UNIVERSITY OF CALIFORNIA

Santa Barbara

Expeditions in Neurocartography:

Mappings between Structural and Functional Pathways

in Artificial and Cognitive Neural Systems

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy

in

Physics

by

Ann M. Hermundstad

Committee in charge:

Professor Jean Carlson, Committee Chair

Professor James Langer

Professor Elizabeth Gwinn

Professor Scott Grafton

September 2012

The dissertation of Ann M. Hermundstad is approved:

Professor James Langer

Professor Elizabeth Gwinn

Professor Scott Grafton

Professor Jean Carlson, Committee Chair

Date

August 2012

Expeditions in Neurocartography:

Mappings between Structural and Functional Pathways

in Artificial and Cognitive Neural Systems

For Mama and Daddy

Acknowledgments

I am deeply grateful for the support of so many people:

For my awesome advisor, Jean Carlson. You have continually challenged me to strive for more, to push myself beyond what I thought I could do. Through your support, I have been able to pursue a broader range of research, education, and outreach endeavors than I thought possible. Your mentorship has been invaluable, and it has been an incredible experience to work with and learn from you.

For the members of the Complex Systems Group, past and present: Kevin, Dani, Sean, Eric, Nada, Lisa, Ahmed, Charles, and Kimberly. You are such an interactive and creative group, and your feedback has shaped this work. Thank you in particular to Lisa, Eric, and Sean for initiating me into the group and to Kevin, Dani, and Eric for being insightful mentors and collaborators.

For my mentors, collaborators, and committee members. For Jim Langer and Scott Grafton, who have continually challenged me to think broadly and fundamentally. Jim, thank you for providing deep insight into the "big" questions. Scott, thank you for fueling my curiosity in the fascinating field of neuroscience research. For Ralph Archuleta, who has provided intellectual and academic support in Earth Science, and for Boris Shraiman, who has provided similar support in Biophysics (and who I am holding fully responsible for any dancing that I may have done across the stage at graduation). For Elizabeth Gwinn, who provided very helpful feedback as a committee member and who is actively involved in education and outreach efforts, and for the many people additionally involved in

outreach, including Deborah Fygenson, Crystal Martin, Andrew Cleland, Wendy Ibsen, Fiona Goodchild, and Arica Lubin. For Omer Blaes, who, in addition to teaching an amazing quantum mechanics course, has fostered a supportive environment for graduate students as the Department Chair, and for Don Marolf, who has provided similar support as the Graduate Student Advisor. And for Everett Lipman, who was kind enough to show me around Santa Barbara when I first visited, and who has since been a wonderful mentor and friend.

For my professors at the Colorado School of Mines who encouraged me to seek unifying descriptions of the world: Nathan Palmer, Ed Cecil, Lincoln Carr, and Toni Lefton. Nathan, in addition to your continued support and friendship, thank you for encouraging me to reconsider the subject that I had previously hated. Ed, you taught me that one can be both a brilliant scientist and a compassionate philanthropist (and still have time for the whimsical things in life). Lincoln, thank you for guiding me through the challenges and successes of theoretical physics research. Toni, you inspired me to find poetry in the everyday–I still cannot sit through a scientific talk without scribbling poems in the margins of my notes.

For my wonderful officemates, past and present: Sean (best funny-story-teller, music- and video-sharer, and googly-eye tagger of inanimate objects), Eric (infectious laugher, sneezer, and delicious beer-maker), Cathy (amazing artist, thought-provoker, and possessor of the best understated sense of humor), Idse (stern basketball strategist, adventure-planner, and proud wearer of orange), and Jim (fro-yo enthusiast and lethargy-induced-

death evader). You have kept me laughing. And for my near-officemates, Kevin Brown (with a wealth of wisdom both scientific and non) and Dani Bassett (with an amazing ability to take on the world without batting an eye), who have spent countless hours working with me on tough problems.

For my fellow classmates, who have shared many a theme party and coffee-fueled all-nighter: Dustin, Cathy, Devin, Tomas, Idse, Kevin, Jason, Chris, Amanda, Hyejin, Jenna, James, Nick, Greg, Jess, Bob, Andre, Asher, Mauricio, Jonathan, and Michael. Thank you for unforgettable memories: piano man, caramel apples, the treacherous roads to Solvang, and many, many more.

For close friends near and far, who have been unconditionally supportive, and to whom I owe immense gratitude. From California: Nada, for wisdom and humor. Jake, for charm. Shiva, for the poignancy of imagery. Dustin, for sarcasm. Devin, for adventures. Tomas, for dancing. Kevin, for the best interruptions. Curtis, for philosophical questions. Kurt, for calm in the face of storms. Ariane, for elegance. And Dan, for endless curiosity (and for oh so many dinosaurs). From Colorado: Dimitri, for inspiration. Erik, for craftiness and singing. John, for honest conversations. David, for open horizons. Mike, for life's soundtracks. Riqui, Chase, and Rachel, for creativity and spontaneity. Jenn, for being my partner in crime. Cathy (better known as C-dub, C-monster, C-weed, C-spot-run...), for being my other half.

For my family. My parents have been the ultimate mentors: thank you for believing in me, for challenging me to be the best version of myself. I would not be here without

your unwavering support. Thank you for letting me make disastrous art projects in our living room until I realized that science was (also) fun. Daddy, you teach me that what is right is not always easy (especially when it comes to good trellis-building). Mama, you teach me to sing my way through life (so as not to get lost in grocery stores). Lulubird, my beautiful and inspiring sister, you teach me strength. I'm sorry for throwing so many of your stuffed animals down so many flights of stairs on so many different occasions.

For my grandparents, who inspire me to find passion and compassion in every day. Granny, thank you for teaching me youthfulness. Grandpa, thank you for teaching me how to ask good questions (and for reminding me that I have far surpassed the age when great scientific discoveries are made). Grandma, my kindred spirit, thank you for teaching my to find my bliss (and for keeping Grandpa in check). I am still working on teleportation and the slowing of time, and I should have more energy to focus on those endeavors now that this thesis is (nearly) finished. We will be together in no time at all.

With all my love and appreciation,

Anna, Annski, Annabanana, A-bomb-diggity, Annathan, Annie, Anita,

and finally (after judicious deliberation by my committee), Dr. Ann

August 2012

Ann M. Hermundstad

CONTACT INFORMATION

Complex Systems Group
Department of Physics
University of California
Santa Barbara, CA
93106-9530 USA

Office: 805.893.4544
Email: ann@physics.ucsb.edu
Web: www.physics.ucsb.edu/∼complex

RESEARCH INTERESTS

My research interests lie in understanding relationships between structure and function in multiscale models of interconnected, nonequilibrium systems. Using both theoretical and computational techniques, I investigate how microscale features interact and enable macroscale processes, specifically identifying tradeoffs that give rise to robust and fragile system-wide behavior. In recent work, I have applied these approaches to the study of plasticity in neural network models and functional connectivity in the human brain.

EDUCATION

University of California, Santa Barbara, California, USA

Ph.D. in Physics, August 2012
Advisor: Jean Carlson
Area of specialization: Condensed Matter Theory
Thesis title: *Expeditions in Neurocartography: Mappings between Structural and Functional Pathways in Articial and Cognitive Neural Systems*

M.A. in Physics, June 2010
Advancement topic: *Learning, memory, and the role of neural network architecture*
Committee: Jean Carlson (advisor), James Langer, Elizabeth Gwinn, Scott Grafton

Colorado School of Mines, Golden, Colorado USA

B.S. in Engineering Physics, Minor in Mathematics, May 2006
Advisor: Lincoln Carr

Research Experience

University of California, Santa Barbara, California, USA

2009-present *Graduate Student Researcher*
Collaborators: Jean Carlson (advisor), Kevin Brown, Danielle Bassett, Scott Grafton
Projects: Network analysis of structural and functional connectivity in the human brain, complexity and tradeoffs in neural network models, influence of architecture on network performance, robustness and plasticity in learning and memory processes, applications to biological and computer sciences.

2009-2010 *Graduate Student Researcher*
Collaborators: Jean Carlson (advisor), James Langer, Ralph Archuleta
Project: Friction and energy dissipation in sheared granular materials, applications to earthquake physics.

2008 *Graduate Student Researcher*
Collaborators: Joan-Emma Shea, Giovanni Bellesia
Project: Molecular dynamics simulations of polypeptides, folding of frustrated proteins, effect of surface patterning on folding.

Colorado School of Mines, Golden, CO USA

2005-2006 *Undergraduate Student Researcher*
Collaborators: Lincoln Carr (advisor), Dimitri Dounas-Frazer
Project: Quantum tunneling of Bose-Einstein condensates, dynamic evolution of an interacting condensate confined to a double well potential.

Publications

Hermundstad, A.M., Brown, K.S., Bassett, D.S., Aminoff, E.M., Clewett, D., Frithsen, A., Freeman, S., Johnson, A., Tipper, C., Miller, M.B., Grafton, S.T., and J.M. Carlson (2012), Structurally-constrained relationships between cognitive states in the human brain, *in preparation*.

Hermundstad, A.M., Bassett, D.M., Brown, K.S., Aminoff, E.M., Clewett, D., Frithsen, A., Freeman, S., Johnson, A., Tipper, C., Miller, M.B., Grafton, S.T., and J.M. Carlson (2012), Structural foundations of resting-state and task-based neural activity in the human brain, *submitted*.

Hermundstad, A.M., Brown, K.S, Bassett, D.S. and J.M. Carlson (2011), Learning, memory, and the role of neural network architecture, *PLoS Comput. Biol.*, 7(6):e1002063.

Hermundstad, A.M., Daub, E.G., and J.M. Carlson (2010), Energetics of strain localization in a model of seismic slip, *J. Geophys. Res.*, 115:B06320-1–9.

Dounas-Frazer, D.R., Hermundstad, A.M., and L.D. Carr (2007), Ultracold bosons in a tilted multilevel doublewell potential, *Phys. Rev. Lett.*, 99(20):200402-1–4.

Conference Proceedings

Hermundstad, A.M., Brown, K.S, Bassett, D.S. and J.M. Carlson (2011), Structural Drivers of Function in Information Processing Networks, *Proceedings of the Forty-Fifth Asilomar Conference on Signals, Systems, and Computers*, 837–841.

Hermundstad, A.M., Brown, K.S, Bassett, D.S. and J.M. Carlson (2011), Architectural constraints on learning and memory function, *BMC Neuroscience*, 12(Suppl. 1):P31.

Awards and Fellowships

2012	Chair's Fellowship, Physics Department, UCSB
2011	Philanthropic Education Organization (PEO) National Scholar Award
2011	Computational Neuroscience Meeting Student Poster Award
2011	Organization of Computational Neurosciences Travel Grant
2007-2010	Government Assistance in Areas of National Need (GAANN) Fellowship, UCSB
2010	Chair's Appreciation Award, Physics Department, UCSB
2009	Let's Explore Applied Physics and Mathematics (LEAPS) Fellowship, UCSB
2008	National Science Foundation (NSF) Fellowship Honorable Mention

	2006	Ryan Sayers Memorial Award for Research in Physics and Mathematics, CSM
	2006	Outstanding Graduating Senior, Physics Department, CSM
	2006	Blackwell Award for Excellence in Creative Expression, CSM
	2002-2006	Presidential Scholarship, CSM

Invited Presentations

2012	Neuroscience Seminar, Janelia Farm Research Campus.
2012	Physics of Living Matter Seminar, University of Pennsylvania.
2012	Neuronal Networks Seminar, Brandeis University.
2012	Systems Biology Seminar, UCSF.
2012	Cognition, Perception, and Cognitive Neuroscience Seminar, UCSB.
2012	Annual Founders' Day Meeting, PEO Chapters of Santa Barbara.
2012	Omidyar Seminar, Santa Fe Institute.
2011	Computational Neurobiology Seminar, Salk Institute.
2011	Biophysics Symposium, Princeton University.
2011	Network Science Seminar, Asilomar Conference on Signals, Systems, and Computers.
2011	NSF IGERT Systems Biology Seminar, UCSB.
2011	SAGE Center for the Study of the Mind Seminar, UCSB.
2011	Physics Colloquium, CSM.
2011	Theoretical Physics Seminar, CSM.
2011	Education Seminar, CSM.

Contributed Presentations

2012	UCSB Visit Day, neuroscience poster presentation.
2012	Institute for Collaborative Biotechnologies Annual Meeting, poster presentation.
2011	Organization for Computational Neurosciences (OCNS) Annual Meeting, poster presentation.

2011	Earth Science Group Seminar, UCSB.
2011	UCSB Visit Day, network science poster presentation.
2010	American Association for the Advancement of Science (AAAS) Annual Conference, poster presentation.
2009	Government Assistance in Areas of National Need (GAANN) Seminar, UCSB.
2009	Southern California Earthquake Community (SCEC) Annual Conference, poster presentation.

Teaching Experience

University of California, Santa Barbara, Santa Barbara, California USA

2010 *Instructor*, School for Scientific Thought

Subject: Quantum Mechanics

I designed and taught a quantum mechanics course for high school students. I chose the scope of the course, wrote all lecture material, prepared demonstrations, and wrote worksheets and review material.

2010 *Guest Instructor*, Summer Institute in Mathematics and Science

Subject: Introductory Mechanics

I taught introductory mechanics for incoming college freshman.

2007-2010 *Teaching Assistant*

Subjects: Classical Mechanics, Laboratory Techniques, Quantum Mechanics

I led weekly lab sections covering topics in projectile motion, wave mechanics, simple harmonic motion, acoustics, basic circuitry, special relativity, and introductory quantum mechanics. During lab and discussion, I held office hours, wrote homework and exam solutions, graded homework and lab reports, and proctored exams.

2010-present *Invited Tutor, Koegel Autism Center*

Subjects: Introductory mechanics, wave mechanics, electricity and magnetism.

2010 *Private Tutor*
Subjects: Wave mechanics, electricity and magnetism, modern physics.

Colorado School of Mines, Golden, Colorado USA

2004-2006 *Workshop Facilitator*
Subject: Introductory Mechanics
I composed a weekly worksheet of written problems covering concepts in introductory mechanics and held weekly workshops to help students complete the worksheets.

2004 *Teaching Assistant*
Subject: Introductory Mechanics
I assisted with laboratory demonstrations and written problems.

