural effect on the control actions through solving the optimal control problem under different boundary conditions. I quantify the efficiency of regions in terms of the energy cost for the brain state transition from the default mode to task modes. This analysis is extended to the perturbation analysis of trajectories and is applied to the comparison
between the group with mild traumatic brain injury (mTBI) and the healthy group. My research is the first to demonstrate how control theory can be used to analyze human brain networks.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENT** ................................................................. iv  
**ABSTRACT** ................................................................. v  
**LIST OF TABLES** ................................................................. xi  
**LIST OF ILLUSTRATIONS** ........................................................ xxiii  
**PREFACE** ................................................................. 1  
**CHAPTER 1: Introduction** .......................................................... 1  
1.1 What is a brain network? ....................................................... 1  
1.2 Why should we use the control theory to view the brain network? ........ 3  
1.3 Mathematical concepts ...................................................... 4  
1.3.1 Network Measures ...................................................... 4  
1.3.2 Controllability Settings ................................................. 8  
1.4 Contribution of my works ................................................... 10  
**CHAPTER 2: The Landscape of Neurophysiological Activity Implicit in Brain Network Structure** ............................................................. 11  
2.1 Introduction ................................................................. 11  
2.2 Models and Materials ...................................................... 15  
2.2.1 Human DSI Data Acquisition and Preprocessing ..................... 15  
2.2.2 Structural Network Construction ..................................... 15  
2.2.3 Resting state fMRI data ............................................... 17  
2.2.4 Defining an Energy Landscape ....................................... 18  
2.2.5 Discovering Local Minima ........................................... 19  
2.2.6 Characterizing Local Minima ........................................ 21
3.3.2 Regional Controllability ................................................. 60
3.3.3 Regional Controllability of Cognitive Systems .................. 74
3.4 Discussion ........................................................................ 84
  3.4.1 Theoretically Predicted Controllability of Large-Scale Neural Circuitry 84
  3.4.2 The Role of Hubs in Brain Control ................................ 85
  3.4.3 The Role of Weak Connections in Brain Control ............... 88
  3.4.4 The Role of Community Structure in Brain Control ........... 89
  3.4.5 Methodological Considerations ..................................... 89
3.5 Conclusion ...................................................................... 93
3.6 Materials ........................................................................ 93
  3.6.1 Human DSI Data Acquisition and Preprocessing ............ 93
  3.6.2 Human DTI Data Acquisition and Preprocessing ............ 94
  3.6.3 Macaque Tract Tracing Data ............................... 95
  3.6.4 Association of Brain Regions to Cognitive Systems ....... 96

CHAPTER 4 : Optimal Trajectories for Brain State Transitions ........ 102
4.1 Introduction .................................................................... 102
4.2 Results .......................................................................... 104
  4.2.1 Task Control .............................................................. 105
  4.2.2 Characteristics of Optimal Control Trajectories ............. 107
  4.2.3 Structurally-Driven Task Preference for Control Regions ... 108
  4.2.4 Robustness of Control in Health and Following Injury .... 109
  4.2.5 Graphical Metrics ..................................................... 110
4.3 Discussion ..................................................................... 113
  4.3.1 Role of structural connectivity in shaping brain functional patterns . 113
  4.3.2 Maladaptive Control in Traumatic Brain Injury ............. 118
  4.3.3 Methodological Considerations ................................... 119
  4.3.4 Future Directions ..................................................... 120
4.4 Models and Materials .................................................... 120
4.4.1 Human DSI Data Acquisition and Preprocessing .......................... 120
4.4.2 Structural Network Construction ........................................... 121
4.4.3 Network Control Theory ....................................................... 123
4.4.4 Optimal Control Trajectories ............................................... 124
4.4.5 Specification of the Initial and Target States ........................... 127
4.4.6 Statistics of Optimal Control Trajectories ............................... 128
4.4.7 Control Efficiency ............................................................. 128
4.4.8 Network Communicability to the Target State ......................... 129
4.4.9 Energetic Impact of Brain Regions on Control Trajectories ....... 129
4.4.10 Target Control Model .......................................................... 130

CHAPTER 5 : Closing Remarks ....................................................... 132
5.1 Improvement on the model ...................................................... 132
5.2 Potential Application ............................................................. 133
5.2.1 Mapping the measures to the functions .................................. 133
5.2.2 Learning Procedure energy cost ............................................ 133
5.2.3 Neurodevelopment .............................................................. 134

BIBLIOGRAPHY ............................................................................ 134
LIST OF TABLES

**TABLE 1:** Robustness of the Correlation between Control Metrics and Graphic Metrics: Pearson correlation coefficients $r$ between rank degree, average controllability (AC), boundary controllability (BC), and modal controllability (MC).

**TABLE 2:** Test-Retest Reliability of Controllability Diagnostics: average controllability (AC), boundary controllability (BC), modal controllability (MC) and global controllability (GC).

**TABLE 3:** The $p$-values for the Permutation Tests of the Graphical Metrics of the Healthy and mTBI Groups. In this table, we list the $p$-values from the permutation tests in distinguishing the mTBI group and the healthy group. All metrics are the averages across regions if they are regionally defined. The $p$-value here is the two-sided $p$-value, which is defined as the minimum of the two one-sided $p$-values with no assumptions on the sign of the difference. For the notations, DEG is short for Degree, CPL is short for Characteristic path length, C-COEF is short for Clustering coefficient, MODU is short for Modularity, L-EFF is short for Local efficiency, G-EFF is short for Global efficiency, and DEN is short for Density.
LIST OF ILLUSTRATIONS

FIGURE 1: Conceptual Network. For a given network constructed with the two basic elements "nodes" and "edges", we can define various types of graphical metrics like the degrees, shortest path lengths and local structure like triangle and communities. From these fundamental metrics, we are able to define more complex measurements to quantify the topological structure of the network.

FIGURE 2: Conceptual Schematic. (A) A weighted structural brain network represents the number of white matter streamlines connecting brain regions. (B) We model each brain region as a binary object, being either active or inactive. (C) Using a maximum entropy model, we infer the full landscape of predicted activity patterns – binary vectors indicating the regions that are active and the regions that are not active – as well as the energy of each pattern (or state). We use this mathematical framework to identify and study local minima in the energy landscape: states predicted to form the foundational repertoire of brain function.
FIGURE 3: **Simulated Activation Rates.** (A) The distribution of distances from the first local minimum to other local minima. Each point and error-bar is calculated across a bin of 30 minima; error bars indicate standard error of the mean over the 30 minima. (B) The probability distribution of the radius of each local minimum is well-fit by a power-law. The radius of a local minimum is defined as its distance to the closest sampled point on the energy landscape. (C) The distribution of the pairwise normalized mutual information between all pairs of local minima. (D) Average activation rates for all 14 *a priori* defined cognitive systems [Power et al. (2011)]. Error bars indicate standard error of the mean across subjects.

FIGURE 4: **Stability of Energy Distribution with respect to the Number of Sample Size.** We plot the heuristic probability distribution of the energy for the first 2000, 4000, 6000 and 8000 samples. We can observe that the shapes are pretty consistent and further KS-test results ensure the stability.

FIGURE 5: **Relation between the Regional Energy and the Activation Rate.** Regional Energy and Activation Rate are positively correlated. Here we want to justify that the activation rate is neither dominated by regional energy nor the weighted degrees.

FIGURE 6: **Distribution of Between- and Within- System Energy Tuples.** We show that the between- and within- system energy provides C) a better separation for different neural systems compared to the degrees A), especially B.D) for the Default Mode network, Fronto-Parietal, and Cingulo-Opercular Task Control.
Validating Predicted Activation Rates in Functional Neuroimaging Data. (A) From resting state BOLD data acquired in an independent cohort, we estimated the true activation rate by transforming the continuous BOLD magnitudes to binary state vectors by thresholding the signals at 0 (see Methods). We use these binary state vectors to estimate the activation rates of each brain region across the full resting state scan. Here we show the mean activation rate of each brain region, averaged over subjects. (B) For comparison, we also show the mean predicted activation rate estimated from the local minima of the maximum entropy model, as defined in Equation 2.4, and averaged over subjects. (C) We observe that the activation rates estimated from resting state fMRI data are significantly positively correlated with the activation rates estimated from the local minima of the maximum entropy model (Pearson’s correlation coefficient $r = 0.18$, $p = 0.0046$). Each data point represents a brain region, with either observed or predicted activation rates averaged over subjects.
FIGURE 8: **Utilization Energies of Cognitive Systems.**

- (A) Average within-system energy of each cognitive system; error bars indicate standard error of the mean across subjects.
- (B) Average between-system energy of each cognitive system; error bars indicate standard error of the mean across subjects.
- (C) The 2-dimensional plane mapped out by the within- and between-system energies of different brain systems. Each data point represents a different brain region, and visual clusters of regions are highlighted with lightly colored sectors. The sector direction is determined by minimizing the squared loss in point density of the local cloud and the width is determined by the orthogonal standard derivation at the center along the sector direction. In this panel, all data points represent values averaged across subjects.
- (D) The percentages of minima displaying preferential activation of each system; each minima was assigned to the system which whom it shared the largest normalized mutual information. Errorbars indicate the differences between the observed percentages and those of the null distribution with random activation patterns across regions.

FIGURE 9: **Conceptual Schematic.** From weighted brain networks (A), we estimate control points (B) whose large-scale regional activity can move the brain into new trajectories that traverse diverse cognitive functions (C). In panel (C), we show the original state of the system (state 0), as well as 4 possible states (indicated by the blue circles) that are equidistant from state 0 in the state space (indicated by the black circular line), and which can be reached by trajectories that are more or less energetically costly (indicated by the height of the purple bars).
FIGURE 10: **Partition Similarity As a Function of the Resolution Parameter** Mean (*left*) and variance (*right*) of the partition similarity estimated using the z-score of the Rand coefficient as a function of the structural resolution parameter $\gamma$, varies from 0.5 to 2 in increments of 0.1, for the 5 spatial scales of the Lausanne atlas [Hagmann et al. (2008a) (rows)].

FIGURE 11: **Effect of Boundary Point Criteria Threshold** Color indicates Pearson correlation coefficient, $r$, between the vectors of boundary controllability values estimated for pairs of $\rho$ values in the range $0.15 - 0.25$ in increments of 0.01.
FIGURE 12: Brain Network Control Properties (A) Average controllability quantifies control to many easily reached states. Here we show controllability values, averaged across persons, and ranked for all brain regions plotted on a surface visualization. Warmer colors indicate larger values of average controllability. (B) Scatter plot of weighted degree (ranked for all brain regions) versus average controllability (Pearson correlation $r = 0.91$, $p = 8 \times 10^{-92}$). (C) Modal controllability quantifies control to difficult-to-reach states. Here we show modal controllability values ranked for all brain regions plotted on a surface visualization. (D) Scatter plot of weighted degree (ranked for all brain regions) versus modal controllability ($r = -0.99$, $p = 2 \times 10^{-213}$). (E) Boundary controllability quantifies control to decouple or integrate network modules. Here we show boundary controllability values ranked for all brain regions plotted on a surface visualization. (F) Scatter plot of weighted degree (ranked for all brain regions) versus boundary controllability ($r = 0.13$, $p = 0.03$). In panels (A), (C), and (E), warmer colors indicate larger controllability values, which have been averaged over both replicates and subjects. These results are reliable over a range of atlas resolutions, and are consistent with findings using a network composed of only cortical circuitry. Note nodes are sorted in ascending order of the weighted degree.

FIGURE 13: Average Controllability Across Spatial Scales Surface visualizations of the ranked average controllability (AC) values over the 5 spatial scales of the Lausanne atlas [Hagmann et al. (2008a)].

FIGURE 14: Modal Controllability Across Spatial Scales Surface visualizations of the ranked modal controllability (MC) values over the 5 spatial scales of the Lausanne atlas [Hagmann et al. (2008a)].
FIGURE 15: **Boundary Controllability Across Spatial Scales**

Surface visualizations of the ranked boundary controllability (BC) values over the 5 spatial scales of the Lausanne atlas [Hagmann et al. (2008a)].

FIGURE 16: **Reliability and Conservation of Brain Network Control Properties**

Brain network control properties are reliable across imaging acquisition and conserved in non-human primates. Scatter plots of weighted degree (ranked for all brain regions) versus (A,D) average controllability (Pearson correlation coefficient $r = 0.88$, $p = 1.0 \times 10^{-18}$; $r = 0.90$, $p = 4.9 \times 10^{-34}$), (B,E) modal controllability ($r = -0.99$, $p = 3.9 \times 10^{-179}$; $r = -0.99$, $p = 1.3 \times 10^{-72}$), and (C,F) boundary controllability ($r = 0.14$, $p = 0.028$; $r = -0.19$, $p = 0.074$) for (A,B,C) human diffusion tensor imaging data and (D,E,F) macaque tract tracing data.
FIGURE 17: **Control Roles of Cognitive Systems** Cognitive control hubs are differentially located across cognitive systems. **(A)** Hubs of average controllability are preferentially located in the default mode system. **(B)** Hubs of modal controllability are predominantly located in cognitive control systems, including both the fronto-parietal and cingulo-opercular systems. **(C)** Hubs of boundary controllability are distributed throughout all systems, with the two predominant systems being ventral and dorsal attention systems. Control hubs have been identified at the group level as the 30 regions with the highest controllability values (averaged over replicates and subjects). Raw percentages of control hubs present in each system have been normalized by the number of regions in the cognitive system. By applying this normalization, systems composed of a larger number of regions have the same chance of housing one of the top 30 control hubs as systems composed of a smaller number of regions.

FIGURE 18: **Differential Recruitment of Cognitive Systems to Network Control** Average controllability (AC), modal controllability (MC), and boundary controllability (BC) hubs are differentially located in default mode **(A)**, fronto-parietal and cingulo-opercular cognitive control **(B)**, and attentional control **(C)** systems. Values are averaged over the 3 replicates for each individual; error bars indicate standard deviation of the mean over subjects.
FIGURE 19: Control Roles of Cognitive Systems. Cognitive control hubs are differentially located across cognitive systems. (Left) Hubs of average controllability are preferentially located in the default mode system. (Middle) Hubs of modal controllability are predominantly located in cognitive control systems, including both the fronto-parietal and cingulo-opercular systems. (Right) Hubs of boundary controllability are distributed throughout all systems, with the two predominant systems being ventral and dorsal attention systems. These anatomical distributions are consistent across different definitions of control hubs as either (Top) the 25 nodes with the highest control values, (Middle) the top 30 nodes with the highest control values, or (Bottom) the 35 nodes with the highest control values. 

FIGURE 20: Differential Recruitment of Cognitive Systems to Network Control. Average controllability (AC), modal controllability (MC), and boundary controllability (BC) hubs are differentially located in default mode (A), fronto-parietal and cingulo-opercular cognitive control (B), and attentional control (C) systems. These results are consistently observed whether we define control hubs as the 25 nodes with the highest control values (Top Row), the 30 nodes with the highest control values (Middle Row), or the 35 nodes with the highest control values (Bottom Row). Values are averaged over the 3 replicates for each individual; error bars indicate standard deviation of the mean over subjects.
FIGURE 21: **Differential Recruitment of Cognitive Systems to Network Control**

Average controllability (AC), modal controllability (MC), and boundary controllability (BC) hubs are differentially located in default mode (A), fronto-parietal and cingulo-opercular cognitive control (B), and attentional control (C) systems. Values are averaged over the 3 replicates for each individual; error bars indicate standard deviation of the mean over subjects.

FIGURE 22: **Numeric Stability of Trajectories** In this figure, we compute on the average connectivity matrix of th subjects and show (A) The trajectory’s maximal distance to the target state decreases as the number of control nodes increases. (B) The trajectory’s maximal distance to the target state decreases as the balance parameter $\rho$ increases and becomes “stable” when $\rho > 0.01$.

FIGURE 23: **Conceptual Schematic.** (A) Diffusion imaging data can be used to infer connectivity from one voxel to any other voxel via diffusion tractography algorithms. (B) From the tractography, we construct a weighted network in which $N = 234$ brain regions are connected by the quantitative anisotropy along the tracts linking them (see Methods). (C) We study the optimal control problem in which the brain starts from an initial state (red) at time $t = 0$ and uses multi-point control (control of multiple regions; blue) to arrive at a target state (yellow) at time $t = T$. 
FIGURE 24: **Optimal Control Trajectories.** (A) We study 3 distinct types of state transitions in which the initial state is characterized by high activity in the default mode system, and the target states are characterized by high activity in auditory (blue), visual (green), or motor (red) systems. (B) The activation profiles of all $N = 234$ brain regions as a function of time along the optimal control trajectory, illustrating that activity magnitudes vary by region and by time. (C) The average distance from the current state $x(t)$ to the target state $x(T)$ as a function of time for the trajectories from default mode system to the auditory, visual, and motor systems, illustrating behavior in the large state space. (D) The average control energy utilized by the control set as a function of time for the trajectories from default mode system to the auditory, visual, and motor systems. See Fig. S2B for additional information on the range of these control energy values along the trajectories. Colors representing target states are identical in panels (A), (C), and (D).

FIGURE 25: **Structurally-Driven Task Preference for Control Regions.** (A) Left Regions with high control efficiency (see Eqn 4.20) across all 3 state transitions: from the default mode to auditory, visual, and motor systems. Right Scatterplot of the control efficiency with the average network communicability to the target regions. (B–D) Regions with high control efficiency for the transition from default mode to (B) motor, (C) visual, and (D) auditory targets (top); scatter plot of control efficiency versus normalized network communicability with targets (bottom).
Theoretically, the brain is fully controllable when every region is a control point, but may not be fully controllable when fewer regions are used to effect control. (B) The regions with the highest values of energetic impact on control trajectories upon removal from the network, on average across subjects and tasks, were the supramarginal gyrus and the inferior parietal lobule. In general, the healthy group and the mTBI group displayed similar anatomical patterns of energetic impact. (C) Magnitude and standard derivation of energetic impact averaged over regions and tasks; boxplots indicate variation over subjects. After removing outliers in the distribution, patients with mTBI displayed significantly lower values of average magnitude of energetic impact (permutation test: $p = 2.5 \times 10^{-4}$) and lower values of the average standard deviation of energetic impact $(p = 9.1 \times 10^{-5})$ than healthy controls.
CHAPTER 1 : Introduction

1.1. What is a brain network?

The word *network* dates back to the famous Latin phrase *opus reticulatum*, a brickwork spreading the diamond-shape bricks over the wall or ground [Briggs (2004)]. Analogous to the network(graph) in graph theory, which started with Leonhard Euler’s paper *Seven Bridges of Königsberg* published in 1736, the intersection points of the bricks form nodes of a graph(network), and the cracks among the bricks become the edges. The network science has expanded dramatically with the social media [Ellison et al. (2007)] and the brain imaging techniques [Lauterbur et al. (1973); Le Bihan et al. (1986); Thulborn et al. (1982)]. Along with the development in understanding the physics and topological properties of the complex systems since the mid 1990’s [Albert and Barabási (2002); Strogatz (2001); Boccaletti et al. (2006)], network science identifies its role as an interdisciplinary science to characterize and analyze the structure and function of network, which turns out to be a fundamental form of the data structure in this big-data era [McAfee et al. (2012)].

The human brain can be viewed as a network. On micro scale, the neuronal elements constitute the complex brain networks, where the neuron cells consist of the nodes and the nerve fibers consist of the edges. Outside these nerve fibers, there are various kinds of glial cells with oligodendrocytes, which constitute the myelin sheaths surrounding the axons. The diffusion MRI [Basser et al. (1994)] tracks the anisotropic property of water displacement in both the fibers and the myelin sheaths [Beaulieu and Allen (1994)], thus making it possible to unseal the underlined anatomical structure and to construct a structural brain network, where the node can either be a voxel or a coherent region and the edge is computed via measuring the anisotropy between voxels or regions. The most widely applied one is the fractional anisotropy [Koay et al. (2006)]. The more recent diffusion imaging methods DSI (Diffusion Spectrum Imaging) improves the Diffusion Tensor Imaging by better resolving crossing fibers and identifying long-term tracts [Wedeen et al. (2005); Schmahmann...
et al. (2007)]. These diffusion imaging methods help construct the structural or anatomical brain networks. On the other hand, with the development of brain imaging techniques of higher temporal resolution, like functional MRI, EEG, MEG, researchers assemble functional networks with the edges estimated as certain statistical relations between regional signals [Biswal et al. (1995); Greicius et al. (2003); Stam et al. (2007)].

On the regional scale, the nature of the brain images makes the graph theory an appropriate approach to view the brain structure [Iturria-Medina et al. (2007, 2008)]. A bundle of graphical measures has been applied to quantify both the local and global properties of the brain network. The brain network displays high clustering in connectivity and short average path length between regions, thus identifies itself as a small-world network [Watts and Strogatz (1998); Bassett and Bullmore (2006); Bassett et al. (2006)]. Locally the node centrality measure quantifies the node’s importance in the network. The nodes with high contributions are recognized as network hubs. It has been consistently shown in different scales and parcellations that the precuneate and frontal regions play an important role in the process of global communication [Iturria-Medina et al. (2008)]. Also, it has been shown that the medial parietal, frontal and insular regions are confirmed to locate at central network positions [van den Heuvel et al. (2010); Li et al. (2013); van den Heuvel and Sporns (2011)].

Another widely discussed property of the brain network is the module structure. The module structures have been discovered in both the anatomical network [Hagmann et al. (2008b); van den Heuvel and Sporns (2011)] and the functional network [Meunier et al. (2009b); Power et al. (2011)]. Not only do the modules respond to the brain’s segregation of functions [Meunier et al. (2009b); Power et al. (2011)], but the dynamics of the module roles has also been discovered to encode human’s high-level cognitive functions [Bassett et al. (2011b, 2015, 2011b)] and the development of brain systems [Power et al. (2010); Gü et al. (2015b)].
1.2. Why should we use the control theory to view the brain network?

Conceptually, cognitive control is analogous to the mathematical notions of control used in engineering, where the state of a complex system can be modulated by energetic input. Networked systems—like the brain—are particularly interesting systems to control because of the role of the underlying architecture, which predisposes certain components to specific control actions. In the brain, neuronal ensembles or regions (nodes) are interlinked by anatomical wires (edges) in a complex architecture that has an impact on neural function, development, disease and rehabilitation. It is plausible that the brain could regulate cognitive function by a transient network-level control process akin to those engineered in technological, social and cyberphysical systems. Yet, an exact understanding of the relationship between mathematical measures of controllability and notions of cognitive control from neuroscience remains elusive.

Practically, now that the brain networks have been identified as a small-world network with module structures, researchers have now turned to quantifying the dynamics of brain systems. Not only are the control sets in control theory analogous to the brain’s control systems, but the methodologies and algorithms developed in control theory enable us to

i) analyze the energy cost and efficiency of the brain system with specific constraints,

ii) quantify controllability of the brain modules and the importance of the control sets,

iii) compare the trajectories with respect to different state transitions and control sets.

Thus, the control theory potentially builds a bridge between the anatomical structure of the brain and its dynamics of the cognitive control and provides the technical tools to analyze the non-spontaneous dynamics of a complex system—the human brain—with controllers.
1.3. Mathematical concepts

We can have a better understanding of the brain’s structure and dynamics when we describe it in mathematical concepts. In this section, we will first summarize the definitions of some basic graphical metrics and then introduce the setting of network control theory.

1.3.1. Network Measures

Based on the type of edges, a network can be described as a binary and undirected network or weighted and directed network. Here we summarize some basic metrics in the following paragraphs. For more inclusive and comprehensive definition and examples, one can refer to the review by M. Rubinov and O. Sporns [Rubinov and Sporns (2010)].
Figure 1: **Conceptual Network.** For a given network constructed with the two basic elements "nodes" and "edges", we can define various types of graphical metrics like the degrees, shortest path lengths and local structure like triangle and communities. From these fundamental metrics, we are able to define more complex measurements to quantify the topological structure of the network.
A graph is denoted as $G = (V, E)$, where $V$ is the set of nodes and $E$ is the set of edges. The associated adjacent matrix $A$ with $G$ is defined as $A = \{a_{ij}\}$ such that $a_{ij} = 1$ if and only if $e_{ij}$ is an existing edge in $E$. The definition for the weighted and directed is similar to the binary and undirected network but exclude the constraint that $a_{ij} = a_{ji}$. The associated matrix $W = \{w_{ij}\}$ denotes the weight of edges $E = \{e_{ij}\}$.

**Degree** Degree of node $i$ is defined as the sum over the connections related to node $i$: 

$$ k_i = \sum_{j \in V} a_{ij}. \quad (1.1) $$

Weighted degree of $i$: $k^w_i = \sum_{j \in V} w_{ij}$

(Directed) out-degree of node $i$: $k^\text{out}_i = \sum_{j \in V} a_{ij}$.

(Directed) in-degree of node $i$: $k^\text{in}_i = \sum_{j \in V} a_{ji}$.

**Shortest path length** Shortest path length between node $i$ and node $j$ is defined as the length of the shortest path starting from node $i$ and ending at node $j$:

$$ d_{ij} = \sum_{a_{uv} \in g_{\{i \leftarrow j\}}} a_{uv}, \quad (1.2) $$

where $g_{\{i \leftarrow j\}}$ is the shortest path (geodesic) between region $i$ and $j$. The weighted and directed shortest path length between node $i$ and node $j$: $d^W_{ij}$ is defined correspondingly with the weighted and directed geodesics as well as the map $w_{ij} \rightarrow f(w_{uv})$ from the weight to distance.

**Characteristic path length** Characteristic path length of the network is defined as average of shortest paths:

$$ L = \frac{1}{n} \sum_{i \in V} L_i, \quad (1.3) $$

where $L_i$ is the average distance between node $i$ and all other nodes. For the weighted and directed version, replace the $L_i$ with the weighted average distance $L^W_i$ and directed
average distance $L_i^D$.

**Global efficiency** Global efficiency of the network is defined as the average of the inversed path lengths:

$$E = \frac{1}{n} \sum_{i \in V} E_i = \frac{1}{n} \sum_{i \in V} \frac{\sum_{j \in V, j \neq i} d_{ij}^{-1}}{n - 1},$$

(1.4)

where $E_i$ is the efficiency of node $i$. For the weighted and directed version, replace the $d_{ij}$ with the weighted distance $d_{ij}^W$ and the directed distance $d_{ij}^D$.

