UNIVERSITY OF CALIFORNIA,
IRVINE

Untangling the Connectional Neuroanatomy of the Language Dominant Cerebral Hemisphere using Diffusion−Weighted Magnetic Resonance Imaging

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Neuroscience

by

Vatche George Baboyan

Dissertation Committee:
Professor Gregory Hickok, Chair
Professor Frithjof Kruggel
Professor Emily Grossman

2020

DEDICATION

A successful graduate experience is driven by emotional support just as much as it is by intellectual support. I therefore dedicate this dissertation to my parents, George and Mariejeanne, and to my sister, Pauline for their unconditional support that serves as the backbone for all of my achievements. Reaching this milestone means the world to my Dad and I hope this vindicates his lifelong commitment to help me secure the best education I could possibly achieve as a first-generation American. Lastly, I wish to dedicate this work to my dear friend Adam Mezher. I hope Freesurfer is treating you well up there pal. This one's for you!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to thank the National Science Foundation (NSF), for generously funding my graduate research through NSF's graduate research fellowship program under Grant No. (DGE-1321846). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Next, this thesis was made possible by the incredible mentors that have guided me intellectually over the years. I would like to thank my PhD advisor, Professor Gregory Hickok, whose support, patience, and theoretical perspective allowed me to flourish despite my unreasonable persistence with studying white matter. Your masterful communication of complex ideas is something I aspire to emulate in my life after graduate school. You've made me appreciate the tremendous level of overlap that exists between seemingly disparate brain systems and their shared organizational principles, and that perspective will stay with me for the rest of my career. Thank you.

I also wish to thank Dr. Nitin Tandon and the entire family of exceptionally dedicated researchers within the Neuroimaging Intracranial Electrophysiology Laboratory at the UT Houston Medical School. My experience working in Dr. Tandon's research lab would form the foundation for my interests in studying human language and in chasing a career in the Neurosciences. The relationships that were made during my time in Houston made the experience a truly memorable one; from Vipulkumar Patel and the MRI team, to Delrick Sanker and the EEG team, and of course to Dr. Cihan Kadipasaoglu and the research team, these wonderful people made living away from home so much easier for me.

A very special thanks goes to Professor Frithjof Kruggel from the Biomedical Engineering department, whose guidance on my dissertation committee allowed me constrain my research projects to within reasonable limits. Thank you for our insightful conversations and for your unparalleled commitment to teaching.

I would also like to thank the Mark Mary Stevens' Neuroimaging and Informatics Institute at the Univerity of Southern California. Working with such a world-class team of scientists before starting graduate school really helped me build the technical skills that would be so instrumental later on. It was here that I learned the value of leveraging high-performance computing resources to scale projects up to otherwise unfeasible levels. This skill eventually led me to working with outstanding HPC resources made available for researchers here at UC Irvine. That being said, i'm grateful for the contributions from Dr. Harry Mangalam for his continued commitment to maintaining these computing resources - our entire campus is indebted to your team.

This brings me to the most important contributors to my dissertation: the data. My

research projects featured a variety of incredible diffusion weighted imaging datasets provided by Dr. Tandon and the McGovern Medical School, Dr. Julius Fridrikson and the University of South Carolina, and the Human Connectome Project. Without the inter-institutional efforts forged by a truly collaborative network of scientists, research would be a futile endeavor.

CURRICULUM VITA

Vatche George Baboyan

EDUCATION

Doctor of Philosophy in Cognitive Neuroscience **2020**
University of California, Irvine *Irvine, California*

Master of Science in Cognitive Neuroscience **2018**
University of California, Irvine *Irvine, California*

Bachelor of Arts in Psychology, Global Studies **2011**
University of California, Santa Barbara *Santa Barbara, California*

RESEARCH EXPERIENCE

Graduate Research Fellow **2016–2019**
Auditory & Language Neuroscience Laboratory *Irvine, California*
University of California, Irvine

Project Assistant **2014–2015**
Mark & Mary Stevens Neuroimaging and Informatics Institute *Marina Del Rey, California*
University of Southern California, Keck School of Medicine

Research Assistant (Programmer) **2012–2014**
Neuroimaging & Intracranial Electrophysiology Laboratory *Houston, Texas*
University of Texas Health Science Center, Houston

TEACHING EXPERIENCE

Teaching Assistant **2015-2016**
University of California, Irvine *Irvine, California*

SELECTED HONORS AND AWARDS

National Science Foundation Graduate Research Fellowship **2016–2019**
National Science Foundation

PUBLICATIONS

***Denotes Co-First Authorship**

Baboyan, VG, DiSano MA, Breier JI, Hickok G, Tandon N. (2020). Brain Pathways Critical For Language Laterality. *In Review: Brain*

Kadipasaoglu, C. M., Conner, C. R., **Baboyan, VG,** Rollo, M., Pieters, T. A., Tandon, N. (2017). Network dynamics of human face perception. *PloS One.*

Hafzalla, GW., Prasad, G., **Baboyan, VG,** Faskowitz, J., Jahanshad, N., McMahon, K. L., ... Thompson, P. M. (2016, March). The heritability of the functional connectome is robust to common nonlinear registration methods. In Medical Imaging 2016: Image Processing (Vol. 9784, p. 97841R). *International Society for Optics and Photonics.*

Guadalupe T, Mathias SR, van Erp TGM, Whelan CD, Zwiers MP, Abe Y, Abramovic L, Agartz I, Andreassen OA, Arias-Vasquez A, Aribisala B, Armstrong NJ, Arlt V, Artiges E, Ayesa-Arriola R, **Baboyan, VG,**...,Thompson PM, Grahn DC, Mazoyer B, Fisher SE, Francks C. (2016) Human Brain Asymmetries in 15, 847 People Worldwide Reveal Effects of Age and Sex. *Brain Imaging & Behavior.*

Kadipasaoglu, CM, Whaley ML, Conner CR, **Baboyan, VG,** Tandon N. (2016) Category-Selectivity in Human Visual Cortex Follows Cortical Topology: A Grouped icEEG Study. *PloS One.*

Daianu M, Mendez MF, **Baboyan, VG,** Jin Y, Melrose RJ, Jimenez EE, Thompson PM. (2015) An Advanced White Matter Tract Analysis in Frontotemporal Dementia and Early-Onset Alzheimer's Disease. *Brain Imaging & Behavior.*

Kadipasaoglu, CM, Forseth K, Whaley ML, Conner CR, Rollo MJ,**Baboyan, VG,** Tandon N. (2015) Development of grouped icEEG for the study of cognitive processing. *Frontiers in Psychology.*

Kadipasaoglu CM, **Baboyan, VG*,** Conner CR, Chen G, Saad ZS, Tandon N. (2014) Surface-Based Mixed Effects Multilevel Analysis of Grouped Human Electro-

Corticography. *NeuroImage*.

PRESENTATIONS

Vatche G. Baboyan, Gregory Hickok. Revisiting the connectional neuroanatomy of pars opercularis and the ventral precentral fiber intersection area using data from the Human Connectome Project. *Society for the Neurobiology of Language*. August 19 – 21. 2019, Helsinki, Finland.

Vatche G. Baboyan, Gregory Hickok, Nitin Tandon. Quantification and Parcellation of Posterior Inferior Frontal Structural Connections to Auditory and Supplementary Motor Area Targets: A diffusion imaging study in Neurosurgical Patients. *Society for the Neurobiology of Language*. August 15 – 18, 2018, Quebec City, Canada.

Vatche G. Baboyan, Adam Mezher, Emily L. Dennis, Madelaine Daianu, Yan Jin, Talin Babikian, Christopher C Giza, Robert Asarnow, Paul M Thompson. Disruptions to White Matter Microstructure of the Default Mode Network in Pediatric Traumatic Brain Injury. *Society for Neuroscience*. Oct 17-21, 2015, Chicago, IL,USA.

Tulio Guadalupe, **Vatche G. Baboyan**, Fabrice Crivello, Barbara Franke, Hans Jörgen Grabe, Derrek Hibar, Neda Jahanshad, Sophie Maingault, Nicholas Martin, Katie McMahon, Sarah Medland, Miguel Renteria, Sanjay Sisodiya, Nathalie Tzourio-Mazoyer, Christopher Whelan, Katharina Wittfeld, Margaret Wright, Greig de Zubicaray, Paul Thompson, Bernard Mazoyer, Simon Fisher, Clyde Francks. Meta-analysis of sex and handedness effects on human subcortical asymmetries: ENIGMA-Lateral- ization. *Organization for Human Brain Mapping annual meeting*. June 14-18 2015, Honolulu, Hawaii.

Emily L. Dennis, **Vatche G. Baboyan**, Xue Hua, Clifford R. Jack, Jr, Michael W. Weiner, Paul M Thompson for the DoD-ADNI Initiative (2015). Volumetric Brain Abnormalities in Vietnam Veterans with PTSD. *Organization for Human Brain Mapping annual meeting*. June 14-18 2015, Honolulu, Hawaii.

Emily L. Dennis, **Vatche G. Baboyan**, Yan Jin, Clifford R. Jack, Jr, Michael W. Weiner, Paul M Thompson for the DoD-ADNI Initiative (2015). Tract-based Analysis of White Matter Disruption in PTSD. *Organization for Human Brain Mapping annual meeting*. June 14-18 2015, Honolulu, Hawaii.

Vatche G. Baboyan, Emily L. Dennis, Yan Jin, Talin Babikian, Christopher C Giza, Robert Asarnow, Paul M Thompson. An Analysis of Temporal Lobe Association Fibers in Post-Acute Traumatic Brain Injury: A HARDI study. *Cognitive Neuroscience Society*. March 28-31, 2015, San Francisco, CA.

Jan R. Wessel, **Vatche G. Baboyan**, Nitin Tandon, Adam R. Aron. Motor slowing following unexpected events: individual roles and functional connectivity of the pre-supplementary motor area and the right inferior frontal cortex. *Society for Neuroscience*. November 15-19, 2014, Washington, DC.

Cihan M. Kadipasaoglu,**Vatche G. Baboyan**, Nitin Tandon. Distributed neural systems for category-specific visual processing. *Society for Neuroscience*. November 15-19, 2014, Washington, DC.

Cihan M. Kadipasaoglu, Christopher R. Conner, **Vatche G. Baboyan**, Nitin Tandon. Non-hierarchical network dynamics of face perception. *Society for Neuroscience*. November 15-19, 2014, Washington, DC.

Nitin Tandon, Cihan M. Kadipasaoglu, **Vatche G. Baboyan**, Christopher R. Conner. Network Dynamics of Functional Connectivity in Category-Specific Visual Naming. *American Society for Stereotactic and Functional Neurosurgery*. May 31 – June 3, 2014, Washington, DC.

Vatche G. Baboyan, Michael A. DiSano, Christopher R. Conner, Cihan M. Kadipasaoglu, Joshua I. Breier, Nitin Tandon. Language Laterality estimated using Tract Based Spatial Statistics: A Correlation with the Wada Test. *American Association of Neurological Surgeons*. April 5-9, 2014, San Fransisco, CA.

Cihan M. Kadipasaoglu, Christopher R. Conner,**Vatche G. Baboyan**, Nitin Tandon. Category-specific temporal and spatial dissociations as revealed by grouped human electrocorticography. *Society for the Neurobiology of Language*. November 6 – 8. 2013, San Diego, CA.

Cihan M. Kadipasaoglu, Thomas A Pieters, **Vatche G. Baboyan**, Christopher R. Conner, Nitin Tandon. Study of the human retrosplenial cortex during auditory and

visual naming through grouped electrocorticography and cortical stimulation mapping. *Society for the Neurobiology of Language*. November 6 – 8, 2013, San Diego, CA.

Thomas A. Pieters, Cihan M. Kadipasaoglu, **Vatche G. Baboyan**, Nitin Tandon. Spatial probability of essential language sties: cortical stimulation density map in a population. *Society for the Neurobiology of Language*. November 6 – 8, 2013, San Diego, CA.

Cihan M. Kadipasaoglu, Christopher R. Conner, **Vatche G. Baboyan**, Nitin Tandon. Electrocorticography of the face and place network connectivity. *Organization for Human Brain Mapping*. June 16 – 20, 2013, Seattle, WA.

PROFESSIONAL MEMBERSHIPS

Society for the Neurobiology of Language (SNL)
Cognitive Neuroscience Society (CNS)
Society for Neuroscience (SFN)

ABSTRACT OF THE DISSERTATION

Untangling the Connectional Neuroanatomy of the Language Dominant Cerebral Hemisphere using Diffusion−Weighted Magnetic Resonance Imaging

By

Vatche George Baboyan

Doctor of Philosophy in Cognitive Neuroscience

University of California, Irvine, 2020

Professor Gregory Hickok, Chair

The present thesis is dedicated to studying the human language connectome by combining diffusion-weighted magnetic resonance imaging (dMRI) and high-dimensional predictive algorithms to isolate its relevant connections amongst the broader white matter feature space. According to the classical "Broca-Wernicke" and contemporary "Dual-Stream" neurobiological models of language, the fluent production and repetition of speech relies on a distinct and lateralized neuroanatomical circuit. Where these models diverge is in their neuroanatomical explanations, such that the former attributes these functions to the posterior inferior frontal gyrus (i.e., Broca's area) and its disconnection from the arcuate fasciculus while the latter emphasizes a cortico-centric "dorsal stream" network where repetition processing occurs by way of a sensorimotor node in sylvian parieto-temporal cortex (called "area Spt"). Experiment 1 of this thesis tested the assumption that the speech production system is scaffolded by a lateralized set of connections by performing whole-brain inter-hemispheric white matter comparisons between left- and right-hemisphere dominant subjects as determined by a clinically-indicated Intracarotid Amytal (Wada) Test. Indeed, this semi-automated analysis reconstructed a lateralized network for speech production that was underscored by its convergence of pathways to Broca's area and whose

white matter properties successfully enabled Wada-concordant classifications using a ridge-logistic regression model. Experiment 2 used dMRI collected by the Human Connectome Project to perform multi-shell fiber tractography of Broca's area in order to segment the underlying sub-circuits connecting to this classical region. Experiment 3 linked connectivity directly with behavior using a connectome-based lesion-symptom mapping approach in a large dMRI sample acquired from stroke patients. Over two-thousand distinct connections were mapped and a LASSO-regularized latent projection-based regression algorithm was implemented to automatically isolate the subset of connections predictive of speech repetition performance. We found that the embedded feature selection algorithm identified a local set of sylvian parieto-temporal connections that could make accurate out-of-sample predictions of repetition performance - a result which fuses the claims made by the "Broca-Wernicke" and "Dual-Stream" models regarding the anatomic delineation of the repetition circuit. Together, these studies demonstrate the utility in using dMRI and machine-learning algorithms to understand the anatomy of eloquent functional networks while accommodating the collinearities that are characteristic of brain data.