2004 *Private Tutor*
Subject: Introductory Mechanics

Service

2011-2012	Invited Panel Member, The Practice of Science Seminar, UCSB
2011	Invited Representative, Physics Department Donor Marketing Committee, UCSB
2011	Invited Tutor, National Society for Black Engineers, UCSB
2011	Invited Panel Member, Early Undergraduate Research and Knowledge Acquisition (Eureka!) Seminar, UCSB
2010-2011	Invited Representative, Physics Diversity Committee, UCSB
2009-2010	Program Development, School for Scientific Thought, UCSB
2007-2010	Volunteer, Physics Circus, UCSB
2008	Recruitment, USC Women in Physics Conference

2008	Member, Graduate Student Life Committee, UCSB
2007	Volunteer Instructor, Instituto Americano, Obrajes, Bolivia
2006	Volunteer, Habitat for Humanity, Chacala, Mexico
2005	Invited Representative, Committee to Increase Retention, CSM
2005	Mentor, Freshman Success Seminar, CSM
2003	Volunteer, Samaritan House, Denver, Colorado

References

Dr. Jean Carlson
Professor
Department of Physics
University of California
Santa Barbara, CA, 93106
(805) 893-8345
carlson@physics.ucsb.edu

Dr. Scott Grafton
Professor
Department of Psychology
University of California
Santa Barbara, CA, 93106
(805) 893-5235
grafton@psych.ucsb.edu

Dr. James Langer
Professor Emeritus
Department of Physics
University of California
Santa Barbara, CA, 93106
(805) 893-7597
langer@physics.ucsb.edu

Dr. Elizabeth Gwinn
Professor
Department of Physics
University of California
Santa Barbara, CA, 93106
(805) 893-2564
bgwinn@physics.ucsb.edu

Abstract

Expeditions in Neurocartography:

Mappings between Structural and Functional Pathways

in Artificial and Cognitive Neural Systems

by

Ann M. Hermundstad

Neural systems are inherently complex and dynamic in nature, exhibiting a vast array of functions that range from low-level cellular interactions to high-level cognitive processes. These functions are supported by anatomical interactions that span a similarly wide range of scales, from the synapses between individual neurons to the extended fiber pathways that traverse the brain. Both this structural architecture and the function that it supports are continually interacting with and adapting to the external environment. An understanding of both the capabilities and limitations of neural systems therefore requires integrative approaches for assessing interactions between structural architecture, dynamic functional activity, and environmental variability.

In this dissertation, we apply theoretical, computational, and data-driven techniques to the study of both artificial and cognitive neural systems with the goal of mapping between patterns of structural and functional connectivity across multiple scales of resolution. Theoretical and computational analyses of small artificial networks provide

insight into the limitations of different architectural motifs in facilitating the performance of low-level functions, while data-driven analyses of large-scale human brain networks provide insight the overlap of high-level functions supported by a common structural architecture.

Computational neural networks provide a powerful framework in which to systematically probe the dependence of system function on underlying architecture. We use neural networks to assess the dependence of competitive learning and memory processes on structural variations. By comparing the performance of parallel and layered network architectures during sequential tasks that require both acquisition and retention of information, we identify tradeoffs between learning and memory processes that arise from variations in underlying structure. During the task of supervised, sequential function approximation, networks produce and adapt representations of external information. Performance is evaluated by statistically analyzing the error in these representations while varying the initial network state, the structure of the external information, and the time given to learn the information. The structure of the underlying error landscape connects functional network performance to complexity in network architecture. We employ sloppy model analysis [1] of parallel and layered network landscape minima to isolate variations in the number, curvature, and eigenvector localization properties of local minima within different network landscapes. We find that these variations in landscape structure give rise to tradeoffs in performance; these include the ability of the network to maximize accuracy versus minimize inaccuracy and produce specific

versus generalizable representations of information. Parallel networks generate smooth error landscapes with deep, narrow minima, enabling them to find highly specific representations given sufficient time. While accurate, however, these representations are difficult to generalize. In contrast, layered networks generate rough error landscapes with a variety of local minima, allowing them to quickly find coarse representations. Although less accurate, these representations are easily adaptable. Importantly, the systematic analysis of performance in small neural network models provides insight into the performance of larger composite systems in which statistical analyses of performance would be intractable. Furthermore, the finding that variations in parallel versus layered network architectures give rise to measurable performance tradeoffs has implications for understanding the behavior of a wide variety of natural and artificial learning systems that share these structural features.

Given that the statistical analysis of computational network performance can inform large-scale models of neural systems, we similarly ask to what extent the data-driven study of the brain can in turn inform computational models of network function. Magnetic resonance imaging enables the noninvasive mapping of both anatomical white matter connectivity and dynamic patterns of neural activity in the human brain. We examine the relationship between the structural properties of white matter tracts (structural connectivity) and the functional properties of correlations in neural activity (functional connectivity) within 84 healthy human subjects both at rest and during the performance of attention- and memory-demanding tasks. We show that structural properties, including

the length, number, and spatial location of white matter tracts, are predictive of and can be inferred from the strength of resting-state and task-based functional interaction between brain regions. Importantly, we show that these relationships are both representative of the entire set of subjects and consistently observed within individual subjects. The observed links between structural and functional pathways in the human brain provide insight into the development of large-scale neural architecture and the functional implications of disruptions to this architecture.

This direct analysis of connectivity uncovers relationships between structural and functional interactions without invoking knowledge of the specific brain regions involved in these interactions. By incorporating this additional knowledge into our analysis, we identify structurally-mediated interactions between putative task-related functional networks that both support and distinguish between cognitive states. We build upon previous studies that have identified resting-state functional networks, denoted task-positive (TP) and task-negative (TN) networks, that are strongly anticorrelated at rest but also involve regions of the brain that routinely increase and decrease in activity during attention processes, suggesting that task-based function is encoded in resting-state activity. By identifying regions within our brain networks that have been implicated in TP and TN networks, we investigate the structural mechanisms that support a functional overlap between resting-state and task-driven activity. We show that strong interactions within and between TP and TN networks, as quantified by an increase in the relative number of anatomical connections that support strong functional correlations, distinguish

resting-state, attention-state, and memory-state brain activity. We map differences in these interactions to a phase-like space in which brain states are characterized by the relative contribution from different network interactions. We probe the features of this phase space across subjects and find sets of ordered relationships between cognitive states. This order enables us to group individual based on their similarity in phase space, and we link groupings with abnormal phase relationships to significant deviations in behavioral performance during attention and memory tasks. This suggests that further characterization of this phase space may help identify structural and functional signatures of altered cognitive states.

Together, these findings uncover robust links between structural architecture and functional activity in small-scale artificial network models and in large-scale human brain networks, which together inform and constrain intermediate-level descriptions of structural and functional interactions. The development of integrative, multiscale descriptions of neural system architecture is crucial for understanding both the capabilities and the constraints imposed by this architecture.

Contents

Chapter 1

Introduction

"There is no scientific study more vital to man than the study of his own brain. Our entire view of the universe depends on it."

–Francis H. C. Crick, from *Scientific American*, 1979

The study of the brain is considered by many to be the next scientific frontier, the great unknown. Made up of more cells than there are stars in our galaxy, exploration of the vast expanses and deep corners of the brain has been compared to the exploration of space. Just as the advent of space travel and powerful telescopes has enabled us to chart the topography of the universe, the speed with which new technological methods are being developed is drastically accelerating our understanding of the topography of the brain, and data is being collected faster than scientists can examine it. However, we are still far from understanding how the intricacies of cognitive function arise from the

cellular building blocks of the brain.

The mere attempt to study the brain raises philosophical questions about its capabilities and limitations. With a similar philosophical leaning as Francis Crick, Edward O. Wilson has said, "Overall, the human brain is the most complex object known in the universe–known, that is, to itself." Not only does our view of the universe depend on an understanding of the lens through which it is seen, our ability to understand this lens is constrained by the machinery that we use to study it–namely, the brain itself.

Yet, the desire to understand the brain is ever compelling and often stems from a deep philosophical curiosity about our own origins. What are we? And perhaps more importantly, how are we what we are? As Hippocrates once said, "Men ought to know that from the brain, and from the brain only, arise our pleasures, joy, laughter and jests, as well as our sorrows, pains, griefs, and tears." How does something so esoteric as joy arise from something so physical as a cell?

Early investigations of the brain were driven by a desire to find the seat of the soul, with the idea that consciousness must arise from a transfer of information somewhere within the body. This search begged the questions: what is this information, how is being transferred, where is it going, and why? Even today, we are still seeking answers to these fundamental questions.

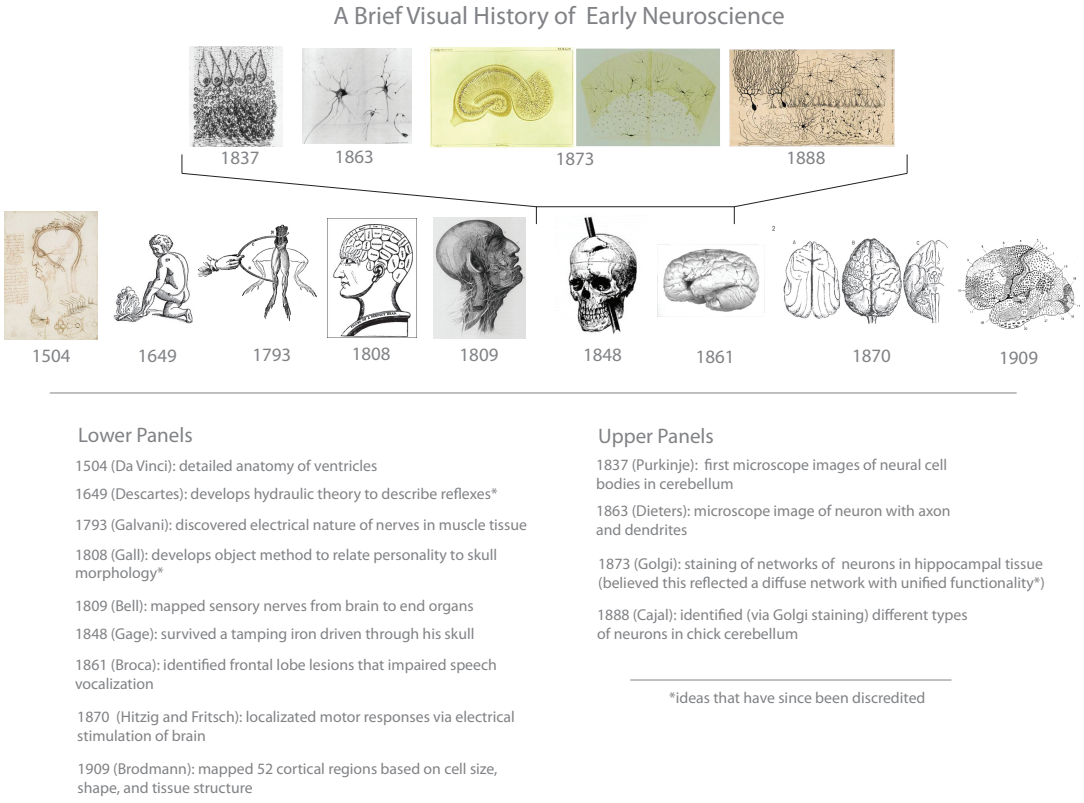

Figure 1.1: Milestones in Neuroscience Research. Lower panels describe significant contributions to the study of large-scale anatomical and functional properties of the brain, while the upper panels describe significant contributions to the study of small-scale neuronal anatomy and function.

1.1 Mapping the Nervous System

Throughout the course of history, scientists have turned to mapmaking as a means of linking the anatomical structure of the body, such as organs, tissues, and bones, to its functional capabilities, such as the movement of a limb, the propensity for good health, or the formation of a thought. In seeking the anatomical foundations of our conscious beings, many candidate organs were proposed as housing the "soul", including the heart, the pineal gland, and the ventricles of the brain. Similarly, many types of information were proposed to mediate the interactions between the soul and the body, including humors, animal spirits, and liquids. However, it wasn't until the 1700's that muscular tissue was found to have electrical properties, and it wasn't until the late 1800's that these electrical properties were linked to the brain. We now know that this electricity is the currency of exchange throughout the entire nervous system, and the brain houses the machinery responsible for controlling this exchange.

Even armed with the knowledge of electricity in the body, the nature of electrical transmission was unknown for many years. What and where were the power plants and wires responsible for the production and exchange of electricity? Experiments in the early 1800's were able to anatomically trace nerves from the brain to different end organs, such as the face or the arm, confirming that there were hard-wired electrical connections between the brain and the body. Experiments in the 1870's further showed that the electrical stimulation of the brain could elicit motor responses from different parts of the body, providing a direct link between the brain as a controller and the body

as a recipient of this control.

Concurrent with the study of these large-scale electrical connections, the invention of the compound microscope in the 1800's enabled scientists to visualize the small-scale properties of these large-scale connections. When examined at small scales, neural tissue was found to contain localized cells with extended projections. There was much debate as to the nature of these cells and their ability to function as discrete units, and many believed that they comprised a diffuse net that supported a single consciousness. It wasn't until the end of the 19th century that scientists began to prescribe to the "neuron doctrine", which stated that the discrete extended cells found in brain tissue, to be called neurons, were the basic unit of the nervous system from which all other structures were built. Information, in the form of electrical signals, traveled in one direction through the neuron, from the input projections (called dendrites) through the cell body to the output projections (called axons). We now know that neurons exhibit a range of different morphologies that enable the exchange of electrical signals between one another.

Two fundamental and related questions arose from the simultaneous studies of the large-scale properties of the brain and the small-scale properties of neurons: how is the brain built from individual neurons, and where does functionality lie? If we were to look for a memory, would we find it in a single cell, in a chunk of brain tissue, or distributed throughout the brain?

1.2 Structural Organization and Functional Localization

The study of structural organization, and therefore the function that it supports, has historically been limited in scope to specific brain regions or localized anatomical structures. The characterization of structural properties, such as the physical shape, size, and location of a cell or organ, has evolved significantly over the past several hundred years, from the first wax mold of ventricles constructed by Da Vinci in 1504 to the mapping of nerves by Bell in 1809 to the description of neuron morphology by Cajal in 1888 to the characterization of cortical cell types by Brodmann in 1909. The identification of such localized structures suggested that function was similarly localized. The difficulty remained, however, in linking specific structural signatures to their functional purpose.

Due to the invasiveness of experimental techniques, early studies of healthy brain function were limited to animals. For example, the experiment that first linked electricity in the brain to movement in the body, conducted in the late 1800's, identified localized regions within a thin strip of brain tissue that controlled involuntary muscle contractions in the legs of dogs. We now know that this mapping between the brain and the body can be extended to a wide range of motor and somatosensory responses, whereby the electric stimulation of a localized brain region can elicit a specific taste or the visualization of different colors.

In comparison, studies of human brain function were often derived from cases in which functionality was disrupted. In one notable 1848 case study, railroad worker Phineas Gage survived an explosion in which a tamping iron was driven through his skull.

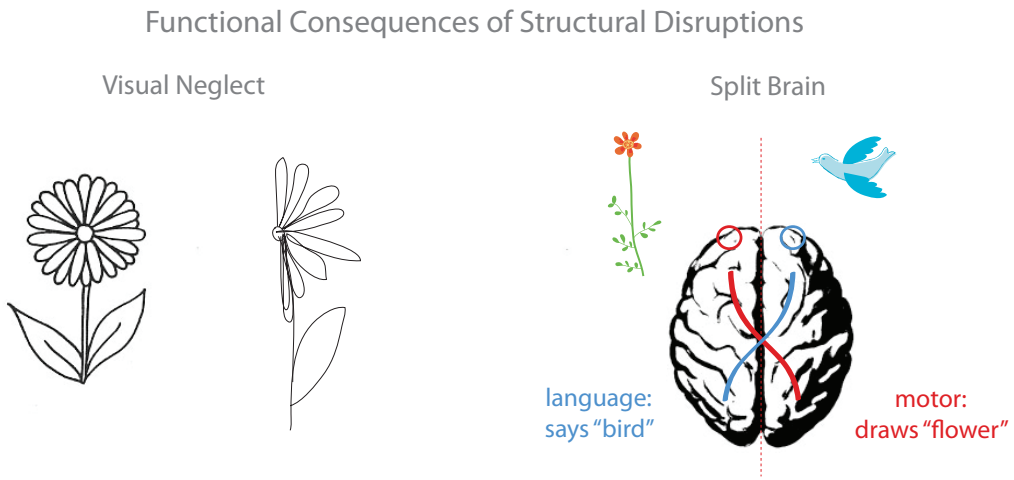

Figure 1.2: Functional Consequences of Structural Disruptions. Visual neglect (left), often the result of a lesion to the right parietal hemisphere, can cause altered or absent perception in the left visual field. Split brain syndrome, a result of a severing of the corpus callosum linking the two cerebral hemispheres, can cause miscommunication between language and motor centers. Objects in the left and right visual fields will be processed by the right and left hemisphere. As language processing is housed in the left hemisphere and right-handed motor control is housed in the right hemisphere, patients presented with two images (one each in the left and right visual fields) will say they see one image but will draw the other image.