**Clustering Coefficient** Clustering coefficient of the network is defined as the ratio of existing triangles to the possible number of triangles:

$$C = \frac{1}{n} \sum_{i \in V} C_i = \frac{1}{n} \sum_{i \in V} \frac{2t_i}{k_i(k_i - 1)},$$

(1.5)

where $C_i$ is the clustering coefficient of node $i$ and $t_i$ is the number of triangles around node $i$. For the weighted version, $C_i^W = \frac{1}{n} \sum_{i \in V} C_i^W = \frac{1}{n} \sum_{i \in V} \frac{2t_i^W}{k_i^W(k_i^W - 1)}$. For the directed version, $C_i^D = \frac{1}{n} \sum_{i \in V} \frac{t_i^D}{(k_i^{out} + k_i^{in})(k_i^{out} + k_i^{in} - 1) - 2 \sum_{j \in V} a_{ij}a_{ji}}$.

**Modularity** Modularity of the network if defined as the normalized sum of the modularity function over the given partition:

$$Q = \frac{1}{2m} \sum_{i,j \in V} (a_{ij} - \frac{k_i k_j}{2m}) \delta_{c_i c_j},$$

(1.6)

where $c_i$ is the cluster label of region $i$ and $2m = \sum_{i,j \in V} a_{ij}$. For the weighted version, $Q_i^W = \frac{1}{2m^W} \sum_{i,j \in V} (w_{ij} - \frac{k_i^W k_j^W}{2m^W}) \delta_{c_i c_j}$. For the directed version, $Q_i^D = \frac{1}{2m^D} \sum_{i,j \in V} (a_{ij} - \frac{k_i^{out} k_j^{in}}{2m^D}) \delta_{c_i c_j}$.

**Closeness Centrality** Closeness centrality of node $i$ is defined as the inverse of the average path lengths:

$$L_i^{-1} = \frac{n - 1}{\sum_{i \in V, j \neq i} d_{ij}}.$$

(1.7)

For the weighted and directed version, replace the $d_{ij}$ with $d_{ij}^W$ and $d_{ij}^D$. 

7
**Participation Coefficient** The participation coefficient of node $i$ is defined as the squared ratio of inter-modular connections to the full connections:

$$c_i = 1 - \sum_{c \in C} \left( \frac{k_i(m)}{k(i)} \right)^2,$$

(1.8)

where the $C$ is the collection of clusters or modules. For the weighted and directed version, replace the $k_i$ with $k_i^W$ and $k_i^{in}$ and $k_i^{out}$.

**Small-worldness** Small-worldness of the network is defined as the quotient of the normalized clustering coefficient and the normalized characteristic path length:

$$S = \frac{C/C_{rand}}{L/L_{rand}}.$$  

(1.9)

For the weighted and directed version, replace the clustering coefficient $C$ with $C^W$ and $C^D$ and the characteristic path lengths with $L^W$ and $L^D$. For a typical small-world network, the measure $S \gg 1$.

### 1.3.2. Controllability Settings

Mathematically speaking, we can study the controllability of a networked system by defining a network represented by the graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ are the vertex and edge sets, respectively. Let $a_{ij}$ be the weight associated with the edge $(i, j) \in \mathcal{E}$, and define the *weighted adjacency matrix* of $G$ as $A = [a_{ij}]$, where $a_{ij} = 0$ whenever $(i, j) \notin \mathcal{E}$. We associate a real value (*state*) with each node, collect the node states into a vector (*network state*), and define the map $x : \mathbb{N}_{\geq 0} \rightarrow \mathbb{R}^n$ to describe the evolution (*network dynamics*) of the network state over time. Given the network and node dynamics, we can use network control theory to quantitatively examine how the network structure constrains the types of control that nodes can exert.

To define the dynamics of neural processes, we draw on prior models linking structural brain networks to resting state functional dynamics. Although neural activity evolves through
neural circuits as a collection of nonlinear dynamic processes, these prior studies have demonstrated that a significant amount of variance in neural dynamics can be predicted from simplified linear models. Based on this literature, we employ a simplified noise-free linear discrete-time and time-invariant network model:

\[ x(t + 1) = Ax(t) + B_K u_K(t), \]  

(1.10)

where \( x : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^N \) describes the state (i.e., a measure of the electrical charge, oxygen level, or firing rate) of brain regions over time, and \( A \in \mathbb{R}^{N \times N} \) is a symmetric and weighted adjacency matrix. In our case, we construct a weighted adjacency matrix whose elements indicate the number of white matter streamlines connecting two different brain regions – denoted here as \( i \) and \( j \) – and we stabilize this matrix by dividing by the mean edge weight. While the model employed above is a discrete-time system, we find that the controllability Gramian is statistically similar to that obtained in a continuous-time system. The diagonal elements of the matrix \( A \) satisfy \( A_{ii} = 0 \). The input matrix \( B_K \) identifies the control points \( K \) in the brain, where \( K = \{k_1, \ldots, k_m\} \) and

\[ B_K = \begin{bmatrix} e_{k_1} & \cdots & e_{k_m} \end{bmatrix}, \]  

(1.11)

and \( e_i \) denotes the \( i \)-th canonical vector of dimension \( N \). The input \( u_K : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^m \) denotes the control strategy. To determine the trajectory from an initial state \( x_0 \) to a target state \( x_T \), we expose the evolution dynamics with boundary condition and energy optimization as following:

\[ \min_u \int_0^T (x_T - x)^T (x_T - x) + \rho u^T u, \]  

s.t. \[ \dot{x}(t) = Ax + Bu, \]  

(1.12)

\[ x(0) = x_0, \]  
\[ x(T) = x_T, \]
where $T$ is the control horizon, and $\rho \in \mathbb{R}_{\geq 0}$. To study the ability of a certain brain region to influence other regions in arbitrary ways we adopt the control theoretic notion of controllability. Controllability of a dynamical system refers to the possibility of driving the state of a dynamical system to a specific target state by means of an external control input. Classic results in control theory ensure that controllability of the network (1.10) from the set of network nodes $\mathcal{K}$ is equivalent to the controllability Gramian $W_{\mathcal{K}}$ being invertible, where

$$W_{\mathcal{K}} = \sum_{\tau=0}^{\infty} A^{\tau} B_{\mathcal{K}} B_{\mathcal{K}}^{T} A^{\tau}.$$  \hspace{1cm} (1.13)

1.4. Contribution of my works

Despite the existing literatures about the controllability of complex networks [Liu et al. (2011); Sorrentino et al. (2007); Yuan et al. (2013)] and the cognitive control of brain systems [Miller (2000); Ochsner and Gross (2005); MacDonald et al. (2000); Ridderinkhof et al. (2004); Power et al. (2011)], we are the first to propose a systematic framework of applying the control theory to analyzing the brain networks and to exploring the possibility of explaining the cognitive control property based on their underlined anatomical structures [Gu et al. (2015a)]. Further, we develop the scheme via proper considerations of the energy issues associated with both of the states and trajectories, thus extend the model from diagonalizing the static networks to quantifying the dynamical properties of multiple state transitions in the brain. Our works establish a new subfield in the network neuroscience as the control theoretic analysis of brain networks.
2.1. Introduction

A human’s adaptability to rapidly changing environments depends critically on the brain’s ability to carefully control the time within (and transitions among) different states. Here, we use the term *state* to refer to a pattern of activity across neurons or brain regions [Tang et al. (2012)]. The recent era of brain mapping has beautifully demonstrated that the pattern of activity across the brain or portions thereof [Mahmoudi et al. (2012)] differs in different cognitive states [Gazzaniga (2013)]. These variable patterns of activity have enabled the study of cognitive function via the manipulation of distinct task elements [Gazzaniga (2013)], the combination of task elements [Szameitat et al. (2011); Alavash et al. (2015)], or the temporal interleaving of task elements [Ruge et al. (2013); Muhle-Karbe et al. (2014)]. Such methods for studying cognitive function are built on the traditional view of mental chronectomy [Donders (1969)], which suggests that brain states are additive and therefore separable in both space and time (although see [Mattar et al. (2015)] for a discussion of potential caveats).

Philosophically, the supposed separability and additivity of brain states suggests the presence of strong constraints on the patterns of activations that can be elicited by the human’s environment. The two most common types of constraints studied in the literature are energetic constraints and structural constraints [Bullmore and Sporns (2012)]. Energetic constraints refer to fundamental limits on the evolution [Niven and Laughlin (2008)] or usage of neural systems [Attwell and Laughlin (2001)], which inform the costs of establishing and maintaining functional connections between anatomically distributed neurons [Bassett et al. (2010)]. Such constraints can be collectively studied within the broad theory of brain function posited by the free energy principal – a notion drawn from statistical physics and information theory – which states that the brain changes its state to minimize the free
energy in neural activity [Friston et al. (2006); Friston (2010)]. The posited preference for low energy states motivates an examination of the time within and transitions among local minimums of a predicted energy landscape of brain activity [Moreno-Bote et al. (2007); Tsodyks et al. (1998)].

While energetic costs likely form critical constraints on functional brain dynamics, an arguably equally important influence is the underlying structure and anatomy linking brain areas. Intuitively, quickly changing the activity of two brain regions that are not directly connected to one another by a structural pathway may be more challenging than changing the activity of two regions that are directly connected [Achard and Bullmore (2007); Bassett et al. (2010)]. Indeed, the role of structural connectivity in constraining and shaping brain dynamics has been the topic of intense neuroscientific inquiry in recent years [Honey et al. (2009, 2010); Deco and Jirsa (2012); Goni et al. (2014); Becker et al. (2016)]. Evidence suggests that the pattern of connections between brain regions directly informs not only the ease with which the brain may control state transitions [Gu et al. (2015a)], but also the ease with which one can externally elicit a state transition using non-invasive neurostimulation [Muldoon et al. (2016b)].

While energy and anatomy both form critical constraints on brain dynamics, they have largely been studied in isolation, hampering an understanding of their collective influence. Here, we propose a novel framework that combines energetic and structural constraints on brain state dynamics in a free energy model explicitly informed by structural connectivity. Using this framework, we map out the predicted energy landscape of brain states, identify local minima in the energy landscape, and study the profile of activation patterns present in these minima. Our approach offers fundamental insights into the distinct role that brain regions and larger cognitive systems play in distributing energy to enable cognitive function. Further, the results lay important groundwork for the study of energy landscapes in psychiatric disease and neurological disorders, where brain state transitions are known to be critically altered but mechanisms driving these alterations remain far from understood.
[Ravizza et al. (2010); Wylie et al. (2010)].
Figure 2: Conceptual Schematic. (A) A weighted structural brain network represents the number of white matter streamlines connecting brain regions. (B) We model each brain region as a binary object, being either active or inactive. (C) Using a maximum entropy model, we infer the full landscape of predicted activity patterns – binary vectors indicating the regions that are active and the regions that are not active – as well as the energy of each pattern (or state). We use this mathematical framework to identify and study local minima in the energy landscape: states predicted to form the foundational repertoire of brain function.
2.2. Models and Materials

2.2.1. Human DSI Data Acquisition and Preprocessing

Diffusion spectrum images (DSI) were acquired for a total of 48 subjects (mean age 22.6±5.1 years, 24 female, 2 left handed) along with a T1 weighted anatomical scan at each scanning session \[\text{Cieslak and Grafton (2014)}\]. Of these subjects, 41 were scanned once, 1 was scanned twice, and 6 were scanned three times, for a total of 61 scans.

DSI scans sampled 257 directions using a Q5 half shell acquisition scheme with a maximum \(b\) value of 5000 and an isotropic voxel size of 2.4mm. We utilized an axial acquisition with the following parameters: \(TR = 11.4s, TE = 138ms, 51\) slices, FoV (231,231,123 mm). All participants volunteered with informed written consent in accordance with the Institutional Review Board/Human Subjects Committee, University of California, Santa Barbara.

DSI data were reconstructed in DSI Studio (www.dsi-studio.labsolver.org) using \(q\)-space diffeomorphic reconstruction (QSDR) \[\text{Yeh and Tseng (2011)}\]. QSDR first reconstructs diffusion weighted images in native space and computes the quantitative anisotropy (QA) in each voxel. These QA values are used to warp the brain to a template QA volume in MNI space using the SPM nonlinear registration algorithm. Once in MNI space, spin density functions were again 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 CSF. Deterministic fiber tracking using a modified FACT algorithm was performed until 100,000 streamlines were reconstructed for each individual.

2.2.2. Structural Network Construction

Anatomical scans were segmented using FreeSurfer \[\text{Dale et al. (1999)}\] and parcellated according to the Lausanne 2008 atlas included in the connectome mapping toolkit \[\text{Hagmann}\].
et al. (2008a)]. A parcellation scheme including 234 regions was registered to the B0 volume from each subject’s DSI data. The B0 to MNI voxel mapping produced via QSDR was used to map region labels from native space to MNI coordinates. To extend region labels through the gray/white matter interface, the atlas was dilated by 4mm. Dilation was accomplished by filling non-labeled voxels with the statistical mode of their neighbors’ labels. In the event of a tie, one of the modes was arbitrarily selected. Each streamline was labeled according to its terminal region pair.

From these data, we built structural brain networks from each of the 61 diffusion spectrum imaging scans. Consistent with previous work [Bassett et al. (2010, 2011a); Hermundstad et al. (2013, 2014); Klimm et al. (2014); Gu et al. (2015a); Muldoon et al. (2016a,b); Sizemore et al. (2015)], we defined these structural brain networks from the streamlines linking $N = 234$ large-scale cortical and subcortical regions extracted from the Lausanne atlas [Hagmann et al. (2008a)]. We summarize these estimates in a weighted adjacency matrix $A$ whose entries $A_{ij}$ reflect the structural connectivity between region $i$ and region $j$ (Fig. 2A).

Following [Gu et al. (2015a)], here we use an edge weight definition based on the quantitative anisotropy (QA). QA is described by Yeh et. al (2010) as a measurement of the signal strength for a specific fiber population $\hat{a}$ in an ODF $\Psi(\hat{a})$ [Yeh et al. (2010); Tuch (2004)]. QA is given by the difference between $\Psi(\hat{a})$ and the isotropic component of the spin density function (SDF, $\psi$) $\text{ISO}(\psi)$ scaled by the SDF’s scaling constant. Along-streamline QA was calculated based on the angles actually used when tracking each streamline. Although along-streamline QA is more specific to the anatomical structure being tracked, QA is more sensitive to MRI artifacts such as B1 inhomogeneity. QA is calculated for each streamline. We then averaged values over all streamlines connecting a pair of regions, and used this value to weight the edge between the regions.
2.2.3. Resting state fMRI data

As an interesting comparison to the computational model, we used resting state fMRI data collected from an independent cohort composed of 25 healthy right-handed adult subjects (15 female), with a mean age of 19.6 years (STD 2.06 year). All subjects gave informed consent in writing, in accordance with the Institutional Review Board of the University of California, Santa Barbara. Resting-state fMRI scans were collected on a 3.0-T Siemens Tim Trio scanner equipped with high performance gradients at the University of California, Santa Barbara Brain Imaging Center. A T2*-weighted echo-planar imaging (EPI) sequence was used (TR=2000 ms; TE=30 ms; flip angle=90; acquisition matrix=64×64; FOV=192 mm; acquisition voxel size = 3×3×3.5 mm; 37 interleaved slices; acquisition length=410s).

We preprocessed the resting state fMRI data using an in-house script adapted from the workflows described in detail elsewhere [Baird et al. (2013); Satterthwaite et al. (2013)]. The first four volumes of each sequence were dropped to control for instability effects of the scanner. Slice timing and motion correction were performed in AFNI using 3dvolreg and FreeSurfer’s BBRegister was used to co-register functional and anatomical spaces. Brain, CSF, and WM masks were extracted, the time series were masked with the brain mask, and grand-mean scaling was applied. The temporal derivative of the original 6 displacement and rotation motion parameters was obtained and the quadratic term was calculated for each of these 12 motion parameters, resulting in a total of 24 motion parameters which were regressed from the signal. The principal components of physiological noise were estimated using CompCor (aCompCor and tCompCor) and these components were additionally regressed from the signal. The global signal was not regressed. Finally, signals were low passed below 0.1 Hz and high passed above 0.01 Hz in AFNI. To extract regional brain signals from the voxel-level time series, a mask for each brain region in the Lausanne2008 atlas was obtained and FreeSurfer was used to individually map regions to the subject space. A winner-takes-all algorithm was used to combine mapped regions into a single mask. The resulting signal for each region was then extracted in FreeSurfer using mrisegstats.
Following data preprocessing and time series extraction, we next turned to extracting observed brain states. Importantly, physiological changes relating to neural computations take place on a time scale much smaller than the time scale of BOLD image acquisition. Thus, we treat each TR as representing a distinct brain state. To maximize consistency between the model-based and data-based approaches, we transformed the continuous BOLD magnitude values into a binary state vector by thresholding regional BOLD signals at 0. From the set of binary state vectors across all TRs, we defined activation rates in a manner identical to that described for the maximum entropy model data.

2.2.4. Defining an Energy Landscape.

We begin by defining a brain state both intuitively and in mathematical terms. A brain state is a macroscopic pattern of BOLD activity across \( K \) regions of the brain (Fig. 2A). For simplicity, here we study the case in which each brain region \( i \) can be either active (\( \sigma_i = 1 \)) or inactive (\( \sigma_i = 0 \)). Then, the binary vector \( \sigma = (\sigma_1, \sigma_2, \ldots, \sigma_K) \) represents a brain state configuration.

Next, we wish to define the energy of a brain state. We build on prior work demonstrating the neurophysiological relevance of maximum entropy models in estimating the energy of brain states in rest and task conditions [Watanabe et al. (2013, 2014)]. For a given state \( \sigma \), we write its energy in the second order expansion:

\[
E(\sigma) = - \frac{1}{2} \sum_{i \neq j} J_{ij} \sigma_i \sigma_j - \sum_i J_i \sigma_i, \tag{2.1}
\]

where \( J \) represents an interaction matrix whose elements \( J_{ij} \) indicate the strength of the interaction between region \( i \) and region \( j \). If \( J_{ij} > 0 \), this edge \((i, j)\) decreases the energy of state \( \sigma \), while if \( J_{ij} < 0 \), this edge \((i, j)\) increases the energy of state \( \sigma \). The column sum of the structural brain network, \( J_i = \sum_j |J_{ij}|/\sqrt{K} \), is the strength of region \( i \). In the thermodynamic equilibrium of the system associated with the energy defined in Eqn 2.1, the entropy of the system is maximized and the probability of the configuration \( \sigma \) is
The choice of the interaction matrix is an important one, particularly as it tunes the relative contribution of edges to the system energy. In this study, we seek to study structural interactions in light of an appropriate null model. We therefore define the interaction matrix to be equal to the modularity matrix [Newman (2006)] of the structural brain network:

\[ J_{ij} = \frac{1}{2m} (A - pp^T/2m)_{ij} \]  

(2.2)

for \( i \neq j \), where \( A \) is the adjacency matrix, \( p_i = \sum_{i=1}^{K} A_{ij} \), and \( 2m = \sum_{j=1}^{K} p_j \). This choice ensures that any element \( J_{ij} \) measures the difference between the strength of the edge \( A_{ij} \) in the brain and the expected strength of that edge in an appropriate null model (here given as the Newman-Girvan null model [Clauset et al. (2004)]). If the edge is stronger than expected, it will decrease the energy of the whole system when activated, while if the edge is weaker than expected, it will increase the energy of the whole system when activated.

2.2.5. Discovering Local Minima.

The model described above provides an explicit correspondence between a brain’s state and the energy of that state, in essence formalizing a multidimensional landscape on which brain dynamics may occur. We now turn to identifying and characterizing the local minima of that energy landscape (Fig. 2C). We begin by defining a local minimum: a binary state \( \sigma^* = (\sigma^*_1, \ldots, \sigma^*_K) \) is a local minimum if \( E(\sigma) \geq E(\sigma^*) \) for all vectors \( \sigma \) satisfying \( ||\sigma - \sigma^*||_1 = 1 \), which means that the state \( \sigma^* \) realizes the lowest energy among its neighboring states within the closed unit sphere. We wish to collect all local minima in a matrix \( \Sigma^* \) with

\[
\Sigma^* = \begin{pmatrix}
\sigma^*_{1,1} & \sigma^*_{1,2} & \cdots & \sigma^*_{1,N} \\
\sigma^*_{2,1} & \sigma^*_{2,2} & \cdots & \sigma^*_{2,N} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma^*_{K,1} & \sigma^*_{K,2} & \cdots & \sigma^*_{K,N}
\end{pmatrix}_{K \times N}
\]  

(2.3)
where \( N \) is the number of local minima and \( K \) is the number of nodes in the structural brain network (or equivalently the cardinality of the adjacency matrix \( A \)).

Now that we have defined a local minimum of the energy landscape, we wish to discover these local minima given the pattern of white matter connections represented in structural brain networks. To discover local minima of \( E(\sigma) \), we first note that the total number of states \( \sigma = (\sigma_1, \ldots, \sigma_K) \) is \( 2^K \), which – when \( K = 234 \) – prohibits an exhaustive analysis of all possibilities. Moreover, the problem of finding the ground state is an NP-complete problem [Cipra (2000)], and thus it is unrealistic to expect to identify all local minima of a structural brain network. We therefore choose to employ a clever heuristic to identify local minima. Specifically, we combine the Metropolis Sampling method [Metropolis et al. (1953)] and a steep search algorithm using gradient descent methods. We identify a starting position by choosing a state uniformly at random from the set of all possible states. Then, we step through the energy landscape via a random walk driven by the Metropolis Sampling criteria (see Algorithm 1). At each point on the walk, we use the steep search algorithm to identify the closest local minimum.

**Algorithm 1:** Heuristic Algorithm to Sample the Energy Landscape to Identify Local Minima.

1. Let \( \sigma^j \) be the vector obtained by changing the value of the \( j \)-th entry of \( \sigma \);
2. for \( t = 1 : N \) do
   3. Randomly select an index \( j \in \{1, \ldots, K\} \);
   4. Set \( \tilde{\sigma}_t = \sigma_t = \sigma^j_{t-1} \) with probability \( p = \min(1, e^{-\beta*(E(\tilde{\sigma}_t) - E(\sigma_{t-1}))}) \) and \( \tilde{\sigma}_t = \sigma_t = \sigma_{t-1} \) otherwise;
   5. while \( \tilde{\sigma}_t \) is not a local minimum do
      6. Set \( j^* = \arg \min_j E(\tilde{\sigma}_t^j) \);
      7. Set \( \tilde{\sigma}_t = \tilde{\sigma}_t^{j^*} \);
   8. Set \( \sigma^*_t = \tilde{\sigma}_t \);

Here, \( \sigma_1, \sigma_2, \ldots, \sigma_N \) are the sampled states, \( \sigma^*_1, \sigma^*_2, \ldots, \sigma^*_N \) are the sampled local minima,
and $\beta$ is the temperature parameter which can be absorbed in $E(\sigma)$. In the context of any sampling procedure, it is important to determine the number of samples necessary to adequately cover the space. Theoretically, we wish to identify a number of samples $N$ following which the distribution of energies of the local minima remains stable. Practically speaking, we choose 4 million samples in this study, and demonstrate the stability of the energy distribution in the Supplement. A second important consideration is to determine the initial state, that is the state from which the random walk begins. Here we choose this state uniformly at random from the set of all possible states. However, this dependence on a uniform probability distribution may not be consistent with the actual probability of states in the energy landscape. We therefore must ensure that our results are not dependent on our choice of initial state. To ensure independence from the initial state, we dismiss the first 30,000 local minima identified, and we demonstrate in the Supplement that this procedure ensures our results are not dependent on the choice of initial state.

2.2.6. Characterizing Local Minima.

Following collection of local minima, we wished to characterize their nature as well as their relationships to one another. First, we estimated the radius of each local minimum as the Hamming distance from the minimum to the closest sampled point on energy landscape. Next, calculated the Hamming distance from each local minimum to the first sampled local minimum, a second quantification of the diversity of the energy landscape that we traverse in our sampling. Finally, we quantify how diverse the observed local minima are by calculating the pairwise normalized mutual information \cite{Manning et al. 2008} of each pair of local minima.

Next, we wished to understand the role of different regions and cognitive systems in the minimal energy states. Cognitive systems are sets of brain regions that show coordinated activity profiles in resting state or task-based BOLD fMRI data \cite{Mattar et al. 2015}. They include the visual, somatosensory, motor, auditory, default mode, salience, fronto-parietal, cingulo-opercular, dorsal and ventral attention systems, as well as subcortical areas. Here,
the specific association of regions of interest to cognitive systems are exactly as listed in 
Gu et al. (2015a) and based originally on results described in Power et al. (2011). We
characterize the roles of these systems in the local minima by assessing their activation
rates, as well as the utilization energies required for communication within and between
systems.

2.2.7. Activation Rates

Intuitively, we define the activation rate of a node \( i \) as the average activation state of that
node over all the local minima. Formally, the activation rate for region \( i \) is defined as

\[
gr_i = \frac{\sum_{l=1}^{N} \sigma^*_{il}}{N},
\]

where \( l \) indexes over states, and recall \( N \) is the number of local minima. The computed
activation rate offers a prediction of which regions are more versus less active across the
local minima (that is, the brain’s locally “stable” states), and can be directly compared
with the resting state activation rate estimated from empirical fMRI data.

2.2.8. Utilization Energies

To complement the information provided by the activation rates, we also defined the en-
ergetic costs associated with utilizing within- and between-systems interactions. We note
that each cognitive system is a subnetwork of the whole brain network. We use the index
set \( I \) to indicate the set of nodes associated with the cognitive system, and thus \( |I| \) gives
the number of nodes in the system. Then, for a given state \( \sigma \), the within-system energy
measures the cost associated with the set of interactions constituting the subnetwork. The
between-system energy measures the cost associated with the set of interactions between
the subnetwork and all other nodes in the whole network. Formally, we define
\[ E^W(\sigma) = -\frac{1}{2|I|(|I|-1)} \left( \sum_{i \neq j, i, j \in I} J_{ij} \sigma_i \sigma_j \right) \]

\[ E^B(\sigma) = -\frac{1}{2|I|(K-|I|)} \left( \sum_{i \in I, j \notin I} J_{ij} \sigma_i \sigma_j \right) \]

where \( E^W \) measures the within-system energy, \( E^B \) measures the between-system energy, and the normalization coefficients \( 1/(|I||I|-1) \), \( 1/(|I|(K-|I|)) \) are chosen by considering the number of the corresponding interactions.