Chapter 1

Background

1.1 Introduction

One of the critical theoretical advances in modern neurosciences has been the shift away from localizationist views of brain function, stressing the importance of communication between brain networks rather than considering brain areas as modular units (Thomas Yeo et al. (2011)). The past two decades of imaging and lesion studies in humans have provided us with a wealth of new information on the human language "network". Today, this exclusively human ability for speech and language is widely conceptualized as being sub-served by the parallel cortical operations of dorsal and ventral processing streams anchored by downstream activity from primary auditory cortex. This model borrows heavily from theoretical principles on neural foundations of the visual system, where dorsal and ventral streams map onto "how" and "what" pathways. In the language domain, it is believed that the dorsal stream is the "how" pathway for driving the speech articulators (Chang et al. (2013)) through sensorimotor feedback while the ventral "what" language pathway supports the acoustic

signal processing needed for the comprehension of speech sounds (Hickok and Poeppel (2007); Hickok (2012)).

A fundamental feature of the dual-stream model of language, and where it diverges from its theoretical contemporaries, is in its assertion that the dorsal sensorimotor stream, but *not* the ventral semantic stream, is heavily asymmetric towards the language-dominant (i.e., mostly left) hemisphere (Hickok et al. (2008)). This claim is bolstered by the ample evidence from the lesion study literature on non-fluent aphasia (i.e., Broca's aphasia) resulting from unilateral brain damage and, conversely, by the lack of evidence showing speech perception deficits from similar unilateral lesions (Hickok et al. (2008)). Speech production therefore appears to have a localizable hemispheric substrate in the language-dominant hemisphere.

1.1.1 The White-Matter Basis for Dual Stream Pathways... or Lack Thereof

A potential criticism of the dual-stream model proposed by Hickok and Poeppel (2007) is that it is under-specified with respect to the contributing white matter pathways enabling each functional stream. Such a deliberate omission may have been primarily driven by the lack of available white matter imaging methods in the early 2000s. Since the advent of diffusion magnetic resonance imaging (dMRI), the precise white matter structures contributing to dual-stream functioning continues to be configured (Dick et al. (2014)). dMRI is an imaging technique that measures the molecular displacement of water across brain tissue along different spatial directions. This powerful technique takes advantage of the biological property of restricted water diffusion along white matter axons of the brain, thereby allowing for mathematical estimations of fiber trajectories based on the eigen-system of the water displacement

matrix (Basser et al. (2000)). Although much progress has been made in its application to language, the literature has become severely limited by a variety of technical limitations. Among them, a pervasive spatial bias accompanies the majority of work conducted in this area. Specifically, statistical analyses are often restricted to predefined fiber bundles to simplify subsequent statistical tests. The arcuate fasciculus, for example, is a bundle that has been implicated in language since Carl Wernicke in the late 1800s as a pathway interconnecting Broca's and Wernicke's areas. Early diffusion imaging studies of the early-to-mid 2000s were then dedicated to relating this pathway to language while ignoring all others (Glasser and Rilling (2008)). Today, with advances in whole-brain statistical analysis techniques of diffusion data paired with high-dimensional machine-learning algorithms that may accomodate collinearities characteristic of brain imaging data, such spatial biases are no longer necessary, and data-driven approaches are now necessary to validate prior findings as well as reveal potentially novel ones. Second, virtual dissections (i.e., tractography) based on diffusion imaging data is both a time-consuming and heavily operator-dependent process, making tractography studies not only limited in sample-sizes, but also in their consistency across studies since considerable disagreement exists in the field regarding the exact terminations of the pathways in question (Bernal and Altman (2010); Tremblay and Dick (2016); Mesulam et al. (2015)).

1.2 Research Aims

This thesis is dedicated to revisiting the connectional neuroanatomy of the language network using modern brain-imaging and machine learning tools. Considering the brain as a massive multivariate feature space with highly correlated spatial components, it becomes essential to use modeling strategies that may accommodate this

structure while, at the same time, minimizing *a priori* theoretical assumptions that may introduce bias to the analyses. The overall objective of this thesis is three-fold: Firstly, if the human brain shows a hemispheric bias for speech, then an inter-hemispheric comparison of white matter properties should reveal the precise neuroanatomical white matter substrate for this property. Are cortico-centric theoretical models justified in their exclusion of specific white matter connections (Hickok and Poeppel (2007); Rauschecker and Scott (2009))? If, for example, the dorsal language stream is indeed unilaterally represented in the language-dominant hemisphere, then it should follow that this functional asymmetry should be accompanied by some structural asymmetry in the brain. These assumptions are tested in experiment #1, where we compare inter-hemispheric asymmetry patterns in a large cohort of patients within whom diffusion imaging data have been collected and shared by the University of Texas Health Science Center. Second, using data from the Human Connectome Project, we evaluate the connectivity patterns of one of the most eloquent regions of the human brain: pars opercularis of the inferior frontal gyrus and the ventral pre-central gyrus. Given the importance of these regions as elements of "Broca's area", a precise understanding of their structural connectivity patterns is surprisingly lacking. This might be explained primarily by the complex organization of white matter fibers projecting to this region. Therefore, assessing its connectivity to the rest of the cerebral cortex in a large group of subjects using multivariate tools may allow us to fully visualize its distributed connectivity patterns and thus provide insight into its functional purpose as a critical node in the dual stream architecture. Third, we integrate findings from experiments #1 and #2 in order to evaluate whether classical models are justified in attributing the arcuate fasciculus as being the so-called "speech repetition pathway". Recent lesion and imaging studies have suggested that speech repetition is subserved by cortical set of brain regions. This relationship will be evaluated in a large clinical cohort of stroke-induced aphasics from whom diffu-

sion imaging and extensive neuropsychological data have been acquired through a collaboration with the Aphasia lab at the University of South Carolina. The proposed experiments are aimed at determining the value in using diffusion weighted imaging and tractography in studying the human language network in a data-driven manner to evaluate whether reliable and accurate predictions could be made using the information provided beneath the brain surface.

Chapter 2

Brain Pathways Critical For Language Laterality: An Inter-Hemispheric White Matter Comparison in Wada-Tested Subjects

2.1 Introduction

Over the past half-century, many claims have been made regarding the structural asymmetry underlying one of the most fundamental properties of the human brain - it's hemispheric preference for language. The most widely reported results have been observed at the cortical level; particularly, in the planum temporale (area PT) region, located posterior to Heschl's gyrus. This asymmetry was first reported in a 1968 pa-

per by Geschwind and Levitsky (Geschwind and Levitsky (1968)), when it was noted by visual examination of post-mortem brain specimens that the surface area of the PT was larger in the left hemisphere. This report, and the dozens of others that have since followed in both imaging and cadaveric studies, have been driven by a strong theoretical bias. Given what was known about perisylvian cortex and it's role in language from classical aphasiology, researchers repeatedly isolated such areas in their research while ignoring the rest of the brain. This same approach has permeated the subcortical white matter domain, such that the most widely reported finding is that the arcuate fasciculus, a pathway classically described as interconnecting Broca's and Wernicke's area, is the critical subcortical substrate of language laterality (Glasser and Rilling (2008); Takaya et al. (2015)). Today, with the development of whole-brain imaging techniques, this bias can easily be resolved by using data-driven analyses across the entire brain and precluding the need for restrictions to specific brain areas. This approach is critical in determining the legitimacy of past claims through careful validation of findings in a data-driven manner. In this study, we sought to apply a data-driven framework on diffusion imaging data in order to determine whether and where in the human brain a relationship between white matter asymmetries and functional language asymmetries may exist. Importantly, we address this topic in a cohort of neurosurgical subjects, whereby the dominant hemisphere has been revealed by a powerful, and clinically indicated procedure, known as the Intra-carotid sodium Amytal Test (IAT - also called the "Wada Test"). This unique sample of patients provides access to an incredibly rare subset of individuals showing a reversed (i.e., Right-hemisphere language dominance) asymmetry pattern. The Wada test is scored as a categorical outcome for each individual patient, Left-Hemisphere Language Dominance (LLD) or Right-Hemisphere Language Dominance (RLD), which captures the ability of that hemisphere to carry out language whilst its contralateral side is put to sleep using an anesthetics injection to its carotid blood supply

(Wada and Rasmussen (1960)). Prior research has suggested that right-hemisphere language dominance presents a homotopic or "mirror" configuration of language networks (Chang et al. (2011)). As such, a comparison of typical LLD and atypical RLD groups may isolate the underlying the neuroanatomical substrate involved and yield a noninvasive imaging biomarker to predict the language-dominant hemisphere in the future. If indeed a region of asymmetry is identified, we can relate this to its underlying fiber pathways using diffusion fiber tractography. This approach will effectively localize the fiber tracts underpinning the whole-brain analyses - in a manner that is agnostic to *a priori* assumptions from neurobiological models of language. Based on the dual stream theoretical framework proposed by Hickok and Poeppel (2007), we hypothesize that this study will reveal that dorsal stream white matter is asymmetrically represented in the language-dominant hemisphere (Hickok and Poeppel (2007).

2.2 Methods

2.2.1 Participants

Ninety-two consecutive neurosurgical patients (44 male, mean age 35.52 years) underwent intra-carotid sodium amytal (IAT) testing to lateralize language (Table 1) prior to neurosurgical intervention for medically intractable epilepsy (n = 85) or a brain tumor (n = 7). Handedness was prospectively assessed using the Edinburgh Handedness Inventory (EHI) (Oldfield (1971)), scored on a scale ranging from negative ten (completely left-handed on all tasks) to positive ten (completely right-handed on all tasks). Each patient provided informed written consent, in accordance with the University of Texas Health Science Center's committee for the protection of human

subjects, for collection and analysis of their data.

2.2.2 Wada Testing

Wada testing was carried out using a standardized protocol for lateralizing language and memory function, detailed previously (Breier et al. (1999)). In brief, hemispheric dominance for language was determined by testing: 1) comprehension of one- and two-step commands; 2) naming of objects or parts of objects presented as line drawings; 3) reading of sentences; and 4) repetition of simple phrases. Performance on each of these tests was scored as either normal, mildly, moderately, or severely deficient. A hemisphere was deemed to support language when it's injection with sodium amytal resulted in the disruption of performance in at least two of the aforementioned tests, with one or more test being rated as having at least moderate disruption or at least three of the four tests being characterized with at least mild disruption. Unilateral language representation was inferred when only one hemisphere met these criteria. To minimize confounding or subjectivity in language lateralization, patients classified as being bilaterally language-dominant were specifically excluded, and subsequent analyses were restricted to those subjects presenting with unambiguous unilateral dominance. Seven patients met the exclusion criteria for bilateral dominance, seventy-two patients were Left-hemisphere Language Dominant (LLD) and thirteen were Right-Hemisphere Language Dominant (RLD).

Table 2.1: Patient Demographics

Patient Demographics

Characteristics	Left-Language Dominant	Right-Language Dominant	Bilaterally-Language Dominant	Totals
Number of subjects (n)	72	13	7	92
Age (yrs)	35.97 (12.43)	37.23 (9.88)	29.29 (8.85)	35.52 (11.91)
Gender	M = 33; F = 39	M = 8; F = 5	M = 3; F = 4	M = 44; F = 48
Handedness (EHI)	R = 46; L= 14; Mix = 12	R = 3; L = 8; Mix = 2	R = 0; L = 6; Mix =1	R = 49; L = 28; Mix = 15
Pathology	Epilepsy = 65; Tumor = 7	Epilepsy = 13; Tumor = 0	Epilepsy = 7;Tumor = 0	Epilepsy = 85; Tumor = 7
Hemisphere of Pathology	L = 36; R = 36	L = 9; R = 4	L = 3; R = 4	L = 48; R = 44

Numbers are means (SD). EHI = Edinburgh Handedness Inventory; M = male; F = female; R = Right; L = Left; Mix = Mixed Mixed Handedness was defined as an EHI score between -5 and 5 (inclusive).

2.2.3 Diffusion Weighted Imaging Acquisition and Processing

Anatomical MRI scans were collected using a 3T whole-body MR Scanner (Philips Medical Systems, Bothell WA) equipped with a 16-channel SENSE head coil. Anatomical images were collected using magnetization-prepared 180-degree radio-frequency pulses and rapid gradient-echo (MP-RAGE) sequence, optimized for gray-white matter contrast, with 1 mm thick sagittal slices and an in-plane resolution of 0.938 x 0.938 mm. DWIs were collected using a Philips 32-direction diffusion encoding scheme (high angular resolution) with the gradient overplus option. One b0 (non-diffusion weighted, b-value = 800 s/mm2) image volume was acquired before the acquisition of one repetition of diffusion-weighted scans. Seventy axial slices were acquired with a 224224 FOV (1.75 1.75 mm pixels) and 2 mm slice thickness.

DWIs were first corrected for eddy current distortion and head motion using a 12-parameter affine registration to the b0 reference volume using the "eddy correct" tool provided by FSL. The B-matrix was rotated using the "fdt rotate bvecs" tool provided within the same distribution. Following brain extraction and quality control of DWIs, FMRIB's diffusion toolbox (http://www.fmrib.ox.ac.uk/fsl/fdt) was used to generate voxel-wise images of fractional anisotropy (FA) by fitting a diffusion tensor model to the data at each brain voxel. The resulting eigen-system images were converted

$$FA = \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_1 - \lambda_3)^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} \tag{2.1}$$

Figure 2.1: Fractional Anisotropy (FA) represented as the standard deviation of dominant eigenvalues. FA is widely used as a proxy for white matter microstructural integrity.

into tensor volumes by DTI-TK (http://www.nitrc.org/projects/dtitk) which was then used for tensor-based spatial normalization of the volumes to an iteratively optimized study-specific template (Zhang et al. (2006)). This algorithm computes a deformable registration to the tensor images, which has been shown to improve DWI co-registration. Each participant's tensor image was then normalized to the study-specic template using one warp that combined afne and diffeomorphic alignments with nal isotropic 1 mm3 resolution, followed by conversion back to an FA map to begin integration with the TBSS workflow (next section).

2.2.4 Whole-Brain Grouped-Statistical Analyses of White Matter Asymmetry

A fundamental challenge when comparing diffusion weighted data across subjects is that significant anatomical variability exists in white matter structures - especially pronounced in the brain's periphery - thus making it difficult to align data across subjects. One solution to this problem is to roughly align (i.e., using nonlinear image registration) all brain images and subsequently generate a so-called white matter "skeleton" representing areas common to all subjects. For each subject, the diffusion data most adjacent to the skeleton are then projected onto the corresponding voxels along the skeleton. This framework, called Tract-Based Spatial Statistics (TBSS, Smith et al. (2007)), allows one to test for differences in diffusion data at each voxel

along the skeleton across large groups of subjects. Here, this strategy will be used to test for differences in asymmetry between the LLD and RLD group along a symmetric version of the skeleton.