Within minutes of the accident, Gage was remarkably able to speak and walk. However, friends reported that his demeanor was altered so severely that he was "no longer Gage." We now know that the region of the brain affected by Gage's accident is related to the control of temperament. In 1861, Paul Broca reported that stroke-induced lesions of a localized region in the left frontal lobe of the brain caused near complete loss of language vocalization, but not of language comprehension, noting that one of his patients could only speak the word "tan." Shortly after, Carl Wernicke reported that lesions to a localized region in the superior temporal gyrus altered language comprehension, observing that patients were able to vocalize words with normal rhythm but were unable to construct meaningful sentences from these words. More complex neurological phenomena that altered visuospatial perception, such as visual neglect and split brain syndrome (Figure 1.2), were similarly linked to structural disruptions between brain regions.

The study of higher-order cognitive processes, such as learning and memory, continues to pose significant challenges, even today. Many now believe that these higher-order functions arise from low-level interactions between localized sensory and motor functions, and a significant amount of current research is aimed at identifying how different brain regions interact with one another to perform coordinated functions. The question of structural and functional localization, restated in the context of current knowledge, could ask: how is diverse functionality, such as the ability to learn and remember information, encoded within single neurons, activity patterns across many neurons, or interactions between whole brain regions?

1.3 Visualizing the Topography of the Brain

A detailed understanding of neural structure and function has historically been limited by an inability to see the brain. Both animal and human dissections have, at various times and in certain cultures, been deemed unethical. If performed, they were limited to postmortem dissections, and therefore any theories of functionality were merely speculative. The invention of the microscope significantly advanced neuroscience research and solidified the cellular basis of the nervous system. Even then, neuron functionality could not be measured directly; rather, the direction of information flow through a single neuron was remarkably inferred from the observed organization of groups of neurons. Similarly, many large scale signatures of brain function were inferred by tracing the neurological effects of electric stimulation. It wasn't until the late 1800's that electrical activity was measured from the brain, rather than directed to the brain, a technique that formed the basis of modern neuroimaging.

Although we are still limited in the degree to which we can directly interact with the brain, recent advances in imaging techniques have accelerated our understanding of its structural architecture and the functionality that this architecture supports (Figure 1.3).[1] For example, electron microscopy has traced, in exquisite detail, the morphological features of and the cellular machinery packed within individual neurons. Two-photon microscopy further enables the imaging of networks of neurons within tissue up to 1mm in depth (for comparison, the human cortex is 2-4mm in thickness). Local functional

[1] The two-photon microscopy image shown in Figure 1.3 is used with permission from Thomas J. Deerinck (http://www.microscopyu.com/featuredmicroscopist/deerinck/deerinckgallery.html).

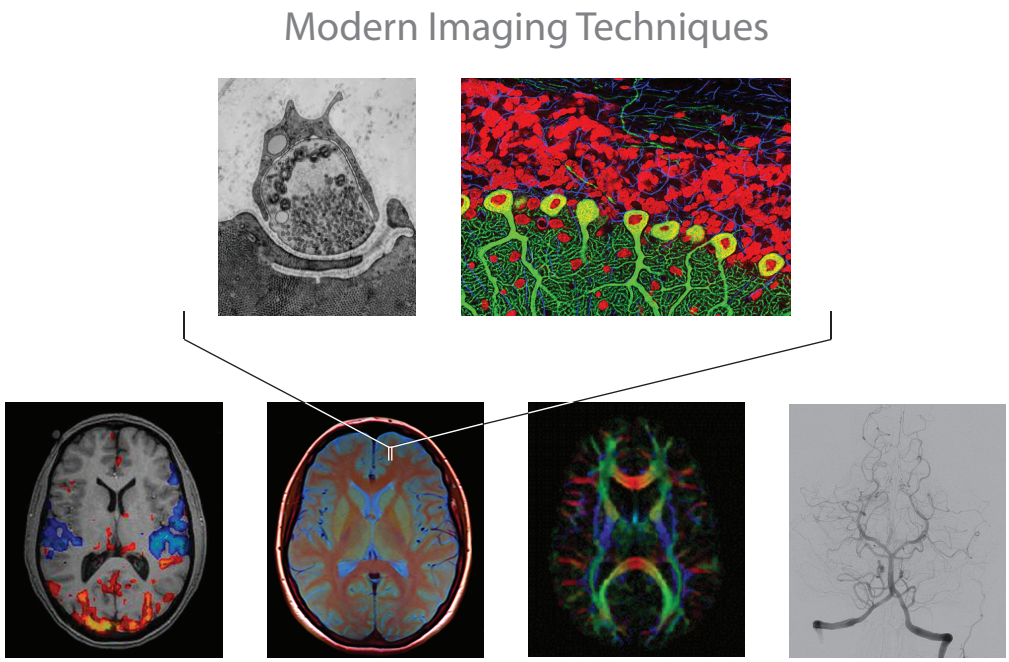

Figure 1.3: Highlights from Modern Imaging Techniques. From top (neuron scale) to bottom (whole brain scale) and from left to right: Images of synapse taken via electron microscopy, Purkinje neurons taken via two-photon microscopy, functional activity taken via fMRI, brain tissue taken via standard MRI, white matter pathways taken via DTI, and blood vessels taken via angiography.

activity can similarly be measured from microelectrode array recordings of the electrical signals generated by neurons or local neural circuits. Together, these techniques for assessing neuron-level properties are direct at the cost of being invasive, and they are therefore largely limited to animal studies.

In comparison, noninvasive neuroimaging techniques do not provide the spatial and temporal resolution of single-cell recordings, but they instead enable the indirect mapping of structural and functional properties on a large scale across the entirety of the brain (see Chapter 4 for further discussion). Magnetic resonance (MR) techniques now enable the noninvasive measurement of a wide range of structural and functional properties. Techniques such diffusion tensor imaging can be used to visualize the physical pathways mapped out by large bundles of neurons, which act as transmission cables between distant regions of the brain. MR techniques can similarly measure local changes in energy consumption, which serve as an indirect measure of slow changes in local neural activity. Techniques such as electroencephalography (EEG) can measure changes in neural activity over shorter timescales by recording the electrical signals that the brain transmits through the skull, and when medically warranted can be applied invasively to record directly from the cortical surface of the brain.

1.4 Multiscale Models of Brain Connectivity

Measurements drawn from this diverse range of modern imaging techniques have shown that the brain exhibits numerous structural motifs exhibited across a range of

spatial scales, from the complicated morphological structure of individual neurons to the large scale structural characteristics of whole brain regions. Similarly, the brain generates a wide range of dynamical responses, from the electrical signals produced by a single neuron to the fluctuations in energy consumption exhibited by a patch of brain tissue. One of the fundamental problems faced by modern neuroscience is the need to connect anatomical and functional interactions across this wide range of scales. Phrased another way, we can ask, "given a pile of 10 billion neurons, how would we build a brain?"

This question has puzzled scientists and engineers alike. Beyond interest in understanding its biological properties, it has been recognized for many years that the brain is capable of amazingly complex computations, suggesting that we should look to the brain for inspiration in designing new technologies. Such ideas sparked the neural network revolution in the 1950's, the aim of which was to design neurally-inspired machines capable of biologically complex functionality.

What are the design principles that would underlie such a machine, and are these the same principles that have guided the evolution of the human brain? Like fingerprints, no two brains are alike. Moreover, the brain is dynamic, not only in function but also in structure. Connections between brain regions are continually building and rebuilding themselves over the course of a lifetime, from the trimming of connections in early childhood to the massive regrowth of connections in adolescence to degradation of connections with age. Even the process of learning a new skill is linked to the fine tuning of structural properties. How can the same structural materials support such variable

functionality, and how in turn can stable functionality be maintained when the structure on which it rests is changing?

Many have suggested that answers to these questions will require a detailed connectome of the human brain, a map that describes the location and function of each cell. To date, such a map has only been constructed for C. elegans, a tiny worm whose functional repertoire spans a few configurations of wiggles. Even given the recent advances in imaging techniques, we are far from being able to reconstruct the structural features intermediate between neurons, local neural circuits, and whole regions of the human brain. We are similarly far from being able to reconstruct the dynamical features intermediate between local neuron signals, large-scale brain wave oscillations, and complex cognitive functions such as language and memory. Furthermore, the wealth of data that is being generated from experimental measurements of brain structure and dynamics will require parallel advances in the computational techniques needed to analyze this data.

Yet, it is exactly this integration of information across spatial and temporal scales that will help us understand how the brain operates. Much work to date has focused on structural and functional specialization. However, the belief that high-level cognitive functions are supported by interactions between many different brain regions suggests that our focus should shift toward identifying integrative approaches for understanding interactions between anatomically and functionally distinct regions. To do this requires a combination of both experimental and computational techniques to simultaneously measure the physical properties of neural systems and explore the implications of varia-

tions in these properties. Furthermore, the combined study of large-scale and small-scale neural systems can provide insight into the capabilities and constraints that different scales impose on one another, from the limited dynamic range of individual neurons to the efficiency of information transfer between whole brain regions.

1.5 Outlook

Motivated by the need to develop such integrative approaches for the study of neural systems, this thesis aims to identify mappings between specific structural features, such as the patterns of connectivity between neurons or brain regions, and the functionality that they support, such as the ability to learn, adapt to, and remember information. Not only does this provide insight into the architectural and functional design principles that constrain neural systems, it also provides insight into the implications of disruptions to this architecture.

In what follows, we investigate relationships between the structural organization and the functional capabilities of neural systems. We combine both computational and experimental techniques to isolate specific structural features that differentially impact functionality on a small scale, within computational networks of neurons, and on a large scale, between localized regions of the human brain.

In Chapters 2 and 3, we use computational models to investigate interactions between learning and memory processes in small network systems. Computational models have the advantage of enabling us to systematically vary the structure of the network, as

described by the pattern of connections between neurons, and the functionality of the network, as described by its ability to accurately perform learning and memory tasks. This enables us to identify specific structural features that support a balance between learning and memory processes. We show, in Chapter 2, that variations in connectivity patterns give rise to tradeoffs in the ability of a network to retain old information while learning new information. In Chapter 3, we investigate extensions of this work across variations in internal network structure and external environmental structure.

In Chapters 4, 5, and 6, we use neuroimaging data to assess human brain networks. We investigate relationships between large scale structural features, as measured by the properties of bundles of neurons in the brain, and functional interactions, as measured by the strength of correlated fluctuations in energy consumption. By distinguishing between neural activity measured in different cognitive states, we identify structural features that support the performance of different cognitive functions. We show that these relationships are representative of many subjects and are consistently maintained within individual subjects. In Chapter 5, we assess these structure-function relationships across variations in our analysis techniques, and we discuss methodological considerations for robustly relating structure and function across a large sample of subjects. In Chapter 6, we introduce information about the anatomical brain regions involved in these structure-function interactions. We show that structurally-mediated interactions between putative task-related brain regions both supports and distinguishes between different cognitive states. We further show that the space of interactions is constrained, such that individual

subjects exhibit a reduced subset of all possible interactions. By quantifying the features of this reduced subset, we identify regions of interaction space that may be predictive of altered behavioral performance during attention and memory tasks.

Together, the work presented in this thesis develops multimodal approaches for mapping structural and functional pathways across a wide range of scales in neural systems. These approaches, and the observed structure-function relationships gleaned from them, provide insight into the architectural design components that can support, constrain, and distinguish between different functions in a wide range of information processing systems.

Chapter 2

Learning and Memory in Neural Network Models

"The brain is a monstrous, beautiful mess. Its billions of nerve cells–called neurons–lie in a tangled web that displays cognitive powers far exceeding any of the silicon machines we have built to mimic it."

–William F. Allman, from *Apprentices of Wonder: Inside the Neural Network Revolution*, 1989

2.1 Introduction

Understanding how neuron-level interactions support higher level functionality is a outstanding problem in the field of neuroscience. Network level models provide one approach for understanding how interactions between different neural components

interact across varying levels of resolution to support complex functionality such as learning and memory.[1]

Learning, the assimilation of new information, and memory, the retention of old information, are competing processes; the first requires flexibility and the second stability in the presence of external stimuli. Varying structural complexity could uncover tradeoffs between flexibility and stability, particularly when comparing the functional performance of structurally distinct learning systems. We use neural networks as model learning systems to explore these tradeoffs in system architectures inspired by both biology and computer science, considering layered structures like those found in cortical lamina [3] and parallel structures such as those used for clustering [4], image processing [5], and forecasting [6]. We find inherent tradeoffs in network performance, most notably between acquisition versus retention of information and between the ability of the network to maximize success versus minimize failure during sequential learning and memory tasks. Identifying tradeoffs in performance that arise from complexity in architecture is crucial for understanding the relationship between structure and function in both natural and artificial learning systems.

Natural neuronal systems display a complex combination of serial and parallel [7] structural motifs which enable the performance of disparate functions [8, 9, 10, 11]. For example, layered [3] and hierarchical [12] architectures theoretically important for

[1] Portions of this chapter originally appeared in Hermundstad *et al.*, "Learning, memory, and the role of neural network architecture," *PLoS Comput. Biol.*, 2011 [2]. Reproduction is in accordance with PLoS Terms of Use (http://www.ploscompbiol.org/static/terms.action) and the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.5/).

sustained limited activity [13] have been consistently identified over a range of spatial scales in primate cortical systems [14]. Neurons themselves are organized into layers, or "lamina," and both intra-laminar [15] and inter-laminar [16] connectivity differentially impact function. Similarly, information processing systems developed by technological innovation rather than natural evolution have structures designed to match their functionality. For example, the topological complexity of very large integrated circuits scales with the function to be performed [17]. Likewise, the internal structure of artificial neural networks can be carefully constructed [18] to enable these systems to learn a variety of complex relationships. While parallel, rather than serial, structures are appealing in artificial neural networks because of their efficiency and speed, variations in structure may provide additional benefits or drawbacks during sequential tasks.

The dependence of functional performance on structural architecture can be systematically examined within the framework of neural networks, where the complexity of both the network architecture and the external information can be varied. In this study, we evaluate the representations of information produced by feedforward neural networks during supervised, sequential tasks that require both acquisition and retention of information. Our approach is quite different from studies in which large, dense networks are given an extended period of time to produce highly accurate representations of information (e.g. [19, 20]). Instead, we investigate the links between structure and function by performing a statistical analysis of the error in the representations produced by small networks during short training sessions, thereby identifying mechanisms which

underlie tradeoffs in performance. Our work therefore has important implications for understanding the behavior of larger, more complicated systems in which statistical studies of performance would be impossible.