### 2.2.9. Permutation Tests for State Association

For a given local minimum configuration \( \sigma^* \), we associate it with system \( i_{\sigma^*} \),

\[ i_{\sigma^*} = \arg \max_i \text{NMI}(\sigma^*, \sigma^{\text{sys}}_i) \]

where \( \sigma^{\text{sys}}_i \) is the configuration pattern of system \( i \) such that the corresponding regions for system \( i \) are activated and the others not and where “NMI” refers to the Normalized Mutual Information [Manning et al. (2008)], which is used to measure the similarities between the two states. To obtain the null distribution, for each local minimum configuration \( \sigma^* \) in the collection \( \Sigma^* \), we permute the configuration at each position of \( \sigma^* \) to achieve a random configuration with the same activation rate, and we then compute the associated percentage in each system. Then we repeat this procedure to generate \( N \) samples and construct the null distribution of the probability of being configured as each system pattern. Considering the large size of the state collection, the variance of the samples in the null distribution will be small. We pick \( N = 50 \) here. See Fig. 8D for the results.
2.2.10. Predicted Functional Connectivity

The approach outlined above offers an assessment of activity states expected in the brain given the pattern of underlying structural connectivity and the expectation that the brain seeks to minimize energy expenditure. In addition to these direct outputs, the approach also offers a means to predict expected functional connectivity. For two regions $i$ and $j$, their predicted functional connectivity is defined as the similarity between their activation patterns through the $N$ local minima. Let $\mathbf{\sigma}_j = (\sigma_{j,1}, \sigma_{j,2}, \ldots, \sigma_{j,N})$ and $k(\cdot, \cdot): \mathbf{\sigma} \times \mathbf{\sigma} \rightarrow \mathbb{R}$ be a pre-defined similarity function (here we use a Pearson correlation). The predicted functional connectivity matrix $\mathbf{C} = (c_{ij})_{K \times K}$ is then defined as $c_{ij} = k(\mathbf{\sigma}_i, \mathbf{\sigma}_j)$. In the body of the results section of this manuscript, we assess the relationship between the functional connectivity predicted by the model, and the functional connectivity observed in resting state fMRI data collected from an independent cohort.

2.3. Results

2.3.1. Local Minima in the Brain’s Energy Landscape

By sampling the energy landscape of each structural connectivity matrix, we identified an average of approximately 450 local minima or low energy brain states: binary vectors indicating the pattern of active and inactive brain regions (see Methods). On average across the 61 scans, 144 brain regions were active in a given local minimum, representing 61.70% of the total (standard deviation: 6.04%). This large percentage suggests that the brain may utilize a broad set of rich and diverse activations to perform cognitive functions [Cha (2016)], rather than activation patterns localized to small geographic areas.

To quantify this diversity, we examined the location of minima on the energy landscape, the size of the basins surrounding the minima, and the mutual information between minima. First, we estimated the distance from the first local minima identified to all subsequent minima (see Methods; Fig. 3A). We observe an order of magnitude change in the distance between the first and second local minima, and the first and last local minima, suggesting
that local minima span a broad geographic domain in the energy landscape. Interestingly, these minima differ not only in their location on the energy landscape, but also in the size of the basins surrounding them. We estimate basin size by calculating the radius of each local minimum (see Methods) and show that the distribution of radii follows a power-law, with the majority of minima displaying a small radius, and only a few minima displaying a large radius (Fig. 3B). More specifically, we fit the function $P(r) = Cr^{-\alpha}$ – where $C$ is a constant – to the data using a statistically principled approach [Clauset et al. (2009); Y and Clauset (2014)]. We identified an $\alpha = 2.6300$ for $r > 6$, and the $p$-value for the goodness of fit was $p < 1 \times 10^{-5}$ indicating that the power law was an accurate fit to the data. As a final quantification of minima diversity, we estimated the normalized mutual information between every pair of local minima, as an intuitive measurement of the similarity between anatomical compositions of the minima. We observe that the probability distribution of normalized mutual information between minima pairs is heavy tailed, indicating that most minima pairs display very dissimilar anatomical compositions, and only a few minima pairs display similar anatomical compositions (Fig. 3C).

From a neurophysiological perspective, it is also important to note that these local minima displayed significant local structure. Specifically, we found that regions within known cognitive systems tended to be active together. The probability that regions were co-active is 48.22%, which was significantly greater than that expected in a null distribution (associated $p$-value was $p < 1 \times 10^{-5}$; see Methods). These results indicate that the structural connectivity between brain regions, and the assumption of energy conservation, predict that regions that belong to the same cognitive system will tend to be co-active with one another during diverse cognitive functions. Indeed, these predictions are consistent with previous studies of functional neuroimaging data demonstrating that groups co-active regions tended to align well with known cognitive systems [Crossley et al. (2013, 2014)].
Figure 3: Simulated Activation Rates. (A) The distribution of distances from the first local minimum to other local minima. Each point and error-bar is calculated across a bin of 30 minima; error bars indicate standard error of the mean over the 30 minima. (B) The probability distribution of the radius of each local minimum is well-fit by a power-law. The radius of a local minimum is defined as its distance to the closest sampled point on the energy landscape. (C) The distribution of the pairwise normalized mutual information between all pairs of local minima. (D) Average activation rates for all 14 *a priori* defined cognitive systems [Power et al. (2011)]. Error bars indicate standard error of the mean across subjects.
2.3.2. Activation Rates of Cognitive Systems

Given the alignment of activation patterns with cognitive systems, we next asked whether certain cognitive systems were activated more frequently than others. To address this question, we studied the activation rate of each cognitive system, which measures how frequently the regions in the cognitive system participated in the set of states identified as local minima. Intuitively, if the activation rate is high, the system is more likely to be active in diverse brain states. We observed that systems indeed showed significantly different activation rates (Fig. 3D). Sensorimotor systems (auditory, visual, somatosensory) tended to display the lowest activation rates, followed by higher order cognitive systems (salience, attention, fronto-parietal, and cingulo-opercular), and subcortical structures. The system with the largest activate rate was the default mode system, suggesting that activation of this system is particularly explicable from structural connectivity and the assumption of energy conservation. The unique role of the default mode system is consistent with predictions from network control theory that highlight the optimal placement of default mode regions within the network to maximize potential to move the brain into many easily reachable states with minimal energetic costs [Gu et al. (2015a)].

It is important to determine whether this activation rate is driven by simple properties of the structural connectivity matrix that do not depend on assumptions of energy conservation. To address this question, we next assessed the relationship between a simple summary statistic of the structural connectivity matrix – the strength, or weighted degree, of a brain region – and the predicted activation rate drawn from the maximum entropy model. We observed that the activation rate was not well predicted by the weighted degree on average over brain regions (see Supplement). These data suggest that the additional assumption of energy conservation produces a set of brain states that cannot be predicted from simple statistics of structural connectivity alone.
2.3.3. Activation Rates of Cognitive Systems

Given the alignment of activation patterns with cognitive systems, we next asked whether certain cognitive systems were activated more frequently than others. To address this question, we studied the activation rate of each cognitive system, which measures how frequently the regions in the cognitive system participated in the set of states identified as local minima. Intuitively, if the activation rate is high, the system is more likely to be active in diverse brain states. We observed that systems indeed showed significantly different activation rates (Fig. 3D). Sensorimotor systems (auditory, visual, somatosensory) tended to display the lowest activation rates, followed by higher order cognitive systems (salience, attention, fronto-parietal, and cingulo-opercular), and subcortical structures. The system with the largest activate rate was the default mode system, suggesting that activation of this system is particularly explicable from structural connectivity and the assumption of energy conservation. The unique role of the default mode system is consistent with predictions from network control theory that highlight the optimal placement of default mode regions within the network to maximize potential to move the brain into many easily reachable states with minimal energetic costs [Gu et al. (2015a)].

It is important to determine whether this activation rate is driven by simple properties of the structural connectivity matrix that do not depend on assumptions of energy conservation. To address this question, we next assessed the relationship between a simple summary statistic of the structural connectivity matrix – the strength, or weighted degree, of a brain region – and the predicted activation rate drawn from the maximum entropy model. We observed that the activation rate was not well predicted by the weighted degree on average over brain regions (see Supplement). These data suggest that the additional assumption of energy conservation produces a set of brain states that cannot be predicted from simple statistics of structural connectivity alone.
**Sufficiency of the Number of Samples** From Figure [4], we can see that when the total number of samples is over 2000, the energy distribution across the configurations of local minimums remain stable with no rejection for KS test ($p = 0.5066$ for the test between Group 1 with 1st-6000th samples and Group 2 with 2001st-8000th samples), which means the total number of samples here is sufficiently large to represent the energy distribution across the configurations of local minimums.

**Correlation between Regional Energy and Activation Rate** From Figure [5], we can see that the simulated activation rate is positively correlated with regional energy but do not follow a simply linear relation. Together with the main text, this suggests that the structural hubs and modularity hubs are more likely but not necessary to work as functional hubs.

**Comparison between Energy and Degree** From Figure [6], we can see that for different functional groups, the degree display a worse separation compared with the between- and within-system energy. On the one hand, as we show in the previous section the structural hubs and the
Figure 4: **Stability of Energy Distribution with respect to the Number of Sample Size.** We plot the heuristic probability distribution of the energy for the first 2000, 4000, 6000 and 8000 samples. We can observe that the shapes are pretty consistent and further KS-test results ensure the stability.
Figure 5: **Relation between the Regional Energy and the Activation Rate.** Regional Energy and Activation Rate are positively correlated. Here we want to justify that the activation rate is neither dominated by regional energy nor the weighted degrees.
Figure 6: Distribution of Between- and Within- System Energy Tuples. We show that the between- and within- system energy provides C) a better separation for different neural systems compared to the degrees A), especially B,D) for the Default Mode network, Fronto-Parietal, and Cingulo-Opercular Task Control.
2.3.4. Relating Predicted Activation Rates to Rates Observed in Functional Neuroimaging Data

Before exercising the model further to explore how energy is utilized in the brain, we wished to quantify the relationships between the theoretically predicted activation rates, and activation rates observed empirically in functional neuroimaging data. Specifically, we studied fMRI data acquired in a separate set of healthy adult human subjects.

Next, we wished to directly quantify similarities between the predicted activation rates and those observed empirically in resting state fMRI. In the resting state data, we observed that highly active regions were located in broad swaths of frontal and parietal cortex, as well as medial prefrontal, precuneus, and cingulate (Fig. 7A). This pattern of high activation is consistent with the so-called “default-mode” of resting state brain function [Raichle et al. (2001)]. In our maximum entropy model, we observed that the areas predicted to have high activation rates show a broad similarity to those observed empirically in the resting state (Fig. 7B). Indeed, we observed that the empirical resting activation rate of brain regions is significantly correlated with the activation rate predicted from the maximum entropy model (Fig. 7C; Pearson correlation coefficient $r = 0.18$, $p = 0.0046$). These results suggest that the modeling framework we use here has significant similarities to observable features of resting state brain dynamics. However, it is important to mention that there are also noticeable differences between the two maps: the predicted activation rates are strong along the medial wall, while the resting state activation rates extend to larger sections of lateral cortices.
Figure 7: **Validating Predicted Activation Rates in Functional Neuroimaging Data.** (A) From resting state BOLD data acquired in an independent cohort, we estimated the true activation rate by transforming the continuous BOLD magnitudes to binary state vectors by thresholding the signals at 0 (see Methods). We use these binary state vectors to estimate the activation rates of each brain region across the full resting state scan. Here we show the mean activation rate of each brain region, averaged over subjects. (B) For comparison, we also show the mean predicted activation rate estimated from the local minima of the maximum entropy model, as defined in Equation [2.4], and averaged over subjects. (C) We observe that the activation rates estimated from resting state fMRI data are significantly positively correlated with the activation rates estimated from the local minima of the maximum entropy model (Pearson’s correlation coefficient $r = 0.18$, $p = 0.0046$). Each data point represents a brain region, with either observed or predicted activation rates averaged over subjects.
Utilization Energies of Cognitive Systems

We next turned to exercising our model to further understand the potential constraints on brain state dynamics. Specifically, we asked how cognitive systems utilized the minimal energy presumably available to them. Intuitively, this question encompasses both how energy is utilized by within-system interactions, and how energy is utilized by between-system interactions. We therefore defined the within-system energy, which measures the cost associated with the set of interactions constituting the cognitive system, and the between-system energy, which measures the cost associated with the set of interactions between cognitive systems. We observed a fairly strong dissociation between these two variables: cognitive systems that display a large within-system energy are not necessarily those that display a large between-system energy (see Fig. 8A and B). Indeed, within- and between-system energies are not significantly correlated across systems (Pearson’s correlation coefficient $r = 0.2287$ and $p = 0.5250$), suggesting that these two variables offer markers of distinct constraints.

Moreover, we observed that the 2-dimensional plane mapped out by the within- and between-system energies of all brain regions revealed the presence of 4 surprisingly distinct clusters (Fig. 8C) that are not explicable by simple statistics such as network degree (see Supplement). Each cluster represents a unique strategy in energy utilization that is directly reflected in its activation pattern; in other words, each cluster offers a distinct balance between the energetic costs of within-system interactions and the energetic costs of between-system interactions. The central cluster, displaying high within-system energies but low between-system energies, is composed of subcortical and fronto-parietal systems. A high between-system energy cone emanating from this central cluster is composed of predominantly primary and secondary sensorimotor cortices in somatosensory, visual, and auditory systems. A second cone emanating from the central cluster with a slightly lower between-system energy is composed predominantly of regions in the default mode system. The final cone emanating from the central cluster with between-system energies near zero is composed
predominantly of regions in the dorsal and ventral attention systems. These results suggest that sensorimotor, default mode, attention, and cognitive control circuits display differential preferences for energy utilization: regions in attentional systems share less energy with other networks than regions in sensorimotor systems, while the default mode maintains an intermediate balance.

The clear differences in the energies utilized by different cognitive systems and by between-system interactions begs the question of whether the brain cares about these energies. Does the brain prefer smaller within-system energies, smaller between-system energies, or some balance between the two? To address this question, we studied the ensemble of local minima, and asked which systems were commonly expressed. Specifically, for each local minimum, we determined which system was most activated, assigned the minimum to that system, and performed this assignment for all minima. We observed that 3 systems were represented at higher percentages than expected in a permutation-based null model (see Methods): the default mode system, the visual system, and the somatosensory system (Fig. 8D). Importantly, these three systems represent the systems with the highest between system energies (Fig. 8C), but are indistinguishable from other cognitive systems in terms of within-system energy. These results suggest that the brain may prefer high integration between systems over low integration, and that the constraint of between-system energies is more fundamental to brain function than the constraint of within-system energies.
Figure 8: Utilization Energies of Cognitive Systems. (A) Average within-system energy of each cognitive system; error bars indicate standard error of the mean across subjects. (B) Average between-system energy of each cognitive system; error bars indicate standard error of the mean across subjects. (C) The 2-dimensional plane mapped out by the within- and between-system energies of different brain systems. Each data point represents a different brain region, and visual clusters of regions are highlighted with lightly colored sectors. The sector direction is determined by minimizing the squared loss in point density of the local cloud and the width is determined by the orthogonal standard derivation at the center along the sector direction. In this panel, all data points represent values averaged across subjects. (D) The percentages of minima displaying preferential activation of each system; each minima was assigned to the system which whom it shared the largest normalized mutual information. Errorbars indicate the differences between the observed percentages and those of the null distribution with random activation patterns across regions.
2.4. Discussion

In this paper, we address the question of how large-scale brain circuitry and distinct energetic constraints produce whole-brain patterns of activity. We build our approach on a maximum entropy model of brain dynamics that is explicitly informed by estimates of white matter microstructure derived from deterministic tractography algorithms. The model allows us to study minimal energy states, which we observe to be composed of co-activity in local spatially-contiguous sets of brain regions reminiscent of cognitive systems. These systems are differentially active, and activity patterns are significantly correlated with the observed activation rate measured in a separate resting state fMRI data set. Finally, we exercise this model to ask how cognitive systems utilize the minimal energy presumably available to them. We find that the energy utilized within and between cognitive systems distinguishes 4 classes of energy utilization dynamics, corresponding to sensorimotor, default mode, attention, and cognitive control functions. These results suggest that diverse cognitive systems are optimized for differential contributions to integrated versus segregated function via distinct patterns of energy utilization. More generally, the results highlight the importance of considering energetic constraints in linking structural connectivity to observed dynamics of neural activity.

2.4.1. The Role of Activation vs. Connectivity in Understanding Brain Dynamics

As the interest in understanding structural brain connectomics has blossomed in the last several years [Sporns et al. (2005); Sporns (2011)], it has not been accompanied by an equally vivid interest in linking subsequent insights to the more traditional notions of brain activation profiles [Bassett et al. (2015)]. Indeed, the fields of systems and cognitive neuroscience have instead experienced a pervasive divide between the relatively newer notions of connectome mapping and the relatively traditional yet highly effective notions of brain mapping [Sporns (2015)], which have led to powerful insights into neural function in the last quarter century [Zeki (2005)]. This divide is at least in part due to the fact that graph theory and network-based methods on which the field of connectomics is based have
few tools available to link node properties (activity) with edge properties (connectivity) [Newman and Clauset (2015)]. While technically explicable, however, the conceptual divide between these fields can only lead to their detriment, and synergistic efforts are necessary to develop a language in which both activity and connectivity can be examined in concert [Bassett et al. (2015)]. This study offers one explicit mathematical modeling framework in which to study the relationships between activation profiles across the whole brain and underlying structural connectivity linking brain regions. Complementary approaches include the model-based techniques formalized in the publicly available resource The Virtual Brain [Schirner et al. (2015); Ritter et al. (2013)].

2.4.2. Co-activation Architecture

In this study, we observed that brain regions affiliated with known cognitive systems – including somatosensory, visual, auditory, default mode, dorsal and ventral attention, frontoparietal, and cingulo-opercular – also tend to be active together with one another in low energy brain states. Indeed, these theoretical results are consistent with previous studies of functional neuroimaging data demonstrating that groups of co-active regions tended to align well with known cognitive systems [Crossley et al. (2013, 2014)]. This correspondence is particularly interesting when one considers how these cognitive systems were initially defined: and that is, based on strong and dense functional connectivity [Power et al. (2011)]. Thus, our results point to a fundamental mapping between activity and connectivity: regions that are active together tend to be functionally connected together. The presence of such a map has been empirically observed in the resting state (though not in task [Bassett et al. (2015)]), in both healthy controls and in people with schizophrenia where the map appears to be fundamentally altered in its nature [Bassett et al. (2012); Zalesky et al. (2012); Yu et al. (2013)]. Here we offer a mechanism for this mapping based on a combined consideration of energy- and connectivity-based constraints.
2.4.3. Critical Importance of Energy Constraints

The quest to understand and predict brain dynamics from the architecture of underlying structural connectivity is certainly not a new one. In fact, there have been concerted efforts over the last decade and more to identify structural predictors of the resting state BOLD signal. Seminal contributions have included the observations of statistically significant correlations between structural connectivity estimated from diffusion imaging data and functional connectivity estimated from fMRI [Honey et al. (2009)], as well as extensions of these correlations that account for long distance paths along white matter tracts [Goñi et al. (2014)] and spectral properties of structural matrices [Becker et al. (2016)]. The question of how brain structure constrains a wide range of brain states (beyond simply the resting state) is a very open area of inquiry. Moreover, this question is particularly challenging to address with empirical data because it is difficult to obtain data from humans in more than a handful of task states [Cole et al. (2014)]. For this reason, computational models play a very important role in offering testbeds for the development of theories linking structure to ensembles of brain states, which can in turn offer testable predictions. Our results suggest that an understanding of the relationship between brain structure and function is perhaps ill-constrained when examining connectivity alone. The additional assumption of energy conservation produces a set of brain states that cannot be simply predicted from statistics of structural connectivity, perhaps offering a mechanism for the large amount of unexplained variance in prior predictions [Honey et al. (2009); Goñi et al. (2014)].

2.4.4. Methodological Considerations

Our results are built on the formalism of the maximum entropy model, which is predicated on pairwise statistics [Bialek et al. (2012)]. However, emerging evidence suggests that some neurophysiological phenomenon are better studied in the framework of simplicial complexes rather than dyads [Giusti et al. (2016)]. For example, integrate and fire neurons exposed to common fluctuating input display strong beyond-pairwise correlations that cannot be captured by maximum entropy models [Leen and Shea-Brown (2015)]. Similar arguments
can also be made for co-activation patterns in BOLD fMRI [Crossley et al. (2013, 2014); Fox et al. (2014)]. It will be interesting in future to determine the role energetic and structural constraints on observed higher-order functional interactions during human cognitive function.

A second important consideration is that the maximum entropy model is appropriate for systems at equilibrium. Therefore, the local minima identified may not accurately represent the full class of states expected to be elicited by daily activity. Instead, the local minima identified here are expected to more accurately represent the set of states expected to appear as a person rests in the so-called default mode of brain function, which is thought to lie near a stable equilibrium [Deco and Jirsa (2012)]. Such an interpretation is consistent with our findings that the activation rates predicted by the maximum entropy model are strongly correlated with the activation rates observed in resting state fMRI data.

2.5. Conclusion

The analyses presented in this study produce information regarding an underlying energy landscape through which the brain is predicted to move. The existence of such a landscape motivates the very interesting question of how the brain transitions between states. In sampling this landscape, we have used a simple random walk in an effort to extract a large ensemble of possible brain states, measured by local minima. However, the question of which walks a healthy (or diseased) brain might take through this landscape remains open. Such walks or dynamical trajectories may be determined by energetic inputs to certain regions of the brain [Gu et al. (2015a)], either by external stimuli or by neuromodulation [Muldoon et al. (2016b)]. In this context, network control theory may offer explicit predictions regarding the optimal dynamic trajectories that the brain may take through a set of states to move from an initial state to a target state with little energetic resources [Gu et al. (2015a); Pasqualetti et al. (2014)]. In addition to inputs to single regions, changes in a cognitive task – for example elicited by task-switching paradigms – may also drive a specific trajectory of brain states. Indeed, it is intuitively plausible that the asymmetric
switch costs observed between cognitively effortful and less effortful tasks [Wu et al. (2015); Davidson et al. (2006)] may be explained by characteristics of the energy landscape defined by structural connections between task-activated brain regions.
CHAPTER 3 : Controllability of Structural Brain Networks

3.1. Introduction

Neuroscientific investigations seek to reveal how neural systems function, how those functions are altered in disease states, and how interventions can redirect these alterations. All three lines of investigation address how neural systems move along trajectories: of cognitive function, disease, or recovery. Fundamental and therefore generalizable mechanisms of these movements have remained elusive.

The complexity of neural dynamics stems in part from the complexity of the underlying anatomy. Different components (neurons, cortical columns, brain areas) are interlinked in complex patterns that enable diverse functions. These structural interactions can be represented as a network, where component parts form the nodes, and where anatomical links form the edges between nodes. The complex architecture of these networks plays a role in neural function \cite{Bullmore2012}, development \cite{Feldt2011}, disease \cite{Bassett2009}, and sensitivity to rehabilitation \cite{Weiss2011}. Yet, how this architecture constrains neural dynamics is far from understood.
Figure 9: **Conceptual Schematic.** From weighted brain networks (A), we estimate control points (B) whose large-scale regional activity can move the brain into new trajectories that traverse diverse cognitive functions (C). In panel (C), we show the original state of the system (state 0), as well as 4 possible states (indicated by the blue circles) that are equidistant from state 0 in the state space (indicated by the black circular line), and which can be reached by trajectories that are more or less energetically costly (indicated by the height of the purple bars).
Here we capitalize on recent theoretical advances in network control theory to quantify the role of structural network organization on the trajectories of neural systems. Controllability of a network refers to the possibility to manipulate certain components to drive the system along a desired trajectory, meaning a path traversing diverse system states. We postulate that the regulation of cognitive function is driven by a network-level control process akin to those utilized in other biological, technological, cyberphysical, and social systems. In this view, particular nodes (brain regions) at critical locations within the anatomical network act as drivers that move the system (brain) into specific modes of action (cognitive functions). While not previously applied to neuroimaging data, network control theory provides a mathematical framework to investigate how structural features of a brain network determine temporal features of cognitive dynamics.

We exploit network control theory to address two questions about how the large-scale circuitry of the human brain constrains its dynamics. First, is the human brain theoretically controllable? Growing evidence from brain computer interfaces [Ruiz et al. (2014)] and neuromodulation [Fox et al. (2012); Shafi et al. (2012)] suggest that exogenously and endogenously induced changes in a region’s activity (as measured by fMRI or EEG) can alter the trajectories of brain function. We therefore hypothesize that the brain is theoretically controllable in the sense defined mathematically by network control theory (see next section). However, since many such dynamic processes impact distributed neural circuits [Turk-Browne (2013); Schmahmann et al. (2008)] rather than single brain regions alone, we conjecture that the brain is difficult to control via localized interventions. Second, which areas of the brain are most influential in constraining or facilitating changes in brain state trajectories? Decades of prior literature demonstrate that brain regions perform different functions [Kanwisher (2010)], and at least some of these functions have been discussed using broad conceptual notions of “cognitive control” [Botvinick and Cohen (2014); Power and Petersen (2013); Corbetta and Shulman (2002)]. Yet, the relationships between such conceptual notions of cognitive control and the mathematical notions of network control have not previously been examined. We aim to directly test for such a relationship in the
To address these questions, we posit that some basic control properties of cognition are determined by *structural* properties of the brain. We build structural brain networks from diffusion spectrum imaging (DSI) data acquired in triplicate from 8 healthy human adults. We perform diffusion tractography to estimate the number of streamlines linking $N = 234$ large-scale cortical and subcortical regions extracted from the Lausanne atlas [Hagmann et al. (2008a)]. We summarize these estimates in a weighted adjacency matrix whose entries reflect the number of streamlines connecting different regions. Finally, we perform a systematic study of the controllability of the dynamical network defined by the weighted adjacency matrix. This construction enables us to examine different controllability measures in individual participants and for different controllability measures.

3.2. Mathematical Models

3.2.1. Network Control Theory

Our understanding of natural systems is intimately related to our ability to control them. Network control theory is a branch of traditional control theory in engineering that addresses the question of how to control a system whose components are linked in a web of interconnections; here, the term *control* indicates perturbing a system to reach a desired state. Answering this question requires (i) knowledge regarding the network connectivity linking system components, and (ii) knowledge regarding how system components act, i.e., their *dynamics*. In turn, the theory provides predictions regarding the system’s function. Critically, in contrast to traditional graph theory which provides descriptive statistics of network structure, network control theory offers mechanistic predictors of network dynamics. The ability to probe mechanistic predictors of brain function is key to move efforts in the human connectome towards an understanding of human cognition.