Once the data were projected onto the symmetric skeleton, inter-hemispheric (i.e., Left-Right) differences in fractional anisotropy were computed between each voxel along the skeleton in the left hemisphere and its contralateral homolog in the right hemisphere. Fractional anisotropy is a metric that is commonly used in diffusion studies to estimate microstructural integrity of white matter fibers. To test for potential differences, we implemented an non-parametric, permutation-based (k = 5,000) general linear model (GLM) approach. An unpaired two-sample design matrix was generated with group as a between-subjects factor while also adding handedness, age, and gender as nuisance covariates given their potential effects on white matter asymmetry. The asymmetry maps were thus modeled using the design matrix below (Eq: 2.2), with treating the response variable, Y, as each voxel in the diffusion maps for each subject, and with the design matrix, X, containing an intercept column, indicator variables for the Wada groupings (LLD vs RLD) in the second and third columns, respectively, and the nuisance covariates in the remaining columns:

$$\mathbf{Y} = \begin{bmatrix} 1 & LLD_1 & RLD_1 & EHI_1 & DxHemi_1 & \\ 1 & LLD_2 & RLD_2 & EHI_2 & DxHemi_2 & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & LLD_{84} & RLD_{84} & EHI_{84} & DxHemi_{84} & \\ 1 & LLD_{85} & RLD_{85} & EHI_{85} & DxHemi_{85} & \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \epsilon \tag{2.2}$$

The non-parametric permutation test shuffles the group memberships 5000 times to generate voxelwise p-values, testing where the asymmetry of the LLD group is

significantly greater than the asymmetry in the RLD group (i.e., where $\beta_1 \quad \beta_2$). P-values were then corrected for multiple comparisons with a specified significance level of $P < 0.01$.

2.2.5 Probabilistic Tractography

In order to relate the differences in white matter asymmetry to their underlying pathways, a *post-hoc* fiber tracking procedure was implemented by inverse transforming the cluster coordinates of the implicated region in each of the subjects in the study. The cluster coordinates were then fed into a probabilistic tractography algorithm (Behrens et al. (2007)), with crossing-fiber estimation as implemented by FSL's *bedpostX* function, to establish the connectivity pattern of pathways intersecting the cluster. The resulting volumetric connectivity maps were projected onto a standardized surface mesh containing a node-to-node correspondence across subjects and thereby enabling direct comparisons in connectivity at each node across subjects. We then statistically compared grouped-differences in connectivity across brain areas of the Human Connectome Project's Multimodal Parcellation atlas (Glasser et al. (2016)), which provides 180 discrete cortical labels per hemisphere.

2.2.6 Hierarchical Clustering & Projection Zone Analysis

To statistically evaluate grouped-differences in connectivity patterns, the tractographic cortical projections were subjected to a separate analysis. After projecting each individual tractogram onto its respective surface-mesh model in native space, the data was kept un-smoothed (Coalson et al. (2018)) and was transformed onto the "fsaverage" template surface. The Human Connectome Project's Multi-Modal Parcellation atlas (Glasser et al. (2016)) was then used to compute averages within

each of the 180 hemispheric parcels provided by the atlas. Of these, only parcels with consistent connections were analyzed (i.e., parcels wherein at least 80% of LLD subjects had nonzero values) and parcels traversing the length of the central sulcus were excluded. Lastly, data from nearby parcels in the atlas were grouped together using a hierarchical clustering algorithm (with "complete linkage" agglomeration). The cluster solution was selected by minimizing the total number of clusters while ensuring spatial coherence of parcels forming a cluster. Statistical comparisons of the connectivity data from each of the resulting clusters - forming cortical "projection zones" of neighboring parcels with shared connectivity patterns - were then carried out with a linear mixed effects model (LME) using the lme4 package in R (Bates et al. (2015)). The LME analysis for each projection zone tested for hemisphere by Wada outcome interactions while controlling for both handedness and diseased hemisphere. All models were fit using maximum likelihood estimation and included random intercepts for both subject ID and for the parcels comprising each projection zone in order to account for inter-individual variability and statistical inter-dependence of these repeated bilateral measurements (e.g., one subject contributing both left and right hemisphere connectivity measurements per parcel). In cases when a projection zone contained a single parcel, the random effect for parcel was excluded. Significant main effects for hemisphere ascertained that the parcels of a given projection zone presented a hemispheric bias across both groups, while significant interaction effects showed a hemispheric bias modulated by Wada outcome. P-values for the parameter estimates were obtained using t-tests with Satterthwaite's method for approximating degrees of freedom, adjusted for multiple comparisons using Bonferroni correction based on the total number of projection zones tested. The connectivity data are displayed as "connectivity fingerprints" (Passingham et al. (2002); Mars et al. (2018)) which describe the full-range of cortical terminations underlying the fiber tracts comprising the TBSS-derived ROI with respect to the relevant HCP atlas parcels.

2.2.7 Binary Classification Analysis Using Regularized Logistic Regression and Nested Cross-Validation

Once identified, the predictive capacity of a supra-threshold cluster(s) was evaluated by classifying the language dominant hemisphere for each subject. FA asymmetry from identified regions along the symmetric skeleton were extracted and used as a feature vector in a logistic regression model (coding RLD as 1, LLD as 0). By using regularization in the form of a shrinkage parameter (λ), the high-dimensionality and collinearity of the feature space could be accommodated during model fitting. This was implemented using the *glmnet* package (Friedman et al. (2010) in R (https://www.R-project.org/) - setting the alpha parameter to zero in order to enforce the so-called "ridge" penalty. Nested cross-validation (CV) was implemented to separately tune the model hyper-parameter (inner loop: repeated stratified 5-fold CV, 3 repeats) and to generate an unbiased estimate of model performance (outer loop: Leave-One-Out-CV). Unequal case weights (proportional to the inverse of the sample proportions) were applied when training the model to prevent bias towards the over-represented class. The classifier was trained using the area under the receiver operating characteristic curve (AUC) as the quality metric, and the entire nested-CV procedure was repeated 20 times to account for the stochasticity associated with repeated resampling within the inner CV loop. The average AUC over the 20 iterations is reported and the proportion of 10,000 class-shuffled permutations yielding an AUC of equal or greater value to this average was reported as the classifier's statistical significance. Resampling, cross validation, and performance evaluation was performed using the *caret* package (Kuhn and Johnson (2013)) in R.

2.3 Results

Tract-Based Spatial Statistics (TBSS) and Exploratory Classification Analysis

The univariate GLM analysis of inter-hemispheric FA asymmetry showed that relative to RLD group, the LLD group has a significantly greater leftward asymmetry in fronto-parietal white matter ($P < .01$), resulting in a single supra-threshold cluster of 449 voxels within this region (Fig. 2.2, Left). This effect was driven by a slightly leftward asymmetry of FA within the LLD group and strong rightward asymmetry of FA within the RLD group. This locus of asymmetry difference is localized under the posterior inferior frontal gyrus and the fronto-parietal operculum (mean voxel coordinates (MNI Space): 131, 128, 93). Over 20 iterations of the nested cross-validation procedure, the regularized logistic regression model resulted in an average AUC of 0.84 ($\pm$ 0.02) ($P < 0.0001$), reflecting the probability that for any random pair of observations drawn from each group, the classifier assigned a higher predicted probability to a subject classified as RLD (coded as 1) relative to LLD (coded as 0) (Fig. 2.3, Right).

Connectivity Analysis

At the group level, fiber tracking seeded from the region-of-interest (ROI) identified by the TBSS analysis revealed a variety of intra- and inter-lobar frontal pathways primarily converging in the posterior inferior frontal cortex (Fig. 2.4B). The grouped pathway trajectories are expressed volumetrically as subcortical group maps and as surface-based projection maps in figure 2.4 (Fig. 2.4). Within the left-hemisphere for both groups, pathways coursed to the posterior middle temporal cortex by way of the

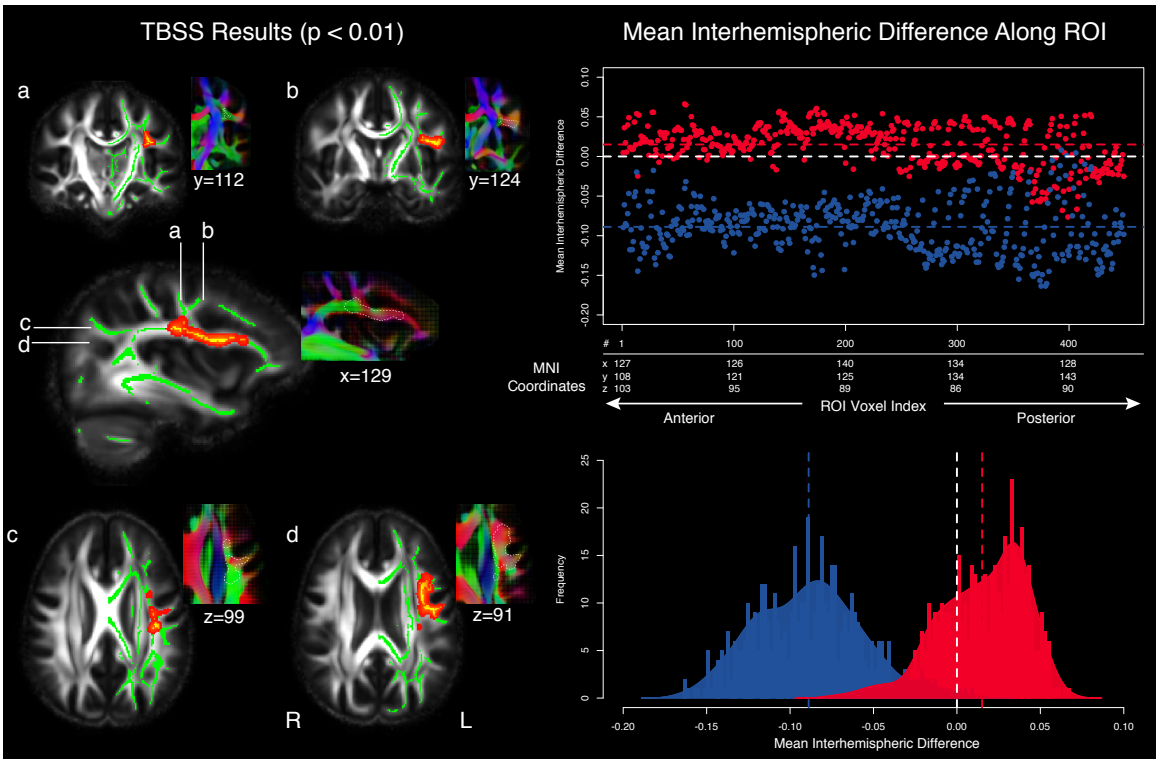

Figure 2.2: Results showing a significant cluster of asymmetry

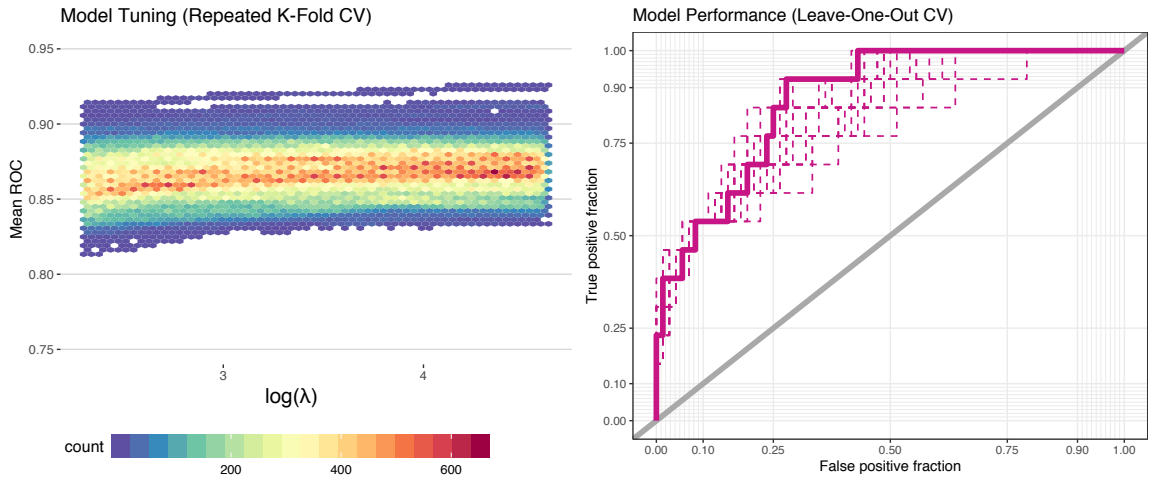

Figure 2.3: Exploratory classification analysis with nested cross-validation. The plots show model performance estimates on the training (left) and on the test set (right).

classical arcuate fasciculus (Fig. 2.4A). Conversely, in the right hemisphere, the posterior pathway terminations were primarily observed in supra-marginal gyrus within the inferior parietal cortex (Fig. 2.4B), by way of the superior longitudinal fasciculus (Fig. 2.4A). Anteriorly, both hemispheres showed cortical connections ventrolater-

ally to pars opercularis of the inferior frontal gyrus (i.e., BA 44), ventral precentral cortex, and insular cortex, and dorsomedially to posterior superior frontal cortex by way of the frontal aslant tract and to the caudal middle frontal gyrus likely through superficial fibers interconnecting inferior and middle frontal gyri (Fig. 2.4A).

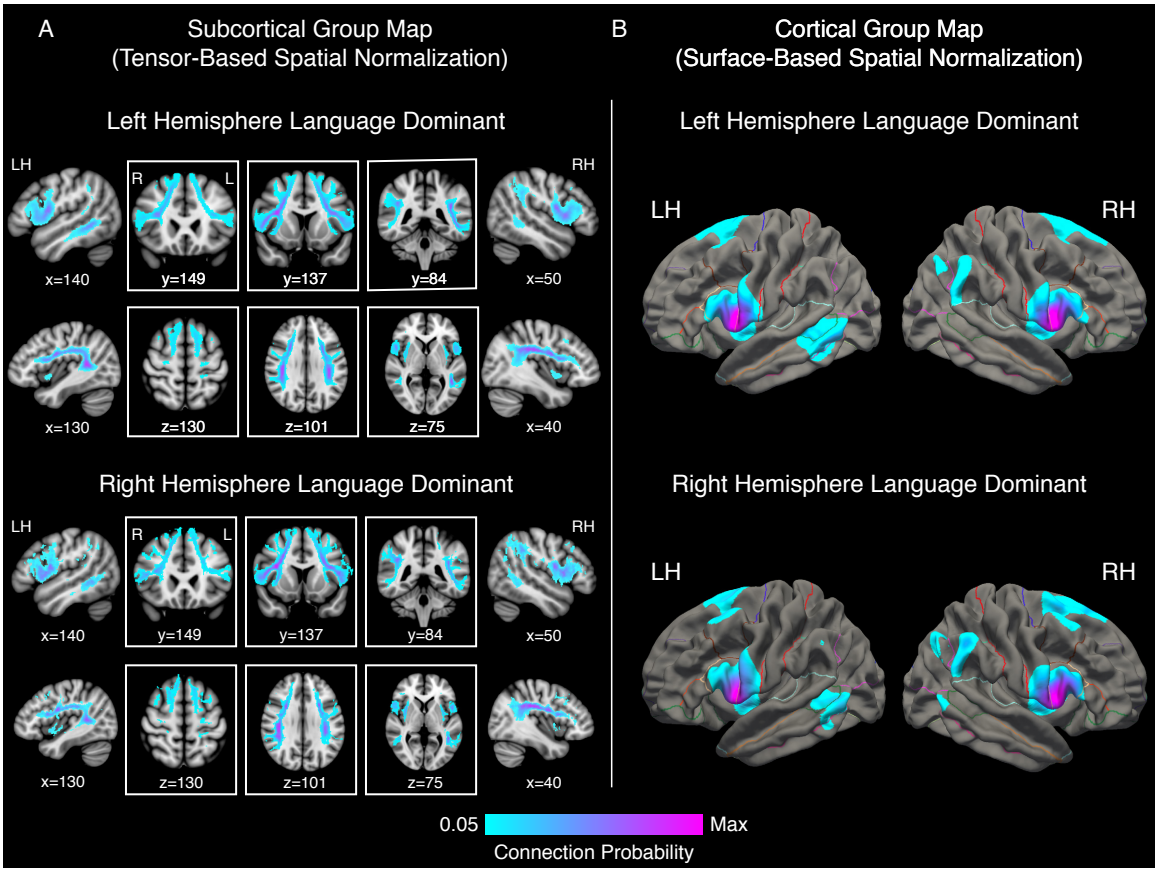

Figure 2.4: Volumetric (A) and surface-based (B) probabilistic maps of the highest probability pathways comprising the TBSS cluster in both LLD and RLD groups. Both representations show the average connection probability across their respective group but thresholded at 5% to remove spurious connections.