In what follows, we provide a brief introduction to the mathematical and computational techniques used to model neurons and networks of neurons.[2] We then discuss the parallel application of artificial neural networks to the study of highly nonlinear computational problems. These two fields, biological and artificial neural network modeling, stemmed from the same realization that networks of neurons perform detailed computations that underlie complicated cognitive processes. Here, we use artificial neural networks to study the biologically-motivated problem of learning and memory interaction. This study takes advantage of the computational machinery provided by artificial neural networks while being simultaneously sensitive to their biological underpinnings and their ability to provide a simplified description of biological processes.

In relating learning and memory process, we assess the extent to which network architectures differ in their ability to both learn and retain information. Across a range of network architectures, we quantify the best, worst, and average performance achieved during sequential tasks that vary in both their duration and complexity. We consider the adaptability of these networks to variable initial states, thereby providing insight into the structure of functional error landscapes. Finally, we explore how landscape variations that arise from structural complexity lead to differences in performance.

[2] For a great reference on modeling neural systems, see Dayan and Abbott's book "Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems" [26]

2.2 Modeling Biological Neurons

Neurons are specialized cells that communicate via the exchange of electrical pulses. These pulses are transmitted throughout the brain and body to support functions ranging from a simple finger tap to the complex recollection of memories.

The extended morphological structure of a neuron consists of dendrites, an axon and axon terminals, and a cell body. The cell membrane is potentiated such that the inside of the cell rests at a lower voltage potential than does the outside of the cell, with a resting value near -70mV. This potential fluctuates in responses to changes in local ion content, such as sodium (Na+) and potassium (K+) content, regulated by ion channels in the cell membrane.

The dynamical building block of neural communication is the action potential, a voltage spike produced by a nonlinearity in chemical dynamics. When the membrane potential of a neuron rises above a certain threshold (near -50mV), chemical nonlinearities drive a rapid increase (depolarization) followed by a rapid decrease and overshoot (hyperpolarization) in the membrane potential. In undergoing this nonlinear process, the neuron is said to have "fired" an action potential. The period of hyperpolarization following the action potential prevents the neuron from firing a second time within a given window of time. The pattern of action potential spikes generated by and exchanged between neurons serves as the basic mode of communication in the brain.

Action potentials are transmitted between neurons across synapses, which can be direct electrical contacts or indirectly mediated via the exchange of chemical neurotrans-

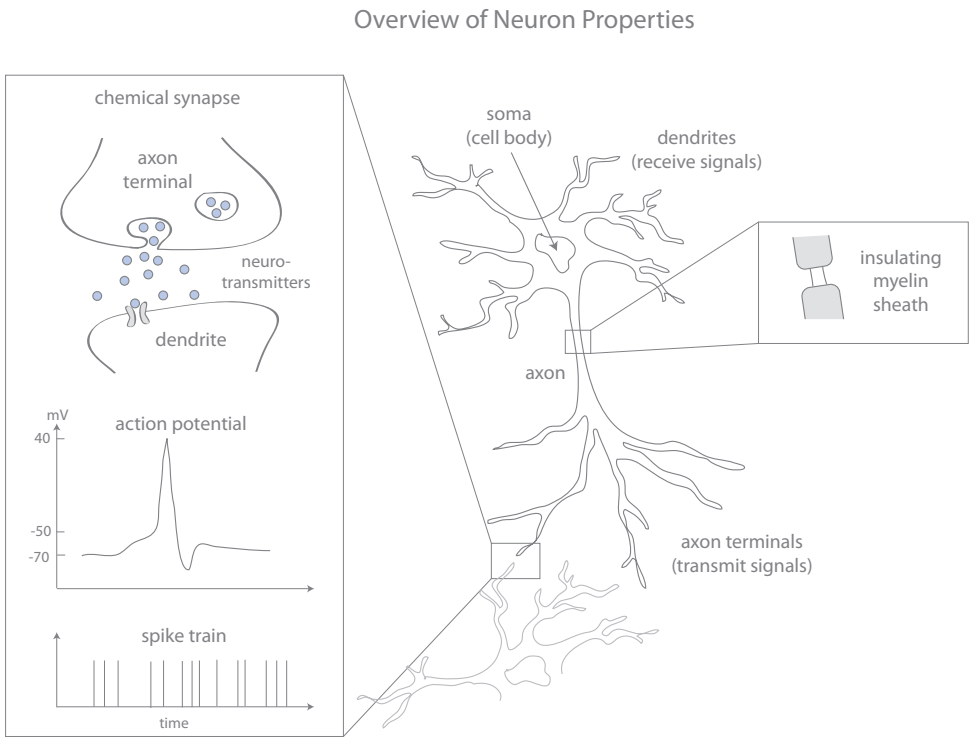

Figure 2.1: Biological Neurons. Neurons consist of a cell body (soma) that collects inputs from other neurons via extended input projections (dendrites) and sends output signals to other neurons along extended output projections (axons). A voltage difference exists across the membrane of the neuron, such that the inside of the neuron rests at a lower potential than does the surrounding medium (resting potential $V \sim -70mV$). Fluctuations in the membrane potential can induce nonlinear responses that trigger a sharp increase (depolarization) and decrease with overshoot (hyperpolarization) of the membrane potential, after which the neuron is said to have fired an action potential. Series of action potentials, called spike trains, encode the information transmitted between neurons. To enhance signal propagation, axons are coated with an insulating layer of myelin, or sheath. Signals are exchanged across axon-dendrite junctions, called synapses. These junctions can be chemical in nature, transmitting action potentials indirectly via the exchange of neurotransmitters, or they can be electrical in nature, enabling the action potential to jump between adjacent membranes.

mitters. Neurons synapse from axon terminals to dendrites, such that a signal passes from the axon terminal of one neuron to the dendrite of a second neuron.

The field of computational neuroscience concerns itself with the mathematical and computational modeling of the electrical and chemical properties of action potential generation, the signal propagation along the extended morphological structure of the neuron, and the network level dynamics produced by the exchange of these electrical and chemical processes. We first provide an overview of the methodological considerations taken for modeling individual neurons, and we return in the next section to discuss the extension of these considerations to network-level models.

2.2.1 Single Neurons

Models of individual neurons can vary significantly in their complexity to describe a range of complicated conductance and morphological properties. One of the first models to describe single neuron dynamics, constructed without a mechanistic description of action potential generation, was the integrate-and-fire model. This model employs a differential equation to describe the dynamics of the sub-threshold membrane potential in terms of a single leakage current, and it records a delta function spike whenever the potential exceeds a threshold value (after which the potential is reset to a sub-threshold value). A more biophysically resolved model was later constructed by Hodgkin and Huxley (for which they were awarded a Nobel Prize in Physiology), which describes fluctuations in the membrane potential in terms of Na+ and K+ currents (in addition

to the integrate-and-fire leakage current). With the appropriate biophysical parameter choices, this model reproduces the dynamics of action potential generation.

In addition to their complex conductance and electrical properties, neurons exhibit complex morphological structures. Single compartment models treat the extended structure of a neuron as a single unit whose properties are described by single values of membrane potential and membrane currents. Multi-compartment models, in comparison, treat the extended morphological structure of a neuron and describe the propagation, decay, and delay of signals along the length of an axon or dendrite.

Together, these models are very useful for describing the behavior of single neurons. However, a typical cortical neuron may synapse with thousands of other neurons, and local circuit activity arises from the sum of these interactions. Network level models are therefore very useful for understanding how individual neuron properties support local circuit dynamics.

2.2.2 Networks of Neurons

A typical neural network model can be constructed by wiring together the model firing neurons discussed above (such as the integrate-and-fire or Hodgkin-Huxley model neurons). While network models of this nature provide highly resolved descriptions of single-neuron dynamics, they are computationally costly. Lower resolution models can be constructed by describing the rate of action potential generation rather than the precise timing of individual action potentials. Such "firing-rate" models, which are less

computationally costly and more naturally support the inclusion of stochasticity in action potential generation, are the focus of the present work.

For tractability, network models often use single-compartment neurons (nodes) wired together via weighted synapses (connections). In a firing-rate model, the functionality of the neuron is described by a transfer function that maps the relationships between the input (received at the dendrites) and output (passed along the axon) firing rates of the neuron. The weighting of the synapse describes the strength of interaction between two neurons in terms of the likelihood that the firing rate of the presynaptic neuron will elicit a response in the postsynaptic neuron. Synapses are characterized as either excitatory or inhibitory, which describes the tendency of the postsynaptic firing rate to respectively increase or decrease with increasing presynaptic firing rates.

Network models are typically characterized as either feedforward or recurrent, which specifies the direction of information flow through the network. Feedforward network models operate as their name would suggest, with information passing through the network in one direction without recurrent or feedback projections. In this case, the output of the network is well defined, and it is straightforward to track the change activity in moving from the input to the output. Recurrent networks, in comparison, allow for the possibility of recurrent or feedback projections. Recurrent networks can produce much more complex dynamics, but their behavior is more difficult to interpret because the information flow is not linear. Furthermore, there is no straightforward definition of the network output, and one instead tracks network dynamics in response to the perturbation

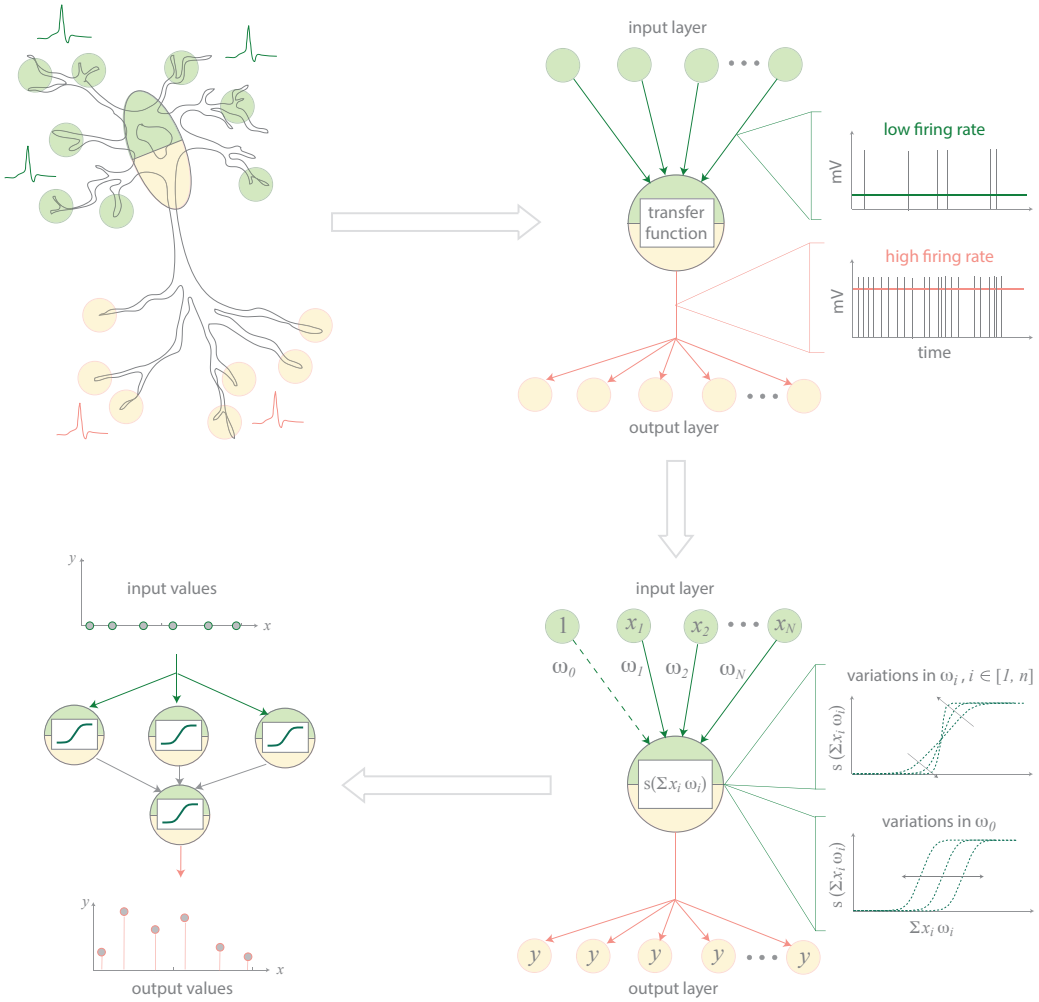

Figure 2.2: Neural Network Representation of Neurons. Neural networks models commonly represent neurons as single nodes in a network and synapses between neurons as connections between nodes. In firing-rate models, the signal passed between nodes represents a steady-state firing rate. Nodes perform a transfer function operation on the collected input signals and pass this output to other nodes in the network. The strengths of connection weights then alter the sensitivity of this transfer function (shown here as a sigmoidal function). Together, this neural network construction enables the representation of external information through the mapping of a set of input values onto a set of output values.

of local network properties.

Network models are particularly useful for studying dynamic processes such as development, learning, and memory. Synaptic plasticity, by which the efficacy of synapses changes in response to the local activity in the circuit, is believed to underlie these processes. Within a network model, synaptic plasticity is implemented via learning rules that describe the change in strength of a network connection in response to local network activity. One such type of biologically realized plasticity, Hebbian plasticity, strengthens the synapse between two neurons if the firing of one neuron is correlated with the firing of the other. Hebbian plasticity has been linked to the activity-dependent potentiation and depression (increased or decreased field potential) of neurons. Non-Hebbian plasticity, thought to be important for learning and developmental processes, alters the strength of the synapse based on either the pre- or post-synaptic firing rates, rather than on the correlation between the two.