Mathematically speaking, we can study the controllability of a networked system by defining a network represented by the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ are the vertex and edge
sets, respectively. Let $a_{ij}$ be the weight associated with the edge $(i, j) \in \mathcal{E}$, and define the *weighted adjacency matrix* of $\mathcal{G}$ as $A = [a_{ij}]$, where $a_{ij} = 0$ whenever $(i, j) \notin \mathcal{E}$. We associate a real value (state) with each node, collect the node states into a vector (network state), and define the map $x : \mathbb{N}_{\geq 0} \to \mathbb{R}^n$ to describe the evolution (network dynamics) of the network state over time. Given the network and node dynamics, we can use network control theory to quantitatively examine how the network structure constrains the types of control that nodes can exert.

Here we are interested in applying network control theory to the human brain. As a quintessentially complex biological system, the human brain offers several contexts in which to think about the notion of “control”: both as a system that implements control and a system to be controlled. For example, control can be thought of as (i) the change in regional BOLD activity produced in response to neurofeedback in real-time fMRI [Christo-pher deCharms et al. (2005)], (ii) the change in regional neural activity elicited by external stimuli [Pagan et al. (2013)], or (iii) the change in regional neural activity provoked by non-invasive brain stimulation [Gratton et al. (2013)]. Each of these mechanisms initially alters the dynamics of single brain regions, but can have consequences for the activity and function of distributed networks. Importantly, this notion of control is based on a very detailed mathematical construct, and is therefore necessarily quite distinct from the cognitive neuroscientist’s common notion of “cognitive control” and the distributed sets of brain regions implicated in its performance [Power and Petersen (2013)]. To minimize obfuscation, we henceforth refer to these two notions as “network control” and “cognitive control”, respectively.

### 3.2.2. Dynamic Model of Neural Processes

To apply network control theory to the human brain, we must define a structural brain network and a model for the dynamics of neural processes. We define both based on prior work in human systems neuroscience. We define structural brain networks by subdividing the entire brain into anatomically distinct brain areas (network nodes), over 5 levels of spatial
resolution from 83 regions to greater than 1000 regions \cite{Cammoun2012}. Consistent with prior work \cite{Bassett2011a,Hermundstad2013,Hermundstad2014}, we connect nodes by the number of white matter streamlines identified by a commonly used deterministic tractography algorithm (for details on the tractography implementation, see \cite{Cieslak2014}). This procedure results in sparse, weighted, undirected structural brain networks for each subject ($N = 8$) and each scanning session ($n = 3$). Properties of this network include high clustering, short path length, and strong modularity, consistent with prior studies of similar network data \cite{Bassett2011a,Hagmann2008a}. The definition of structural brain networks based on tractography data in humans follows from our primary hypothesis that control features of neural dynamics are in part determined by the structural organization of the brain’s white matter tracts.

To define the dynamics of neural processes, we draw on prior models linking structural brain networks to resting state functional dynamics \cite{Honey2009,Honey2010,Abdelnour2014}. Although neural activity evolves through neural circuits as a collection of nonlinear dynamic processes, these prior studies have demonstrated that a significant amount of variance in neural dynamics as measured by fMRI can be predicted from simplified linear models. (See Methodological Considerations for additional discussion on the strengths and weaknesses of the linear model approach.) Based on this literature, we employ a simplified noise-free linear discrete-time and time-invariant network model:

$$ x(t+1) = Ax(t) + Bu(t), \quad (3.1) $$

where $x: \mathbb{R}_{\geq 0} \to \mathbb{R}^N$ describes the state (i.e., a measure of the electrical charge, oxygen level, or firing rate) of brain regions over time, and $A \in \mathbb{R}^{N \times N}$ is a symmetric and weighted adjacency matrix. In our case, we construct a weighted adjacency matrix whose elements indicate the number of white matter streamlines connecting two different brain regions – denoted here as $i$ and $j$ – and we stabilize this matrix by dividing by the mean edge weight. While the model employed above is a discrete-time system, we find that the controllability
Gramian is statistically similar to that obtained in a continuous-time system.

More generally, we note that the network control theory framework is agnostic to the exact type of “activity” that the system produces. However, the model we write down above is a simplified “activity” dynamics that has previously been used to model both neural activity \cite{Galán2008} and regional BOLD activity \cite{Honey2009}. In the context of our work in this paper, we use these dynamics to model fMRI BOLD magnitudes and their coherence across brain regions, but future work may address the utility of this same construct in understanding different temporal scales of brain dynamics.

The diagonal elements of the matrix $A$ satisfy $A_{ii} = 0$. The input matrix $B_K$ identifies the control points $K$ in the brain, where $K = \{k_1, \ldots, k_m\}$ and

$$B_K = \begin{bmatrix} e_{k_1} & \cdots & e_{k_m} \end{bmatrix},$$

and $e_i$ denotes the $i$-th canonical vector of dimension $N$. The input $u_K : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^m$ denotes the control strategy.

3.2.3. Network Controllability

To study the ability of a certain brain region to influence other regions in arbitrary ways we adopt the control theoretic notion of controllability. Controllability of a dynamical system refers to the possibility of driving the state of a dynamical system to a specific target state by means of an external control input \cite{Kalman1963}. Classic results in control theory ensure that controllability of the network (1.10) from the set of network nodes $K$ is equivalent to the controllability Gramian $W_K$ being invertible, where

$$W_K = \sum_{\tau=0}^{\infty} A^\tau B_K B_K^T A^\tau.$$  

We utilize this framework to choose control nodes one at a time, and thus the input matrix $B$ in fact reduces to a one-dimensional vector.
Besides ensuring controllability, the eigenvalues of the controllability Gramian are a quantitative measure of the magnitude of the control input that drives a network to a desired target state [Kailath (1980)], and the structure of the Gramian itself provides systematic guidelines for the selection of control areas that can theoretically optimize cognitive functions. While the magnitude of the control input may not be the unique feature to take into account when controlling brain dynamics [Kumar et al. (2014)], it allows us to better understand the relationship between the structural organization of the brain and its dynamics, and opens the door to the development of novel diagnostics and opportunities for intervention.

**Average Controllability** Average controllability of a network equals the average input energy from a set of control nodes and over all possible target states [Marx et al.] (2004); Shaker and Tahavori (2012)]. As a known result, average input energy is proportional to $\text{Trace}(W^{-1}K)$, the trace of the inverse of the controllability Gramian. Instead, we adopt $\text{Trace}(WK)$ as a measure of average controllability for two main reasons: first, $\text{Trace}(W^{-1}K)$ and $\text{Trace}(WK)$ satisfy a relation of inverse proportionality, so that the information obtained from the two metrics are correlated with one another and, second, $W_K$ is typically very ill-conditioned (see paragraph “Global Controllability”) even for coarse network resolutions, so that $\text{Trace}(W^{-1}K)$ cannot be accurately computed even for small brain networks. It should be noted that $\text{Trace}(WK)$ encodes a well-defined control metric, namely the energy of the network impulse response or, equivalently, the network $H_2$ norm [Kailath (1980)]. Regions with high average controllability are, on average, most influential in the control of network dynamics over all different target states.

**Modal Controllability** Modal controllability refers to the ability of a node to control each evolutionary mode of a dynamical network [Hamdan and Nayfeh (1989)], and can be used to identify states that are difficult to control from a set of control nodes. Modal controllability is computed from the eigenvector matrix $V = [v_{ij}]$ of the network adjacency matrix $A$. By extension from the PBH test [Kailath (1980)], if the entry $v_{ij}$ is small, then
the $j$-th mode is poorly controllable from node $i$. Following [Pasqualetti et al. (2014)], we define $\phi_i = \sum_{j=1}^{N} (1 - \lambda_j^2(A))v_{ij}^2$ as a scaled measure of the controllability of all $N$ modes $\lambda_1(A), \ldots, \lambda_N(A)$ from the brain region $i$. Regions with high modal controllability are able to control all the dynamic modes of the network, and hence to drive the dynamics towards hard-to-reach configurations.

**Boundary Controllability** Boundary controllability measures the ability of a set of control nodes to decouple the trajectories of disjoint brain regions. To evaluate the boundary controllability of different brain regions, we proceed as follows. First, we compute a *robust partition* of the brain network as described in [Bassett et al. (2013)], and we identify the set of $N_1$ boundary nodes. We assign to these boundary nodes the boundary controllability value of 1. Second, following [Pasqualetti et al. (2014)], we determine the two-partition of the least controllable subnetwork from its Fiedler eigenvector [Fortunato (2010); Fiedler (1973)], and we identify the additional boundary nodes. We assign to these boundary nodes the boundary controllability value of $(N - N_1)/N$. Finally, we iterate this process until all nodes have been assigned a boundary controllability value.

This method is largely based on the algorithm proposed in [Pasqualetti et al. (2014)]. However, for the application to brain networks derived from diffusion tractography, we have made two important modifications to more accurately estimate the initial partition and constrain the boundary point criteria as described in detail below.

The first modification concerns the definitions of the first level subnetworks for which we compute a two-partition based on the Fiedler eigenvector. In initial work, Pasqualetti et al. [Pasqualetti et al. (2014)] suggest computing the Fiedler eigenvector of the adjacency matrix to create first level subnetworks defined by a two-partition. In contrast, we define this first level of subnetworks as composed of network communities, identified by maximizing the modularity quality function [Newman (2006)] using a Louvain-like [Blondel et al. (2008)] locally greedy algorithm [Jutla et al. (2011–2012)]. Our choice is based on extensive recent literature demonstrating that the brain is composed of many subnetworks (not just 2)
Meunier et al. (2010); Bassett and Siebenhuhner (2013)], which can be extracted using modularity maximization approaches [Meunier et al. (2009a); Chen et al. (2008); Bassett et al. (2011a)], and which correspond to sets of brain areas performing related functions [Power et al. (2011); Chen et al. (2008)].

The modularity quality function provides an estimate of the quality of a hard partition of the $N \times N$ adjacency matrix $A$ into network communities (whereby each brain region is assigned to exactly one network community) [Newman and Girvan (2004); Newman (2004, 2006); Porter et al. (2009); Fortunato (2010)]

$$Q_0 = \sum_{ij} [A_{ij} - \gamma P_{ij}] \delta(g_i, g_j),$$  \hspace{1cm} (3.4)

where brain region $i$ is assigned to community $g_i$, brain region $j$ is assigned to community $g_j$, $\delta(g_i, g_j) = 1$ if $g_i = g_j$ and it equals 0 otherwise, $\gamma$ is a structural resolution parameter, and $P_{ij}$ is the expected weight of the edge connecting node $i$ and node $j$ under a specified null model. Maximization of $Q_0$ yields a hard partition of a network into communities such that the total edge weight inside of communities is as large as possible (relative to the null model and subject to the limitations of the employed computational heuristics, as optimizing $Q_0$ is NP-hard [Porter et al. (2009); Fortunato (2010); Brandes et al. (2008)].

Because the modularity quality function has many near-degeneracies, it is important to perform the optimization algorithm multiple times [Good et al. (2010)]. We perform 100 optimizations of the Louvain-like locally greedy algorithm [Jutla et al. (2011–2012)] for each adjacency matrix corresponding to a single scan. To distill a single representative partition, we create a consensus partition from these 100 optimizations based on statistical comparison to an appropriate null model [Bassett and Siebenhuhner (2013)].

In a final consideration, we choose a value for the structural resolution parameter $\gamma$. The choice $\gamma = 1$ is very common, but it is important to consider multiple values of $\gamma$ to examine community structure at multiple scales [Reichardt and Bornholdt (2006); Porter...
et al. (2009); Onnela et al. (2012)]. Indeed, recent work has demonstrated that in some networks, a structural resolution parameter value that accurately captures the underlying community structure can be identified by the $\gamma$ value at which the 100 optimizations produce similar partitions [Bassett and Siebenhühner (2013)]. To quantitatively estimate similarity in partitions, we adopt the $z$-score of the Rand coefficient [Traud et al. (2011)]. For each pair of partitions $\alpha$ and $\beta$, we calculate the Rand $z$-score in terms of the total number of pairs of nodes in the network $M$, the number of pairs $M_\alpha$ that are in the same community in partition $\alpha$, the number of pairs $M_\beta$ that are in the same community in partition $\beta$, and the number of pairs of nodes $w_{\alpha\beta}$ that are assigned to the same community both in partition $\alpha$ and in partition $\beta$. The $z$-score of the Rand coefficient comparing these two partitions is

$$z_{\alpha\beta} = \frac{1}{\sigma_{w_{\alpha\beta}}}w_{\alpha\beta} - \frac{M_\alpha M_\beta}{M},$$

where $\sigma_{w_{\alpha\beta}}$ is the standard deviation of $w_{\alpha\beta}$. Let the mean partition similarity denote the mean value of $z_{\alpha\beta}$ over all possible partition pairs for $\alpha \neq \beta$. Let the variance of the partition similarity denote the variance of $z_{\alpha\beta}$ over all possible partition pairs for $\alpha \neq \beta$.

Empirically, we calculated a group adjacency matrix by averaging the adjacency matrices of all subjects and scans. We optimized the modularity quality function 100 times and we computed the mean and variance of the partition similarity for a range of $\gamma$ values and for all 5 spatial resolutions. Across all atlases, we observed that the mean partition similarity was high and the variance of the partition similarity was low for values of $\gamma$ ranging between 1.5 and 2. For Scale 125 (the atlas for which we report results in the main manuscript), we observed a maximum mean partition similarity and a minimum variance of partition similarity at $\gamma = 1.6$. We therefore chose to set $\gamma = 1.6$ for the remainder of the analysis in this study.

The second modification concerns the definition of a boundary point. After calculating the Fiedler eigenvector of a subnetwork to determine a partition of the subnetwork into two communities, we must identify “boundary points”, which are nodes that contain con-
Figure 10: **Partition Similarity As a Function of the Resolution Parameter** Mean (left) and variance (right) of the partition similarity estimated using the $z$-score of the Rand coefficient as a function of the structural resolution parameter $\gamma$, varies from 0.5 to 2 in increments of 0.1, for the 5 spatial scales of the Lausanne atlas [Hagmann et al. (2008a)] (rows).
nections to both communities. In the original work by Pasqualetti and colleagues, it was suggested that a boundary point was a node with any number of connections to both communities. However, in weighted brain networks we suggest that a more stringent definition is more appropriate for the following reason: practically all nodes in the brain have non-zero weighted connections to both identified communities. Therefore, we instead set a threshold ratio $\rho$ to identify boundary points. Considering the adaptivity to the local measure, we set a threshold ratio $\rho$ instead of a global threshold value. In detail, for a network $G = (V, E)$ with partition $P = (V_1, \cdots, V_n)$, a node $i \in V_k$ is called a boundary node if

$$\sum_{l \neq k} a_{kl} \geq \rho \cdot \max(A) \quad (3.6)$$

where $A$ is the adjacency matrix. Here, $\max(A)$ can be replaced with other statistics and $\rho$ needs to be chosen carefully. If $\rho$ is too small, there will be no effect and the algorithm tends to add the total subnetwork as the set of boundary points. If $\rho$ is too large, there will be only a few points recognized as the boundary points.

In the results described in the main manuscript, we set the threshold ratio to $\rho = 0.2$. To determine whether our results are robust to this choice, we calculate boundary controllability values across all regions in the Scale 125 atlas, for each scan using $\rho$ values that vary between 0.15 and 0.25 in increments of 0.01. We then asked how similar regional control values were for different choices of $\rho$. Specifically, for any pair of $\rho$ values, we computed the Pearson correlation coefficient between the vectors of regional control values for the two $\rho$ values. We show the results of this analysis in Fig. 11. We observe that the boundary control values are highly similar across choices of $\rho$ (minimum Pearson correlation approximately 0.68, corresponding to a $p = 0$, indicating that our results are robust to small variation in the boundary point criteria threshold.
Figure 11: **Effect of Boundary Point Criteria Threshold** Color indicates Pearson correlation coefficient, $r$, between the vectors of boundary controllability values estimated for pairs of $\rho$ values in the range $0.15 - 0.25$ in increments of 0.01.
Thus, the final algorithm used in the calculation of boundary controllability in this paper can be summarized as follows. We begin with the application of a community detection method to the brain network to extract a partition of brain regions into network communities. We then recursively apply a Fiedler bipartition to add boundary nodes within communities, with the goal of improving the local controllability of the network. At each stage of the algorithm, we define the boundary nodes of the network as the nodes that maintain edges to nodes in other communities. Algorithmically, we can write:

Algorithm 2: Algorithm for the Selection of Boundary Control Nodes

Data: Network $G = (V, E)$ with adjacency matrix $A = (a_{ij})$, Number of control nodes $m$, threshold ratio $\rho$;

Result: Control Nodes Index Set $K$;

1. Define an empty set of control nodes $K = \emptyset$;
2. Initialize the partition $P$ with the result of a community detection algorithm and initialize the boundary nodes set $B = \emptyset$;
3. Add the boundary points of the initial partition;
4. while $|K| < m$ do
   5. Select least controllable community $l = \arg \min \{\lambda_{\min}(W_{i,\infty}), i = 1, ..., |P|\}$;
   6. Compute Fiedler two-partition $P_f$ of $l$-th community;
   7. Compute boundary nodes $B_f$ of $P_f$ with the given threshold ratio $\rho$;
   8. Update partition $P$ with $P_f$;
   9. Update control nodes with boundary nodes $K = K \cup B_f$;
5. end
6. return $K$.

Average, modal, and boundary controllability each provide a scalar value for each brain region. To enable direct comparison between controllability diagnostics and across different subjects, we perform ranking and normalization steps. In particular, for each of the controllability diagnostics we (i) rank the scalar values for each subject, and (ii) average the ranked values across the subjects.
3.2.4. Correlation Between Degree and Average Controllability

In the main manuscript, we describe a strong correlation between node degree and average controllability for the networks that we study. Here we provide a possible explanation for this effect. Due to the property $\text{Trace}(ABC) = \text{Trace}(BCA)$, the average controllability with a single control node $j$ equals the $j$-th diagonal elements of $(I - A^2)^{-1}$. Since $A$ is stable, a first order approximation yields

$$(I - A^2)^{-1} \approx I + A^2, \quad (3.7)$$

and the $j$-th diagonal element of $(I - A^2)^{-1}$ is $1 + \sum_{i=1}^{N} A_{ij}^2$. Since the degree of the $j$-th node equals $d_j = \sum_{i=1}^{N} A_{ij}$, a positive correlation between node degree and average controllability is mathematically expected in the networks that we study here.

3.2.5. Lower Bound on the Largest Eigenvalue of the Controllability Gramian

In the main manuscript, we show that the smallest eigenvalue of the controllability Gramian is in fact much smaller than its largest counterpart. In fact, the largest eigenvalue of the controllability Gramian is lower bounded by 1. To see this, notice that

$$\lambda_{\max}(W_K) = \lambda_{\max} \left( \sum_{\tau=0}^{\infty} A^\tau B_{K} B_{K}^T A^\tau \right) \geq \lambda_{\max} \left( \sum_{\tau=0}^{0} A^\tau B_{K} B_{K}^T A^\tau \right) = \lambda_{\max}(B_{K} B_{K}^T) = 1, \quad (3.8)$$

where the inequality follows from the fact that $A^\tau B_{K} B_{K}^T A^\tau$ is positive semi-definite for all $\tau$.

3.2.6. Additional Details for Control Methods

Let the network be controllable in $T$ steps, and let $x_f$ be the desired final state in time $T$, with $||x_f||_2 = 1$. Define the energy of the control input $u_K$ as

$$E(u_K, T) = ||u_K||_2^2, \quad (3.9)$$
where $T$ is the control horizon. The unique control input that steers the network state from $x(0) = 0$ to $x(T) = x_f$ with minimum is [Kailath (1980)]

$$u^K_*(t) = B^T(A^T)^{T-t-1}W^{-1}_{K,T}x_f$$

(3.10)

with $t \in \{0, \ldots, T-1\}$. Then it can be seen that

$$E(u^K_*, T) = \sum_{\tau=0}^{T-1} ||u^K_*(\tau)||^2 = x_f^TW^{-1}_{K,T}x_f \leq \lambda^{-1}_{\min}(W_{K,T}),$$

(3.11)

where equality is achieved whenever $x_f$ is an eigenvector of $W_{K,T}$ associated with $\lambda_{\min}(W_{K,T})$.

Because the control energy is limited in practical applications, controllable networks featuring small Gramian eigenvalues cannot be steered to certain states.

3.3. Results

3.3.1. Global Controllability

We first sought to address the question: “Is the human brain theoretically controllable?”. To answer this question, we computed the eigenvalues of the controllability Gramian for each brain region as a control node. We observed that the smallest eigenvalues were consistently greater than 0, indicating that the system is theoretically controllable through a single region. However, these values were small (mean $2.5 \times 10^{-23}$, STD $4.8 \times 10^{-23}$) with respect to the largest eigenvalues (always greater or equal to 1), indicating that in practice the system is extremely hard to control through a single region.

The minute nature of the values of global controllability motivates a more thorough examination of their reliability. In the supplementary materials, we show that exact values of the smallest eigenvalues of the Gramian are not reproducible across multiple scanning sessions (see Table S3: Test-Retest Reliability of Controllability Diagnostics, where the $p$-values for global controllability are consistently greater than 0.05). However, we do observe that the approximate order of these eigenvalues is indeed reproducible, and varies monotonically
over spatial scales of the parcellation of the brain into regions of interest (see Table S1 in the supplementary materials). Together, these results indicate that while the brain is consistently theoretically controllable through a single region, individual differences in global controllability cannot be accurately measured using these techniques.

### 3.3.2. Regional Controllability

We next sought to address the question: “Which areas of the brain are most influential in constraining or facilitating changes in trajectories?” To address this question, we employed 3 diagnostics of regional controllability: the average, modal, and boundary controllability. Each of these diagnostics captures a different control goal [Pasqualetti et al. (2014)]. Average controllability identifies brain areas that, on average, can steer the system into different dynamical states with little effort (i.e., input energy). Loosely speaking, regions with high average controllability can move the brain to many easily reachable states. Modal controllability identifies brain areas that can effectively modulate all dynamical features of the network, and push the brain into difficult-to-reach states, that is, states that are reachable but require substantial input energy. Boundary controllability identifies brain areas that lie at the boundaries between network communities, and that can therefore control the integration between cognitive systems. For mathematical definitions of these diagnostics, see Materials and Methods.

**Average Controllability**  Average controllability identifies brain areas that can steer the system into many different states. The average controllability is greatest in precuneus, posterior cingulate, superior frontal, paracentral, precentral and subcortical structures (Fig. 12A). Strikingly similar to the structural “core” of the human cerebral cortex [Hagmann et al. (2008a)], these regions are “hubs”, having high network degree defined as the average weight of edges emanating from that region. Indeed, the average controllability is strongly correlated with weighted degree (also known as node strength; Pearson correlation $r = 0.91$, $p = 8 \times 10^{-92}$; Fig. 12B). In addition to being structural hubs, we note that these regions also form the anterior and posterior medial portions of the default mode.
system (which we explicitly test in the following section).

Modal Controllability  Modal controllability identifies brain areas that can steer the system into difficult-to-reach states. The modal controllability is greatest in postcentral, supramarginal, inferior parietal, pars orbitalis, medial orbitofrontal, and rostral middle frontal cortices (Fig. 12C). Areas with high modal controllability are not hubs of the network but instead have low degree. The modal controllability is strongly anti-correlated with weighted degree (Pearson correlation $r = -0.99$, $p = 2 \times 10^{-213}$; Fig. 12D), consistent with the notion that difficult-to-reach states require the control of sparsely connected areas.

Boundary Controllability  Boundary controllability identifies brain areas that can steer the system into states where different cognitive systems are either coupled or decoupled. This control goal complements but differs from those of average and modal controllability. The boundary controllability is greatest in rostral middle frontal, lateral orbitofrontal, frontal pole, medial orbitofrontal, superior frontal, and anterior cingulate cortices (Fig. 12E). In contrast to areas with high average or modal controllability, areas with high boundary controllability are neither hubs nor non-hubs. The boundary controllability of all brain regions is not strongly correlated or anti-correlated with weighted degree (Pearson correlation $r = 0.13$, $p = 0.03$; Fig. 12E).

Reproducibility of Controllability Diagnostics Across Spatial Scales  In the main manuscript, we show the anatomical distribution of the 3 controllability diagnostics over the $N = 234$ brain regions of the Scale 125 atlas. In Fig. 13 of this supplement, we show that the anatomical distribution of average controllability is visually similar across all 5 spatial scales assessed with the entire Lausanne atlas family. In Fig. 14 and Fig. 15 we show a similar reproducibility of the anatomical distribution of modal and boundary controllability, respectively.
Figure 12: **Brain Network Control Properties**  
(A) Average controllability quantifies control to many easily reached states. Here we show controllability values, averaged across persons, and ranked for all brain regions plotted on a surface visualization. Warmer colors indicate larger values of average controllability.  
(B) Scatter plot of weighted degree (ranked for all brain regions) versus average controllability (Pearson correlation $r = 0.91$, $p = 8 \times 10^{-92}$).  
(C) Modal controllability quantifies control to difficult-to-reach states. Here we show modal controllability values ranked for all brain regions plotted on a surface visualization.  
(D) Scatter plot of weighted degree (ranked for all brain regions) versus modal controllability ($r = -0.99$, $p = 2 \times 10^{-213}$).  
(E) Boundary controllability quantifies control to decouple or integrate network modules. Here we show boundary controllability values ranked for all brain regions plotted on a surface visualization.  
(F) Scatter plot of weighted degree (ranked for all brain regions) versus boundary controllability ($r = 0.13$, $p = 0.03$).  
In panels (A), (C), and (E), warmer colors indicate larger controllability values, which have been averaged over both replicates and subjects. These results are reliable over a range of atlas resolutions, and are consistent with findings using a network composed of only cortical circuitry. Note nodes are sorted in ascending order of the weighted degree.
Figure 13: **Average Controllability Across Spatial Scales** Surface visualizations of the ranked average controllability (AC) values over the 5 spatial scales of the Lausanne atlas [Hagmann et al. (2008a)].
Figure 14: Modal Controllability Across Spatial Scales Surface visualizations of the ranked modal controllability (MC) values over the 5 spatial scales of the Lausanne atlas [Hagmann et al. (2008a)].
Figure 15: **Boundary Controllability Across Spatial Scales** Surface visualizations of the ranked boundary controllability (BC) values over the 5 spatial scales of the Lausanne atlas [Hagmann et al. (2008a)].
Reproducibility of Degree-Controllability Correlations Across Spatial Scales  

In the main manuscript, we observed that for Scale 125 ($N = 234$) the degree was strongly positively correlated with the average controllability, strongly negatively correlated with the modal controllability, and neither strongly positively nor strongly negatively correlated with the boundary controllability. In Table 1 we report the correlations between degree and the 3 controllability diagnostics as a function of spatial resolution: from Scale 33 ($N = 83$) to Scale 500 ($N = 1015$). We observe that the degree-controllability correlations reported for Scale 125 are reproducibly observed across the remaining 4 spatial scales, comprising both higher and lower spatial resolutions.
Table 1: Robustness of the Correlation between Control Metrics and Graphic Metrics: Pearson correlation coefficients $r$ between rank degree, average controllability (AC), boundary controllability (BC), and modal controllability (MC).