Hierarchical Clustering into Projection Zones and Linear Mixed Effects Modeling

Of the 180 parcels in the HCP atlas, 32 parcels reached the criterion for further analysis (Fig. 2.5). The hierarchical clustering analysis of their connection probabilities

produced 13 projection zones (i.e., clusters) that are shown in figure 2.5. Each zone corresponds to a local collection of HCP atlas parcels sharing similar connectivity patterns within inferior frontal (zones iF1, iF2, iF3), superior frontal (zones sF1, sF2, and sF3), insular/opercular (zones INS, OP1, and OP2), inferior parietal cortex (zone iP), parieto-temporal (zone PT), and lateral temporal cortex (zones T1 and T2). The connectivity fingerprints demonstrate the primary projection sites of the connections emanating from the TBSS-derived ROI, grouped according to zone membership (Fig. **??**, Bottom). At a corrected level of significance, the LME analysis revealed a significant main effect of hemisphere in 5 out of the 13 zones tested (Table 2). These effects, sorted by effect size magnitude (Cohen's d), were found in lateral temporal zone T2 containing parcels STSdp, STSvp, TE1p, and PHT (Beta = -0.48, S.E = 0.02, p = 1.44 x 10-72, d = -1.22), inferior parietal zone iP containing parcels IP2, PF, and PFm (Beta = 0.29, S.E = 0.03, p = 1.06 x 10-9, d = 0.78), superior frontal zone sF2 containing parcels 8Ad and area s6-8 (Beta = 0.17, S.E = 0.04, p = 6.16 x 10-21, d = 0.5), the zone between inferior parietal and lateral temporal cortex in the planum temporale (zone PT) containing parcels PSL, RI, and PFcm (Beta = 0.14, S.E = 0.03, p = 1.16 x 10-8, d = 0.44), and posterior inferior frontal zone iF1 containing parcels 6v, IFJp, and area 43 (Beta = -0.08, S.E = 0.02, p = 1.16 x 10-8, d = -0.22). Positive coefficients for the main effect of hemisphere reflect clusters that showed increased connectivity in the right hemisphere relative to the left hemisphere across both groups (i.e., rightward asymmetry), while negative coefficients reflect decreased connectivity in the right hemisphere relative to the left hemisphere across both groups (i.e., leftward asymmetry). As such, zone T2 located in lateral temporal cortex and zone iF1 in posterior inferior frontal cortex showed a leftward asymmetry across both LLD and RLD groups. Conversely, clusters located within the planum temporale (zone PT), inferior parietal cortex (zone iP), and superior frontal cortex (sF2) showed a rightward asymmetry across both groups. For hemisphere by Wada

outcome interactions, two zones showed significant interaction effects, but only 1 remained significant after multiple comparisons correction; specifically, the insular zone (zone INS) containing parcels FOP3 and MI (Beta = 0.17, S.E = 0.08, uncorrected p = 0.0359, d = 0.57) trended toward significance. The only significant interaction surviving multiple comparisons correction was observed in zone iF2 - containing parcels 44, 45, area 6r, and FOP4 (Beta = 0.13, S.E = 0.07, p = 0.01, d = 0.59). Positive interaction effects reflect clusters where an increase in connectivity is seen in the right hemisphere in the RLD relative to the left hemisphere in the LLD group, and this pattern was observed in zones INS and iF3 (Fig. 2.6, bottom).

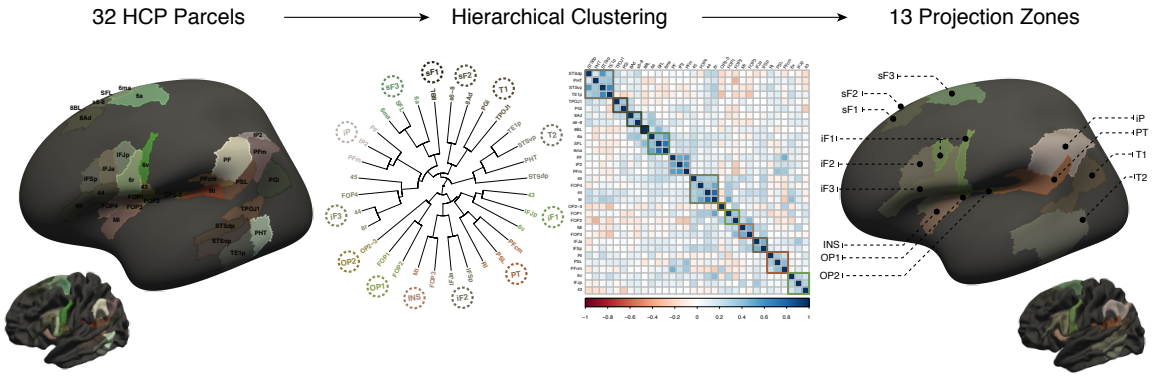

Figure 2.5: Hierarchical clustering of individual parcels into projection zones. (A) (Left) Relevant Human Connectome Project (HCP) atlas parcels were identified and clustered using a hierarchical clustering algorithm (middle). The dendrogram and correlation matrix heat-map illustrate the clustering solution by color-coding individual parcels according to cluster membership. (Right) This approach resulted in 13 clusters - referred to as "projection zones" - representing proximal parcels that shared similar connectivity patterns.

2.4 Discussion

We identified a highly conserved and localized asymmetry in the white matter under the posterior inferior frontal gyrus and the fronto-parietal operculum that closely indexes language lateralization. Tractographic connectivity of this region (mediated

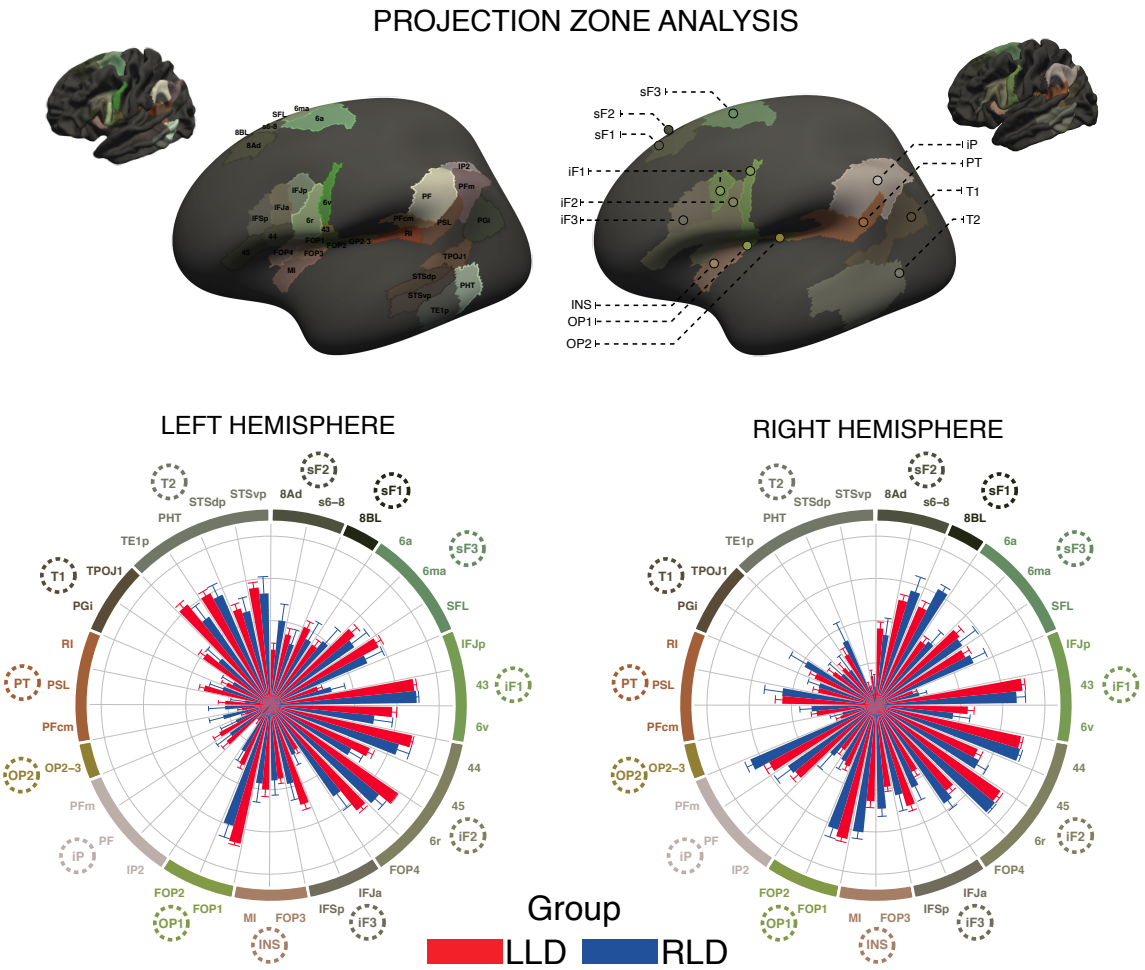

Figure 2.6: (Top) Inflated and mid-thickness brain surfaces showing the individual HCP parcels (left) as well as their projection zone representation (right). (Bottom) Grouped structural connectivity fingerprints plot the average connection probabilities of the relevant HCP parcels, ordered and colored by their projection zone membership. The 13 projection zones were then subjected to separate linear mixed effects models testing for hemisphere by Wada outcome interactions in the connection probabilities of the parcels contained in each zone. Zones were named according to their locations in superior frontal (sF1, sF2, sF3), inferior frontal (iF1, iF2, iF3), insular (INS), opercular (OP1, OP2), inferior parietal (iP), parieto-temporal (PT), and temporal (T1, T2) cortex.

primarily by the arcuate fasciculus and the frontal aslant tract) specifically elaborated all canonical language cortex. This work reveals distinctions in the microstructural integrity of this region in Left-Language Dominant (LLD) and Right-Language Dominant (RLD) groups with a slight leftward asymmetry of fronto-parietal fractional

Table 2.2: Linear mixed effects (LME) models evaluating the hemispheric associations in connectivity between projection zones and Wada outcome.

Projection Zone(Parcel Names)	**Main Effect of Hemisphere**			**Hemisphere by Wada Interaction Effect**		
	Beta (SE)	p value	Effect size (d)	Beta (SE)	p value	Effect size (d)
PT (PSL, RI, PFcm)	0.14 (0.03)	0.000**	0.45	0.05 (0.07)	0.439	0.17
sF3 (SFL, 6ma, 6a)	-0.04 (0.03)	0.183	-0.11	0.04 (0.07)	0.558	0.12
iF1 (6v, IFJp, 43)	-0.08 (0.02)	0.000*	-0.22	-0.02 (0.06)	0.660	-0.07
sF2 (8Ad, s6-8)	0.17 (0.04)	0.000**	0.5	-0.05 (0.09)	0.612	-0.13
sF1 (8BL)	0.13 (0.05)	0.018	0.38	0.2 (0.14)	0.159	0.59
iF2 (44, 45, 6r, FOP4)	-0.04 (0.01)	0.018	-0.16	0.13 (0.04)	0.001*	0.59
iF3(IFJa, IFSp)	0.08 (0.03)	0.011	0.25	0.06 (0.08)	0.443	0.2
OP2 (OP2-3)	0.06 (0.03)	0.057	0	-0.15 (0.09)	0.098	-0.66
INS (MI, FOP3)	0.06 (0.03)	0.056	0.2	0.17 (0.08)	0.036	0.57
OP1 (FOP1, FOP2)	0 (0.02)	0.662	-0.03	0.05 (0.06)	0.359	0.18
T2 (STSdp, STSvp, TE1p, PHT)	-0.48 (0.2)	0.000**	-1.22	0.04 (0.06)	0.463	0.11
T1 (TPOJ1, PGi)	-0.03 (0.04)	0.408	-0.09	0.08 (0.1)	0.399	0.25
IP (IP2, PF, PFm)	0.29 (0.03)	0.000**	0.78	0.13 (0.07)	0.067	0.35

Note: All LME models controlled for handedness and disease hemisphere by including these variables as covariates. In addition, all models included random intercepts for both subject ID and for the parcels comprising each cluster in order to account for inter-individual variability and for statistical inter-dependence of these repeated bilateral measurements in connectivity (e.g., one subject contributing left and right hemisphere connectivity measurements per parcel). In cases when a cluster contained a single parcel, the random effect for parcel was excluded.
*Significant at p <0.05 (corrected)
**Significant at p <0.001 (corrected)

anisotropy (FA) in the LLD group and a much stronger rightward asymmetry of FA in the RLD group. The rare group of RLD subjects was critical in revealing regions important for speech in the more typical LLD group. These findings have import to the establishment of the language dominant hemisphere, solely using a data-driven approach that ascertains inter-hemispheric differences in white matter microstructure. The probabilistic tractography analysis, done unconstrained by prior knowledge of the language literature, revealed pathways of great interest to language researchers: the classical arcuate fasciculus (AF), connections between the posterior inferior frontal cortex (pIFC) and the superior frontal gyrus (SFG) by way of the frontal aslant tract (FAT), and connections medial to the insula - mediated by either the inferior fronto-occipital fasciculus (IFOF) or fronto-insular pathway. Furthermore, a quantitative asymmetry analysis of cortical connectivity revealed significant grouped-differences in connectivity patterns to anatomical Broca's area (BA 44/45), to pre-supplementary motor area (pre-SMA), as well as to caudal middle temporal cortex. Lastly, using the TBSS-derived region of asymmetry as a feature in a machine-learning classifier, we

were able to effectively classify the language-dominant hemisphere in the majority of subjects in each group.