The implementation of synaptic plasticity via learning rules enables the use of neural networks for the study of various learning processes. Within the context of neural network models, learning can be understood as the process by a which a network trains itself to produce an internal representation, stored within local dynamics or connection strengths, of some external information. Learning studies are typically classified into three types, unsupervised, reinforcement, and supervised, based on the degree of feedback that the network receives from the external environment throughout this training process. During unsupervised learning, the response of the network to input information is

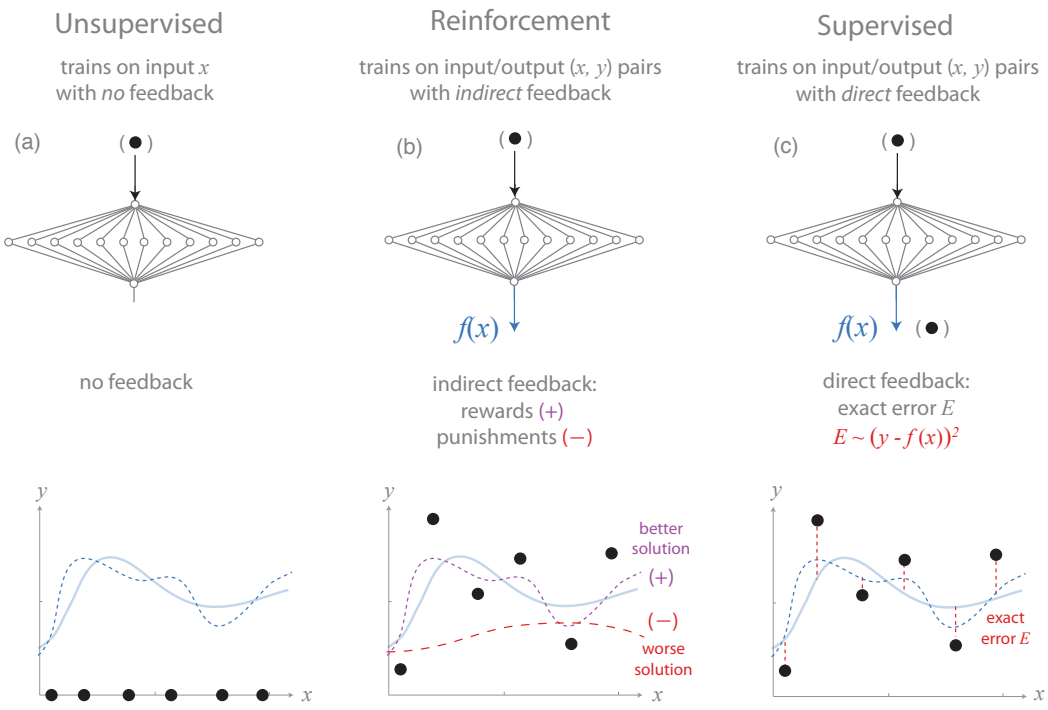

Figure 2.3: Types of Learning in Neural Network Models. Learning is typically categorized into three types, unsupervised, reinforcement, and supervised learning, depending on the degree of feedback provided by the external environment. We use use the example of one-dimensional function approximation to illustrate these three learning types. In unsupervised learning (left column), the network is trained on inputs alone, and it produces a representation of those inputs based on internal plasticity rules without feedback on the accuracy of such a representation. In reinforcement learning (middle column), the network trains on input-output pairs, and it receives indirect feedback via rewards $(+)$ and punishments $(-)$ regarding the accuracy of its representation of those input values. If the internal representation produced by the network is moving closer to the desired output (marked by the proximity of purple dotted line to the black points), the network receives a reward, and if the representation is moving away from the desired output (marked by the proximity of the red dotted line to the black points), the network received a punishment. In supervised learning (right column), the network trains on input-output pairs, and it receives direct feedback E (vertical red lines) regarding the inaccuracy of its internal representation of the input values.

controlled solely by internal network structure and dynamics, with no feedback from the environment regarding the accuracy of this response. In reinforcement learning, the network receives feedback in the form of rewards and punishments that inform the network of its relative success or failure in producing a desired input-output relationship. Finally, in supervised learning, the network receives direct feedback about the degree to which it is accurately representing a specific input-output relationship. In the second two cases, the "environment" can be an abstract teacher (as in the case of artificial neural network studies, discussed in the following section), or it can be a biologically realistic "teacher network" that provides dynamic feedback to the "pupil network" during training.

2.3 Artificial Neural Networks and the Perceptron

In addition to their biological significance, it was realized that neurons were very successful in their ability to perform computational functions, an observation that led to the creation and rapid growth of the field of artificial neural networks ([21, 22]). Artificial networks were constructed based on coarse simplifications of biological neurons, with neurons represented as nodes and synapses as connections between nodes, for the performance of complex computational tasks. As in biologically-inspired networks, neuron functionality is approximated by a transfer function that maps the input (received at the dendrites) to the output (passed along the axon) of the neuron. Connections between nodes, which can be undirected, unidirectional, bidirectional, can be modified in strength to alter the input-output relationship produced by the network.

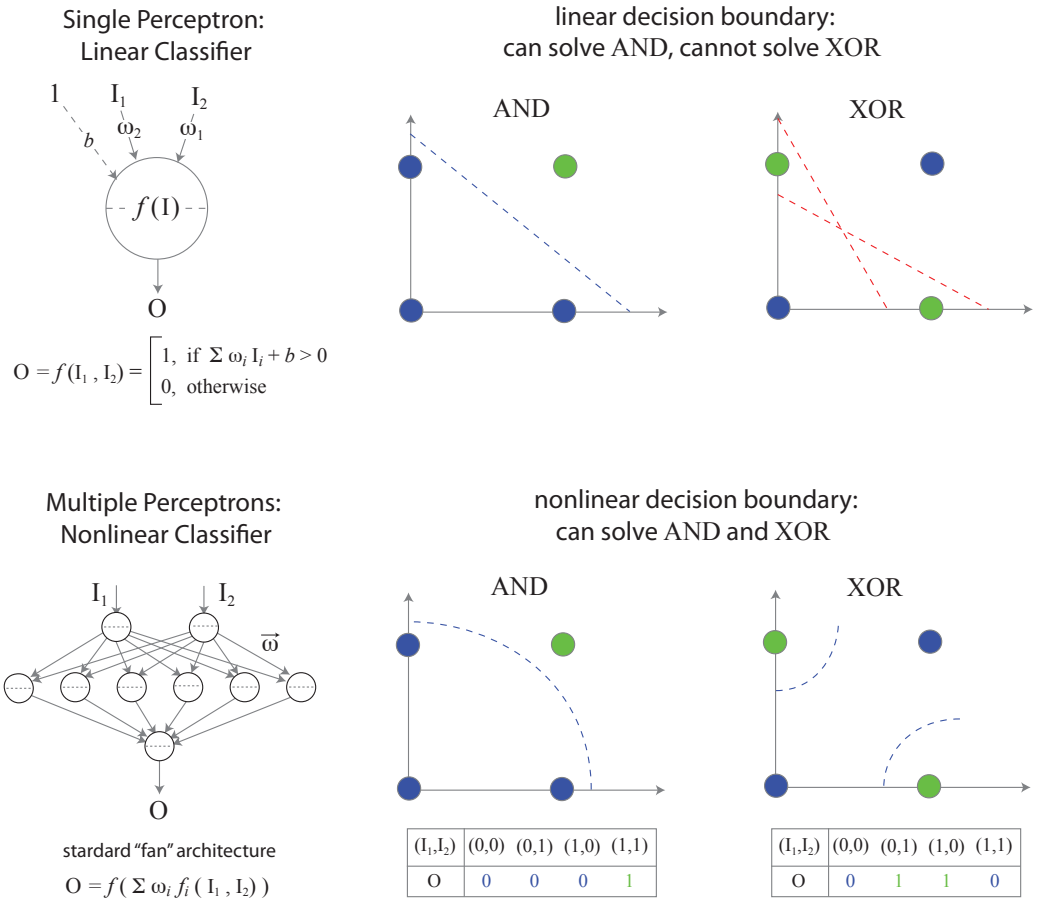

Figure 2.4: The Perceptron as a Linear Classifier. The single perceptron, one of the first and simplest examples of a feedforward neural network, is a binary linear classifier. It is able to solve the linearly-separable AND problem by finding the equation for a line that separates an output of 1(produced by an input of (1,1)) from the remaining outputs of 0 (produced by inputs of (0,0), (1,0), and (0,1)). However, the perceptron cannot solve the XOR problem, because there is no linear classification boundary that well separates an output of 0 from an output of 1. Multi-layer perceptron networks, however, can produce nonlinear classification boundaries and can therefore solve both the AND and the XOR problem (in addition to many other, highly nonlinear problems).

Although biological and artificial neural networks share many common features (given that the latter was inspired by the former), artificial neural networks are typically constructed for the performance of highly nonlinear computational problems, and they need not conform to biological constraints. Given the wide range of tunable properties, neural networks can be designed to perform very specific computational functions through controlled variations in network connectivity, the weighting of network connections, and the transfer function of individual nodes (see Figure 2.6 for examples of neural network applications).

A single-node, binary linear classifier, the perceptron, was one of the first successes in the field of artificial neural networks ([23]). Through adjustments of connection weights, the perceptron could find a linear classification boundary that separated different input-output relationships. For example, the perceptron could solve the AND problem by drawing a linear boundary that separates a output of "1" (produced by inputs (1,1)) from an output of "0" (produced by inputs (0,0), (1,0), and (0,1)). This linear classification, however, is of limited utility, and the failure of the perceptron to solve the XOR problem motivated the construction of multiple perceptron network models that could find nonlinear classification boundaries.

In the common construction of a multiple perceptron network, a single input layer fans out into one "hidden layer" and recombines into a single output layer. With minimal increase in structure compared to the single perceptron, the use of multiple perceptrons significantly increases the representational power of the network and enables the finding

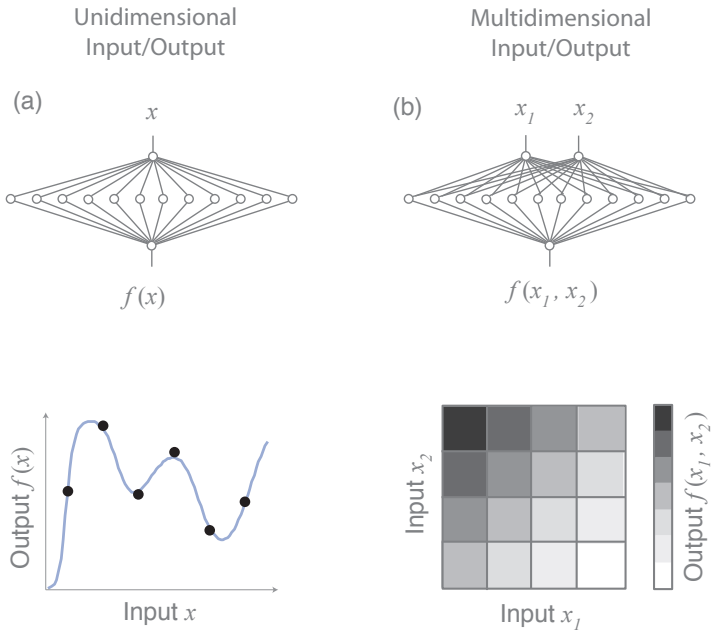

Figure 2.5: Types of Training Functions. The number of nodes in the input and output layers of neural networks correspond to the dimensionality of the input and output space of the training function. (a) A network with a single input and single output node can represent a set of (x, y) points by finding a functional approximation $f(x)$ of the 1D input space x. The input node accepts a value x (horizontal displacement), and the output node produces a value $f(x)$ that approximates the desired output y (vertical displacement). (b) In comparison, a network with two input nodes and one output node can represent a greyscale image by finding a functional approximation $f(x_1, x_2)$ of a 2D input space (x_1, x_2). In this case, the network accepts two inputs, x_1 and x_2 (pixel coordinates), and produces a single output $f(x_1, x_2)$ that approximates the desired output z (pixel intensity).

of nonlinear classification boundaries. Furthermore, the parallel one-layer construction of this fan architecture is very efficient.

The number of nodes in the input and output layers of multiple perceptron networks corresponds to the dimensionality of the problem at hand. For the XOR problem, the binary classification of (x,y) inputs is mapped onto a network with two input nodes (one each for the x and y values) and one output node (whose binary response assigns the

input values to one of two possible groupings). However, artificial neural networks are by no means limited to small input/output spaces, binary responses, or classification problems. Higher-dimensional problems, such as image processing, can be addressed by constructing networks whose input/output space matches the dimensionality of the image. For example, a network with two input nodes (one each for (x,y) coordinates of image pixels) and one output node (for the pixel intensity) could be designed to reproduce the image, where the output could be binary or continuous depending on whether the network was reproducing a binary or greyscale image. Alternatively, for an image classification problem (as is used for handwriting analysis), a network could be designed with a set of input nodes totaling the number of pixels in the image and with a set of binary output nodes totaling the number of classification groups (e.g. 26, one for each letter, in handwriting analysis).

With the proper considerations, neural networks can be tailored to fit the goals and constraints of a wide range of problems (e.g. Figure 2.6). Because of this, many years of research have been dedicated to the optimal design of neural networks for the accurate performance of very specific computational problems. In particular, given sufficient numbers of nodes, neural networks can approximate any function within a prescribed error tolerance ([24, 25]). However, given a fixed number of nodes and connections (or analogously neurons and synapses), the degree to which different network constructions vary in their performance of competitive functions is not understood. We therefore take a very different approach to the study of neural networks than is traditionally considered,

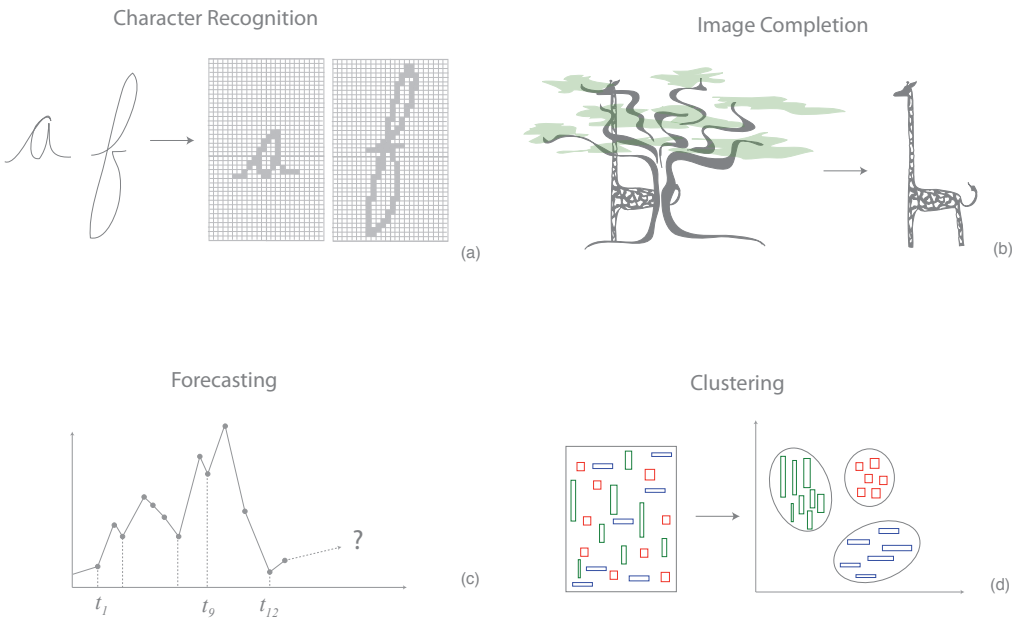

Figure 2.6: Types of Studies for Which Neural Networks are Commonly Used. Neural networks are commonly used for a variety of highly nonlinear problems, such as (a) character recognition, e.g. used by the Post Office for the automated recognition of handwriting, (b) image reconstruction, e.g. used in image processing to infer obstructed objects, (c) clustering, e.g. used to predict the buying preferences of a demographic from information such as age, gender, ethnicity, and income, and (d) forecasting, e.g. used to predict stock market trends from past history. Neural networks are advantageous for these studies because they can be designed to collect a wide range of different input features. The transfer functions of individual nodes and the connectivity between nodes can be optimized to robustly perform very specific functions.

whereby we investigate the extent to which variations in network properties impact the ability of the network to balance several different functions. We use a common and well-studied artificial neural network model, the feedforward backpropagation network trained via conjugate gradient descent, for the biologically-motivated study of sequential learning and memory processes. By varying the network topology, the information presented the network, and the training time allowed to process this information, we identify specific network properties that support and constrain a balance between the acquisition versus retention of information. This approach is crucial for isolating the effects of biological constraints, such as layered topologies, noisy environments, and limited processing time, that are not commonly addressed in traditional artificial neural network studies.

2.4 Feedforward Neural Network Model

Feedforward networks are constructed in layers of nodes, and nodes in adjacent layers are connected via variable, unidirectional weights. It is customary to refer to the first and last layers as the "input" and "output" layers, and all remaining intermediate layers as "hidden" layers. The number of nodes in each layer need not be the same across layers, nor do the patterns of connections linking nodes in adjacent layers. As discussed previously, however, the number of nodes in the input and output layers correspond respectively to the dimensionality of the input and output spaces.