<table>
<thead>
<tr>
<th>Scale 33</th>
<th>Degree</th>
<th>AC</th>
<th>BC</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>1.0000</td>
<td>0.9764</td>
<td>0.5225</td>
<td>-0.9923</td>
</tr>
<tr>
<td>AC</td>
<td>0.9764</td>
<td>1.0000</td>
<td>0.6302</td>
<td>-0.9688</td>
</tr>
<tr>
<td>BC</td>
<td>0.5225</td>
<td>0.6302</td>
<td>1.0000</td>
<td>-0.5120</td>
</tr>
<tr>
<td>MC</td>
<td>-0.9923</td>
<td>-0.9688</td>
<td>-0.5120</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale 60</th>
<th>Degree</th>
<th>AC</th>
<th>BC</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>1.0000</td>
<td>0.9429</td>
<td>0.4733</td>
<td>-0.9912</td>
</tr>
<tr>
<td>AC</td>
<td>0.9429</td>
<td>1.0000</td>
<td>0.6262</td>
<td>-0.9320</td>
</tr>
<tr>
<td>BC</td>
<td>0.4733</td>
<td>0.6252</td>
<td>1.0000</td>
<td>-0.4806</td>
</tr>
<tr>
<td>MC</td>
<td>-0.9912</td>
<td>-0.9320</td>
<td>-0.4806</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale 125</th>
<th>Degree</th>
<th>AC</th>
<th>BC</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>1.0000</td>
<td>0.9205</td>
<td>0.1385</td>
<td>-0.9937</td>
</tr>
<tr>
<td>AC</td>
<td>0.9205</td>
<td>1.0000</td>
<td>0.1461</td>
<td>-0.9125</td>
</tr>
<tr>
<td>BC</td>
<td>0.1385</td>
<td>0.1461</td>
<td>1.0000</td>
<td>-0.1270</td>
</tr>
<tr>
<td>MC</td>
<td>-0.9937</td>
<td>-0.9125</td>
<td>-0.1270</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale 250</th>
<th>Degree</th>
<th>AC</th>
<th>BC</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>1.0000</td>
<td>0.9114</td>
<td>0.3310</td>
<td>-0.8626</td>
</tr>
<tr>
<td>AC</td>
<td>0.9114</td>
<td>1.0000</td>
<td>0.4785</td>
<td>-0.7822</td>
</tr>
<tr>
<td>BC</td>
<td>0.3310</td>
<td>0.4785</td>
<td>1.0000</td>
<td>-0.2968</td>
</tr>
<tr>
<td>MC</td>
<td>-0.8626</td>
<td>-0.7822</td>
<td>-0.2968</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale 500</th>
<th>Degree</th>
<th>AC</th>
<th>BC</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>1.0000</td>
<td>0.9122</td>
<td>0.2709</td>
<td>-0.9962</td>
</tr>
<tr>
<td>AC</td>
<td>0.9122</td>
<td>1.0000</td>
<td>0.3637</td>
<td>-0.9042</td>
</tr>
<tr>
<td>BC</td>
<td>0.2709</td>
<td>0.3637</td>
<td>1.0000</td>
<td>-0.2366</td>
</tr>
<tr>
<td>MC</td>
<td>-0.9962</td>
<td>-0.9042</td>
<td>-0.2366</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Reliability of Controllability Diagnostics

With any new technique, it is critical to evaluate the reliability of the estimated diagnostics. The regional controllability diagnostics that we report and utilize here are highly reliable across multiple scanning sessions, indicating their potential use in explaining individual differences in cortical function. Moreover, the anatomical distribution of controllability diagnostics is consistent across 5 parcellation schemes segregating the brain into 83, 129, 234, 463, and 1015 regions of interest, suggesting that these measures are robust quantifications of brain dynamics.

In addition to reliability across spatial resolutions and multiple scanning sessions, we next asked whether our results could be reliably reproduced using different imaging acquisitions and different subject cohorts. To address this question, we constructed 234-region structural brain networks from diffusion tensor imaging data acquired on an independent sample of 85 healthy human adult subjects [Hermundstad et al. (2013, 2014)]. Consistent with our previous results, these data display a strong positive correlation between average controllability and weighted degree (Pearson correlation coefficient $r = 0.88$, $p = 2.5 \times 10^{-80}$; see Fig. 16A), a strong negative correlation between modal controllability and weighted degree ($r = -0.99$, $p = 1.2 \times 10^{-184}$; see Fig. 16B), and a weaker relationship between boundary controllability and weighted degree ($r = 0.0084$, $p = 0.90$; see Fig. 16C). These data support the claim that the architecture of structural brain networks differentially impacts the putative role of brain regions in different control strategies.

Test-Retest Reliability of Controllability Diagnostics

When proposing a new diagnostic of brain network architecture, it is critical to determine the reliability of those diagnostic values across iterative measurement. Here we capitalize on the fact that the same 8 subjects whose data are reported in the main manuscript were imaged over 3 different days. We utilize these iterative scans to assess the test-retest reliability of the 3 controllability diagnostics.

To compare the results among different scans and subjects, we consider the average correlation. Suppose we have $n$ subjects and for each of them we have $K$ scans with corresponding
Figure 16: Reliability and Conservation of Brain Network Control Properties
Brain network control properties are reliable across imaging acquisition and conserved in non-human primates. Scatter plots of weighted degree (ranked for all brain regions) versus (A,D) average controllability (Pearson correlation coefficient $r = 0.88$, $p = 1.0 \times 10^{-78}$; $r = 0.90$, $p = 4.9 \times 10^{-34}$), (B,E) modal controllability ($r = -0.99$, $p = 3.9 \times 10^{-179}$; $r = -0.99$, $p = 1.3 \times 10^{-72}$), and (C,F) boundary controllability ($r = 0.14$, $p = 0.028$; $r = -0.19$, $p = 0.074$) for (A,B,C) human diffusion tensor imaging data and (D,E,F) macaque tract tracing data.
controllability values $c_i^1, \ldots, c_i^K$. The averaged correlation between controllability diagnostic values for subject $i$ and subject $j$ is defined as

$$R_{ij}^B = \frac{\sum_{s=1}^{K} \sum_{t=1}^{K} \text{corr}(c_s^i, c_t^j)}{K^2}$$

(3.12)

for subject $i \neq j$ and where $s$ and $t$ index scanning sessions, and $\text{corr}$ indicates the calculation of a Pearson correlation coefficient. The average correlation between controllability diagnostic values for the same subject across scanning sessions is defined as

$$R_{ii}^W = \frac{\sum_{s \neq t} \text{corr}(c_s^i, c_t^i)}{K(K-1)}$$

(3.13)

for $i = j$. We refer to the quantity $R_{ij}^B$ as the average between-subject correlation and to the quantity $R_{ii}^W$ as the average within-subject correlation.

We report the within- and between-subject correlations for all 3 controllability diagnostics and for global controllability across all 5 spatial scales of the Lausanne atlas family in Tab. 2. We observe that all 3 controllability diagnostics display significantly greater within-subject correlation than between-subject correlation, indicating that these diagnostics are statistically reproducible across scanning sessions and significantly different across individuals. The average and modal controllability display a relatively high mean $R$ (approximately 0.90) and relatively low standard error. While still statistically reproducible across scanning sessions, the boundary controllability displays a lower mean $R$ than the average and modal controllability, and a higher standard error. The global controllability is not reproducible across scanning sessions. These observations are consistently observed across the 5 spatial scales of the Lausanne atlas family of parcellations.

**Conservation of Controllability-Topology Relationship Across Species** Finally, we asked whether the relationship between controllability and topology was conserved in non-human primates. Using a data set drawn from CoCoMac \cite{Kotter2004} that delineated 2402 projections between 95 cortical and subcortical areas \cite{KaiserHilgetag2006}, we
Table 2: **Test-Retest Reliability of Controllability Diagnostics**: average controllability (AC), boundary controllability (BC), modal controllability (MC) and global controllability (GC).

<table>
<thead>
<tr>
<th>Scale 33</th>
<th>AC</th>
<th>BC</th>
<th>MC</th>
<th>GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Within</td>
<td>0.9642</td>
<td>0.7250</td>
<td>0.9708</td>
<td>0.0674</td>
</tr>
<tr>
<td>Mean Between</td>
<td>0.8966</td>
<td>0.3436</td>
<td>0.9191</td>
<td>0.0501</td>
</tr>
<tr>
<td>STE Within</td>
<td>0.0222</td>
<td>0.1279</td>
<td>0.0119</td>
<td>0.0527</td>
</tr>
<tr>
<td>STE Between</td>
<td>0.0227</td>
<td>0.2014</td>
<td>0.0168</td>
<td>0.0426</td>
</tr>
<tr>
<td>p-value</td>
<td>5.7e-11</td>
<td>2.5e-6</td>
<td>9.2e-12</td>
<td>0.4626</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale 60</th>
<th>AC</th>
<th>BC</th>
<th>MC</th>
<th>GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Within</td>
<td>0.9510</td>
<td>0.6146</td>
<td>0.9519</td>
<td>0.0662</td>
</tr>
<tr>
<td>Mean Between</td>
<td>0.8449</td>
<td>0.3465</td>
<td>0.8544</td>
<td>0.0501</td>
</tr>
<tr>
<td>STE Within</td>
<td>0.0283</td>
<td>0.1552</td>
<td>0.0199</td>
<td>0.0590</td>
</tr>
<tr>
<td>STE Between</td>
<td>0.0333</td>
<td>0.1351</td>
<td>0.0261</td>
<td>0.0387</td>
</tr>
<tr>
<td>p-value</td>
<td>4.0e-12</td>
<td>2.8e-6</td>
<td>9.5e-15</td>
<td>0.3076</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale 125</th>
<th>AC</th>
<th>BC</th>
<th>MC</th>
<th>GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Within</td>
<td>0.9404</td>
<td>0.5147</td>
<td>0.9348</td>
<td>0.0782</td>
</tr>
<tr>
<td>Mean Between</td>
<td>0.8036</td>
<td>0.1954</td>
<td>0.7900</td>
<td>0.0527</td>
</tr>
<tr>
<td>STE Within</td>
<td>0.0298</td>
<td>0.1311</td>
<td>0.0234</td>
<td>0.0508</td>
</tr>
<tr>
<td>STE Between</td>
<td>0.0383</td>
<td>0.1638</td>
<td>0.0350</td>
<td>0.0449</td>
</tr>
<tr>
<td>p-value</td>
<td>5.3×10^{-14}</td>
<td>1.8×10^{-6}</td>
<td>1.1×10^{-16}</td>
<td>0.1442</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale 250</th>
<th>AC</th>
<th>BC</th>
<th>MC</th>
<th>GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Within</td>
<td>0.9320</td>
<td>0.5192</td>
<td>0.9230</td>
<td>0.0481</td>
</tr>
<tr>
<td>Mean Between</td>
<td>0.7751</td>
<td>0.2208</td>
<td>0.7451</td>
<td>0.0391</td>
</tr>
<tr>
<td>STE Within</td>
<td>0.0256</td>
<td>0.2012</td>
<td>0.0227</td>
<td>0.0333</td>
</tr>
<tr>
<td>STE Between</td>
<td>0.0332</td>
<td>0.2077</td>
<td>0.0267</td>
<td>0.0286</td>
</tr>
<tr>
<td>p-value</td>
<td>4.5e-19</td>
<td>4.1e-6</td>
<td>7.8e-26</td>
<td>0.4268</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale 500</th>
<th>AC</th>
<th>BC</th>
<th>MC</th>
<th>GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Within</td>
<td>0.9090</td>
<td>0.4990</td>
<td>0.8982</td>
<td>0.0395</td>
</tr>
<tr>
<td>Mean Between</td>
<td>0.7261</td>
<td>0.2367</td>
<td>0.6909</td>
<td>0.0229</td>
</tr>
<tr>
<td>STE Within</td>
<td>0.0280</td>
<td>0.1373</td>
<td>0.0263</td>
<td>0.0327</td>
</tr>
<tr>
<td>STE Between</td>
<td>0.0338</td>
<td>0.1290</td>
<td>0.0224</td>
<td>0.0241</td>
</tr>
<tr>
<td>p-value</td>
<td>1.0e-21</td>
<td>1.4e-6</td>
<td>4.6e-33</td>
<td>0.0864</td>
</tr>
</tbody>
</table>
again observed consistent results, including a strong positive correlation between average controllability and weighted degree (Pearson correlation coefficient $r = 0.90, p = 4.9 \times 10^{-34}$; see Fig. 16D), a strong negative correlation between modal controllability and weighted degree ($r = -0.99, p = 1.3 \times 10^{-72}$; see Fig. 16E), and a non-significant correlation between boundary controllability and weighted degree ($r = -0.19, p = 0.074$; see Fig. 16F). These data indicate that the role of brain hubs and non-hubs in different control strategies is conserved across human and non-human primates.
Figure 17: **Control Roles of Cognitive Systems** Cognitive control hubs are differentially located across cognitive systems. *(A)* Hubs of average controllability are preferentially located in the default mode system. *(B)* Hubs of modal controllability are predominantly located in cognitive control systems, including both the fronto-parietal and cingulo-opercular systems. *(C)* Hubs of boundary controllability are distributed throughout all systems, with the two predominant systems being ventral and dorsal attention systems. Control hubs have been identified at the group level as the 30 regions with the highest controllability values (averaged over replicates and subjects). Raw percentages of control hubs present in each system have been normalized by the number of regions in the cognitive system. By applying this normalization, systems composed of a larger number of regions have the same chance of housing one of the top 30 control hubs as systems composed of a smaller number of regions.
3.3.3. Regional Controllability of Cognitive Systems

After confirming reliability and conservation of our findings, we asked the question “Are control regions differentially located in or between known cognitive systems?”. Drawing from the literature, we formulate 3 specific hypotheses addressing this question. First, based on the fact that average controllability identifies areas of the brain that may be important in steering the system into many easily reachable states, we hypothesize that areas of high average controllability would map on to areas active in the brain’s baseline or “default” state (the resting state), from which the brain smoothly moves to multitudinous task states. In contrast, modal controllability identifies areas of the brain that may be important in steering the system to difficult-to-reach states. We hypothesize that areas of high modal controllability would therefore map on to areas responsible for the brain’s transitions between difficult tasks, specifically executive areas involved in cognitive control. Finally, boundary controllability identifies areas of the brain that can steer the system into states where different cognitive systems are either decoupled or integrated. Because these areas mathematically sit at the boundaries between network communities or putative functional modules, we expect that these areas would map relatively uniformly onto all cognitive systems: each system having a few boundary nodes that might play a role in linking that system to another. However, we also postulate a particular enrichment of the attention systems, based on their role in feature selection, gating, orienting and multi-tasking which constrain integration across other cognitive systems.

To test these hypotheses, we assigned the 234 regions of the Lausanne atlas to the following large-scale cortical networks, which we refer to as “cognitive systems”: auditory, visual, sensorimotor, ventral attention, dorsal attention, default mode, fronto-parietal, and cingulo-opercular. This set of cognitive systems, and the association of regions to these cognitive systems, has previously been extracted from resting state data using a network-based clustering approach [Power et al. (2011)] and has been widely applied to examine the roles of cognitive systems in task-based and resting-state connectivity [Cole et al. (2013); Power]
We find that regions of high controllability are differentially associated with the 8 cognitive systems (Fig. 17). We define the set of high control hubs as the 30 regions with the largest controllability values (averaged over all scans), and we calculate the percent of hubs present from each of the 8 cognitive systems. To correct for system size, we normalize the raw percentage of hubs present in a given cognitive system by the number of regions in a cognitive system. By applying this normalization, systems composed of a larger number of regions do not have an increased normalized probability of housing one of the top 30 control hubs than systems composed of a smaller number of regions. Consistent with our hypotheses, 30% of average control hubs lie in the default mode system, 32% of modal control hubs lie in the fronto-parietal and cingulo-opercular cognitive control systems, and 34% of boundary control hubs lie in the ventral and dorsal attention systems. Our results are qualitatively similar if we choose a larger or smaller set of control hubs, in different imaging acquisition schemes including diffusion tensor imaging, and in a large independent subject cohort.
Figure 18: **Differential Recruitment of Cognitive Systems to Network Control** Average controllability (AC), modal controllability (MC), and boundary controllability (BC) hubs are differentially located in default mode (A), fronto-parietal and cingulo-opercular cognitive control (B), and attentional control (C) systems. Values are averaged over the 3 replicates for each individual; error bars indicate standard deviation of the mean over subjects.
These results suggest the presence of a controllability-by-system interaction: certain types of controllability may be utilized or enabled by different cognitive systems. To directly test for this interaction, we extract control hubs for each scan, determine their association with the three hypothesized control systems (default mode, fronto-parietal and cingulo-opercular cognitive control, and attentional control), and quantify the mean controllability value for all hubs in each system (Fig 18). We observe that regions of the default mode system form strong average controllability hubs but weaker modal and boundary controllability hubs. Regions of the cognitive control networks (fronto-parietal and cingulo-opercular) form strong modal controllability hubs and regions of the attentional control networks (ventral and dorsal) form strong boundary controllability hubs. To statistically validate this finding, we perform a repeated measures 2-way Analysis of Variance with cognitive system and controllability diagnostic as categorical factors, and with scan replicate as a repeated measure. The main effect of system is significant ($F(9) = 42.40; p = 0$), the main effect of diagnostic is significant ($F(2) = 22.25, p = 0.0013$), and the interaction between system and diagnostic is also significant ($F(18) = 39.81; p = 0$). These statistics indeed suggest that structural differences between the default mode, cognitive control, and attentional control systems may facilitate their distinct roles in controlling trajectories of brain network function. These results are robustly observed in different imaging acquisition schemes including both with diffusion tensor imaging and in a large independent subject cohort.

Reproducibility of Control Roles of Cognitive Systems In the main text, we observed that a 30% of average control hubs lie in the default mode system, 32% of modal control hubs lie in the fronto-parietal and cingulo-opercular cognitive control systems, and 34% of boundary control hubs lie in the ventral and dorsal attention systems. Here we demonstrate that these results are qualitatively reproduced for different definitions of control hubs: namely, the 25 nodes with the highest control values (out of a possible 234 nodes), the 30 nodes with the highest control values (as shown in the main manuscript), or the 35 nodes with the highest control values. When control hubs are defined as the top 25 nodes,
we observe that 32% of average control hubs lie in the default mode system, 33% of modal control hubs lie in the fronto-parietal and cingulo-opercular systems, and 33% of boundary control hubs lie in the ventral and dorsal attentional systems. When control hubs are defined as the top 35 nodes, we observe that 28% of average control hubs lie in the default mode system, 31% of modal control hubs lie in the fronto-parietal and cingulo-opercular systems, and 32% of boundary control hubs lie in the ventral and dorsal attentional systems. These results demonstrate that the presence of a controllability-by-system interaction is robust to small variation in the choice of the size of the control hub set.

Reproducibility of Differential Recruitment of Cognitive Systems to Network Control

In the main text, we observed the presence of a controllability-by-system interaction, and interpreted this as indicative of the possibility that certain types of controllability may be utilized or enabled by different cognitive systems. In particular, we observed that regions of the default mode system form strong average controllability hubs but weaker modal and boundary controllability hubs. Regions of the cognitive control networks (fronto-parietal and cingulo-opercular) form strong modal controllability hubs and regions of the attentional control networks (ventral and dorsal) form strong boundary controllability hubs. Here we demonstrate that these results are qualitatively reproduced for different definitions of control hubs: namely, the 25 nodes with the highest control values, the 30 nodes with the highest control values (as shown in the main manuscript), or the 35 nodes with the highest control values; see Fig. 19.

In the main text, we validate this finding by performing a repeated measures 2-way Analysis of Variance (ANOVA) with system and controllability diagnostic as categorical factors, and with scan replicate as a repeated measure. Here, we performed the same ANOVA for the case in which control nodes are defined as the 25 nodes with the highest control values or the 35 nodes with the highest control values, and found similar results in both cases: (i) for top 25 nodes, the main effect of system is $F(9) = 43.7716 \ (p = 0)$, the main effect of diagnostic is $F(2) = 16.5413 \ (p = 2.0553e - 4)$, and the interaction between system and
Figure 19: Control Roles of Cognitive Systems. Cognitive control hubs are differentially located across cognitive systems. (Left) Hubs of average controllability are preferentially located in the default mode system. (Middle) Hubs of modal controllability are predominantly located in cognitive control systems, including both the fronto-parietal and cingulo-opercular systems. (Right) Hubs of boundary controllability are distributed throughout all systems, with the two predominant systems being ventral and dorsal attention systems. These anatomical distributions are consistent across different definitions of control hubs as either (Top) the 25 nodes with the highest control values, (Middle) the top 30 nodes with the highest control values, or (Bottom) the 35 nodes with the highest control values.
diagnostic is $F(18) = 42.1475 \ (p = 0)$; (ii) for the top 35 nodes, the main effect of system is $F(9) = 34.3787 \ (p = 0)$, the main effect of diagnostic is $F(2) = 9.7420 \ (p = 0.0022)$, and the interaction between system and diagnostic is $F(18) = 36.9762 \ (p = 0)$. Consistent with the results reported in the main manuscript, these statistics indeed suggest that structural differences between the default mode, cognitive control, and attentional control systems may facilitate their distinct roles in controlling dynamic trajectories of brain network function. We observe that the weaker control hubs we include in the analysis (i.e., larger number of control hubs), the less significant the relationship to cognitive systems. This suggests that the strong control hubs are significantly associated with cognitive systems but that weak control nodes may not be.

Robustness of Results to Alternative Weighting Schemes  There is currently no accepted weighting scheme for constructing anatomical networks from diffusion imaging data. Weighting connections between ROIs based on the number of streamlines connecting them (as estimated by diffusion tractography algorithms) is the most commonly utilized scheme. However, it has been argued that these estimates can be biased by variation in the sizes of the regions under study [Hagmann et al. (2008a)]: large regions may have a higher probability of displaying more streamlines than smaller regions. While this potential bias does not appear to drastically alter large-scale topological properties of anatomical networks, its local effects are not well characterized [Bassett et al. (2011a)]. Our results, based on the number of streamlines, are unlikely to be affected by this potential bias for one key reason: the Lausanne atlas family purposefully attempts to equalize region size, particularly in the higher scales [Cammoun et al.] (2012)]. Nevertheless, to confirm that our results were robust to an alternative weighting scheme that accounts for region size, we divided each $ij^{th}$ element in the adjacency matrix $A$ in Scale 125 ($N=234$) by the sum of the sizes of the two regions that it connects to create an alternative adjacency matrix $A'$. We then computed the controllability diagnostics and rank degree of $A'$ for each scan. Similar to our results obtained with the original weighting scheme, we observed that (i) the mean average controllability across the scans is strongly and positively correlated with mean rank
Figure 20: **Differential Recruitment of Cognitive Systems to Network Control.** Average controllability (AC), modal controllability (MC), and boundary controllability (BC) hubs are differentially located in default mode (A), fronto-parietal and cingulo-opercular cognitive control (B), and attentional control (C) systems. These results are consistently observed whether we define control hubs as the 25 nodes with the highest control values (*Top Row*), the 30 nodes with the highest control values (*Middle Row*), or the 35 nodes with the highest control values (*Bottom Row*). Values are averaged over the 3 replicates for each individual; error bars indicate standard deviation of the mean over subjects.
degree ($r = 0.97, p = 1.43 \times 10^{-150}$), (ii) the mean modal controllability across the scans is strongly and negatively correlated with mean rank degree ($r = -0.96, p = 2.51 \times 10^{-130}$), and (iii) the mean boundary controllability is not significantly correlated with mean rank degree ($r = -0.01, p = 0.32$). Furthermore, the three network controllability diagnostics are again differentially recruited to known cognitive systems in the same manner as they were for the original weighting scheme (Compare Figure 21 to Figure 4 in the main manuscript). Together, these findings indicate that the results reported in the main manuscript are robust to variations in weighting scheme that include a correction for region size.
Figure 21: Differential Recruitment of Cognitive Systems to Network Control Average controllability (AC), modal controllability (MC), and boundary controllability (BC) hubs are differentially located in default mode (A), fronto-parietal and cingulo-opercular cognitive control (B), and attentional control (C) systems. Values are averaged over the 3 replicates for each individual; error bars indicate standard deviation of the mean over subjects.
3.4. Discussion

The brain is a networked dynamical system that moves between diverse cognitive states to enable complex behaviors. Fundamental principles constraining these trajectories have remained elusive. Here we use network control theory to offer a mechanistic explanation for how the brain moves between cognitive states drawn from the network organization of white matter microstructure. Our results indicate that densely connected areas are theoretically expected to facilitate the movement of the brain to many easily-reachable states and we show that these areas are preferentially located in the default mode system. Weakly connected areas, predominantly located in cognitive control systems, are theoretically expected to facilitate the movement of the brain to difficult-to-reach states. Finally, areas located on the boundary between network communities, predominantly located in attentional control systems, are theoretically expected to facilitate the integration or segregation of diverse cognitive systems. As a whole, this body of work suggests that structural network differences between the default mode, cognitive control, and attentional control systems dictate their distinct roles in controlling trajectories of brain network function.