To conclude, this data-driven analysis applied to a unique cohort of Wada-tested patients validates prior claims about the human language network but the crucial insight provided here lies in the confluence of connections to anatomical Broca's area as being a neuroanatomic substrate in the hemispheric dominance for speech. This claim is consistent with prior imaging studies that show functional and structural properties of anterior language sites being more reliably predictive of hemispheric language dominance, in contrast to those observed in posterior perisylvian cortex. We recognize that the data and subjects used in this study are limited by their clinical nature but we believe these findings to be robust. A recent voxel-based analysis (VBA) study showed comparable asymmetry patterns between Wada-tested patients with epilepsy and healthy controls (Keller et al. (2018)); a demographic representing nearly all of the subjects in the present work. Moreover, a study that implemented graph theoretic analyses on high-angular resolution diffusion imaging (HARDI) data have implicated a strikingly similar network architecture that appears to degenerate in the non-fluent variant of PPA (Mandelli et al. (2016)). Lastly, we highlight the possibility that having speech in the right-hemisphere influences subtle structural properties of perisylvian networks. Specifically, while asymmetric temporal lobe projections of the AF in the left-hemisphere appear to be a critical feature of left-hemisphere language dominance, rightward fronto-parietal microstructure and structural connectivity is an anatomical substrate for right-hemisphere language dominance. Thus, considering Broca's area as a convergence zone between the AF/SLF fiber complex, the frontal aslant tract, and fronto-insular tracts better explains the constellation of symptoms associated with lesions to this area (Mandelli et al. (2016)); namely, with agrammatism (Wilson et al. (2012)), reduction in fluency (Mandelli et al. (2014)), and with phonologically paraphasic production (Duffau (2015)).

Chapter 3

Dissociating the Structural Networks Underlying Brodmann Area 44 and Ventral Premotor Cortex using Partial Least Squares Discriminant Analysis: A Human Connectome Project Study

3.1 Introduction

Our first experiment showed that the posterior inferior frontal cortex - including the ventral premotor cortex (vPMC) - is a crucial convergence zone for a variety of pathways critical for the hemispheric dominance for speech. Lesion studies have high-

lighted that cerebral infarctions to these cortical areas result in a well-documented neuropsychological syndrome called "Broca's aphasia", presenting with a profile of agrammatic and non-fluent pattern of speech output (Dronkers et al. (2007)). Similarly, a specific sub-type of primary progressive aphasia (PPA) which selectively targets this area of the brain - the non-fluent variant of PPA (nvPPA) - results in symptoms closely relating to those seen in Broca's aphasia (Gorno-Tempini et al. (2011)). It is thus within reason to expect that the constellation of symptoms arising from these syndromes might be best explained by the differential patterns of disruption to discrete white matter structures converging to this region. In the literature, the overwhelming focus of research has been dedicated to relating language deficits to fiber pathways connecting to posterior peri-sylvian areas (i.e., the "canonical language areas"). Recently, with advanced diffusion imaging and virtual tract reconstruction techniques, a greater emphasis is being placed on studying pathways connecting so-called "noncanonical language areas" (Catani et al. (2012, 2013)). As a result, the connectivity patterns of these central hubs - namely, of BA 44 and vPMC - is now being revisited (Mandelli et al. (2014)). Unfortunately, these studies rely on *a priori* specifications of their cortical endpoints in order to extract the pathways in question. This approach becomes problematic if the exact endpoints are ill-defined or if the contributing networks are unknown. To date, no study investigates the broad cortico-cortical connectivity patterns of BA44 and vPMC directly within a large sample of subjects, preventing a full appreciation of their overall relationships with the rest of the brain. Here using a sample of data from 100 human subjects afforded by the Human Connectome Project (HCP), we sought to characterize the predominant connectivity patterns of these eloquent areas separately and also draw comparisons between them. Without having to pre-specify "target" regions-of-interest (ROIs), we can use a data-driven approach to evaluate and compare the underlying connectivity patterns in a spatially unbiased manner.

Theoretically, the constellation of symptoms that are seen in Broca's aphasia and the nonfluent variant of PPA are being explained as "dorsal stream" deficits (Fridriksson et al. (2018)). This stream, as part of the dual-stream model proposed by Hickok and Poeppel (Hickok and Poeppel (2007)), is thought to enable disparate linguistic processes ranging from the sensorimotor aspects of speech (Isenberg et al. (2012)), to phonological working memory (Buchsbaum et al. (2011)), and syntax (Wilson et al. (2012)). The underlying fiber pathways enabling this cortico-centric model of speech has received less attention. By examining the connectivity of two dorsal stream hubs in BA 44 and vPMC, we are in effect studying pathways comprising this stream. Our approach is twofold: First, we carry out a fiber tracking procedure seeded in each of the sub-regions to assess connectivity to the rest of the brain. An exploratory principal components analysis is then applied to the grouped results from each sub-region to generate brain eigen-images representing the regions explaining the most variance in connectivity patterns. Next, we apply a supervised machine-learning classification algorithm, partial least squares- discriminant analysis (PLS-DA), to identify a latent projection of the data onto an axis which maximally separates the sub-regions. By virtue of the weights assigned to the optimal class-separating projection of the data, we can learn about the exact anatomical networks distinguishing these adjacent areas.

3.2 Methods

3.2.1 Participants

One hundred unprocessed T1-weighted scans along with their corresponding pre-processed diffusion-weighted scans Glasser et al. (2013) were downloaded through the Human Connectome Project (HCP) database. Of these data, all were males from the

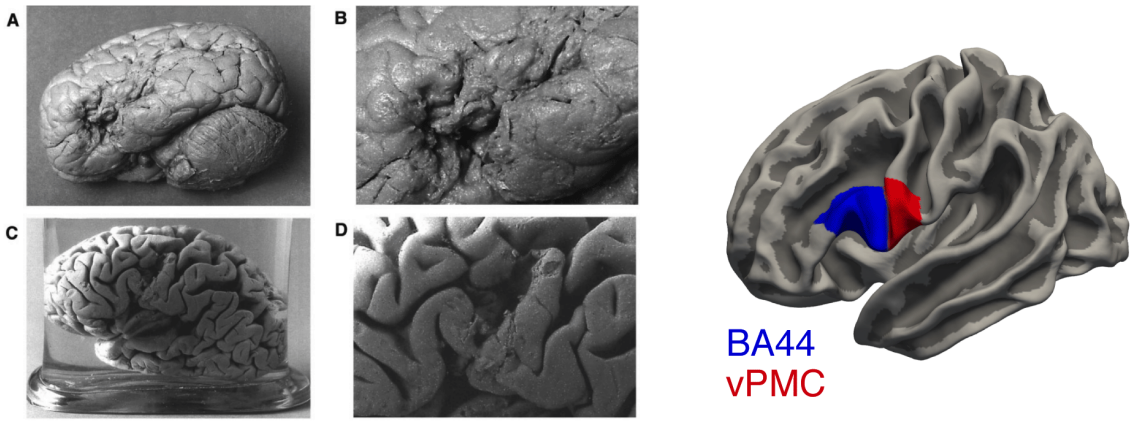

Figure 3.1: (Left) Photographs of the brains of Leborgne and Lelong, Paul Broca's first two aphasic patients. Notice the focal brain damage in the second patient (panels C and D), restricted to the inferior aspect of the precentral sulcus; specifically, to the junction between BA44 and vPMC. (Right) The ROIs of BA44 and vPMC used in the present study.

"HCP young adult" study and ages ranged from 22 to 35 years (mean age 27.5 years).

3.2.2 Anatomical Parcellation and Segmentation

We constructed surface-based models from the T1-weighted structural image of each individual using FreeSurfer (http://surfer.nmr.mgh.harvard.edu). From FreeSurfer's automated procedure (Dale et al. (1999)), we obtained an anatomical parcellation and labeling of 74 cortical and subcortical sulci and gyri for each hemisphere (Destrieux et al. (2010)). The brain parcellation into discrete sulci and gyri was then used to extract seed ROI masks for the different pIFC sub-regions. For BA 44, the gyrus label corresponding to the pars opercularis of the inferior frontal gyrus was used. For the vPMC, the entirety of the precentral gyrus label was modified to match pars opercularis in vertical extent, hence restricting the precentral gyrus to generate an ROI of solely it's ventral portion.

3.2.3 Probabilistic Tractography

Next, probabilistic tractography was initiated from each subregion separately using the *probtrackX* algorithm in FSL. Importantly, to restrict the streamlines to purely intra-hemispheric fibers, exclusion masks of the brain stem, corpus callosum, thalamus, ventricles, and basal ganglia were added. An exclusion mask was also added in the temporal stem region to prevent recording of streamlines from ventral stream pathways; namely, from the uncinate fasciculus and inferior fronto-occipital fasciculus (Saur et al. (2008); Weiller et al. (2011); Friederici (2015)). Once obtained, the tractography data were then projected onto a standardized surface model which contained a node-to-node correspondence across subjects (Dale et al. (1999)), allowing for direct statistical comparisons at each node along the mesh across all subjects.

3.2.4 Descriptive Connectivity Fingerprints and Exploratory Principal Components Analysis

Two different atlases were used to summarize the connectivity patterns of each subregion. The first was the recently developed Glasser HCP Multi-modal Parcellation atlas (Glasser et al. (2016)) which subdivides each hemisphere into 181 cortical regions (Fig.3.2, Right) based on multiple imaging modalities. The second atlas comes from a study by Yeo et al.(2011), which subdivides the brain on the basis of its intrinsic connectivity patterns (Fig.3.2, Left) using resting-state functional MRI (rs-fMRI) (Thomas Yeo et al. (2011)). This atlas embeds the brain into seven different functional networks: The Default Mode, Dorsal Attention, Fronto-parietal, Limbic, Somatomotor, Ventral Attention, and Visual Networks.

Principal Components Analysis (PCA) / Singular Value Decomposition (SVD) is

an unsupervised dimensionality reduction tool that seeks to seeks to generate linear combinations of a data matrix (i.e., feature space) while maximally preserving as much information - or variance - available in the data. In this setting, the SVD will be performed on the list of connectivity data at each vertex of the surface-mesh while across each of the 100 subjects in the study. This approach will be used to generate brain eigenimages and subject-specific projections (principal scores) onto each brain eigenimage. The resulting eigenimages will be mapped onto the brain surface with loading at each node representing the relative contribution to the overall connectivity pattern that explains most of the variability across subjects. This approach will be used as an exploratory technique to visualize the major loci of variation in connectivity patterns for each of the sub-regions (BA44 and vPMC), individually, before drawing comparisons between them (next section).

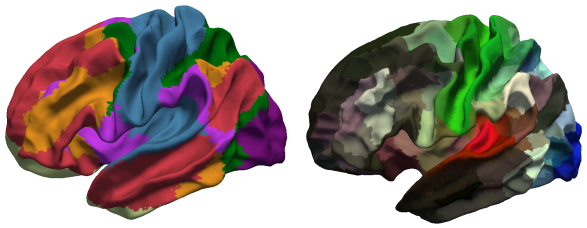

Figure 3.2: The Yeo (Left) and Glasser HCP (Right) atlases were used to characterize the feature space with respect to both functional and multimodal parcellation of the human cerebral cortex.

3.2.5 Partial Least Squares-Discriminant Analysis: Data Preprocesing & Analysis

Data Preprocessing

Rather than using node-wise connectivity matrix as features (40k x 200), a series of pre-processing steps were needed. First, each individual's node-wise tractogram was thresholded to include only the top 10% of connections. This step ensured that only

the strongest (i.e., not spurious) connections would be fed into the classifier. Next, the connectivity data was summarized across each of the 181 ROIs defined by the Glasser atlas by computing the average number of streamlines intersecting each ROI. This effectively reduces the dimensionality of the data from 40,000 features to just 181 features. Last, certain ROIs will not have received any projections from either vPMC nor BA44, leading to columns of zero-variance features across both matrices. These columns were thus removed, resulting in a reduced data matrix of 99 features representing the average connectivity to each of the 99 ROIs from the Glasser atlas.

Using Partial Least Squares Projection

The objective here is to establish a linear combination of features which maximally differentiates class-membership (i.e., vPMC vs BA44). Given that many of the ROIs will exhibit correlations amongst each other, a model using latent variable projections is ideally suitable. To this aim, we subjected the data matrices to a classification algorithm, called partial least squares discriminant analysis (PLS-DA) (Barker and Rayens (2003)), that would project the data onto a subspace spanned by the first PLS component best discriminating the sub-regions on the basis of their distributed connectivity patterns. The appeal behind PLS-DA is that it allows one to interpret the weights of the loadings going into the PLS scores - as opposed to being purely a "black box" algorithm (Brereton and Lloyd (2014)). Figure 3.3B provides a conceptual understanding of how the projections differ between PCA and PLS-DA. PCA projects the data matrix (X) onto axes representing the direction of maximum variance (pink arrow), while PLS-DA projects the data matrix (X) onto a direction which maximizes the covariance with the class labels (Y) (orange arrow), and can therefore be seen as the supervised counterpart to PCA. Given that BA44 and vPMC will likely exhibit distinct connectivity patterns, the feature weights onto the first PLS

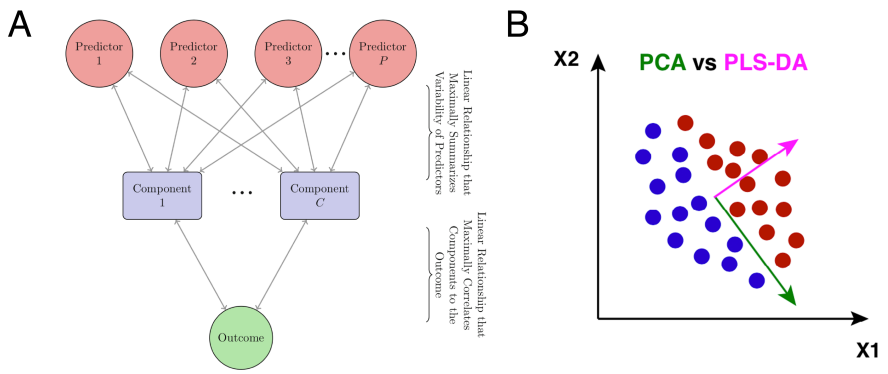

Figure 3.3: A) Conceptual diagram depicting the structure of a Partial Least Squares model. PLS finds components that simultaneously summarize variation of the predictors while being optimally correlated with the outcome. B) A simple example illustrating the differences in latent projections between PCA (green arrow) and PLS-DA (pink arrow).

component should provide informative information into the cortical projections most relevant to the classification rule, thereby providing insight into the broad cortico-cortical network of each sub-region. As noted in the prior section, the feature space will be comprised of the average connectivity values in each of the 99 ROIs of the Glasser atlas. Each subject will have two rows in the overall data matrix (X): One row for that subject's connectivity from BA44 to each of the 99 ROIs and the other for the vPMC connectivity to each of the same 99 ROIs. Overall, the dimensions of the data matrix will be 200 subjects x 99 ROIs and the response vector (Y) will contain the class-membership of each row; namely, the pIFC sub-region from which each row was generated (i.e., BA44 or vPMC). Using k-fold (k=5) cross-validation, the PLS-DA model was tuned over the number of PLS components to retain when making classifications and the optimal model was selected by picking the model with the smallest number of PLS components yielding the highest area under the ROC curve. Given that the classifier will have received data from distinct anatomical circuits, we expect the algorithm to select a minimal number of components to optimally classify the groups.