We will focus our discussion on networks with a single input and single output node,

for which the network generates a one-dimensional output space as a function of the one-dimensional input. We will discuss extensions to higher dimensional problems in Chapter 3.[3]

Each node is described a transfer function $g_n(x)$ that maps its collected input onto a single output. Biological transfer functions can vary from cell to cell, many of which are nonlinear and saturate at very high or very low input values. In the analysis presented here, we choose the sigmoidal transfer function:

$$g_n(x) = s(x) = \frac{1}{(1 + e^{(-x)})}, \tag{2.1}$$

a function that is known to describe, for example, the change in field potential produced by neurons in the visual cortex [26].

The output y of a given node is a function of the weighted sum of its inputs x_p, given by $y = s(\sum_{p=1} \omega_p x_p - \theta)$. The weight ω_p of the p^{th} input connection represents the strength of the synapse from node p onto the given node. The variable threshold θ is treated as an additional weight connected to a constant input ($\theta = \omega_0 x_0$) and represents the activity-dependent firing threshold of a neuron. Representing the threshold as $\theta = \omega_0 x_0$, where $x_0 = 1$ for all nodes, allows us to organize all adjustable parameters into a single, N_p-dimensional weight vector $\boldsymbol{\omega}$. Together, these weights contain all information about the network representation of a given input.

The signals x and y passed between nodes represent steady-state firing rates of neurons,

[3] For a detailed description of FFBP neural network models, see Rojas' book "Neural Networks - A Systematic Introduction" [27]

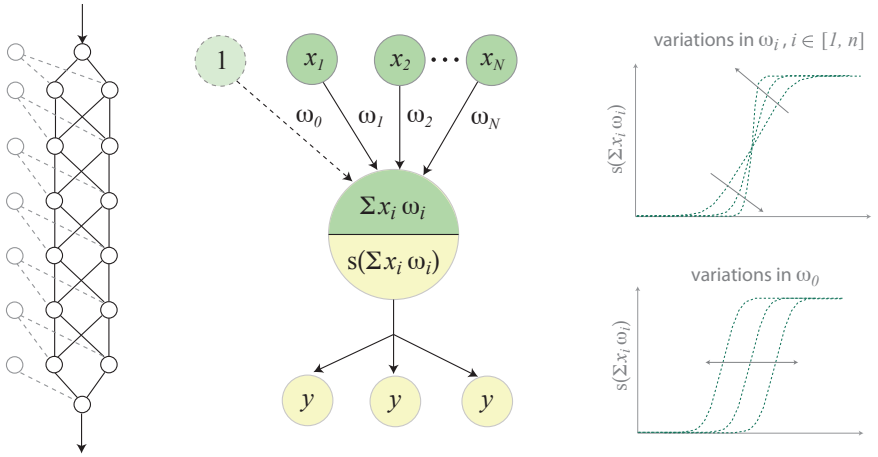

Figure 2.7: Schematic of a Feedforward Neural Network. Feedforward neural networks are designed such that signals are transmitted in one direction (by definition, the 'forward' direction) through the network, from nodes in layer ℓ to nodes in layer $\ell + 1$, with no recurrent connections between nodes in a given layer. The first layer is defined to be the 'input layer'. Connections feeding into the input layer are typically unweighted and transmit raw input values into the network. The last layer is defined to be the 'output layer'. The signal transmitted at the output layer can be compared to the desired output, and an error can be assigned to the difference between the network and desired outputs. All intermediate layers are termed 'hidden layers' because they do not have direct access to the input or output information. Hidden layer connectivity is often chosen such that all nodes in layer ℓ are connected to all nodes in layer $\ell + 1$, for which the two layers are said to be 'fully-connected'. Connections between nodes can be weighted by continuous or binary weights. Typically, hidden and output nodes (but not input nodes) receive an additional 'bias input' of value 1 connected by a variable 'bias weight', which provides a variable linear bias to the input of the given node. The transfer function of each node can be chosen to fit the specific problem at hand. Typical transfer functions include the sigmoidal transfer function used here, which performs a nonlinear 'squashing' of the input that saturates at 0 and 1 for very low and very high input values. Sigmoidal responses have been observed in certain types of neurons. The values of the input weights then serve to vary the sensitivity of the node (or neuron) by altering the nonlinearity of the sigmoid. The bias weight shifts the sigmoid to the right or left and is analogous to an activity-dependent firing threshold in a neuron.

which change iteratively during training. There is then no notion of time other than the iteration index, which represents computational rather than dynamic time. The use of temporally-varying signals is not standard for feedforward networks and introduces significant complexity into the model. We therefore focus on the analysis of steady-state networks. In the subsequent sections, we discuss applications of feedforward networks to sequential learning, for which time is represented in the sequence of training sessions.

2.4.1 Training Methods

Supervised training of a network is accomplished via learning rules that dictate how local or global variables, such as the connection weight $\boldsymbol{\omega}$, are adjusted in order to produce a desired input-output relationship. For the task of one-dimensional function approximation, the network is presented with a training pattern of N_d pairs of input x_d and target y_d values, denoted $(\boldsymbol{x}, \boldsymbol{y})$. We restrict the input x space to the range $(0, 1)$, and the sigmoid transfer function restricts the output y space to the range $(0, 1)$. The set of variable weights $\boldsymbol{\omega}$ is iteratively updated in order to minimize the output error $E(\boldsymbol{\omega})$. We use offline training for which $E(\boldsymbol{\omega})$ is the sum of squared errors between the network output $y(\boldsymbol{\omega})$ and target output y calculated after all N_d points are presented to the network:

$$E(\boldsymbol{\omega}) = \frac{1}{2} \sum_{d=1}^{N_d} \left(y_d(\boldsymbol{\omega}) - y_d \right)^2 . \tag{2.2}$$

This error gives us a direct measure of the success of the network in approximating

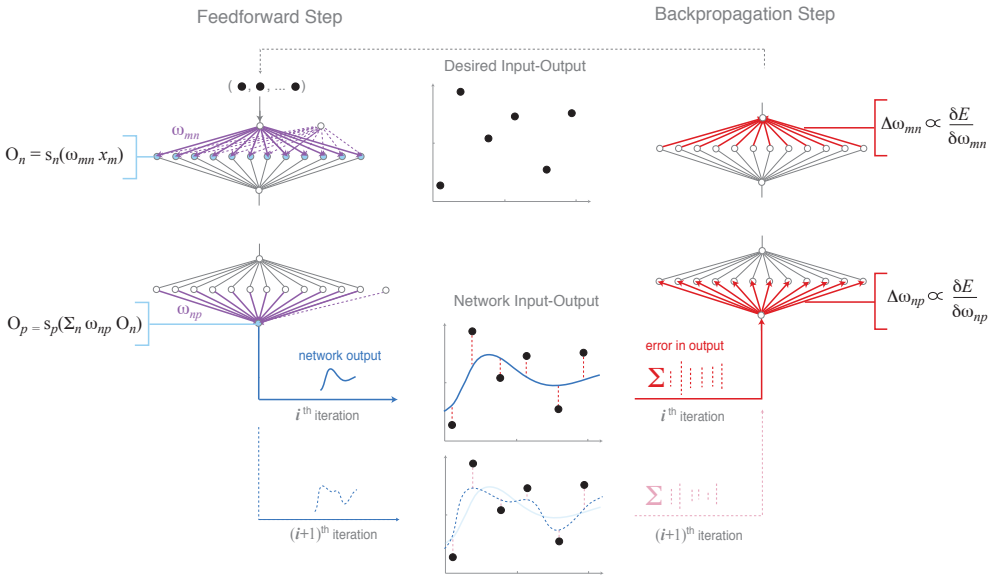

Figure 2.8: Illustration of the Feedforward Backpropagation Algorithm. Consider a single layer network training on a set of (x, y) training points (black circles). In the feedforward sweep (left column), the x values of these points are fed via the input node into the single hidden layer of the network. The output O_n of nodes n in the hidden layer is computed as a function of weighted inputs $\omega_{mn}x_m$. The values O_n are then collected at the output node, which returns a value O_p as a function of summed inputs $\sum \omega_{np}O_n$. This output, when computed across the full range of input values, yields a functional representation of the training points (solid blue line). The error E is computed as the least squares deviation of the network output from the desired output (training points). In the backpropagation sweep (right column), the error is fed backwards through the network to compute the gradients $\delta E/\delta\omega_{np}$ and $\delta E/\delta\omega_{mn}$, which are then used to update the weights ω_{np} and ω_{mn}. The process of updating the network weights results in a more accurate functional representation (dotted blue line) with a correspondingly lower error. This process, one feedforward and one backpropagation sweep, is then repeated iteratively until the error falls below a given threshold value.

the input and thereby allows us to "supervise" the network during training.

As $E(\boldsymbol{\omega})$ is an N_p-dimensional function of the connections weights $\boldsymbol{\omega}$, the process of updating $\boldsymbol{\omega}$ to minimize error can be understood as searching an N_p-dimensional error landscape for a global error minimum (see [28]). Standard gradient descent algorithms search this landscape by updating in the negative gradient direction $-\boldsymbol{g} = -\partial E(\boldsymbol{\omega})/\partial\boldsymbol{\omega}$, and such methods are extremely useful for navigating symmetric basins. We use a variation of this, the conjugate gradient descent method, that is more efficient for searching asymmetric basins [29] (see Figure 2.9). By this method, the weights $\boldsymbol{\omega}_k$ at iteration k are updated according to the rule:

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k + \lambda_{k+1}\boldsymbol{u}_{k+1}, \tag{2.3}$$

where λ_{k+1} is an adaptive step size. The conjugate gradient direction $\boldsymbol{u}_{k+1}$ combines information from the gradient direction $-\boldsymbol{g}_k$ and the step direction $\boldsymbol{u}_k$ from the previous iteration:

$$\boldsymbol{u}_{k+1} = -\boldsymbol{g}_k + \beta_{k+1,k}\boldsymbol{u}_k, \tag{2.4}$$

where $\boldsymbol{u}_0 = -\boldsymbol{g}_0$. We choose the Polak-Ribiere update $\beta_{k+1,k}$, given by:

$$\beta_{k+1} = \frac{\boldsymbol{g}_{k+1}^T \cdot (\boldsymbol{g}_{k+1}^T - \boldsymbol{g}_{k+1})}{\boldsymbol{g}_k^T \cdot \boldsymbol{g}_k}. \tag{2.5}$$

The steps of the feedforward backpropagation algorithm are illustrated schematically in Figure 2.8. Additional details regarding the conjugate gradient method and the Polak-

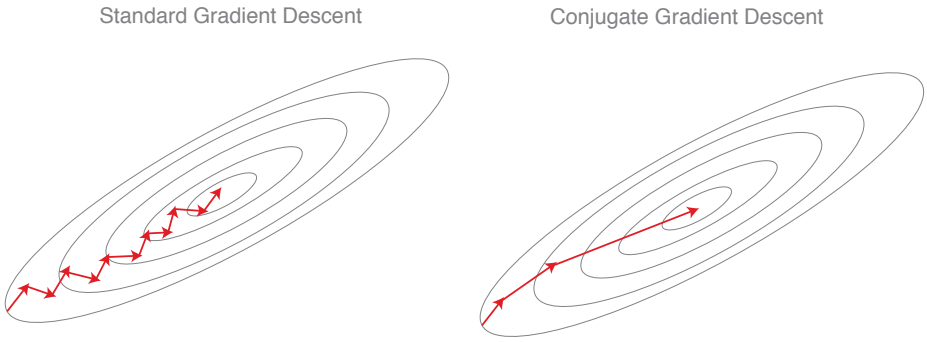

Figure 2.9: Advantage of Conjugate versus Direct Gradient Descent Approaches. In standard gradient descent methods (left), the navigation of long, narrow minima can require many successive steps. In comparison, the conjugate gradient method more efficiently enables networks to find minima through fewer numbers of larger steps.

Ribiere update can be found in references [29], [30], and [31]. The update process repeats for a finite number of iterations or until the error plateaus or falls below a desired threshold, thereby marking the end of a training session.

2.4.2 Catastrophic Forgetting and the Role of Rehearsal

Standard neural network studies optimize training on a single dataset. However, to study the integration of new and old information, as is needed to study the interaction between learning and memory processing, information must be presented to the network sequentially. This is often accomplished by coupling several training sessions, in which the network fully trains on a single pattern before being presented with a new pattern. The set of weights produced during the first training session forms the starting point for the second training session and retains the memory of the original pattern.

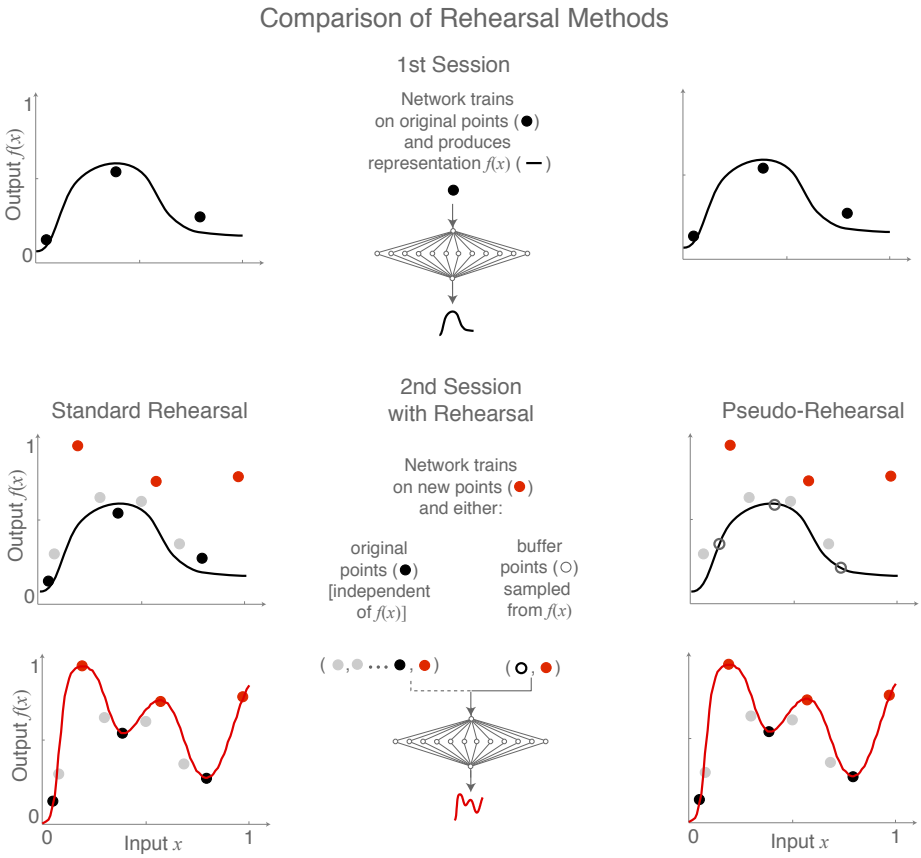

Figure 2.10: Illustration of the Pseudorehearsal Method. We compare sequential training of networks using standard rehearsal (left column) and pseudorehearsal (right column). During the first training session, the network learns a representation of a set of original points (black filled circles). To retain this representation during subsequent training sessions, standard rehearsal requires that the network retrain with both the original and the new points (red filled circles). However, as the number of training session increases, this method requires that the network retain all previous training points (grey filled circles), which is both biologically unrealistic (as it requires the network to retain a exact history of the information that it has seen) and computationally expensive (as the number of previous point grows with the number of sequential training sessions). The biologically-motivated pseudorehearsal method addresses both of these problems. With this method, the network preserves information from prior training sessions by retraining with a set of points sampled from its own internal representation (black open circles). This method relies solely on the network representation of all previous training points, rather than the points themselves, and therefore the computational expense does not grow with the number of training sessions.