3.4.1. Theoretically Predicted Controllability of Large-Scale Neural Circuitry

The relationship between any mathematical measure of controllability, and what it means for a brain to be in control is unknown. Nevertheless, network controllability diagnostics provide theoretical predictions regarding the controllability of large-scale neural circuitry. Using the smallest eigenvalues of the controllability Gramian, we provide evidence that structural brain network architecture is theoretically controllable, but practically very difficult to control. The theoretical possibility of controlling the brain from a single region is consistent with a large body of scientific evidence stemming from (i) patient studies that demonstrate that lesions to single brain areas can have dramatic effects on regional activity, inter-regional connectivity, and by extension cognitive function and behavior [Grefkes and Fink (2014)], and (ii) real-time fMRI studies of neuromodulation that demonstrate that subjects can control the activity of single brain regions to modulate pain perception.
These findings are consistent with the theoretical expectation that it is possible – with a single input – to move the brain to a single target state. However, our prediction that the brain is pragmatically difficult to control indicates that it is practically impossible to move the brain to any target state that we might desire with little control action. This predicted difficulty is consistent with the fact that even complex combinations of drugs, brain stimulation, and cognitive therapies [Pallanti and Hollander (2014)] can still fail to correct cognitive function when it has gone wrong. These findings suggest that some trajectories are extremely difficult and practically impossible to accomplish using an input to single brain areas: for example, from a disease state to a healthy state. The complexity of cognitive function and its underlying mechanisms, illustrated by control difficulty, calls for new tools to quantify and understand which trajectories are amenable to control, thereby informing the use of targeted therapies including brain stimulation [Johnson et al. (2013); Schiff (2012)].

3.4.2. The Role of Hubs in Brain Control

We can study brain controllability either globally, as described above, or locally. Network control theory posits the diagnostic of average controllability as a quantification of a node’s role in moving the system to many easily reachable states. We show that brain regions with high average controllability tend to be areas with a large number of white matter streamlines connecting them to the rest of the network – that is, network hubs. These hubs tend to be located in areas of the default mode system. This suggests that the brain has a baseline, resting state organization which is optimized to allow the brain to move to a large number of easily reachable states. Regions of high average controllability, which have the greatest predicted influence in moving the brain to this plethora of states, are highly active at rest, particularly in anterior and posterior medial portions of the default mode network. If we assume that the brain has been optimized over evolutionary time scales to maximally enable a complex functional battery [Bassett et al. (2010); Bullmore and Sporns (2012)], these results suggest the intriguing possibility that the large majority of complex functions
performed by the brain are easily reachable from the default mode state. The few functions which might be difficult to reach from the default mode state may utilize alternative control mechanisms, including modal and boundary control. Moreover, the regional specificity of our findings suggest a regional differentiation of the default mode system in terms of control roles: anterior and posterior medial portions of the default mode system may be more critical in average control strategies than temporal and parietal portions of the same system.

It is interesting to consider the role of the default mode system as average controller in relation to its activation profiles across various task scenarios. Extensive prior literature has demonstrated that the default mode system is activated during so-called “rest” conditions [Gusnard et al. (2001); Raichle and Snyder (2007)], and largely deactivated during many task conditions [Fox et al. (2005)]. In concert with our findings that the default mode (particularly anterior and posterior medial) regions are optimally placed to push the brain into many easily reachable states, these data suggest that the default mode is a type of pluripotent “ground state”, which can move the brain into many task-based activation profiles (“excited states”; which can have little similarity to the ground state [Bassett et al. (2015)]), and to which the brain relaxes back after the task has been performed, readying the brain to move to new task states, when the cycle will repeat. Importantly, these dynamic notions of brain function are predicated on the underlying structure of the white matter pathways facilitating cognitive processes.

Indeed, a growing body of literature has begun to highlight new features of brain network hubs that might contribute to the dynamic functional role outlined by network control theory here. The so-called rich-club organization of the human connectome [van den Heuvel and Sporns (2011)] refers to the fact that many brain network hubs are densely interconnected to one another [Zhou and Mondragón (2004); Colizza et al. (2006); Opsahl et al. (2008)]. This organization is evident across species [de Reus and van den Heuvel (2013); Towlson et al. (2013); Scholtens et al. (2014); van den Heuvel et al. (2015)], changes over time scales of human development [Cao et al. (2014); Ball et al. (2014); Kim et al. (2014)], and is al-
tered in neurological disorders and psychiatric disease \cite{van_den_Heuvel_2013, Collin_2014, Gong_2015, Wang_2015}. The rich club is thought to play important roles in information integration \cite{van_den_Heuvel_and_Sporns_2011}, facilitating the functional dynamics necessary for cognitive functions. The fact that hubs in many different cognitive systems are linked together potentially provides a structural substrate for the movement of the brain between cognitive processes. Supporting this hypothesis, work by Senden and colleagues uses a spin glass model of neural networks for simulating stable configurations of cortical activity, and shows that networks with rich-club architecture display functional dynamics characterized by a larger set of attractors (and hence greater diversity of the functional repertoire) than that expected in scale-free networks devoid of rich clubs \cite{Senden_2014}. Our results provide a theoretical mechanism for these empirical findings: hubs nodes in the brain tend to have high average controllability, indicating that they are critical for moving the brain into many easily reachable states (attractors), thereby facilitating a great diversity of functional dynamics.

The fact that structural hubs, particularly in the default mode network, play such a striking role in brain network controllability may further help to explain the growing body of evidence indicating that disease states can preferentially target hub areas \cite{Bassett_and_Bullmore_2009, Crossley_2014}. \textit{In silico} studies suggest that lesions to highly structurally connected areas have a greater impact on ensuing functional connectivity than lesions to sparsely connected areas \cite{Alstott_2009}. Moreover, alterations to default mode hubs are associated with drastic changes in cognitive function associated with normative aging \cite{Tomasi_and_Volkow_2012} and neurodegenerative disorders like Alzheimer’s \cite{Buckner_2009}. Our results provide a mechanistic explanation for these findings by suggesting that hubs form key control points in brain networks; alterations to hub regions can therefore have disproportionately high impacts on system function.
3.4.3. The Role of Weak Connections in Brain Control

While our results demonstrate that hubs are theoretically implicated in moving the brain to many easily reachable states, weakly connected areas are critical for moving the brain to difficult-to-reach states. We observe that these modal control points, while distributed across the brain, tend to be predominantly located in cognitive control systems including the fronto-parietal and cingulo-opercular networks. These two systems are characterized by different functional connectivity patterns at rest [Power et al. (2011)] and are thought to support distinct functional roles within the general area of cognitive control [Elton and Gao (2014)]; task-switching [Stoet and Snyder (2009)] and task-set maintenance [Shenhav et al. (2013)], respectively. Our results suggest a fundamental underlying mechanism of cognitive control: brain regions sparsely interconnected with the rest of the brain are critically important for moving the system into difficult-to-reach states. This theoretical hypothesis is consistent with the increased engagement of the cognitive control system in highly effortful tasks [Lifshitz et al. (2013)].

More generally, the fact that weak connections play a critical role in system dynamics is one that has traditionally received little attention [Schneidman et al. (2006); Granovetter (1983)]. However, recent work has begun to demonstrate the relevance of weak connections for both cognitive function and psychiatric disease. For example, the topology of weak connections in resting state fMRI could be used to classify healthy volunteers versus schizophrenia patients, while the topology of strong connections could not [Bassett et al. (2012)]. Moreover, in healthy individuals, the topology of weak connections more accurately correlates with intelligent quotients than the topology of strong connections [Cole et al. (2012); Santarnecchi et al. (2014)]. These findings challenge the traditional view of a prominent role of strong connections in brain dynamics. Our results provide a mechanistic rationale for the importance of weak connections, which are theoretically critical in enabling a system to move to difficult-to-reach states, which may include high performance states (such as measured by IQ) or altered performance states (such as those present in psychiatric
3.4.4. The Role of Community Structure in Brain Control

In addition to the two mechanisms that enable trajectories to (i) many easily reachable states, and (ii) a few difficult to reach states, networked systems often utilize a third mechanism – boundary controllability – which enables the segregation or integration of network modules. Modular structure has been reported in structural [Chen et al. (2008)], functional [Meunier et al. (2009a)], and dynamic [Bassett et al. (2011a)] brain networks. In resting state connectivity studies, these modules have been linked to known cognitive systems [Power et al. (2011)]. Our results suggest that a widely distributed set of brain areas across all of these systems enables segregation and integration of putative cognitive modules. We also observe a particular enrichment of boundary control hubs in dorsal and ventral attentional systems, suggesting that attentional control may be implemented by boundary control strategies integrating or segregating disparate cognitive systems. Such a theoretical prediction is supported by evidence that attentional control integrates different cognitive functions [Corbetta and Shulman (2002); Pessoa (2009)], and that disconnection of attentional networks is accompanied by extensive cognitive deficits [Corbetta and Shulman (2011); Parks and Madden (2013)].

3.4.5. Methodological Considerations

Graph theory has proven to be an extremely productive framework in which to understand the structure and function of large-scale brain circuits [Bullmore and Sporns (2009)] and their implications for human cognition [Medaglia et al. (2015)]; alternative approaches that build on this framework – like network control theory – necessarily require sceptical evaluation to clearly delineate value added. Graph theory specifically and network science more generally have provided a toolbox of diagnostics to describe the organization of graphs or networks. Yet, the relationships between this organization and the system’s function remain speculative at worst and correlative at best. Groundbreaking new discoveries will necessi-
tate a fundamental turn from descriptive statistics towards mechanistic predictions. What are the mechanisms by which network structure affects functional dynamics? And how could one intervene in a network to push the system dynamics towards a specific, targeted goal? To address these questions, we must have a framework that incorporates not just brain network structure, but also models neural dynamics. Network control theory offers exactly such a framework, along with a toolbox for selecting control nodes to effect specific control strategies (e.g., average, modal, and boundary). In the context of this study, the advantages are clear: using graph theory, we can identify regions of high degree, while using network control theory, we can understand the functional role of these regions as being critical for guiding the movement of the brain into many easy-to-reach states. Similarly, using graph theory we can identify regions of low degree, while using network control theory, we can understand the functional role of these regions as being critical for guiding the movement of the brain into difficult-to-reach states. More generally, network control theory offers invaluable theoretically-validated tools to inform explanations of brain function (e.g., cognitive processes and computations), perturbations of brain function (via non-invasive stimulation paradigms), and predictions of brain function (e.g., in altered or engineered neural architectures).

As noted in the Mathematical Models section, our approach is built on a simplified linear model of neural dynamics. Neural dynamics are intrinsically nonlinear as described by decades of research, and therefore it is imperative to delineate the strengths and weaknesses of the linear model approach. First, we note that nonlinear behavior may be accurately approximated by linear behavior in certain scenarios (see for example \cite{Honey2009}). In support of this theoretical fact, reference \cite{FernandezGalan2008} proposes a linearized model for the nonlinear neural dynamics described in \cite{Honey2009}, and reference \cite{Honey2009} shows that predictions of function from structure can be obtained with both linear and nonlinear models. Second, we note that the controllability of a linearized model has implications for the controllability of the original nonlinear model. Specifically, if the linearized system is controllable, then the nonlinear system is \textit{locally controllable}.
Third, linear models of a system accurately approximate nonlinear models in a neighborhood of the operating point. For example, in \textit{gain scheduling}, several linear controllers are used to control a nonlinear system, where each controller is designed based on a linearization of the system around a certain operating point \cite{Leith2000}, and an observable parameter is then used to switch between controllers. Gain scheduling has been successfully applied in many different application areas, including flight and process control, proving that controllers based on linearized dynamics can be effectively used for the control of nonlinear dynamics. Thus, while neural dynamics are inherently nonlinear, the study of linear models of neural dynamics can offer fundamental insights into system function.

In addition to the simplified linear model, our approach is also built on diffusion imaging data and associated tractography methods. It is therefore important to be aware of the current limitations of these techniques, and to understand how those limitations might impact on observed results. Important limitations of current tractography algorithms include (i) the inability to determine the precise origin/termination of connections in the cortex, (ii) difficulty in distinguishing branching from merging or kissing axons, making long-range connections hard to interpret, and (iii) inability to distinguish afferents from efferents \cite{Jbabdi2011}. These limitations motivate ongoing methodological development in combination with postmortem validation \cite{Dell’Acqua2012}, and they also constrain the types of interpretations we can draw from tractography data. Here we have used diffusion spectrum imaging data \cite{Wedeen2005} acquired with 257 directions using a Q5 half shell acquisition scheme, and on which we applied a \textit{q}-space diffeomorphic reconstruction \cite{Yeh2011} (see Supplementary Materials). This approach provides significantly more thorough data for tractography than the more common 30-direction DTI acquisition and modeling schemes \cite{Jones2013}, particularly in estimates of longer fibers, and fibers located away from the medial wall. In the future, should accurate estimates of directionality be available, it will be interesting to examine the nuances added to the controllability profiles of brain regions based on the
polarity of their connections.

It is important to note that we have taken an explicitly quantitative approach to controllability that differs from prior qualitative approaches. In important prior work, Liu and colleagues adopt a binary notion of controllability (as originally described in Kalman et al. (1963)) that is agnostic to the difficulty of the control task. In contrast, we ask how difficult the system is to control. In practice, these two questions can provide very different insights. Although a network may be generically controllable by any single node (as estimated in Kalman et al. (1963) and adopted in Liu et al. (2011)), the actual control input may not be implementable due to actuator constraints and limitations (Liu et al. (2011)). A second important distinction between the two approaches is that structural controllability (as adopted in Liu et al. (2011)) does not inform the design of realistic control algorithms. In contrast, we explore 3 controllability notions leading to the design of control strategies posited in the literature (see Pasqualetti et al. (2014) and the references therein) and ask how they relate to structural features of human brain anatomical networks. Our choice to focus on control strategies leads to a third important distinction between the two approaches. Namely, that the results presented in Liu et al. (2011) are generic in the sense that they hold for almost every choice of network parameters Wonham (1985) but they may fail to hold if certain symmetries or constraints are present (Reinschke (1988), Section 15; and Parlangeli and Notarstefano (2012)). In contrast, the 3 control strategies utilized here depend strongly on the properties of the network under study, and are therefore sensitive to biologically relevant information present in the system.

Finally, we note that in this work, we have focused on examining the controllability of single brain areas in order to quantitatively examine their independent roles in network control. However, future work may provide additional insights by studying controllability of sets of brain regions, perhaps within specific cognitive systems.
3.5. Conclusion

A fundamental understanding of the principles by which the brain transitions between diverse cognitive states enabling behavior would necessarily have far-reaching implications for basic cognitive neuroscience and applications in myriad clinical domains [Cocchi et al. (2013)]. Our results suggest that macro-scale structural design could underlie basic cognitive control processes, via the fundamental mechanism of network controllability. These findings lay the groundwork for future studies examining relationships between individual differences in network controllability diagnostics and behavioral, cognitive, clinical, and genetic variables.

3.6. Materials

All participants volunteered with informed consent in accordance with the Institutional Review Board/Human Subjects Committee, University of California, Santa Barbara. We examine 3 diagnostics of controllability utilized in the network control literature: average controllability, modal controllability, and boundary controllability.

3.6.1. Human DSI Data Acquisition and Preprocessing

Diffusion spectrum images (DSI [Wedeen et al. (2005)]) were acquired for a total of 8 subjects in triplicate (mean age 27 ± 5 years, 2 female, 2 left handed) along with a T1 weighted anatomical scan at each scanning session [Cieslak and Grafton (2014)]. DSI scans sampled 257 directions using a Q5 half shell acquisition scheme with a maximum $b$ value of 5000 and an isotropic voxel size of 2.4mm. We utilized an axial acquisition with the following parameters: $TR = 11.4s$, $TE = 138ms$, 51 slices, FoV (231,231,123 mm). All participants volunteered with informed consent in accordance with the Institutional Review Board/Human Subjects Committee, University of California, Santa Barbara.

DSI data were reconstructed in DSI Studio (www.dsi-studio.labsolver.org) using q-space diffeomorphic reconstruction (QSDR) [Yeh and Tseng (2011)]. QSDR first reconstructs...
diffusion weighted images in native space and computes the quantitative anisotropy (QA) in each voxel. These QA values are used to warp the brain to a template QA volume in MNI space using the SPM nonlinear registration algorithm. Once in MNI space, spin density functions were again reconstructed with a mean diffusion distance of 1.25mm using three fiber orientations per voxel. Fiber tracking was performed in DSI Studio with an angular cutoff of 55°, step size of 1.0mm, minimum length of 10mm, spin density function smoothing of 0.0, maximum length of 400mm and a QA threshold determined by DWI signal in the CSF. Deterministic fiber tracking using a modified FACT algorithm was performed until 100,000 streamlines were reconstructed for each individual.

Anatomical scans were segmented using FreeSurfer [Dale et al. (1999)] and parcellated according to the Lausanne 2008 atlas included in the connectome mapping toolkit [Gerhard et al. (2011); Hagmann et al. (2008a)]. A parcellation scheme including 234 regions was registered to the B0 volume from each subject’s DSI data. The B0 to MNI voxel mapping produced via QSDR was used to map region labels from native space to MNI coordinates. To extend region labels through the gray/white matter interface, the atlas was dilated by 4mm. Dilation was accomplished by filling non-labeled voxels with the statistical mode of their neighbors’ labels. In the event of a tie, one of the modes was arbitrarily selected. Each streamline was labeled according to its terminal region pair.

3.6.2. Human DTI Data Acquisition and Preprocessing

All scans were acquired at 3 T with a Siemens Tim Trio MRI scanner with a 12 channel phased array head coil using an echo-planar diffusion-weighted technique acquired with iPAT and an acceleration factor of 2. The timing parameters of the pulse sequence were TE/TR = 94/8400 ms, 30 diffusion directions with a maximal b-value of 1000 s/mm$^2$ and two averages. Two b0 images were acquired. The matrix size was $128 \times 128$ and the slice number was 60. The field of view was $230 \times 230$ mm$^2$ and the slice thickness 2 mm. Acquisition time was 9:08 min per DTI scan. In addition to diffusion scans, a three dimensional (3D) high-resolution T1-weighted sagittal sequence image of the whole brain was obtained by
a magnetization prepared rapid acquisition gradient-echo (MPRAGE) sequence with the following parameters: TR = 15.0 ms; TE = 4.2 ms; flip angle = 9 degrees, 3D acquisition, FOV = 256 mm; slice thickness = 0.89 mm, matrix = 256 × 256.

Following prior work [Bassett et al. (2011a); Hermundstad et al. (2013, 2014)], motion artifact and image distortions caused by eddy-currents were corrected in FMRIB’s Diffusion Toolbox in FSL software by applying an affine alignment of each diffusion-weighted image to the b0 image. In the current study, we did not correct for EPI distortions. In this Siemens scanner, the geometric distortion for diffusion imaging from EPI was found in prior tests to be less than 2 mm (i.e., less than a single voxel), and mainly along the anterior posterior (phase encoding) direction. Because the resolution of the diffusion images was larger than the magnitude of the distortion, no correction was required.

3.6.3. Macaque Tract Tracing Data

To address the question of whether the relationship between controllability diagnostics and network topology (as measured by weighted degree) was conserved in non-human primates, we used a data set drawn from CoCoMac [Kötter (2004)] that delineated 2402 projections between 95 cortical and subcortical areas [Kaiser and Hilgetag (2006)]. These connectivity data were based on three extensive neuroanatomical compilations [Carmichael and Price (1994); Felleman and Van Essen (1991); Lewis and van Essen (2000)] that collectively cover large parts of the cerebral cortex. Although these data may be partially incomplete, particularly for connections of motor, auditory, and somatosensory areas [Kaiser and Hilgetag (2006)], they represent an extensive effort in tract tracing, and therefore have been used extensively in studies of primate connectivity [Kaiser and Hilgetag (2006)]. The CoCoMac database contains information on studies that report the source and target site of tracer injections, thereby specifying the specific presence or absence of anatomical projections between brain regions. The database contains no clear information on the sex or age of the monkeys, and therefore this data likely represents connectivity information from monkeys of both sexes [Scholtens et al. (2014)].
3.6.4. Association of Brain Regions to Cognitive Systems

To examine the relationship between controllability diagnostics and cognitive systems, we developed a map of brain areas to a set of cognitive systems previously defined in the literature: the fronto-parietal, cingulo-opercular, dorsal attention, ventral attention, default mode, motor and somatosensory, auditory, visual, subcortical systems [Power et al. (2011)]. Such a mapping was inspired by a recent paper from Power et al. (2012) who associated 264 brain areas to these cognitive systems, defined by a clustering technique applied to functional brain networks [Power et al. (2011)]. Similar to previous work [Power et al. (2011)], our association of areas to systems is a gross approximation and it should not be interpreted as indicating that areas have single functions. We use this association only as a pragmatic means to assess whether controllability diagnostics are differentially identified in distributed neural circuits.

The 234 areas examined in the main manuscript were drawn from 42 cortical structures. Here we associate these 42 structures to the set of 9 cognitive systems:

- **Lateral Orbitofrontal** In the Power et al. (2012) decomposition, portions of lateral orbitofrontal cortex (or BA 47) are assigned to default mode, salience, and ventral attention systems. To choose a single association for this region, we turned to the wider literature. In a recent meta-analysis, Zald and colleagues examined the role of medial and lateral orbitofrontal cortex in widespread functional networks [Zald et al. (2014)]. The lateral orbitofrontal cortex showed co-activations with prefrontal regions and areas involved in cognitive functions including language and memory but not with areas of the default mode, autonomic, and limbic systems. Rothkirch et al. (2012) similarly demonstrated that lateral orbitofrontal cortex appears to be modulated by implicit motivational value, rather than salience [Rothkirch et al. (2012)], arguing against its inclusion in the salience system. Anderson and colleagues suggest that lateral orbitofrontal cortex provides a specificity in top-down control of attention in collaboration with dorsolateral prefrontal cortex [Anderson et al. (2010)]. Cognitive
system assignment: “Ventral Attention”.

• **Pars Orbitalis** In the Power et al. (2012) decomposition, portions of pars orbitalis (or BA 47) are assigned to default mode, salience, and ventral attention systems. To choose a single association for this region, we turned to the wider literature. The pars orbitalis is a part of the ventrolateral prefrontal cortex, and is known to play a role in cognitive control processes [Xiang et al. (2010)], particularly in conflict adaptation [Egner (2011)], inhibition [Enriquez-Geppert et al. (2013)], which differ significantly from those enabled by the fronto-parietal network [Elton and Gao (2014)]. Cognitive system assignment: “Cingulo-Opercular”.

• **Frontal Pole** In this parcellation scheme, the frontal pole corresponds to portions of BA 9 and 10. These areas form hubs of the fronto-parietal cognitive control system [Tu et al. (2013)]. Cognitive system assignment: “Fronto-parietal”.

• **Medial Orbitofrontal.** The medial frontal cortex is one of the key hubs of the fronto-parietal network [Tu et al. (2013); Stern et al. (2012)]. Cognitive system assignment: “Fronto-parietal”.

• **Pars Triangularis** In this parcellation scheme, the pars triangularis corresponds to portions of BA 45, and therefore maps to the fronto-parietal cognitive control system [Elton and Gao (2014)]. Cognitive system assignment: “Fronto-parietal”.

• **Pars Opercularis** The pars of opercularis (corresponding roughly to BA 44) forms a hub of the cingulo-opercular cognitive control system [Elton and Gao (2014)]. Cognitive system assignment: “Cingulo-Opercular”.

• **Rostral Middle Frontal** The rostral middle frontal cortex, corresponding roughly to BA 10, forms a hub of the cingulo-opercular cognitive control system [Elton and Gao (2014)]. Cognitive system assignment: “Cingulo-Opercular”.

• **Superior Frontal.** In the Power et al. (2012) decomposition, portions of the superior
frontal cortex are predominantly affiliated with the default mode system, consistent with previous literature [Xu et al. (2013); Lukoshe et al. (2013); Sun et al. (2013); Fang et al. (2013)]. Cognitive system assignment: “Default Mode”.

- **Caudal Middle Frontal** The caudal middle frontal cortex is a prefrontal cortical structure broadly associated with executive function [Marqués-Iturria et al. (2014); Lopez-Larson et al. (2011)], top-down control [Durazzo et al. (2011)], and secondary motor processes [Verstraete et al. (2011); Duffield et al. (2013)]. Cognitive system assignment: “Fronto-parietal”.

- **Precentral** The precentral cortex is part of the somatosensory system. Cognitive system assignment: “Somatosensory”.

- **Paracentral** The paracentral cortex is part of the somatosensory system. Cognitive system assignment: “Somatosensory”.

- **Rostral Anterior Cingulate** The anterior cingulate is a hub of the cingulo-opercular network [Ninaus et al. (2013); Gratton et al. (2013); Vaden et al. (2013); Becerril and Barch (2013); Sestieri et al. (2014); Tu et al. (2012)]. Cognitive system assignment: “Cingulo-Opercular”.

- **Caudal Anterior Cingulate** The anterior cingulate is a hub of the cingulo-opercular network [Ninaus et al. (2013); Gratton et al. (2013); Vaden et al. (2013); Becerril and Barch (2013); Sestieri et al. (2014); Tu et al. (2012)]. Cognitive system assignment: “Cingulo-Opercular”.

- **Posterior Cingulate**. The posterior cingulate is a known hub of the default mode system [Stern et al. (2012); Khalsa et al. (2013); Leech and Sharp (2014)]. Cognitive system assignment: “Default Mode”.

- **Isthmus Cingulate** The isthmus cingulate is thought to be a hub of the default mode system [Zhu et al. (2013)] and of the limbic system [Leow et al. (2013)]. Cognitive
• **Post Central** The postcentral cortex is part of the somatosensory system. Cognitive system assignment: “Somatosensory”.

• **Supramarginal** The supramarginal gyrus appears to play a role in the dorsal [Schmidt et al. (2013)] and ventral [Burianová et al. (2012)] attention networks, and executive function more broadly [Yin et al. (2012); Vasconcelos et al. (2014)]. In the Power et al. (2012) decomposition, this area was assigned to the cingulo-opercular system [Power et al. (2011)]. Cognitive system assignment: “Cingulo-Opercular”.

• **Superior Parietal** The superior parietal cortex plays a role in both the dorsal attention system [Xu et al. (2014); Sestieri et al. (2013)] and the somatosensory-motor system [Fabbri et al. (2014)]. Cognitive system assignment: “Dorsal Attention”.

• **Inferior Parietal**. The inferior parietal cortex is one of the key hubs of the fronto-parietal network [Tu et al. (2013); Stern et al. (2012)]. Cognitive system assignment: “Fronto-parietal”.