3.3 Results

3.3.1 Exploratory PCA/SVD

As evidenced by the scree plots for both vPMC and BA44 (Fig. 3.4B), the first principal component (PC) explained nearly 70% of total variance in the data. No other PC explained greater than 2.5% of the total variance in either experiment, indicating that the decomposition was able to retain the majority of the information in the connectivity data just from using a single component representation. The resulting eigen-images from the first PC for each sub-region, shown in figure 3.4, color-code the spatial sources of variability. These exploratory plots are effective at visualizing the overall trends. BA44 shows projections primarily to the pre-supplementary motor area, anterior insular cortex, and posterior temporal cortex. Conversely, vPMC shows distinct projections to supplementary motor area, posterior insular cortex, and posterior temporal cortex albeit to a lesser extent than BA44. vPMC appears to have a greater set of projections to inferior parietal cortex which is pattern otherwise not observed in the BA44 eigen-image.

3.3.2 Sparse Partial Least Squares Discriminant Analysis

The optimal PLS-DA classifier, selected through the k-fold cross-validation procedure, was the model that used a single PLS component which achieved perfect classification accuracy. This result was expected given that the classifier received as inputs data from distinct structural networks. Our goal in using the PLS-DA algorithm was not simply to achieve accurate classification accuracy but primarily to interpret the feature weights assigned to the PLS component used in distinguishing the different sub-regions. The weights assigned to the optimal class-separating projection of the

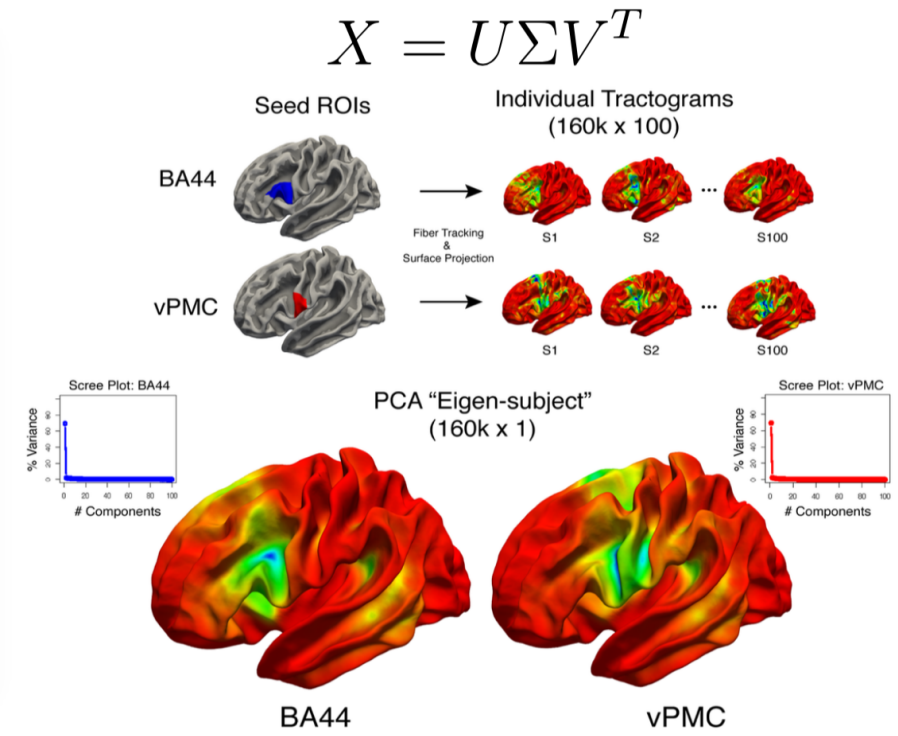

Figure 3.4: PCA reveals predominant eigen-images for each subregion. (Top) Seed ROIs of BA44 and vPMC, along with the surface-projected tractograms of individual subjects. (Bottom) Separate SVD analyses reduce dimensionality from 160k vertices by 100 subject, to 160k vertices by 1 principal component explaining 69% of the variance for each sub-region. Vertices are color-coded by their loadings onto the first singular vector (i.e, PC1).

data are plotted in Figure 3.5A), where each horizontal bar represents the weight magnitudes for a given ROI and are color-coded according to the networks defined by the Yeo atlas. ROIs given negative weights were those that loaded onto BA44 while those given positive weights loaded on vPMC. To visualize these weighting patterns on the brain, the entire Glasser atlas ROIs was then colorized by assigning an RGB triplet based on the magnitude of its ROI weightings, such that bright red areas were those that were most strongly weighted toward vPMC and bright blue areas were those that were most strongly weighted toward BA44. Regions in black reflect ROIs given near-zero weights while regions in white reflect ROIs that were not included in the analysis because neither sub-region projected to these areas based on

our tracking criterion. Based on the different patterns of brain regions assigned to the classification rule, it is evident that vPMC ROIs cluster around a core sensorimotor network while BA44 ROIs cluster around more higher order cognitive networks by way of its connections to the dorsolateral prefrontal cortex, pre-supplementary motor area, and posterior temporal cortex (Fig. 3.4B).

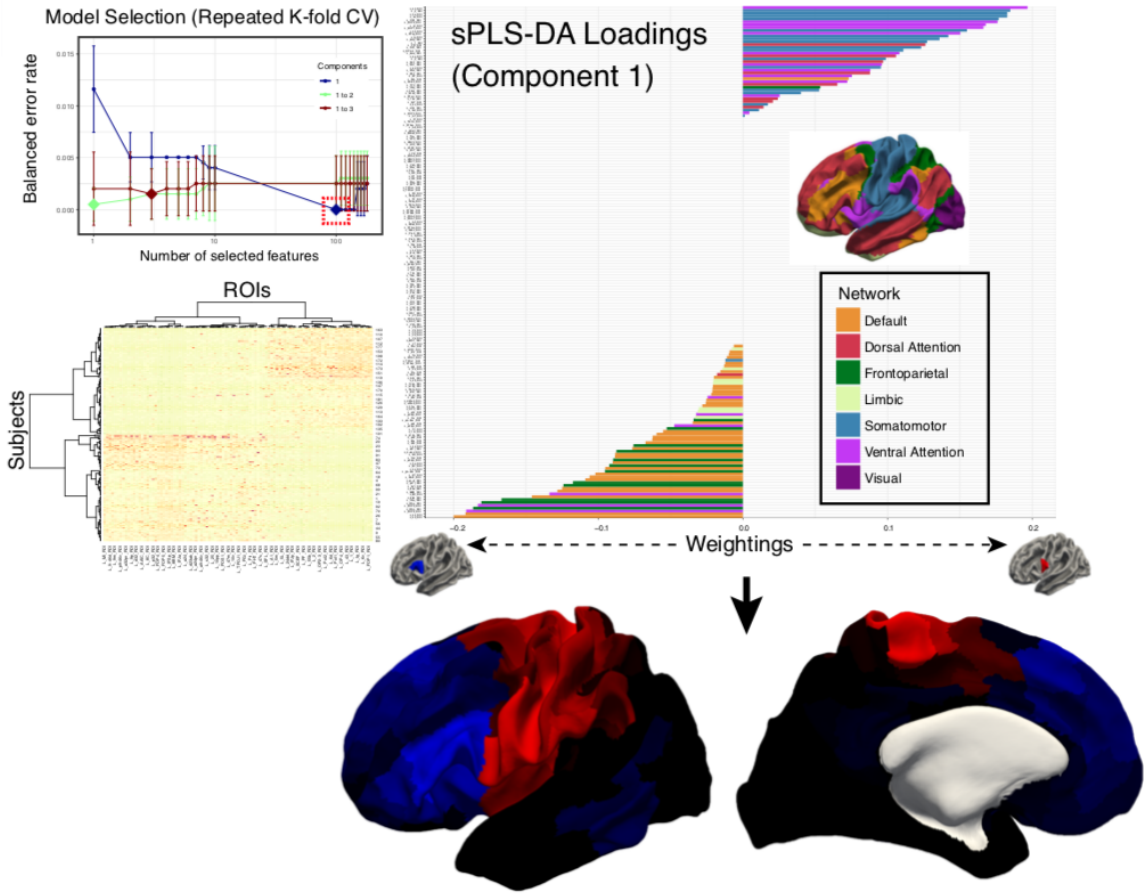

Figure 3.5: (Top Right) PLS-DA optimal model (with 1 PLS component) weights sorted by magnitude and colored by their respective locations in the Yeo atlas. (Bottom) Visualizing optimal class-separating Glasser atlas ROIs projected onto the brain surface: ROIs in bright load onto vPMC, Blue ROIs load onto BA44

3.4 Discussion

The present anatomical experiment was dedicated toward taking a multivariate approach in evaluating the cortico-cortical connectivity patterns of the posterior inferior frontal cortex (pIFC); specifically, of brodmann area 44 (BA44) and ventral premotor cortex (vPMC). Prior studies of this region have often been limited by a spatial bias, such that only connections to pre-specified cortical areas have been investigated while ignoring the rest of the cerebral cortex Catani et al. (2013); Wilson et al. (2012); Brown et al. (2014); Takaya et al. (2015); Kinoshita et al. (2015). Latent variable projection routines like principal components analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) applied on brain connectivity data provide powerful and complimentary approaches in describing how these distributed brain networks are organized across large pools of human subjects. By applying them here, our results provide insight into the structural correlates of various functional sub-networks involved in language. Here we showed the each of BA44 and vPMC are central nodes embedded into both sensorimotor and higher-order aspects of language. Our data provide strong anatomical support for the "neuroanatomical pathway model" proposed by Friederici and colleagues (Friederici (2015)). This model suggests that the dorsal-stream of speech processing is likely subdivided into parallel perisylvian circuits: a sensorimotor circuit that interconnects the inferior parietal cortex to the vPMC and a syntactic processing circuit that interconnects BA44 with the posterior middle temporal cortex. A unique contribution of this study, however, is that each of BA44 and vPMC also receive extensive intra-frontal and fronto-insular connections. This exact finding was shown in the neurosurgical subjects in experiment #1 and has now been replicated within a healthy population but using a much higher-resolution diffusion imaging dataset from the Human Connectome Project. Although purely speculative, these intra-frontal and fronto-insular networks likely interact closely with the parallel

dorsal streams, as noted above, through some motor control mechanism which, in turn, appears to have its own functional specializations. In other words, given that BA44 connects to the pre-supplementary motor area make it highly likely it constitutes a functionally distinguishable network relative to the connections from vPMC to the supplementary motor area (Figures 3.4- 3.5). Ford et al. (2010) were the first to show that these sets of fiber bundles actually belong to much larger set interconnecting nearly the entire span of medial and lateral frontal cortex Ford et al. (2010), and the idea that they may each serve different functional roles appears plausible. Indeed, this idea has recently gained traction from neurosurgical stimulation studies in humans (Corrivetti et al. (2019); Rech et al. (2019)). Corrivetti et al. (2019) presented evidence suggesting an anterior-to-posterior functional gradient existing among them. Intra-operative electrical white matter stimulation to posterior aspects led to transient motor-speech deficits ranging from complete speech-arrest to stuttering and vocalizations. Conversely, stimulation of more anterior aspects produced impairments in naming, implicating its role in lexical-semantic aspects of language (Corrivetti et al. (2019)). Furthermore, Mandelli *et al. (2016)* showed that, in patients with the nonfluent-variant of primary progressive aphasia, cortical atrophy starting from a specific epicenter in BA 44 leads to longitudinal pathological changes to the networks described here (Mandelli et al. (2016)). The functional role of these intra-frontal networks is discussed and addressed in Experiment #3 (Next Chapter).

Chapter 4

Isolating the Speech Repetition Network using Connectome-based Lesion-Symptom Mapping in People with Stroke Induced Aphasia

4.1 Introduction

The dual stream framework for speech processing posits that a distinct functional network for sensorimotor control is subserved by a left-lateralized structural architecture. Localization of the cortical substrates underlying a fundamental sensorimotor ability - speech repetition - have been circumscribed to various posterior perisylvian structures - primarily, to supramarginal gyrus of the inferior parietal cortex, poste-

rior superior temporal cortex, and their adjoining parieto-temporal junction in the planum temporale in an area called "Spt" (Hickok and Poeppel (2007); Buchsbaum et al. (2011)). Damage to these structures is thought to impair speech repetition, as is seen in "conduction aphasics" who present with repetition deficits and paraphasic speech output. This type of aphasia is distinct from expressive and receptive aphasia, in that the impairments appear to selectively disrupt the ability to repeat speech sounds while fluency and comprehension remain intact, and whose output is characterized by frequent phonologically paraphasic errors likely driven by damage to an auditory-motor integration network that serves as an internal feedback control mechanism (Buchsbaum et al. (2011)).

Mapping the underlying white-matter network for speech repetition has been a more challenging endeavor for a variety of reasons. First, cerebral infarctions tend to produce widespread lesions to perisylvian cortical areas, making it difficult to track the underlying pathways using traditional tractography approaches. Second, these lesions are highly variable, making it difficult to draw exact correlations between structure and function. Third, studies in healthy controls that localize speech repetition sites using task-based fMRI are heavily driven by certain theoretical assumptions, such that activation blobs are often selectively chosen for subsequent analysis based on *a priori* knowledge from the literature while ignoring everything else (Isenberg et al. (2012); Saur et al. (2008)). Here, we address these limitations by mapping the structural connectome in a relatively large sample of people with stroke-induced aphasia while using data-driven feature selection algorithms in order to automatically localize the connections most predictive of speech repetition scores. This will, for the first time, isolate the underlying network sub-serving this ability in a purely data-driven manner, consistent with the objectives of this thesis.

4.2 Methods

4.2.1 Population and Neuropsychological Assessment of Word Repetition

The data for this project was obtained from an archival database in the Aphasia Lab, University of South Carolina and Medical University of South Carolina. Participants had sustained a single-event stroke to the left hemisphere at least 6 months prior to study inclusion and were tested either as part of an aphasia treatment study or strictly for the purpose of lesion-symptom mapping research. A total of 72 participants we're enrolled in this study and had both imaging a neuropsychological assessments performed. For this study, two tests of speech repetition were administered and averaged for each participant: The Repetition subtest of the Western Aphasia Battery (Kertesz (2006)) and the Philadelphia Repetition Test (Dell et al. (2007)).