Sequential training in this manner, however, leads to the problem of "catastrophic forgetting," in which a network learns new information at the cost of forgetting old information. This is a longstanding problem in the sequential training of neural networks and has been addressed with several types of rehearsal methods [32, 33, 34]. Standard rehearsal involves training the network with both the original and new information during sequential training sessions. However, this requires that the network have access to all previous training patterns, a requirement that is both computationally expensive and biologically unrealistic. We use a more biologically motivated approach, the pseudorehearsal method [35], in which the network trains with a *representation* of the original information. Pseudorehearsal has been shown to prevent catastrophic forgetting in both feedforward and recurrent networks and does not require extensive storage of examples [35, 36] (see Figure 2.10 for a comparison of standard versus pseudo rehearsal methods).

Consider two coupled training sessions, denoted 1 and 2, in which the network is presented with two different sets of points, (x_1, y_1) and (x_2, y_2). Prior to the first training sessions, network weights are randomly chosen. Following the first training session, the set of weights ω_1 generated by the network forms a functional representation of the original points (x_1, y_1). To integrate information about the second set of points, (x_2, y_2), the network must retrain on this set of points. Rather than randomizing the network weights, however, the network begins in the configuration specified by ω_1. Retraining on (x_2, y_2) alone leads to catastrophic forgetting. Retraining on the full set of points $\{(x_1, y_1), (x_2, y_2)\}$, a standard rehearsal method, alleviates this problem but requires

that the network have access to both (x_1, y_1) and (x_2, y_2). In the case of two coupled training sessions, this may not be problematic, but as the number of sessions increases, the feasibility of storing all previously encountered patterns becomes increasingly problematic. Assume, instead, that the network samples its own representation by selecting a set of (x_b, y_b) pairs that are produced by ω_1. These points are distinct from the training points (x_1, y_1), but they contain information about the network approximation of (x_1, y_1). Retraining with the set of points $\{(x_b, y_b), (x_2, y_2)\}$ similarly alleviates catastrophic forgetting, but it does not rely on the original pattern (x_1, y_1). Rather, it relies only upon the representation stored in the current state of the network (in this case, ω_1). As the number of training sessions increases, the size of (x_b, y_b) remains constant, while the size of (x_1, y_1) grows linearly with the number of sessions.

2.4.3 Task Implementation

To consider variations in network performance across a range of network structures, we consider the five distinct architectures shown in Figure 2.11. Each network is composed of N_n hidden nodes arranged into h layers of ℓ nodes per layer. The parallel "fan" and layered "stacked" networks are both fully connected and have the same total number of connections. The connectivities of the "intermediate" networks, which have slightly greater numbers of connections than the fan and stacked networks, were chosen in order to roughly maintain the same total number of adjustable parameters per network, N_p, noted in Figure 2.11.

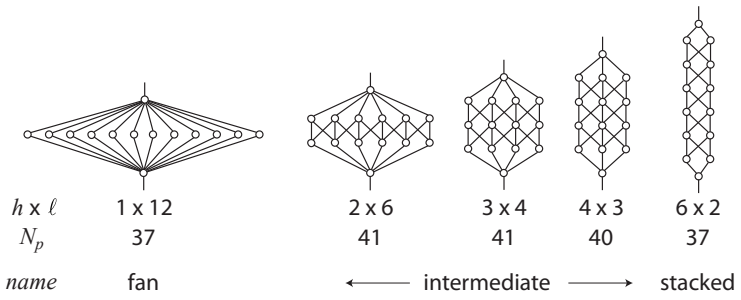

Figure 2.11: Network Architectures Considered in this Study. We consider five network architectures that vary in the ratio of their height h to their width ℓ, which correspond respectively to the number of hidden layers per network and the number of nodes per hidden layer. Indicated below each network is the number N_p of parameters (given by the total number of connection and bias weights) and the name by which we refer to the network. The parallel "fan" and layers "stacked" networks are fully connected. The connectivity of the intermediate networks, which are not full connected, is chosen to roughly maintain a fixed total number of parameters across networks.

We train each network over two sequential sessions, using identical training methods across all networks. The steps of the sequential training process are shown schematically in Figure 2.12 and are described below, where $U(a, b)$ denotes a continuous uniform probability distribution over the interval (a, b):

First Training Session

Step 1.1 - Initialize: Network weights are randomly chosen from $U(-5, 5)$. We refer to this state of the network as the "randomly initialized state".

Step 1.2 - Train: The network trains on six "original" points $(\boldsymbol{x}^{(o)}, \boldsymbol{y}^{(o)})$ whose values remain fixed for all simulations. The original points are chosen to be evenly spaced in x ($\boldsymbol{x}^{(o)} = (.1, .26, .42, .58, .74, .9)$) and random in y ($\boldsymbol{y}^{(o)} = (.55, .92, .53, .78, .33, .49)$). Similar behavior is observed for different choices, including permutations, of the specific

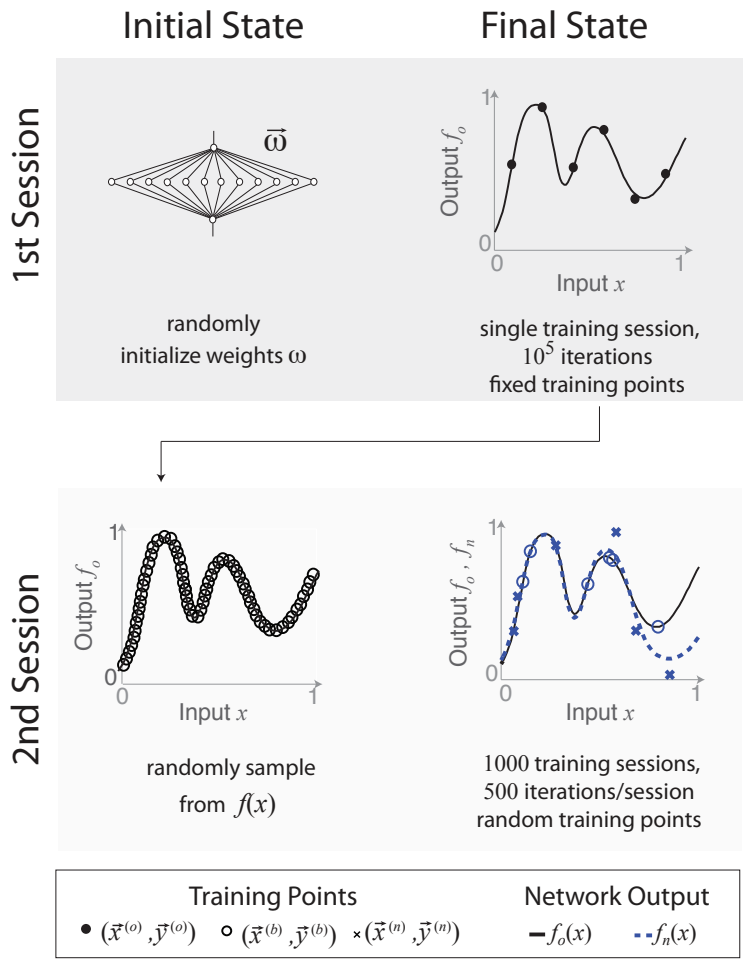

Figure 2.12: Sequential Learning and Memory Task. Illustration of the sequential learning task described in the text applied to the fan network. Network weight are randomly initialized prior to the first training session. During the first training session, the network is given an extended training session of 10^5 iterations to generate a representation $f_o(x)$ of six randomly chosen $(x^{(o)}, y^{(o)})$ points. Prior to the second training session, the network randomly samples six buffer points $(x^{(b)}, y^{(b)})$ from f_o. During the second training session, the network is given a short training session of 500 iterations to produce a representation $f_n(x)$ of both new $(x^{(n)}, y^{(n)})$ and buffer points. The second session is repeated 1000 times to generate a statistical distribution of network representations.

values used here (see Chapter 3). The original points represent the information we wish the network to remember during subsequent training. The network is given 10^5 iterations to generate a functional representation f_o of $(\boldsymbol{x}^{(o)}, \boldsymbol{y}^{(o)})$ (see upper right panel in Figure 2.12), and training ceases if the error plateaus ($\Delta E < 10^{-5}$ for 1000 iterations). We refer to this situation as allowing "unlimited" training time because in practice, the network finds a solution before reaching the maximum number of iterations.

Second Training Session

Step 2.1 - Sample: The set of weights that produce f_o forms the starting point for the second training session. We refer to this state of the network as the "sampled state" in order to distinguish it from the randomly initialized state chosen prior to the first training session. In this state, the network randomly samples a pool of 1000 buffer points $(x^{(b)}, y^{(b)})$ from f_o (see lower left panel in Figure 2.12.) This is accomplished by (*i*) randomly choosing input $x^{(b)}$ values from $U(0, 1)$ and (*ii*) computing the corresponding output $y^b = f_o(x^{(b)})$ values using the set of network weights that produce f_o. Subsets of buffer points, which lie along the functional representation f_o of the original points, are used in the following step to simulate memory rehearsal.

Step 2.2 - Re-train: The network re-trains on six new points $(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)})$ and six buffer points $(\boldsymbol{x}^{(b)}, \boldsymbol{y}^{(b)})$ (see fourth panel of Figure 2.11(b)). New points are chosen by randomly selecting six independent $x^{(n)}$ and $y^{(n)}$ values from $U(0, 1)$. Buffer points are chosen by randomly selecting, with uniform probability, six $(x^{(b)}, y^{(b)})$ pairs from the pool of the buffer points generated in *Step 2.1*. Training on the same number of new

and buffer points places equal emphasis on learning and memory rehearsal. Because the new points are randomly chosen and poorly constrained, we repeat the second training session 1000 times to generate a distribution of solutions $\{f_n\}$ (see lower right panel of Figure 2.12). Both the new and buffer points vary from session to session, but the buffer points are always sampled from the same original function f_o. We restrict the training time of each session to 500 iterations, thereby giving the network "limited" time to learn.

Notation: We use the super and subscripts "o" and "n" to refer respectively to the "original" and "new" points, $(\boldsymbol{x}^{(o)}, \boldsymbol{y}^{(o)})$ and $(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)})$, and functional approximations, f_o and f_n. Each function f_o produces a single error value $E_o^{(o)}$ measured with respect to $(\boldsymbol{x}^{(o)}, \boldsymbol{y}^{(o)})$. Each set of functions $\{f_n\}$ produces two sets of error values, $\{E_n^{(o)}\}$ and $\{E_n^{(n)}\}$, measured with respect to $(\boldsymbol{x}^{(o)}, \boldsymbol{y}^{(o)})$ and $(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)})$, respectively.

2.5 Tradeoffs in Learning and Memory Tasks

We train the five networks shown in Figure 2.11(a), first considering the differences between the boundary fan (parallel) and stacked (layered) networks. Given the large number of adjustable parameters N_p relative to the small number of training points N_d, we expect all five networks to fit the points with high accuracy. Instead, the networks show significant differences in performance both within individual training sessions and measured statistically over many sessions. These results, discussed below, show the same qualitative features for larger networks and for permuted training points (see Chapter 3).

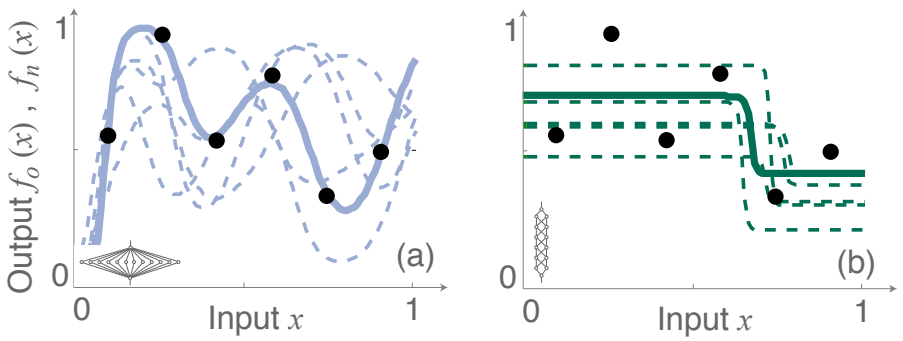

Figure 2.13: Solutions Produced by Fan and Stacked Networks. Panels (a) and (b) show solutions produced respectively by the fan and stacked networks, indicating for each network the approximation f_o (solid curve) of the original points (point markers) and a subset of approximations $\{f_n\}$ (dashed curves) of the new and buffer points. In this realization, the fan network fits the original points with a high order polynomial, while the stacked network produces a largely linear fit. Subsequent approximations $\{f_n\}$ retain these features of f_o.

2.5.1 Fan and Stacked Architectures

Examples of the solutions f_o and $\{f_n\}$ produced by the fan and stacked networks are shown in Figures 2.13. Each set $\{f_n\}$ is characterized by errors $\{E_n^{(o)}\}$ and $\{E_n^{(n)}\}$, which measure the ability of the network to retain and learn information, respectively. The complementary cumulative distribution functions (cCDFs) of these errors are shown in Figures 2.14(a) and 2.14(b), where the cCDF(E) gives the probability that the network produces an error value greater than E for every value of E.

The fan and stacked networks produce qualitatively different types of solutions f_o and $\{f_n\}$. While the specific functional form of f_o depends on the randomly initialized network state (see the following section), the f_o solutions shown here have errors that are representative of the average network performance over a range of randomly initialized

states. The stacked solution f_o averages over the variation in the original points (Figure 2.13(b)). In contrast, the fan solution f_o accurately fits all six original points with a high order polynomial (Figure 2.13(a)). In both networks, subsequent solutions $\{f_n\}$ retain the features of f_o. Because the sigmoid transfer function (see Models) is identical for all nodes, the differences between the fan and stacked solutions arise solely from variations in network architecture. As the sigmoid function maps an infinite input space to a finite output space bounded between 0 and 1, successive applications of sigmoids produced by serial (stacked) computations tend to result in linear outputs, while a sum of sigmoids produced by parallel (fan) computations tends to result in highly variable outputs.

The interference between the two training sessions results in the deviation of $\{f_n\}$ from f_o, which tends to increase $\{E_n^{(o)}\}$ relative to $E_o^{(o)}$. We find that in its best case, the stacked network shows no deviation in $\{E_n^{(o)}\}$ from $E_o^{(o)}$. In contrast, the fan network shows a minimum deviation of 130% and a higher deviation on average compared to the stacked network. This deviation measures the ability of the network to retain the original representation f_o, regardless of how erroneous that representation may be. Although the stacked network generates a higher error representation of the original points during the first training session, it can more accurately retain this representation when presented with new points.