• **Precuneus** The precuneus is a hub of the default mode system [Cavanna and Trimble (2006); Sheline and Raichle (2013)]. Cognitive system assignment: “Default Mode”.

• **Cuneus** The cuneus is a part of the visual system [Xu et al. (2014); Delli Pizzi et al. (2014); Collignon et al. (2013)]. Cognitive system assignment: “Visual”.

• **Pericalcarine** The pericalcarine is a part of the visual system [Gaetz et al. (2012); Delli Pizzi et al. (2014)]. Cognitive system assignment: “Visual”.

• **Lateral Occipital** The lateral occipital cortex is a part of the visual system [Bedny et al. (2012)]. Cognitive system assignment: “Visual”.

• **Lingual** The lingual gyrus is a part of the visual system [Delli Pizzi et al. (2014); Boldt et al. (2014)]. Cognitive system assignment: “Visual”.

...
- **Fusiform** The lingual gyrus is a part of the visual system [Xu et al. (2014)]. Cognitive system assignment: “Visual”.

- **Parahippocampal** The parahippocampal cortex has been associated with many cognitive processes including visuospatial processing and episodic memory [Aminoff et al. (2013)]. Cognitive system assignment: “Other”.

- **Entorhinal cortex** The entorhinal cortex encodes visual information [Killian et al. (2012)]. Cognitive system assignment: “Visual”.

- **Temporal Pole** The temporal pole plays a role in language processing, including naming [Semenza (2011)], and in social and emotional processing [Olson et al. (2007)]. Cognitive system assignment: “Other”.

- **Inferior Temporal** The inferior temporal cortex is associated with visual processing [Hirabayashi and Miyashita (2014)], emotion perception of visual objects [Sabatinelli et al. (2011)], and shape recognition [Tompa and Sáry (2010)]. Cognitive system assignment: “Visual”.

- **Middle Temporal** The middle temporal cortex is associated with cognitive control processes [Noonan et al. (2013)], theory of mind [Rodrigo et al. (2014)], and social cognition [Hayashi et al. (2014)]. Cognitive system assignment: “Other”.

- **Bank of the Superior Temporal Sulcus** The bank of the superior temporal sulcus forms a part of the early cortical auditory network [Kilian-Hütten et al. (2011)]. Cognitive system assignment: “Auditory”.

- **Superior Temporal** The superior temporal cortex forms a part of the auditory system [Woods and Alain (2009)]. Cognitive system assignment: “Auditory”.

- **Transverse Temporal** The transverse temporal cortex forms a part of the auditory system [Simon et al. (2013)]. Cognitive system assignment: “Auditory”.

100
• **Insula.** The insula is one of the key hubs of the fronto-parietal network [Tu et al. (2013); Stern et al. (2012)]. Cognitive system assignment: “Fronto-parietal”.

• **Thalamus.** Cognitive system assignment: “Subcortical”.

• **Caudate.** Cognitive system assignment: “Subcortical”.

• **Putamen.** Cognitive system assignment: “Subcortical”.

• **Pallidum.** Cognitive system assignment: “Subcortical”.

• **Nucleus Accumbens.** Cognitive system assignment: “Subcortical”.

• **Hippocampus.** Cognitive system assignment: “Subcortical”.

• **Amygdala.** Cognitive system assignment: “Subcortical”.

• **Brainstem.** Cognitive system assignment: “Other”.

CHAPTER 4 : Optimal Trajectories for Brain State Transitions

4.1. Introduction

Neuroscientific investigations seek to reveal how neural systems perform complex functions, how those systems and functions are altered in disease states, and how therapeutic interventions can be used to redirect these alterations. In essence, all three lines of investigation seek to address how neural systems move along dynamic trajectories: of cognitive function, disease, or recovery. However, fundamental and therefore generalizable mechanisms of these movements have remained elusive.

The complexity of neural dynamics stems in part from the architectural complexity of the underlying anatomy. Different components (neurons, cortical columns, brain areas) are linked with one another in complex spatial patterns that enable diverse neural functions. These structural interactions can be represented as a graph or network, where component parts form the nodes, and where anatomical links form the edges between nodes [Bullmore and Sporns (2009)]. The architecture of these networks displays fascinating heterogenous features that play a role in neural function [Medaglia et al. (2015)], development [Di Martino et al. (2014)], disease [Braun et al. (2015)], and sensitivity to rehabilitation [Weiss et al. (2011)]. Despite these recent discoveries, how these architectural features constrain neural dynamics in any of these phenomena is far from understood.

One simple and intuitive way to formulate questions about how neural dynamics are constrained by brain network architecture is to define a state of the brain by the $1 \times N$ vector representing magnitudes of neural activity across $N$ brain regions, and further to define brain network architecture by the $N \times N$ adjacency matrix representing the number of white matter streamlines linking brain regions. Building on these two definitions, we can ask how the organization of the white matter architecture constrains the possible (or observed) brain states. Moreover, building on decades of cognitive neuroscience that have carefully delineated the role of regional activation in cognitive functions [Gazzaniga (2013);
Szameitat et al. (2011); Alavash et al. (2015), we can then map brain states to cognitive processes, and extend our question to: how does the organization of white matter architecture constrain cognitive states [Hermundstad et al. (2013, 2014)], and the processes that enable us to move between those states [Cocchi et al. (2013)]?

These types of questions can be addressed by drawing on recent insights from the application of network control theory [Pasqualetti et al. (2014)] to neuroimaging data [Gu et al. (2015a)]. In this framework, biologically-informed mathematical models of brain dynamics can be used to infer how the topology of white matter architecture constrains how the brain may effect (or control) transitions between brain states. Prior work using these techniques has assessed the average or global role of white matter emanating from single brain regions in explaining state transitions [Gu et al. (2015a)]. In contrast, here we examine transitions that are elicited via the collective control of region sets. This choice is motivated by our growing understanding of large-scale brain dynamics, as measured by functional magnetic resonance imaging (fMRI). Pervasive evidence suggests that sets of brain regions are recruited in concert with one another to support intrinsic cognitive function [Power et al. (2011); Yeo et al. (2011)], and to effect action [Bassett et al. (2011a); Bassett and Siebenhuhner (2013); Bassett et al. (2015)]. Thus, extending our focus from single-point control to multi-point control is consistent with our growing understanding of the biology.

In a second important departure from previous work, we examine how the brain may move from a specified initial state to a specified target state in finite time, rather than on average over an infinite number of unspecified transitions [Gu et al. (2015a)]. The choice to focus on specific trajectories of brain function is motivated by the fact that the brain, in its baseline condition, displays high activity in a small set of regions, predominantly located in precuneus, posterior cingulate, medial and lateral temporal, and superior frontal cortex [Raichle (2015); Raichle and Snyder (2007); Raichle et al. (2001)]. This so-called default mode forms a natural initial state, from which we are interested in transitions to selected target states. The selection of target states here is more realistic compared to averaging
over the full state spaces because in practice, the brain explores only a limited part of its complete state space; many model-allowed states are never visited. While many such target states might be of interest in various contexts, we focus this first study on examining transitions into target states of high activity in primary sensorimotor cortex: specifically visual, auditory, and motor cortices. These states represent the simplest and most fundamental targets to transition to from the default mode: for example, transitioning from the default mode to visual states might represent an immediate response to a surprising stimuli. Similarly, the transition from the default mode to motor states might represent the simple transition from rest to action. Moreover, these transitions are of particular interest in many clinical disorders because the cognitive functions performed by these target areas are often altered in stroke [Carter et al. (2012)] and traumatic brain injury [Nudo (2006); Lee et al. (2011)], significantly altering quality of life [Kalpinski et al. (2013)].

Using network control theory, we examine the optimal trajectories from an initial state (composed of high activity in the default mode system) to target states (composed of high activity in sensorimotor systems) with finite time and limited energy. In this optimal control context, we investigate the role of white matter connectivity between brain regions in constraining dynamic state transitions by asking three interrelated questions. First, we ask which brain regions are theoretically predicted to be most energetically efficient in eliciting state transitions. Second, we ask whether these state transitions are best elicited by one of three well-known control strategies [Gu et al. (2015a)]. Third, we ask how stable these state transitions are if we remove nodes from the network, and we compare this theoretical stability between a group of healthy adults and a group of patients with mild traumatic brain injury.

4.2. Results

Our fundamental goal in this work was to build a mathematical framework to model brain state transitions from the perspective of network control theory. To begin, we set the initial state of the brain to be an activation pattern consistent with those empirically observed
in the brain’s baseline condition. More specifically, we set the initial state such that the regions of the default mode network had activity magnitudes equal to 1 (“on”), while all other regions had activity magnitudes equal to 0 (“off”). Furthermore, we examined 3 distinct target states such that regions of the (i) auditory, (ii) visual, or (iii) motor systems had activity magnitudes equal to 1 (“on”), while all other regions had activity magnitudes equal to 0 (“off”). In this context, we sought to understand characteristics of the transitions between initial and target states that could be performed with minimal energy, minimal time, and along short trajectories in state space by multiple control regions (multi-point control; see Fig. 23C and Methods).

4.2.1. Task Control

The trajectory of a given task is a curve from the initial state to the target state. When the system is fully controlled, the distance to the target state decreases monotonically along the trajectory. Usually, the distance to the target state may increase within some periods of time and reaches zero finally. In our model, the maximal distance to the target state (extreme distance) is affected by both the number of control nodes and the value of the balance parameter $\rho$. Intuitively, the larger control set you have, the lower extreme distance you would get because of the better distribution of the input energy (Fig. 22A). For the impact of $\rho$, when it is too small, the optimal trajectory would tend to arriving at the target instantly, which requires a huge amount of energy within a very short period of time. This is theoretically workable. However, in practice, the over-big input would cause the numeric instability in the trajectory and divergent extreme distance. As we show in Fig. 22B, the trajectory becomes stable when $\rho > 0.01$. In this article, we set $\rho = 1$ in the stable range to given equal weight for the distance term and energy term.
Figure 22: **Numeric Stability of Trajectories** In this figure, we compute on the average connectivity matrix of th subjects and show (A) The trajectory’s maximal distance to the target state decreases as the number of control nodes increases. (B) The trajectory’s maximal distance to the target state decreases as the balance parameter $\rho$ increases and becomes “stable” when $\rho > 0.01$. 
4.2.2. Characteristics of Optimal Control Trajectories

To effect the three state transitions from default mode to auditory, visual, and motor systems (Fig. 24A), we take a hypothesis-driven approach and assign dorsal and ventral attention [Posner and Petersen (1989)], fronto-parietal, and cingulo-opercular cognitive control regions [Gu et al. (2015a)] as the so-called “control set”: that is, this set of 87 regions will utilize control energy using a multi-point control strategy. (For a data-driven assessment of optimal control sets in a somewhat related control problem, see [Betzel et al. (2016)].) In the optimal control trajectories for each of these three state transitions, we observe that brain regions display time-varying activity magnitude that differ from one another (Fig. 24B). The optimal trajectories themselves display multiple peaks in the distance from the target state as a function of time, and are altered very little by whether the target state is the auditory, visual, or motor system (Fig. 24C). As is intuitive, the time-dependent energy utilized by the control set is inversely related to the distance between the current state and the target state: when the current state is far from target state, little control energy is utilized, while as the current state moves closer to the target state, a larger magnitude of control energy is utilized (Fig. 24D).

It is important to note that these general characteristics of the optimal control trajectories are dependent on our choice of the control set (which here we guide with biologically motivated hypotheses), as well as on a penalty on the time required for the transition (ρ in Equation 4.3; see Methods). In the supplement, we examine the effect of alternative choices for both the control set and ρ. First, we find that when the control set includes every node in the network, the distance to the target state decreases monotonically to zero along the trajectory (Fig. 22A). Second, we consider the effect of the temporal penalty term, ρ. For the results presented here, we fix ρ to be equal to 1. However, we explore a wide range of rho values, and show that when ρ is small, the optimal control trajectory is largely driven by a minimization of the integrated squared distance to the target. In contrast, when ρ is large, the optimal control trajectory is largely driven by the magnitude of the utilized
Importantly, we did not perform a full sweep of $\rho$ from 0 to infinity because very small values of $\rho$ cause numeric instabilities in the calculations.

### 4.2.3. Structurally-Driven Task Preference for Control Regions

We next ask the question whether certain brain regions are located at specific points in the structural network that make them predisposed to play consistent and important roles in driving optimal control trajectories. To answer this question, we choose control sets of the same size as the brain’s hypothesized cognitive control set; recall that in the previous section, we defined the brain’s cognitive control set to consist of the 87 nodes of the dorsal and ventral attention, fronto-parietal, and cingulo-opercular systems following [Gu et al. (2015a)]. Here, we choose the 87 regions of these new control sets uniformly at random from the set of nodes that were not in the initial state (default mode system) or in any of the final states (auditory, visual, or motor systems). Using these “random” control sets, we computed the optimal control trajectory for each of the three state transitions and for each subject separately. Then, we rank the random control sets in descending order according to the energy cost of the trajectory and we assign every region participating in an $r$-ranked control set with rank-value $r$. Next, we define the control efficiency of a brain region to be the average of its rank values in all of the random control sets divided by the total number of regions in a set. Intuitively, a region with a high control efficiency is one that exerts control with little energy utilization.

In general, we observe that a region’s preference for being an optimal controller (exerting control with little energy utilization) is positively correlated with its network communicability to the regions of high activity in the target state (Spearman correlation $r = 0.27$, $p < 4.8 \times 10^{-4}$; see Fig. 25A). We recall that network communicability is a measurement of the distance from one region to another taking into account walks of all possible lengths (see Methods). Interestingly, we observed this same correlation between control efficiency and network communicability across optimal control trajectories for all three state transitions, from the default mode to the auditory ($r = 0.36$, $p = 1.4 \times 10^{-8}$),
visual \( (r = 0.51, p = 1.1 \times 10^{-16}) \), or motor \( (r = 0.42, p = 2.1 \times 10^{-11}) \) systems (Fig. 25B-D). Together, these results indicate that regions that are close (in terms of walk lengths) to regions of high activity in the target state are efficient controllers for that specific state transition.

The general role that network proximity to the target state plays for control regions ensures that regions that are proximate to all three target states (auditory, visual, and motor) will be consistent controllers, while regions that are proximate to only one of the target states will be task-specific controllers. To better understand the anatomy of efficient controllers, we transformed control efficiency values to \( z \)-scores and defined an efficient control hub to be any region whose associated \( p \)-value was less than 0.025. Across all three state transitions, we found that the supramarginal gyrus and the inferior parietal lobule consistently acted as efficient control hubs. The consistent control role of these regions is consistent with their anatomical location ventral to the angular gyrus where primary sensory inputs converge [Kandel et al., (2000)]. The areas that are more specific to the three state transitions include medial parietal cortex (motor transition), orbitofrontal and inferior temporal (visual transition), and superior temporal (auditory transition).

4.2.4. Robustness of Control in Health and Following Injury

Theoretically, if every region of the brain is treated as a control region, then the brain can be fully controlled (see Supplement). However, how the system responds to the removal of nodes from the control set gives us an intuition regarding the robustness of control. Indeed, for different networks, removing a small number of control nodes can decrease the controllability either minimally or dramatically (Fig. 26A). To directly assess control robustness from a theoretical point of view, we iteratively remove nodes from the network and compute the energetic impact of each region on the optimal trajectory as the resulting increase in the log value of the energy cost (see Eqn 4.22 in Methods). Intuitively, regions with high energetic impact are those whose removal from the network causes the greatest increase in the energy required for the state transition. Across all subjects and all tasks, we
observe that the regions with the highest energy impact are the supramarginal gyrus and the inferior parietal lobule, the same regions that emerged as consistent efficient controllers in Fig. 24A.

In addition to studying the energetic impact of a single region, we can also examine the energetic impact of regions across the whole brain in one individual versus another individual. Intuitively, a brain with low average energetic impact of regions can be said to display robust control, being little affected by point perturbations. Counterintuitively, this robustness is often detrimental to complex systems as it indicates that the network may have high redundancy and decreased specificity in connectivity patterns across nodes, significantly limiting the dynamic range of states available to the system.

We used these notions to assess the impact of mild traumatic brain injury (mTBI) on brain connectivity. First, we observed that individuals with mTBI displayed anatomically similar patterns of energetic impact on control trajectories as regions are removed from the network (Fig. 26B). However, the average magnitude and variability of the energetic impact differed significantly between the two groups, with individuals having experienced mTBI displaying significantly lower values of average magnitude of energetic impact (permutation test: \( p = 2.5 \times 10^{-4} \)) and lower values of the average standard deviation of energetic impact (\( p = 9.1 \times 10^{-5} \)). These results indicate that the mTBI patients have greater robustness of control, consistent with a homogenization of network connectivity and increase in redundancy. We note that common graphic metrics including the degree, path length, clustering coefficient, modularity, local efficiency, global efficiency, and density were not significantly different between the two groups, suggesting that this effect is specific to control.

4.2.5. Graphical Metrics

In the previous section, we show that the mild Traumatic Brain Injury (mTBI) group have significantly lower removal controllability (low energy impact by the removal) than the healthy group. We want to know whether the difference between the two groups can also
be recognized by the commonly used graphical measures as well. Seven metrics, Degree, Characteristic path length, Clustering Coefficient, Modularity, Global efficiency, and Density are computed on each subject of the two groups and permutation tests are performed to calculate the $p$-values of the hypothesis test $H_0$-the mTBI and the healthy groups have same means for the given metric vs $H_1$-the mTBI and the healthy groups have different means for the given metric. It turns out that none of these graphical metrics display any significant distinguish between the two groups(see Table. 4.2.5). Thus the controllability measures actually quantify something more than the listed measures above.
Table 3: The \( p \)-values for the Permutation Tests of the Graphical Metrics of the Healthy and mTBI Groups. In this table, we list the \( p \)-values from the permutation tests in distinguishing the mTBI group and the healthy group. All metrics are the averages across regions if they are regionally defined. The \( p \)-value here is the two-sided \( p \)-value, which is defined as the minimum of the two one-sided \( p \)-values with no assumptions on the sign of the difference. For the notations, DEG is short for Degree, CPL is short for Characteristic path length, C-COFF is short for Clustering coefficient, MODU is short for Modularity, L-EFF is short for Local efficiency, G-EFF is short for Global efficiency, and DEN is short for Density.

<table>
<thead>
<tr>
<th></th>
<th>DEG</th>
<th>CPL</th>
<th>C-COFF</th>
<th>MODU</th>
<th>L-EFF</th>
<th>G-EFF</th>
<th>DEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )</td>
<td>0.3635</td>
<td>0.1791</td>
<td>0.4338</td>
<td>0.4208</td>
<td>0.1460</td>
<td>0.3220</td>
<td>0.3635</td>
</tr>
</tbody>
</table>
4.3. Discussion

Here we ask whether structural connectivity estimated by white matter streamlines forms a fundamental constraint on how the brain may move between diverse cognitive states. To address this question, we capitalize on groundbreaking recent advances in the field of network control theory to identify and characterize optimal trajectories from an initial state (composed of high activity in the default mode system) to target states (composed of high activity in sensorimotor systems) with finite time, limited energy, and multi-point control. Using structural brain networks estimated from diffusion imaging data acquired in a large cohort of 48 healthy individuals and 11 patients with mild traumatic brain injury, we show that these optimal control trajectories are characterized by time-varying activity magnitudes across brain regions, the patterns of which differ from one region to another. We show that the regions critical for eliciting these state transitions differ depending on the target state, but that heteromodal association hubs – predominantly in supramarginal gyrus and inferior parietal lobule – are consistently recruited for all three transitions. Finally, we study the sensitivity of optimal control trajectories to the removal of nodes from the network, and we demonstrate that brain networks from individuals with mTBI display maladaptive control capabilities suggestive of a limited dynamic range of states available to the system. Together, these results offer initial insights into how structural network differences between individuals impact their potential to control transitions between cognitive states.

4.3.1. Role of structural connectivity in shaping brain functional patterns

A growing body of literature on the interrelationship of brain structure and function has demonstrated that the brain’s network of anatomical connections constrains the range of spontaneous [Deco et al. (2011)] and task-related [Hermundstad et al. (2013)] fluctuations in brain activity. Evidence for such structural underpinnings comes from two distinct lines of research. On one hand, empirical studies have demonstrated that structural insults in the form of lesions result in acute reorganization of the brain’s pattern of functional coupling [Johnston et al. (2008); O’Reilly et al. (2013)]. These observations are further buttressed by
Figure 23: **Conceptual Schematic.** (A) Diffusion imaging data can be used to infer connectivity from one voxel to any other voxel via diffusion tractography algorithms. (B) From the tractography, we construct a weighted network in which \( N = 234 \) brain regions are connected by the quantitative anisotropy along the tracts linking them (see Methods). (C) We study the optimal control problem in which the brain starts from an initial state (red) at time \( t = 0 \) and uses multi-point control (control of multiple regions; blue) to arrive at a target state (yellow) at time \( t = T \).
Figure 24: **Optimal Control Trajectories.** (A) We study 3 distinct types of state transitions in which the initial state is characterized by high activity in the default mode system, and the target states are characterized by high activity in auditory (blue), visual (green), or motor (red) systems. (B) The activation profiles of all $N = 234$ brain regions as a function of time along the optimal control trajectory, illustrating that activity magnitudes vary by region and by time. (C) The average distance from the current state $x(t)$ to the target state $x(T)$ as a function of time for the trajectories from default mode system to the auditory, visual, and motor systems, illustrating behavior in the large state space. (D) The average control energy utilized by the control set as a function of time for the trajectories from default mode system to the auditory, visual, and motor systems. See Fig. S2B for additional information on the range of these control energy values along the trajectories. Colors representing target states are identical in panels (A), (C), and (D).
Figure 25: **Structurally-Driven Task Preference for Control Regions.** (A) Left Regions with high control efficiency (see Eqn 4.20) across all 3 state transitions: from the default mode to auditory, visual, and motor systems. Right Scatterplot of the control efficiency with the average network communicability to the target regions. (B–D) Regions with high control efficiency for the transition from default mode to (B) motor, (C) visual, and (D) auditory targets (top); scatter plot of control efficiency versus normalized network communicability with targets (bottom).
Figure 26: Robustness of Control in Health and Following Injury. (A) Theoretically, the brain is fully controllable when every region is a control point, but may not be fully controllable when fewer regions are used to effect control. (B) The regions with the highest values of energetic impact on control trajectories upon removal from the network, on average across subjects and tasks, were the supramarginal gyrus and the inferior parietal lobule. In general, the healthy group and the mTBI group displayed similar anatomical patterns of energetic impact. (C) Magnitude and standard derivation of energetic impact averaged over regions and tasks; boxplots indicate variation over subjects. After removing outliers in the distribution, patients with mTBI displayed significantly lower values of average magnitude of energetic impact (permutation test: $p = 2.5 \times 10^{-4}$) and lower values of the average standard deviation of energetic impact ($p = 9.1 \times 10^{-5}$) than healthy controls.
simulation studies in which structural connectivity has been used to constrain interactions among dynamic elements in biophysical models of brain activity [Honey et al. (2007, 2009); Adachi et al. (2011)] and network communication models [Goñi et al. (2014); Abdelnour et al. (2014); Mišić et al. (2015)]. Though this forward modeling approach has proven fruitful in predicting observed patterns of functional connectivity, the precise mapping of brain structure to function remains unclear.

The present study builds on this body of work, using a dynamical model of how brain activity propagates over a network in order to gain insight into what features of that network facilitate easy transitions from a baseline (default mode) state to states where the brain’s primary sensorimotor systems are activated. We use this model to demonstrate that brain regions are differentially-suited for particular control tasks, roles that can be predicted on the basis of how well-connected they are to regions in the target state. We further demonstrate that this mapping of brain structure to specific functions is altered in individuals with TBI, suggesting that injury may alter control profiles of individual brain regions.

4.3.2. Maladaptive Control in Traumatic Brain Injury

Traumatic brain injury is a common source of brain dysfunction, affecting more than 200,000 individuals per year in the United States alone. Injuries – often caused by motor vehicle and sports accidents – result in damage to neuronal axons, particularly in long-distance white matter fiber bundles [Johnson et al. (2013)]. The pattern of injury is often widespread and variable across individuals [Kinnunen et al. (2011); Sidaros et al. (2008); Hellyer et al. (2013)], challenging comprehensive predictors and generalizable interventions. Recent evidence directly addresses these challenges by suggesting that this widespread damage to white matter tracts critically impacts large-scale network organization in the human brain, as measured by diffusion imaging tractography [Kinnunen et al. (2011); Fagerholm et al. (2015)], thereby directly altering cognitive function [Sharp et al. (2014)]. Indeed, machine learning techniques can be used to distinguish mTBI from controls with 93.4% prediction accuracy using graph-based statistics, which separately correlated with cognitive perfor-
mance including information processing speed, executive function, and associative memory [Fagerholm et al. (2015)]. The fundamental drivers of these predictions are located at regions of the brain characterized by high eigenvector centrality, including cingulate cortex and caudate, which display decreased eigenvector centrality following injury. Interestingly, in a separate literature from the dynamical systems community in theoretical physics (and with some validated applicability to neuroimaging data [Khambhati et al. (2016)]), networks with decreased heterogeneity of eigenvector centrality have been shown to be more synchronizable [Rad et al. (2008); Jalili et al. (2008); Dadashi et al. (2010); Chun-Hsien and Suh-Yuh (2014)]. Our results bridge these two literatures by showing that individuals with mTBI display structural brain networks that are more controllable, perhaps due to the known increases in structural network homogeneities [Fagerholm et al. (2015)] that may lead to an increased ease in synchronizing default mode and primary sensorimotor systems. However, it is interesting to note that other physical systems display dynamics in which controllability and synchronizability are antagonistically related [Whalen et al. (2015)], and it would be interesting in future to assess the role of structural connectivity in predicting patterns of functional coherence in TBI more broadly across executive, attention, memory, and reward-related circuits.