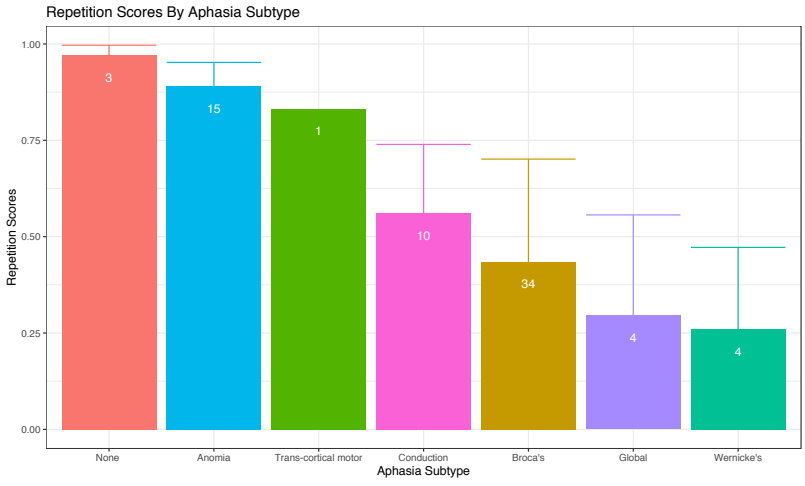

Figure 4.1: The barplots represent average performance over the repetition measures. Patients with Anomia had the highest scores while patients with Wernicke's had the lowest. Conduction Aphasics performed at just over chance levels. The number of participants per subtype are denoted on each individual bar with Broca's aphasics representing the largest subgroup in this sample.

4.2.2 Tractography Protocol and Preprocessing

The structural connectome contains the connectivity values for each pair of 192 regions-of-interest (ROIs) defined by the AICHA atlas (Joliot et al. (2015)), resulting in a 192x192 symmetric matrix per subject (36,864 total elements). These values represent the probabilistic fiber counts, normalized both by the distance between the ROIs and the volume of the two ROIs. The lower triangular of the matrix was extracted for each subject and transformed to a single 1-dimensional vector containing 18,240 elements with each value representing a particular pairwise connection between two AICHA atlas ROIs. These vectors were then concatenated across subjects to form a 72 x 18,240 matrix to be used as the feature space for subsequent analysis. Lastly, to further eliminate near-zero-variance features or features with low connectivity values across the entire population, the matrix was thresholded to remove any features with a average connectivity value of less than one, thereby reducing the feature space to a 72 x 2363 matrix.

4.2.3 Sparse Partial-Least Squares Regression and Variable Importance

We previously described the PLS formulation (subsection 3.2.5) as a supervised counterpart to principal components analysis, wherein the covariance between the data matrix (X) and the response (Y) is maximized when performing the projection onto the latent PLS components (Lê Cao et al. (2008)). In Chapter 3, this sparsity-inducing algorithm was used to perform a binary classification (i.e., sPLS-DA). In this chapter, its regression counterpart (i.e., sPLS-R) will be used to predict a continuous outcome: speech repetition scores derived from a neuropsychological test battery. The goal in this setting is to allow the algorithm to automatically implement feature selection

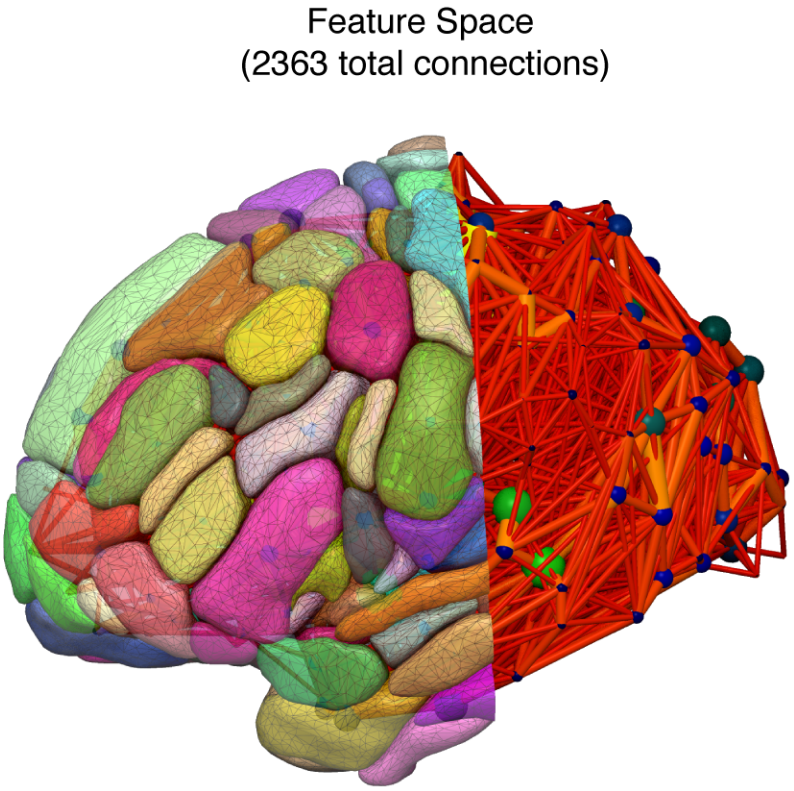

Figure 4.2: All supra-threshold connections used in the connectome-based lesion-symptom mapping analysis with each edge representing a unique connection and each node representing a region defined by the AICHA atlas.

using the LASSO penalty (Rohart et al. (2017)) and subsequently evaluate the importance of the retained features in the model (i.e., variable importance) by ranking the features on the basis of their predictive value with the repetition scores. Prior to model tuning, 25% of the data were held out and the remaining 75% was used to tune the sPLS model's complexity, defined as the number of features to retain during model fitting. A range of values were tested for this complexity parameter, 5-100, and the optimal number of features was determined using repeated k-fold cross validation using mean absolute error as the quality metric of the fit at each iteration. Second, the number of PLS components to be retained was fixed at 2 and these components effectively represent multivariate connectional signatures whose linear combinations

possess predictive value for the repetition scores.

4.2.4 Feature Ranking using Variable Importance in Projection and Out-of-Sample Predictions

In order to rank the retained features on the basis of their predictive value, the variable importance in projection (VIP) method was used (Farrés et al. (2015)). The VIP coefficient essentially represents the impact of a given variable into the construction of the PLS components while weighting the impact by the variance accounted for by each of the components. The VIP score for the jth variable is given as:

$$VIP_j = \sqrt{\frac{\sum_{f=1}^{F} w_{jf}^2 \cdot SSY_f \cdot J}{SSY_{total} \cdot F}} \tag{4.1}$$

where w_{jf} is the weight value for j variable and f component and SSY_f is the sum of squares of explained variance for the f^{th} component and J number of X variables. SSY_{total} is the total sum of squares explained of the dependent variable, and F is the total number of components. The w_{jf}^2 gives the importance of the j^{th} variable in each fth component, and VIP_j is a measure of the global contribution of j variable in the complete PLS model. The VIP coefficients therefore represent a weighted sum of the PLS loadings, which take into account the explained variance of each of the connectional brain signatures (i.e., PLS Component) at each iteration and is directly related to the nonzero features in the optimal model. Since the average of the squared VIP scores equals 1, 'greater than one rule' is generally used as a criterion for variable selection (Chong and Jun (2005)).

The aforementioned model tuning procedure was subjected to a bootstrap resampling analysis, wherein the data was re-sampled (with replacement) 500 times and the VIP scores for the retained features at each iteration was recorded. The median VIP scores across the 500 iterations was then used as a measure of feature importance. Lastly, the highest ranked features were used to fit a PLS model on the full training set and performance was evaluated by making predictions on the test set (i.e., the 25% of samples that were held out during model tuning). The predicted repetition scores were compared with the actual repetition scores using kendall's tau metric, which is a non-parametric correlation coefficient to evaluate the strength of an association.

4.3 Results

The 2363 connections surviving the connectivity threshold are displayed in Figure 4.2 as a connectomics graph representation with respect to the AICHA atlas parcellation. These connections serve as the the entire feature space for the sPLS-R feature selection algorithm. After 500 iterations of the bootstrapped resampling procedure, the median VIP scores were computed and only 7 features presented with a median VIP score greater than one (Fig. 4.3, Left).

The feature with the strongest median VIP score was revealed as a local connection within the Intraparietal Sulcus: "S_Intra-Parietal_3" to "S_Intra-Parietal_2". The majority of the remaining connections identified by the feature selection algorithm were found in close anatomical vicinity within parieto-temporal cortical areas; specifically, connections between superior temporal, supramarginal, and angular gyrus appeared most prominent. The collection of features identified by the analysis are shown as edges on the AICHA atlas in figure (Fig. 4.3, Right), with the edge thickness scaled according to the VIP score magnitude. This plot shows that among the features iden-

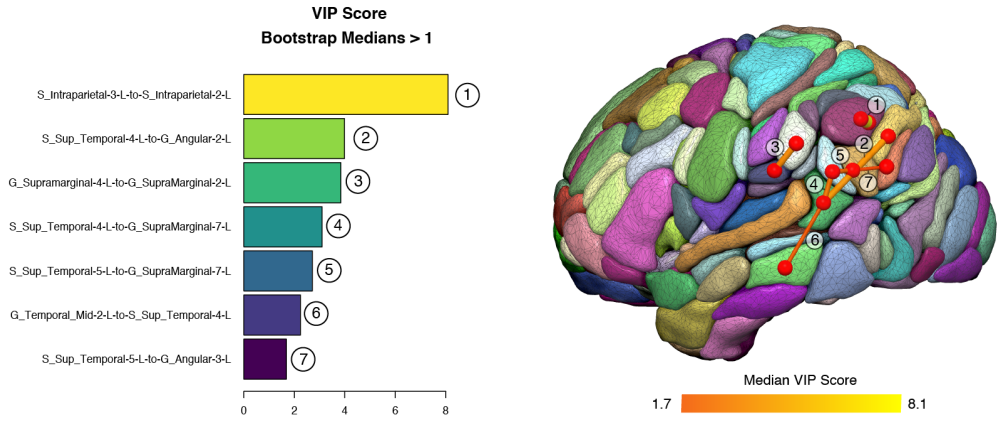

Figure 4.3: The results from the 500 bootstrap resamples of the model tuning procedure are aggregated, sorted, and displayed using both the mean and median of the distribution of VIP scores, respectively.

tified by the analysis, parietal connections had the strongest predictive value relative to the temporal connections. After fitting the PLS model using the 7 parieto-temporal connections, the model predictions were evaluated on the test set and achieved a Kendall's tau value of 0.35 ($p = 0.02$). The scatter-plot of actual vs fitted repetition scores illustrates their strong association across all aphasia subtypes present in the test set (Fig. 4.4).

4.4 Discussion

The neuroanatomical basis for speech repetition has been a point of contention ever since the classical neurobiological models of language proposed in the late 19th century. So-called "disconnectionist" accounts explaining repetition deficits have explained such disruptions to the accurate fascicles, preventing communication between anterior and posterior language sites. Such accounts are still widely taught in medical schools today yet modern evidence in support of this claim is surprisingly lacking

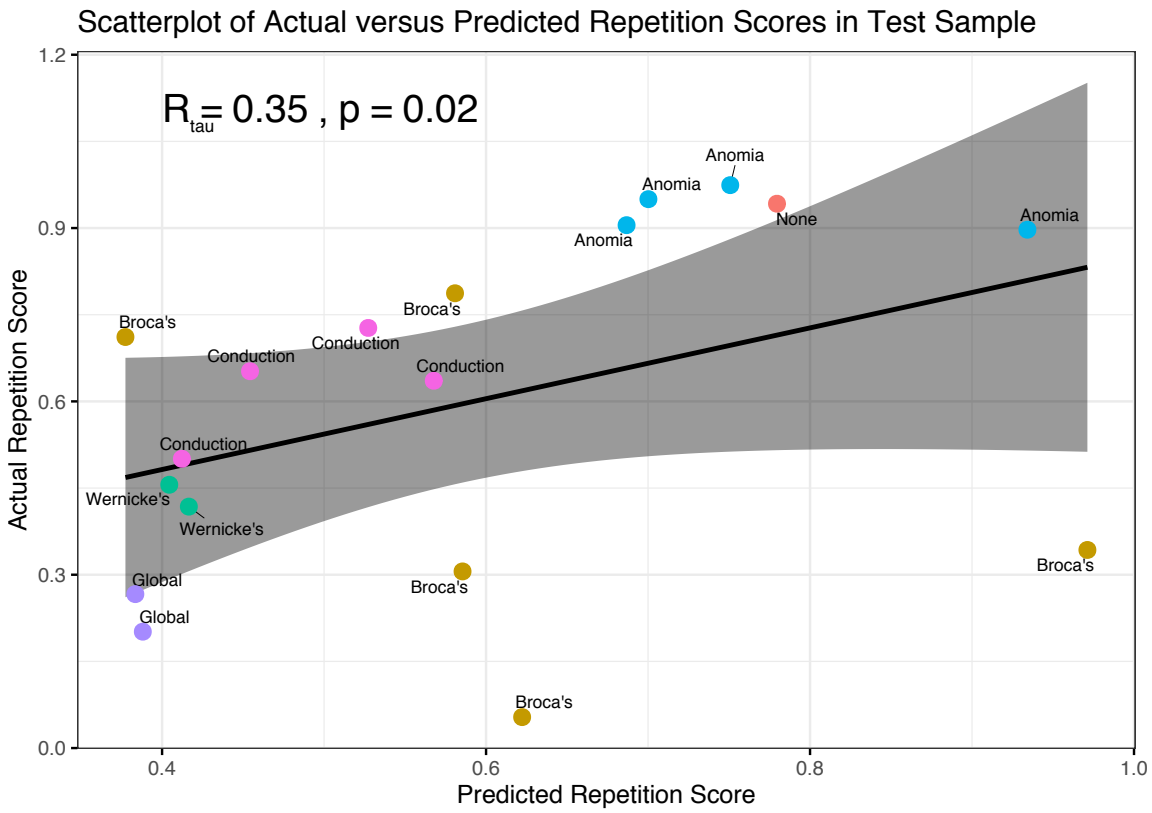

Figure 4.4: A scatterplot of the actual vs predicted repetition scores from the generated PLS model, with points colored by aphasia sub-type.

(Moritz-Gasser and Duffau (2013)). Here, for the first time, we mapped the structural connectome in an aphasic population within whom speech repetition scores were collected and correlations with the connectome were drawn. Our data-driven feature selection framework localized a small network of connections in the parieto-temporal junction as being essential for predicting repetition scores in out-of-sample participants. This novel finding pertaining to subcortical white matter correlates of repetition is remarkably in line with more recent studies evaluating its *cortical* substrates. It has thus been hypothesized that lesions to this temporo-parietal junction directly cause conduction aphasia - a syndrome characterized by repetition deficits (Buchsbaum et al. (2011); Isenberg et al. (2012); Rogalsky et al. (2015)). It therefore appears that speech repetition is, by-and-large, a disruption to the cortical areas associated with auditory-motor integration considering that this ability has been localized to

the exact regions identified in the present work (Rogalsky et al. (2015); Lukic et al. (2019)). Similarly, studies in primary progressive aphasia populations have identified a variant of the neurodegenerative disease, named the logopenic variant of PPA (lvPPA), wherein areas in perisylvian cortex are selectively impacted, leading to impairments with repetition abilities. One such study evaluated the relation between cortical thickness and repetition performance, and it was found that poor repetition of long phrases was associated with cortical thinning in left temporo-parietal areas, especially pronounced in subject with lvPPA (Lukic et al. (2019)). The cortical maps revealed in that study strikingly overlap with the connections identified here, indeed suggesting that avSTM impairments are related either to local cortical (Buchsbaum et al. (2011); Lukic et al. (2019)) or subcortical mechanisms in the parieto-temporal cortex rather than to the disconnection of the long-rage association pathway of the classical arcuate fasciculus.

Chapter 5

Conclusion

This thesis was dedicated to testing several of the most widely accepted assumptions relating to the structural architecture of the human language network. With an emphasis on white matter connectomics, the present work implemented several data-driven approaches to lessen the reliance on a priori theoretical approaches that often contaminate the literature. With such a framework in mind, several novel insights were obtained that would have otherwise been unfeasible using classical approaches.

First, we asked whether there were any white matter asymmetries that indexed hemispheric language dominance as recreated by the clinical gold standard Wada procedure. Experiment 1 tested the hypothesis using a whole-brain spatially unbiased approach. Is the accurate fasciculus the sole arbiter of language dominance, as classical models propose? Or could the analysis reveal a more nuanced picture in which a greater subset of pathways are implicated in this fundamental property of the human brain (i.e., it's hemispheric preference for speech)? Indeed, we found that several non-canonical pathways were identified by the analysis, all of which shared a common projection zone in so-called "Broca's area" in the posterior inferior frontal cortex. In

addition to the classical accurate The implicated pathways involved classical perisylvian pathways (i.e., the superior longitudinal fasciculus and it's arcuate branch) as well as pathways connecting to the insular cortex and supplementary motor cortex in the superior frontal gyrus. These latter pathways - the fronto-insular tract and frontal aslant tract - have been studied in isolation with relevance for speech production. To our knowledge, this is the first analysis to implicate these connections in a data-driven manner. Together, experiment 1 revealed a novel result pinpointing to Broca's area as being a critical convergence zone for a variety of pathways critical for speech. A somewhat controversial finding was that the accurate fasciculus - defined as white matter interconnecting inferior frontal and middle temporal cortex - was left lateralized across both groups of participants (i.e., even in right-hemisphere dominant subjects). Despite this, our classification analysis using fractional anisotropy from within fronto-parietal white matter enabled successful classification of the dominant hemisphere across both groups. This result points to the possibility that right hemisphere dominance can occur in the presence of a left-lateralized arcuate.

Second, experiment 2 was dedicated to using state-of-the-art diffusion imaging data from the Human Connectome Project in order to take a closer look at the connectivity patterns of 2 central nodes within the speech production network: Pars Opercularis of the inferior frontal gyrus (Brodmann area 44) and ventral precentral cortex. The connectivity of the latter area - the vPMC - is widely understudied due to its complex underlying fibre configuration but is increasingly being recognized as a higher-order language region. By seeding a tractography algorithm in each of these areas, we found that each subregion of Broca's area possessed distinct connectivity patterns with the rest of the brain. Specifically, BA44 was densely connected with superior frontal, insular, and middle temporal cortex; a network closely resembling the architecture identified from experiment 1. Conversely, vPMC was densely connected with a sensorimotor core network that was by-and-large supra-Sylvian. The vPMC had notably

stronger connections to the inferior parietal and supramarginal areas, both of which are known to be associated with phonological working memory together with the vPMC. Together, these findings shed light on the distributed connectivity patterns of these critical nodes in the dominant hemisphere.

Third, experiment 3 was dedicated to uncovering the pathways important for speech repetition. Classically, it was proposed that this ability was carried out by the connections of the arcuate and whose disconnection would lead to conduction aphasia (i.e., an impairment with repetition). More recent evidence suggests that conduction aphasia is actually a cortical impairment rather than a subcortical disconnection. Using a connectome-based lesion symptom mapping approach, we mapped the structural connectome of a population of stroke patients, and tested whether there was a subset of connections predictive of speech repetition scores (as measured by the Western Aphasia Battery). This study is by far the most comprehensive analysis available in the literature, by evaluating over 2000 different connections and their relationship with repetition ability. Using a novel feature selection framework made possible using the sparse partial least squares regression algorithm, we identified a local cluster of connection in parieto-temporal cortex. This result gives credence to the idea that speech repetition is a deficit with a focal set of brain regions responsible for auditory verbal working memory (avWM). Damage to this focal region impairs the phonological buffer that is crucial to storing information in memory prior to repetition. The finding that fronto-temporal connections were not identified by the analysis provides evidence that the classical disconnectionist model is inadequate in explaining speech repetition deficits. Rather, local cortical areas and their and local subcortical connections are fundamental in successful verbatim repetition by enabling access to phonological stores. Our data provides support for the dual stream model for language, which emphasized in area in parieto-temporal cortex, area Spt, as being the locus for sensorimotor integration for speech-related actions.

Bibliography

Barker, M. and Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(3):166–173.

Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). In vivo fiber tractography using dt-mri data. *Magnetic resonance in medicine*, 44(4):625–632.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using 700 lme4. *Journal of Statistical Software*, 701.

Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage*, 34(1):144–155.

Bernal, B. and Altman, N. (2010). The connectivity of the superior longitudinal fasciculus: a tractography dti study. *Magnetic resonance imaging*, 28(2):217–225.

Breier, J. I., Simos, P., Zouridakis, G., Wheless, J., Willmore, L., Constantinou, J., Maggio, W., and Papanicolaou, A. (1999). Language dominance determined by magnetic source imaging: a comparison with the wada procedure. *Neurology*, 53(5):938–938.

Brereton, R. G. and Lloyd, G. R. (2014). Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics*, 28(4):213–225.

Brown, E. C., Jeong, J.-W., Muzik, O., Rothermel, R., Matsuzaki, N., Juhász, C., Sood, S., and Asano, E. (2014). Evaluating the arcuate fasciculus with combined diffusion-weighted mri tractography and electrocorticography. *Human brain mapping*, 35(5):2333–2347.

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'esposito, M., and Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory–an aggregate analysis of lesion and fmri data. *Brain and language*, 119(3):119–128.

Catani, M., Dell'Acqua, F., Vergani, F., Malik, F., Hodge, H., Roy, P., Valabregue, R., and De Schotten, M. T. (2012). Short frontal lobe connections of the human brain. *cortex*, 48(2):273–291.

Catani, M., Mesulam, M. M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C., Thompson, C. K., Thiebaut de Schotten, M., Dell'Acqua, F., Weintraub, S., et al. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. *Brain*, 136(8):2619–2628.

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., and Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, 110(7):2653–2658.

Chang, E. F., Wang, D. D., Perry, D. W., Barbaro, N. M., and Berger, M. S. (2011). Homotopic organization of essential language sites in right and bilateral cerebral hemispheric dominance. *Journal of neurosurgery*, 114(4):893–902.

Chong, I.-G. and Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and intelligent laboratory systems*, 78(1-2):103–112.

Coalson, T. S., Van Essen, D. C., and Glasser, M. F. (2018). The impact of traditional neuroimaging methods on the spatial localization of cortical areas. *Proceedings of the National Academy of Sciences*, 115(27):E6356–E6365.

Corrivetti, F., de Schotten, M. T., Poisson, I., Froelich, S., Descoteaux, M., Rheault, F., and Mandonnet, E. (2019). Dissociating motor–speech from lexico-semantic systems in the left frontal lobe: insight from a series of 17 awake intraoperative mappings in glioma patients. *Brain Structure and Function*, pages 1–15.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194.

Dell, G. S., Martin, N., and Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, 56(4):490–520.

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53(1):1–15.

Dick, A. S., Bernal, B., and Tremblay, P. (2014). The language connectome: new pathways, new concepts. *The Neuroscientist*, 20(5):453–467.

Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T., and Cabanis, E. A. (2007). Paul broca's historic cases: high resolution mr imaging of the brains of leborgne and lelong. *Brain*, 130(5):1432–1441.

Duffau, H. (2015). Stimulation mapping of white matter tracts to study brain functional connectivity. *Nature Reviews Neurology*, 11(5):255.

Farrés, M., Platikanov, S., Tsakovski, S., and Tauler, R. (2015). Comparison of the variable importance in projection (vip) and of the selectivity ratio (sr) methods for variable selection and interpretation. *Journal of Chemometrics*, 29(10):528–536.

Ford, A., McGregor, K. M., Case, K., Crosson, B., and White, K. D. (2010). Structural connectivity of broca's area and medial frontal cortex. *Neuroimage*, 52(4):1230–1237.

Fridriksson, J., den Ouden, D.-B., Hillis, A. E., Hickok, G., Rorden, C., Basilakos, A., Yourganov, G., and Bonilha, L. (2018). Anatomy of aphasia revisited. *Brain*, 141(3):848–862.

Friederici, A. D. (2015). The neuroanatomical pathway model of language: Syntactic and semantic networks. In *Neurobiology of Language*, pages 349–356. Elsevier.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1.

Geschwind, N. and Levitsky, W. (1968). Human brain: left-right asymmetries in temporal speech region. *Science*, 161(3837):186–187.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171.

Glasser, M. F. and Rilling, J. K. (2008). Dti tractography of the human brain's language pathways. *Cerebral cortex*, 18(11):2471–2482.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., et al. (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124.

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., Ogar, J. M., Rohrer, J., Black, S., Boeve, B. F., et al. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, 76(11):1006–1014.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2):135.

Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K., Barde, L., and Grant, A. (2008). Bilateral capacity for speech sound processing in auditory comprehension: evidence from wada procedures. *Brain and language*, 107(3):179–184.

Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5):393.

Isenberg, A. L., Vaden Jr, K. I., Saberi, K., Muftuler, L. T., and Hickok, G. (2012). Functionally distinct regions for spatial processing and sensory motor integration in the planum temporale. *Human brain mapping*, 33(10):2453–2463.

Joliot, M., Jobard, G., Naveau, M., Delcroix, N., Petit, L., Zago, L., Crivello, F., Mellet, E., Mazoyer, B., and Tzourio-Mazoyer, N. (2015). Aicha: An atlas of intrinsic connectivity of homotopic areas. *Journal of neuroscience methods*, 254:46–59.

Keller, S. S., Roberts, N., Baker, G., Sluming, V., Cezayirli, E., Mayes, A., Eldridge, P., Marson, A. G., and Wieshmann, U. C. (2018). A voxel-based asymmetry study of the relationship between hemispheric asymmetry and language dominance in wada tested patients. *Human brain mapping*, 39(7):3032–3045.

Kertesz, A. (2006). Western aphasia battery-revised (wab-r). *Austin, Texas: Pro-Ed*.

Kinoshita, M., de Champfleur, N. M., Deverdun, J., Moritz-Gasser, S., Herbet, G., and Duffau, H. (2015). Role of fronto-striatal tract and frontal aslant tract in movement and speech: an axonal mapping study. *Brain Structure and Function*, 220(6):3399–3412.

Kuhn, M. and Johnson, K. (2013). *Applied predictive modeling*, volume 26. Springer.

Lê Cao, K.-A., Rossouw, D., Robert-Granié, C., and Besse, P. (2008). A sparse pls for variable selection when integrating omics data. *Statistical applications in genetics and molecular biology*, 7(1).

Lukic, S., Mandelli, M. L., Welch, A., Jordan, K., Shwe, W., Neuhaus, J., Miller, Z., Hubbard, H. I., Henry, M., Miller, B. L., et al. (2019). Neurocognitive basis of repetition deficits in primary progressive aphasia. *Brain and language*, 194:35–45.

Mandelli, M. L., Caverzasi, E., Binney, R. J., Henry, M. L., Lobach, I., Block, N., Amirbekian, B., Dronkers, N., Miller, B. L., Henry, R. G., et al. (2014). Frontal white matter tracts sustaining speech production in primary progressive aphasia. *Journal of Neuroscience*, 34(29):9754–9767.

Mandelli, M. L., Vilaplana, E., Brown, J. A., Hubbard, H. I., Binney, R. J., Attygalle, S., Santos-Santos, M. A., Miller, Z. A., Pakvasa, M., Henry, M. L., et al. (2016). Healthy brain connectivity predicts atrophy progression in non-fluent variant of primary progressive aphasia. *Brain*, 139(10):2778–2791.

Mars, R. B., Passingham, R. E., and Jbabdi, S. (2018). Connectivity fingerprints: From areal descriptions to abstract spaces. *Trends in cognitive sciences*.

Mesulam, M.-M., Thompson, C. K., Weintraub, S., and Rogalski, E. J. (2015). The wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. *Brain*, 138(8):2423–2437.

Moritz-Gasser, S. and Duffau, H. (2013). The anatomo-functional connectivity of word repetition: insights provided by awake brain tumor surgery. *Frontiers in human neuroscience*, 7:405.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1):97–113.

Passingham, R. E., Stephan, K. E., and Kötter, R. (2002). The anatomical basis of functional localization in the cortex. *Nature Reviews Neuroscience*, 3(8):606.

Rauschecker, J. P. and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature neuroscience*, 12(6):718.

Rech, F., Herbet, G., Gaudeau, Y., Mézières, S., Moureau, J.-M., Moritz-Gasser, S., and Duffau, H. (2019). A probabilistic map of negative motor areas of the upper limb and face: a brain stimulation study. *Brain*, 142(4):952–965.

Rogalsky, C., Poppa, T., Chen, K.-H., Anderson, S. W., Damasio, H., Love, T., and Hickok, G. (2015). Speech repetition as a window on the neurobiology of auditory–motor integration for speech: A voxel-based lesion symptom mapping study. *Neuropsychologia*, 71:18–27.

Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixomics: An r package for 'omics feature selection and multiple data integration. *PLoS computational biology*, 13(11):e1005752.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., et al. (2008). Ventral and dorsal pathways for language. *Proceedings of the national academy of Sciences*, pages pnas–0805234105.

Smith, S. M., Johansen-Berg, H., Jenkinson, M., Rueckert, D., Nichols, T. E., Miller, K. L., Robson, M. D., Jones, D. K., Klein, J. C., Bartsch, A. J., et al. (2007). Acquisition and voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics. *Nature protocols*, 2(3):499.

Takaya, S., Kuperberg, G., Liu, H., Greve, D., Makris, N., and Stufflebeam, S. M. (2015). Asymmetric projections of the arcuate fasciculus to the temporal cortex underlie lateralized language function in the human brain. *Frontiers in neuroanatomy*, 9:119.

Thomas Yeo, B., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106(3):1125–1165.

Tremblay, P. and Dick, A. S. (2016). Broca and wernicke are dead, or moving past the classic model of language neurobiology. *Brain and language*, 162:60–71.

Wada, J. and Rasmussen, T. (1960). Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance: experimental and clinical observations. *Journal of neurosurgery*, 17(2):266–282.

Weiller, C., Bormann, T., Saur, D., Musso, M., and Rijntjes, M. (2011). How the ventral pathway got lost–and what its recovery might mean. *Brain and language*, 118(1-2):29–39.

Wilson, S. M., Galantucci, S., Tartaglia, M. C., and Gorno-Tempini, M. L. (2012). The neural basis of syntactic deficits in primary progressive aphasia. *Brain and language*, 122(3):190–198.

Zhang, H., Yushkevich, P. A., Alexander, D. C., and Gee, J. C. (2006). Deformable registration of diffusion tensor mr images with explicit orientation optimization. *Medical image analysis*, 10(5):764–785.