4.3.3. Methodological Considerations

A few methodological points are worthy of additional considerations. First, in this study we examined structural brain networks derived from diffusion imaging data and associated tractography algorithms. These algorithms remain in their relative infancy, and can still report spurious tracts or fail to report existing tracts [Thomas et al. (2014); Reveley et al. (2015); Pestilli et al. (2014)]. Formal validation in animal studies remains the gold standard for these types of data. Second, following [Gu et al. (2015a); Betzel et al. (2016); Muldoon et al. (2016b)], we employ a linear dynamical model, consistent with prior empirical studies demonstrating their ability to predict features of resting state fMRI data [Galán (2008); Honey et al. (2009)]. This choice is to some degree predicated on the well-developed
theoretical and analytical results in the engineering and physics literature examining the relationship between control and network topology [Liu et al. (2011); Müller and Schuppert (2011); Yan et al. (2012)]. Moreover, it is plausible that even these results using simple linear models may offer important intuitions for controlling nonlinear models of brain function. Indeed, theoretical work over the last several decades has demonstrated the utility of describing non-linear systems in terms of a linear approximation in the neighborhood of the system’s equilibrium points [Luenberger (1979)].

4.3.4. Future Directions

An interesting hypothesis generated by the current framework is that control capabilities may be altered dimensionally across traditionally separated diagnostic groups that display dysconnectivity in network hubs, as measured by regions of high eigenvector centrality. Indeed, mounting evidence suggests that the overload or failure of brain network hubs may be a common neurophysiological mechanism of a range of neurological disorders including Alzheimer’s disease, multiple sclerosis, traumatic brain injury and epilepsy [Stam (2014)]. Alterations in these hubs can also be used to predict the progression of psychiatric disorders such as schizophrenia [Collin et al. (2015)]. In both neurological and psychiatric disorders, these changes to network hubs may alter the control capabilities of the individual, challenging the normal executive functions required for daily living. It is also intuitively plausible that normal variation in hub architecture may play a role in individual differences in control capabilities in healthy individuals, impacting on the speed with which they transition between cognitive states. These topics will form important provender for future work.

4.4. Models and Materials

4.4.1. Human DSI Data Acquisition and Preprocessing

Both T1 weighted anatomical images and diffusion spectrum images (DSI) were acquired from 59 human adults with 72 scans in total, among which 61 scans were acquired from 48 healthy subjects (mean age 22.6 ± 5.1 years, 24 female, 2 left handed) and 11 were acquired
from individuals with mild traumatic brain injury \cite{Cieslak2014}(mean age 33.8 ± 13.3 years, 4 female, handedness unclear). Participants in the mild traumatic brain injury group were recruited by advertisement and recruitment from local neurologists. Inclusion criteria were age ≥18, a history of community acquired concussion in the preceding 10–180 days with any persistent cognitive complaints based on self report. The concussions were primarily secondary to motor vehicle, bicycle, skateboard and horseback riding accidents. Subjects with an abnormal CT (eg., skull fracture or hemorrhage) or a preexisting neurologic condition were excluded. All participants volunteered with informed written consent in accordance with the Institutional Review Board/Human Subjects Committee, University of California, Santa Barbara.

DSI scans sampled 257 directions using a Q5 half shell acquisition scheme with a maximum \( b \) value of 5000 and an isotropic voxel size of 2.4mm. We utilized an axial acquisition with the following parameters: \( TR = 11.4 \)s, \( TE = 138 \)ms, 51 slices, FoV (231,231,123 mm). DSI data were reconstructed in DSI Studio (www.dsi-studio.labsolver.org) using \( q \)-space diffeomorphic reconstruction (QSDR) \cite{Yeh2011}. QSDR first reconstructs diffusion weighted images in native space and computes the quantitative anisotropy (QA) in each voxel. These QA values are used to warp the brain to a template QA volume in MNI space using the SPM nonlinear registration algorithm. Once in MNI space, spin density functions were again 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 CSF. Deterministic fiber tracking using a modified FACT algorithm was performed until 100,000 streamlines were reconstructed for each individual.

4.4.2. Structural Network Construction

In addition to diffusion scans, a three-dimensional high-resolution T1-weighted sagittal sequence image of the whole brain was obtained at each scanning session by a magnetization-
prepared rapid acquisition gradient-echo sequence with the following parameters: TR=15.0 ms; TE=4.2ms; flip angle=9 degrees, 3D acquisition, FOV=256mm; slice thickness=0.89mm, matrix=256 ×256. Anatomical scans were segmented using FreeSurfer [Dale et al. (1999)] and parcellated according to the Lausanne 2008 atlas included in the connectome mapping toolkit [Hagmann et al. (2008b)]. A parcellation scheme including 234 regions was registered to the B0 volume from each subject’s DSI data. The B0 to MNI voxel mapping produced via QSDR was used to map region labels from native space to MNI coordinates. To extend region labels through the gray/white matter interface, the atlas was dilated by 4mm. Dilation was accomplished by filling non-labeled voxels with the statistical mode of their neighbors’ labels. In the event of a tie, one of the modes was arbitrarily selected. Each streamline was labeled according to its terminal region pair.

From these data, we built structural brain networks from each of the 72 diffusion spectrum imaging scans. Consistent with previous work [Bassett et al. (2010, 2011a); Hermundstad et al. (2013); Klimm et al. (2014); Gu et al. (2015a); Muldoon et al. (2016a,b); Sizemore et al. (2015)], we defined these structural brain networks from the streamlines linking \( N = 234 \) large-scale cortical and subcortical regions extracted from the Lausanne atlas [Hagmann et al. (2008b)]. We summarize these estimates in a weighted adjacency matrix \( A \) whose entries \( A_{ij} \) reflect the structural connectivity between region \( i \) and region \( j \) (Fig. 23A).

Following [Gu et al. (2015a)], here we use an edge weight definition based on the quantitative anisotropy (QA). QA is described by Yeh et. al (2010) as a measurement of the signal strength for a specific fiber population \( \hat{a} \) in an ODF \( \Psi(\hat{a}) \) [Yeh et al. (2010); Tuch (2004)]. QA is given by the difference between \( \Psi(\hat{a}) \) and the isotropic component of the spin density function (SDF, \( \psi \)) \( \text{ISO}(\psi) \) scaled by the SDF’s scaling constant. Along-streamline QA was calculated based on the angles actually used when tracking each streamline. Although along-streamline QA is more specific to the anatomical structure being tracked, QA is more sensitive to MRI artifacts such as B1 inhomogeneity. QA is calculated for each streamline. We then averaged values over all streamlines connecting a pair of regions, and used this
value to weight the edge between two regions.

4.4.3. Network Control Theory

Next, we consider the general question of how the brain moves between different states, where a state is defined as a pattern of activity across brain regions or voxels. In particular, we are interested in studying how the activity in individual brain regions affects the trajectory of the brain as it transitions between states; here, we define a trajectory as a set of states ordered in time. To address this question, we follow [Gu et al. (2015a); Muldoon et al. (2016a); Betzel et al. (2016)] by adopting notions from the emerging field of network control theory, which offers a theoretical framework for describing the role of network nodes in the control of a dynamical networked system.

Network control theory is predicated on the choice of both a structural network representation for the system, and a prescribed model of node dynamics. In the context of the human brain, a natural choice for the structural network representation is the graph on $N$ brain regions whose $ij$th edge represents the QA between node $i$ and node $j$. The choice for the model of node dynamics is perhaps less constrained, as many models are available to the investigator. These models range in complexity from simple linear models of neural dynamics with few parameters to nonlinear neural mass models with hundreds of parameters [Gu et al. (2015a); Muldoon et al. (2016a)].

In choosing a model of neural dynamics to employ, we consider multiple factors. First, although the evolution of neural activity acts as a collection of nonlinear dynamic processes, prior studies have demonstrated the possibility of predicting a significant amount of variance in neural dynamics as measured by fMRI through simplified linear models [Galán (2008); Honey et al. (2009); Gu et al. (2015a)]. On the basis of this literature, we employ a simplified noise-free linear continuous-time and time-invariant network model

$$\dot{x}(t) = Ax(t) + Bu(t),$$

(4.1)
where $x : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^N$ describes the state of brain regions over time, and $A \in \mathbb{R}^{N \times N}$ is a symmetric and weighted adjacency matrix. The diagonal elements of the matrix $A$ satisfy $A_{ii} = 0$. The input matrix $B_K$ identifies the control nodes $K$ in the brain, where $K = \{k_1, \ldots, k_m\}$ and
\[
B_K = [e_{k_1}, \ldots, e_{k_2}]
\] (4.2)
and $e_i$ denotes the $i$-th canonical vector of dimension $N$. The input $u_K : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^m$ denotes the control strategy. Intuitively, this model enables us to frame questions related to brain state trajectories in a formal mathematics. Moreover, it allows us to capitalize on recent advances in network control theory [Pasqualetti et al. (2014)] to inform our understanding of internal cognitive control [Gu et al. (2015a); Betzel et al. (2016)] and to inform the development of optimal external neuromodulations using brain stimulation [Muldoon et al. (2016b)].

4.4.4. Optimal Control Trajectories

Given the above-defined model of neural dynamics, as well as the structural network representation extracted from diffusion imaging data, we can now formally address the question of how the activity in individual brain regions effects the trajectory of the brain as it transitions between states.

We begin by defining an optimization problem to identify the trajectory between a specified pair of brain states that minimizes a given cost function. We define a cost function by the weighted sum of the energy cost of the transition and the integrated squared distance between the transition states and the target state. We choose this dual-term cost function for two reasons. First, theoretically, the energy cost term constrains the range of the time-dependent control energy $u(t)$. In practice, this means that the brain cannot use an infinite amount of energy to perform the task (i.e., elicit the state transition), a constraint that is consistent with the natural energetic restrictions implicit in the nature of all biological systems but particularly neural systems [Niven and Laughlin (2008); Laughlin et al. (1998)].
Second, the term of the integrated distance term provides a direct constraint on the trajectory. Mathematically, this constraint penalizes trajectories that traverse states that are far away from the target state, based on the intuition that optimal transitions between states should possess reasonable lengths rather than being characterized by a random walk in state space. Together, these two terms in the cost function enable us to define an optimal control model from which we expect to find trajectories (from a given initial state to a specified target state) characterized by a balance between energy cost and trajectory length.

In the context of the optimization problem defined above, we wish to determine the trajectory from an initial state $x_0$ to a target state $x_T$. To do so, it suffices to solve the variational problem with the constraints from Equation 4.1 and the boundary conditions for $x(t)$, i.e. $x(0)$ is the initial state and $x(T)$ is the target state. Mathematically, the variational problem is formulated as

\begin{equation}
\min_u \int_0^T (x_T - x)^T(x_T - x) + \rho u^T u, \tag{4.3}
\end{equation}

\[s.t. \quad \dot{x}(t) = Ax + Bu,\]

\[x(0) = x_0,\]

\[x(T) = x_T,\]

where $T$ is the control horizon, and $\rho \in \mathbb{R}_{>0}$.

To compute an optimal control $u^*$ that induces a transition from the initial state $x_0$ to the target state $x_T$, we define the Hamiltonian as

\[H(p, x, u, t) = x^T x + \rho u^T u + p^T (Ax + Bu).\] \(4.4\)

From the Pontryagin minimum principle [Boltyanskii et al. (1960)], if $u^*$ is an optimal solution to the minimization problem with corresponding state trajectory $x^*$, then there
exists \( p^* \) such that

\[
\frac{\partial H}{\partial x} = -2(x_T - x^*) + A^T p^* = -\dot{p}^*,
\]

(4.5)

\[
\frac{\partial H}{\partial u} = 2\rho u^* + B^T p^* = 0.
\]

(4.6)

which reduces to

\[
\begin{bmatrix}
\dot{x}^* \\
\dot{p}^*
\end{bmatrix} = \begin{bmatrix}
A & -(2\rho)^{-1}BB^T \\
-2I & -A^T
\end{bmatrix} \begin{bmatrix}
x^* \\
p^*
\end{bmatrix} + \begin{bmatrix}
0 \\
2x^T
\end{bmatrix}
\]

(4.7)

Next, we denote

\[
\tilde{A} = \begin{bmatrix}
A & -(2\rho)^{-1}BB^T \\
-2I & -A^T
\end{bmatrix},
\]

(4.8)

\[
\tilde{x} = \begin{bmatrix}
x^* \\
p^*
\end{bmatrix},
\]

(4.9)

\[
\tilde{b} = \begin{bmatrix}
0 \\
2x^T
\end{bmatrix},
\]

(4.10)

then Eqn \( 4.7 \) can be written as

\[
\dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{b},
\]

(4.11)

from which we can derive that

\[
\tilde{x} + \tilde{A}^{-1}\tilde{b} = e^{\tilde{A}t}\tilde{c},
\]

(4.12)

where \( \tilde{c} \) is a constant to be fixed from the boundary conditions. Let \( \tilde{b} = \begin{bmatrix}
\tilde{b}_1 \\
\tilde{b}_2
\end{bmatrix} = \tilde{A}^{-1}\tilde{b}, \)
\[ e^{-At} = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix} \]

and plug in \( t = 0, T \) with the corresponding \( x_0 \) and \( x_T \), we have

\[
\begin{bmatrix} x(0) \\ p(0) \end{bmatrix} + \begin{bmatrix} \bar{c}_1 \\ \bar{c}_2 \end{bmatrix} = \begin{bmatrix} \bar{c}_1 \\ \bar{c}_2 \end{bmatrix}, \quad (4.13)
\]

\[
\begin{bmatrix} x(T) \\ p(T) \end{bmatrix} + \begin{bmatrix} \bar{c}_1 \\ \bar{c}_2 \end{bmatrix} = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}^{-1} \begin{bmatrix} \bar{c}_1 \\ \bar{c}_2 \end{bmatrix}. \quad (4.14)
\]

Note that from Equation \(4.13\), we can solve for \( \bar{c}_1 \), where

\[
\bar{c}_1 = x(0) + \bar{b}_1. \quad (4.15)
\]

Finally, with \( \bar{c}_1 \) on hand from Eqn \(4.14\), we can compute \( p(T) \), where

\[
p(T) = E_{12}^{-1}(\bar{c}_1 - E_{11} \bar{b}_1 - E_{12} \bar{b}_2 - E_{11} x(T)) \quad (4.16)
\]

with which we can finally get \( \bar{c}_2 \), where

\[
\bar{c}_2 = E_{21} x(T) + E_{22} p(T) + E_{21} \bar{b}_1 + E_{22} \bar{b}_2 \quad (4.17)
\]

and further the \( u(t) \) and \( x(t) \) from Equation \(4.12\).

### 4.4.5. Specification of the Initial and Target States

In contrast to the model explored in \cite{Gu et al. 2015a}, the model we describe above is able to examine arbitrary transitions: that is, transitions from any given initial state to any given target state. An important question then is which sets of transitions are biologically relevant for a human brain. As in many complex systems, it is intuitively plausible that not all possible state transitions are required or even healthy \cite{Cornelius et al. 2013}. Here, we choose to focus on trajectories whose initial state is the brain’s baseline condition: states
of high activity in the default mode system, predominantly located in precuneus, posterior cingulate, and superior frontal cortex [Raichle (2015); Raichle and Snyder (2007); Raichle et al. (2001)]. We further constrain ourselves to studying the simplest of target states, in which high activity is present in primary sensorimotor cortex: specifically visual, auditory, and motor cortices, which form fundamental drivers for basic human function. Admittedly, many empirically observed brain states are more complex than those we study here, and in fact many activation patterns cannot be clearly partitioned pre-specified cognitive systems. Yet, these simplified settings are reasonable characterizations of a few fundamental activation patterns, and enable us to consider the energetic issues related to the associated control tasks.

4.4.6. Statistics of Optimal Control Trajectories

After calculating the optimal trajectories between initial and final states, we next sought to address the question of whether these trajectories differed in their energetic requirements between individual’s whose brains were healthy and normally functioning, and individuals who had experience a mild traumatic brain injury and had presented with complaints of mild cognitive impairment. To address this question, we computed the energy cost of a trajectory, integrated over time $T$, as

$$E(K, x_0, x_T) = \int_0^T u_{K,x_0,x_T}^2 dt$$ (4.18)

where $u_{K,x_0,x_T}$ is the associated control input with the given control set $K$, initial state $x_0$ and the target state $x_T$. We treat this energy as a simple statistic that can be compared across trajectories and subject groups, as an indirect measure from which we may infer optimality of cognitive function.

4.4.7. Control Efficiency

The control efficiency is defined for each region to quantify its efficiency in effecting the transition from the default mode state to the three target states. Mathematically, suppose
we have $N$ randomly chosen control sets, each indexed by $\mathcal{K}_1, \ldots, \mathcal{K}_N$, for the target states $x^j_T$, $j = 1, 2, 3$, we calculate the corresponding optimal trajectory with respect to $\mathcal{K}_k$ and denote the energy cost of the trajectory as $E(\mathcal{K}_k, x_0, x^j_T)$. The tiered value of the control set $\mathcal{K}_k$ for target $x^j_T$ is then defined as

$$t_{kj} = \sum_{l=1}^{N} \mathbb{1}(E(\mathcal{K}_l, x_0, x^j_T) > E(\mathcal{K}_k, x_0, x^j_T))$$ (4.19)

where lower energy costs imply higher tiered values. The control efficiency for node $i$ in task $j$ is then

$$\zeta_{ij} = \frac{\sum_{k=1}^{N} \mathbb{1}(i \in \mathcal{K}_k) \cdot t_{ij}}{\sum_{k=1}^{N} \mathbb{1}(i \in \mathcal{K}_k)}.$$ (4.20)

or intuitively, the average of these tiered values.

4.4.8. Network Communicability to the Target State

For a given weighted network $\mathbf{A}$, the network communicability $\mathbf{G}$ quantifies the extent of indirect connectivity among nodes. Here we adopt the generalized definition in [Crofts and Higham (2009)] and define the network communicability as $\mathbf{G} = \exp(\mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2})$, where $\mathbf{D}$ is the diagonal matrix with the diagonal element $D_{ii} = \sum_i A_{ij}$. For a given target state $x_T$, denote the set of active regions as $\mathcal{I}_{x_T}$, the communicability to the target states ($G_{T_i}$) is then defined as the sum of communicability to all of the target regions, i.e.

$G_{T_i} = \sum_{j \in \mathcal{I}_{x_T}} G_{ij}$. Further, the normalized network communicability to the target regions ($\zeta_i$) is then defined as

$$\zeta_i = \frac{G_{T_i}}{\sum_j G_{T_j}}.$$ (4.21)

All results reported in this study are based on the normalized network communicability.

4.4.9. Energetic Impact of Brain Regions on Control Trajectories

To quantify the robustness of controllability of a node when it is removed from the control set consisting of all nodes, we iteratively remove nodes from the network and compute the energetic impact of each region on the optimal trajectory as the resulting increase in the
log value of the energy cost. Intuitively, regions with high energetic impact are those whose removal from the network causes the greatest increase in the energy required for the state transition. Mathematically, denote $K_0$ as the control set of all nodes and $K_i$ as the control set without node $[K]_i$, the energetic impact of node $i$ for target $x_T$ is defined as

$$I_{ij} = \log \frac{E(K_i, x_0, x_T^j)}{E(K_0, x_0, x_T)}$$

which intuitively measures robustness controllability.

4.4.10. Target Control Model

If we only care about the transformation on the target regions, we can modify the control problem 4.3 to the following:

$$\min_u \int_0^T (x_T - x)^T (x_T - x) + \rho u^T u, \quad s.t. \quad \dot{x}(t) = Ax + Bu,$$

$$x(0) = x_0, \quad x^I(T) = x^T_I,$$

where $I$ is the index set of the target regions, $I^c$ the complement, and $x^T(T)$, $x^T_I$ are the constraints of the vector to the corresponding regions of $I$. Similar as the derivation for the optimal control model, we can get Eqn 4.13 and Eqn 4.14 but with constraints $p^{T^c}(T) = 0$ and $x^{T^c}(T) = x^{T^c}_I$. Without loss of generality, we can assume that $x(t) = \begin{bmatrix} x^T(t) \\ x^{T^c}(t) \end{bmatrix}$. Otherwise, we can easily reorder the regions before all the computation. Plug the constraints into Eqn 4.16, we get

$$\begin{bmatrix} p^{T}(T) \\ p^{T^c}(T) \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} - \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} x^T(T) \\ x^{T^c}(T) \end{bmatrix}$$

(4.24)
where \( v = E^{-1}_{i2} (\tilde{c}_1 - E_{11} \tilde{b}_1 - E_{12} \tilde{b}_2) \) and \( M = E^{-1}_{i2} E_{11} \). From Eqn\[4.24\], we can get

\[
\begin{align*}
x^T (T) &= M_{22}^{-1} (v_2 - M_{21} x^T (T) - p(T)^x) \quad (4.25) \\
p^T (T) &= v_1 - M_{11} x^T (T) - M_{12} x^T (T) \quad (4.26)
\end{align*}
\]

With \( p(T), x(T) \) on hand, we can follow Eqn\[4.17\] to compute \( \tilde{c}_2 \) and then \( \tilde{x}(t) \).
CHAPTER 5 : Closing Remarks

Through the course of this Ph.D work, we have developed a number of further ideas about both the improvement on the adaptability for more complex situations and the potential applications in explaining biological phenomenas.

5.1. Improvement on the model

There are two main limitations in the current model. One is the definition of brain states and the other is the assumption on the transform rules. On the model side, we could define the brain states as the abstract representation of brain’s working patterns and interpret them as the electrical and magnetic signals, but the truth remains undetectable [Klausberger et al. (2003)]. Although the statistical correlations do exist between the structural and functional connection [Honey et al. (2009); Greicius et al. (2009)], the fluctuation of functional series are not necessarily to follow the structural patterns, let alone the variance in signal types.

To loose the constraints due to the state definition, we seek to build a functional connectivity model to quantify the controllability in the fluctuations directly. The approach consists of two big steps. The first step is fitting the dynamics without the control input. Considering the complexity and noisy level of the brain signals like fMRI and EEG, we propose two possible plans. We can either adapt those time series model in Finance and Economics to our case [Babu and Reddy (2014); Lindquist et al. (2014)], or fit the whole brain with neural based oscillating networks [Kitzbichler et al. (2009); Cumin and Unsworth (2007)].

After the fitting step, we add the control nodes and compute the controllability metrics. Take the Gramian as an example, for the linear time-series model, the traditional controllability Gramian could be constructed through the temporal dependence [Sakthivel et al. (2012)]; if the dynamics is unknown and probably nonlinear, we are still able to define the empirical Gramian [Himpe and Ohlberger (2015)]. Then we could further talk about the controllability stuffs on the fitted functional networks.
5.2. Potential Application

In the thesis work, we focus on the construction of the theoretical framework of the brain controllability. We compare our model’s prediction with the functional imaging data and obtain significant correlations. However, there still lacks of systematic investigation of the potential application of our model to explaining and predicting a more wide range of cognitive behaviors. For the future research in validation and application, here we list three possible directions.

5.2.1. *Mapping the measures to the functions*

Our model is able to quantify the controllability of a single region, a given set and the whole system with or without the specifics for the performed tasks. Thus we propose that we could build the mapping relation between the cognitive performance and the controllability metrics. That is, we calculate the controllability metrics for each region/system/network, and look for possible link between the metrics and the cognitive performance. Besides, in our model, we can predict which strategy each system prefers in general and for specific tasks. Thus we could further compare the theoretical trajectories from the model and the true trajectories from the data to validate our hypothesis. Moreover, through constructing and analyzing the control model on the structural and functional networks of the same subject, we could broaden our knowledge on the structure-induced constraints on brain functions.

5.2.2. *Learning Procedure energy cost*

In Bassett’s series paper of learning procedure [Bassett et al. (2011b); Cole et al. (2014); Bassett et al. (2015)], they show that the brains that are more easier to tune among multiple tasks tend to learn faster. Next step we plan to further explore the effective factors for learning and understand the state transformation from the aspect of energy. For example, we want to solve the dilemma of the energy cost in the brain activity. On the one hand, the low-barrier brain is more likely to transfer faster thus to behave in high efficiency due to the
relatively lower cost of energy for the state transition; on the other hand, the costly brain is also arguable to be more efficient because the high cost may come from its ability to use more energy, and result in the complexity in structure or the versatility in functions. The exploration on these problems would strengthen our understanding for the brain learning.

5.2.3. Neurodevelopment

We have the knowledge that both the brain structure and function change significantly throughout the development and display a strong temporal dependence across the life [Jack et al. (2008); Bullmore and Sporns (2009); Fair et al. (2008)]. Yet how the structural and functional changes fit with each other lacks clear explanation. We attempt to build a bridge between these two parts. Via applying our controllability framework into the series of structural networks, we could quantify the brain-level, region-level and system-level controllability change, which will help us understand how the brain varies its efficiency in the energy cost and reachability to multiple states. Further, combining the multi-scale controllability measures with the mapping from controllability measures to the cognitive functions, we could investigate whether the maturation and degeneration in functions are related to the corresponding structural changes, especially people’s controllability of certain brain systems.
BIBLIOGRAPHY


M. Bedny, A. Pascual-Leone, S. Dravida, and R. Saxe. A sensitive period for language in
the visual cortex: distinct patterns of plasticity in congenitally versus late blind adults. 

R. F. Betzel, S. Gu, J. D. Medaglia, F. Pasqualetti, and D. S. Bassett. Optimally controlling


B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde. Functional connectivity in
the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in

V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of commu-
nities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10:
P1000, 2008.

S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks:


V. G. Boltyanskii, R. V. Gamkrelidze, and L. S. Pontryagin. The theory of optimal pro-

M. M. Botvinick and J. D. Cohen. The computational and neural basis of cognitive control:


U. Braun, S. F. Muldoon, and D. S. Bassett. On human brain networks in health and

keithbriggs.info/network.html

R. L. Buckner, J. Sepulcre, T. Talukdar, F. M. Krienen, H. Liu, T. Hedden, J. R. Andrews-
Hanna, R. A. Sperling, and K. A. Johnson. Cortical hubs revealed by intrinsic functional


C. Giusti, R. Ghrist, and D. S. Bassett. Two’s company, three (or more) is a simplex: Algebraic-topological tools for understanding higher-order structure in neural data. *Trends in Cognitive Sciences*, Under Consideration, 2016.


M. Jalili, A. A. Rad, and M. Hasler. Enhancing synchronizability of weighted dynamical


D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet.


A. Lukoshe, T. White, M. N. Schmidt, A. van der Lught, and A. C. Hokken-Koelega. Diver-


C. Thomas, Q. Y. Frank, M. O. Irfanoglu, P. Modi, K. S. Saleem, D. A. Leopold, and C. Pierpaoli. Anatomical accuracy of brain connections derived from diffusion mri trac-


M. P. van den Heuvel, O. Sporns, G. Collin, T. Scheewe, R. C. Mandl, W. Cahn, J. Goñi,