The minimum and maximum values of $\{E_n^{(o)}\}$ measure the best success and worst failure of the network in retaining old information while avoiding interference from new information. While the bounded output space limits the maximum error, linear

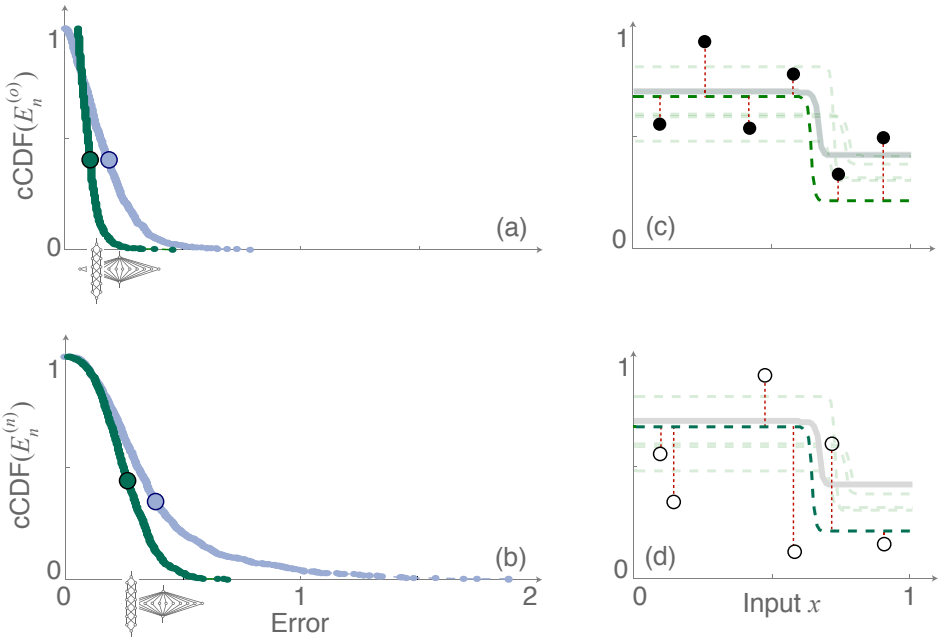

Figure 2.14: Error Distributions Produced by Fan and Stacked Networks. Panels (a) and (b) respectively show the cCDFs of $\{E_n^{(o)}\}$ and $\{E_n^{(n)}\}$, with the average value of each distribution marked by a filled circle and with the network responsible for producing the distribution indicated beneath the distribution average. Panels (c) and (d) illustrate the calculation of error values $E_n^{(o)}$ and $E_n^{(n)}$ from a representation f_n, where error is measured with respect to (c) original versus (d) new points. (a) The fan network achieves a lower minimum but higher maximum error on the original points than does the stacked network, resulting in a wider distribution with a higher average error. (b) Both networks produce low minimum errors on the new points, but the fan network again produces higher average and maximum errors than does the stacked network.

solutions tend to further restrict these bounds. As a result, the stacked network has a lower maximum error at the cost of having a higher minimum error, as shown in Figure 2.14a. In contrast, the fan network can retain the original information more accurately by achieving a lower minimum error, but it can also fail more catastrophically with a higher maximum error.

Similar features are observed in the distributions of $\{E_n^{(n)}\}$ shown in Figure 2.14b. The minimum and maximum values of $\{E_n^{(n)}\}$ measure the best success and worst failure of the network in learning new information while attempting to retain old information. While both networks achieve low minimum error at their best, the fan network produces a much larger maximum error than does the stacked network. In addition to achieving more extreme best and worst cases, the fan network also has higher average error values $\langle\{E_n^{(o)}\}\rangle$ and $\langle\{E_n^{(n)}\}\rangle$.

2.5.2 Intermediate Architectures: Tradeoffs in Performance

We extend our analysis to the intermediate architectures shown in Figure 2.11, organizing results based on the degree of serialization h/ℓ (a purely geometrical factor).

Tradeoffs in performance are observed across the range of architectures. For example, in Figure 2.15(a), we see a tradeoff between the minimum and maximum values of $\{E_n^{(o)}\}$. As h/ℓ increases, the network does not fail as badly in its worst case but also does not succeed as well in its best case. Furthermore, we find a tradeoff in performance between the first and second sessions, shown in Figure 2.15(b). Increasing h/ℓ worsens

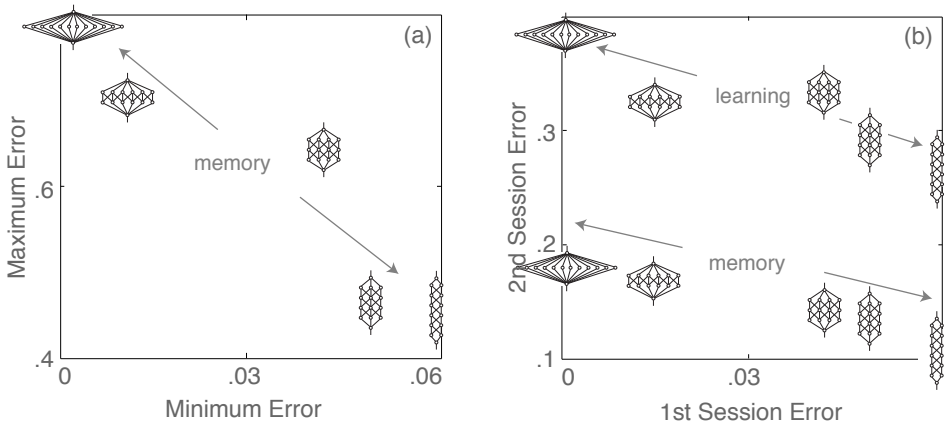

Figure 2.15: Tradeoffs in Performance Exhibited Across Network Architectures. Tradeoffs in network performance are measured with respect to solutions f_o and $\{f_n\}$ produced by the five networks shown in Figure 2.11(a). With respect to solutions $\{f_n\}$ produced during the second training session, increasing h/ℓ (a) decreases the maximum value of $\{E_n^{(o)}\}$ at the cost of increasing its minimum value. (d) Furthermore, increasing h/ℓ increases $E_o^{(o)}$ achieved during the first session at the cost of decreasing $\langle\{E_n^{(n)}\}\rangle$ and $\langle\{E_n^{(o)}\}\rangle$ achieved during the second session.

performance during the first session by increasing $E_o^{(o)}$ but improves average performance during the second session by decreasing both $\langle\{E_n^{(n)}\}\rangle$ and $\langle\{E_n^{(o)}\}\rangle$, suggesting a tradeoff between the accuracy and generalizability of network solutions. The fan network, which produces a very accurate, specific representation of the original points, shows a much higher average error when it tries to generalize this representation. In contrast, the coarser representation produced by the stacked network is better able to incorporate new information.

We further find that layering improves several aspects of performance. Figure 2.16(a) shows that increasing h/ℓ decreases the maximum error in both $\{E_n^{(o)}\}$ and $\{E_n^{(n)}\}$, indicating that the stacked architecture is best suited for minimizing failure in both

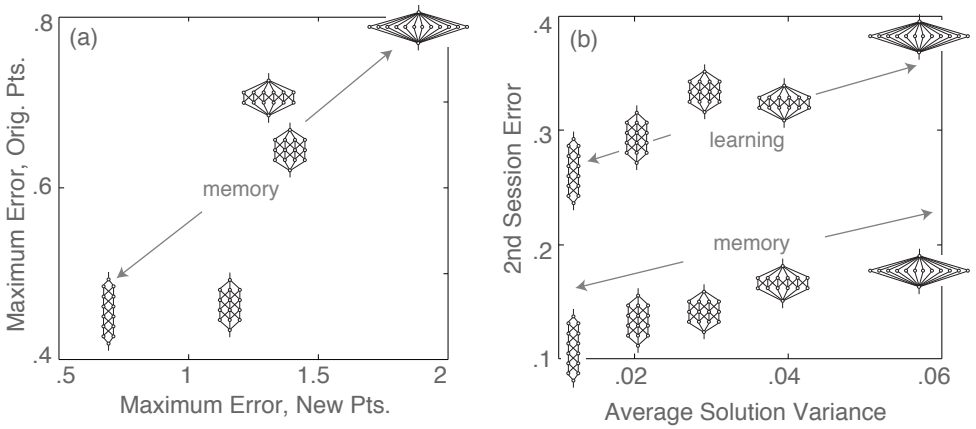

Figure 2.16: Impact of Layering on Improved Performance. Variations in network performance are measured with respect to solutions f_o and $\{f_n\}$ produced by the five networks shown in Figure 2.11(a). With respect to solutions $\{f_n\}$ produced during the second training session, increasing h/ℓ (a) decreases the maximum error in both $\{E_n^{(n)}\}$ and $\{E_n^{(o)}\}$, and (b) decreases the average solution variance $\langle\{(\Delta f_n)^2\}\rangle$ and the average errors $\langle\{E_n^{(n)}\}\rangle$ and $\langle\{E_n^{(o)}\}\rangle$. This shows that layering improves performance by decreasing the average and maximum network error, which corresponds to a decrease in average solution variance.

learning and memory. Figure 2.16(b) shows that increasing h/ℓ decreases both the average solution variance $\langle\{(\Delta f_n)^2\}\rangle$ and the average errors $\langle\{E_n^{(n)}\}\rangle$ and $\langle\{E_n^{(o)}\}\rangle$. While we might naively expect that high solution variance (fan) would indicate a flexible network able to accurately fit nonlinear data, we instead find that high variance leads to high average error. In contrast, low variance, linear solutions (stacked) tend to minimize average error.

2.6 Adaptation to Variable Learning Conditions

Both natural and artificial systems can be found in a variety of states when presented with new information. The success in learning this information may depend both on

the initial state of the system and on the learning conditions. We explore these possible dependencies by varying both the randomly initialized network state and the training conditions.

2.6.1 Variable Initialized States

Because the conjugate gradient descent algorithm is deterministic, the randomly initialized state determines f_o, which then influences subsequent solutions $\{f_n\}$.

To study the influence of random initialization on f_o, we train all five networks on the original points with 500 sets of randomly chosen weights, each of which correspond to a specific starting location within the error landscape. By allowing "unlimited" training time, each network deterministically locates error minima with errors $\{E_o^{(o)}\}$ and corresponding solutions $\{f_o\}$. In this manner, we systematically probe the features of the underlying error landscape.

The cCDF of $\{E_o^{(o)}\}$, shown in Figure 2.17(a), reveals that the fan network consistently finds zero error solutions, while all other networks find solutions with a wide range of error values. The networks can collectively produce both zero error and high error solutions and do so with probabilities that respectively decrease and increase as h/ℓ increases. The discontinuities in the stacked error distribution may indicate that the error landscape is composed of localized sets of minima with distinct depths. In comparison, the intermediate distributions show greater continuity in error, suggesting the presence of a larger number of connected minima with variable depths.

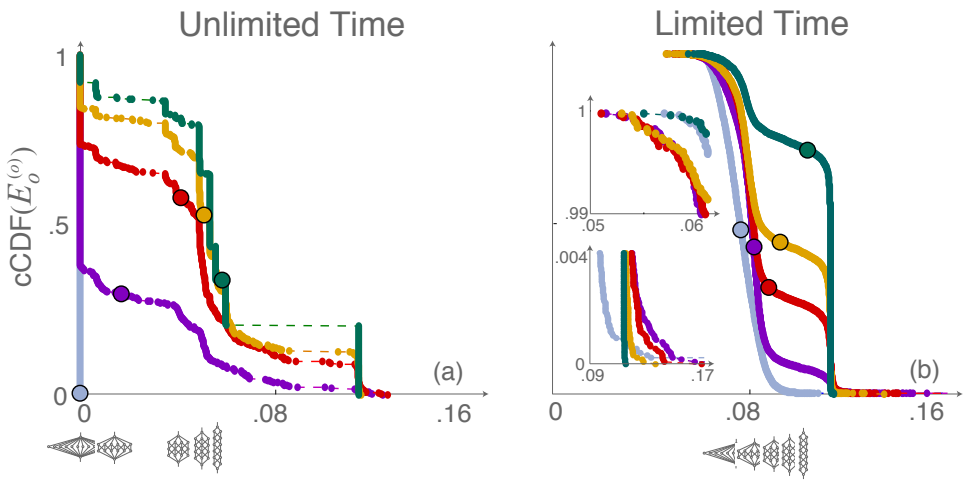

Figure 2.17: Distribution of Minima within Error Landscapes cCDFs of $\{E_o^{(o)}\}$ are shown given (a) unlimited and (b) limited training time for the five networks shown in Figure 2.11(a), where the average value of each distribution is marked by a filled circle and the network responsible for producing the distribution is indicated beneath the distribution average. (a) The fan network consistently finds zero error solutions, while all other networks find solutions with a range of error values. (b) Intermediate networks find lower error solutions than do the fan and stacked networks (upper inset). Increasing h/ℓ significantly decreases the both the maximum error and the frequency of high error solutions (lower inset). In both (a) and (b), increasing h/ℓ increases $\langle\{E_o^{(o)}\}\rangle$ (filled circles).

The distributions are more heavily weighted toward high error as h/ℓ increases, thereby increasing the average error $\langle\{E_o^{(o)}\}\rangle$. For a given architecture, the average number of training iterations decreases with increasing solution error, indicating an inherent tradeoff between speed and accuracy. While able to produce solutions with the same degree of accuracy as the fan network, the intermediate and stacked networks can also quickly produce coarse solutions. However, the intermediate networks require fewer iterations than the stacked network to reach solutions of similar error, suggesting that the presence of additional connections may facilitate faster performance.

If we inspect the solutions produced by each network, we find that low, medium, and high error solutions correspond respectively to fitting all, some, or none of the points with a high order polynomial and fitting the remaining points with a horizontal line. To emphasize differences in network performance, the solutions f_o used to generate the results shown in Figures 2.13 and 2.15 were chosen because their error was representative of the distribution averages shown in Figure 2.17(a).

2.6.2 Temporal Constraints

In natural systems, the time allowed to gather information from the environment is often limited, and a highly specific representation of information may not be desirable or even attainable. To investigate the effect of temporal constraints, we train the five networks on the original points with 5000 sets of randomly chosen weights, now terminating training after 100 iterations. While the case of unlimited training time considered in

the previous subsection enables us to identify the distribution of minima within each network error landscape, limiting training time highlights the ability of each network to efficiently locate these minima. The increased number of randomly initialized states allows us to better resolve the edges of the error distributions shown in Figure 2.17(b).

Once training time is limited, all distributions shift toward higher error values, again revealing a tradeoff between speed and accuracy. As before, $\langle\{E_o^{(o)}\}\rangle$ increases as h/ℓ increases. Discontinuities in the distributions are also removed, indicating that the networks do not have sufficient time to consistently find distinct sets of minima.

The dynamic range of performance decreases as h/ℓ increases, resulting in significant differences between the edges of each distribution. At the rightmost edge, both the frequency of high error solutions and the maximum error value increase as h/ℓ increases. The stacked network shows an abrupt cutoff near the minimum error achieved by fitting the original points with a horizontal line. All other distributions extend beyond this value. In contrast to the case of unlimited training time, the fan network shows the least consistency in performance and produces several catastrophic errors, thereby revealing the greatest sensitivity to changes in training time. At the leftmost edge of the distributions, the intermediate networks find lower minimum error values than do the fan and stacked networks. This is similar to the behavior observed with unlimited training time, where the intermediate networks found comparable solutions to the fan and stacked extremes in fewer iterations. It may therefore be interesting in the future to verify the dependence of performance on the number of network connections.

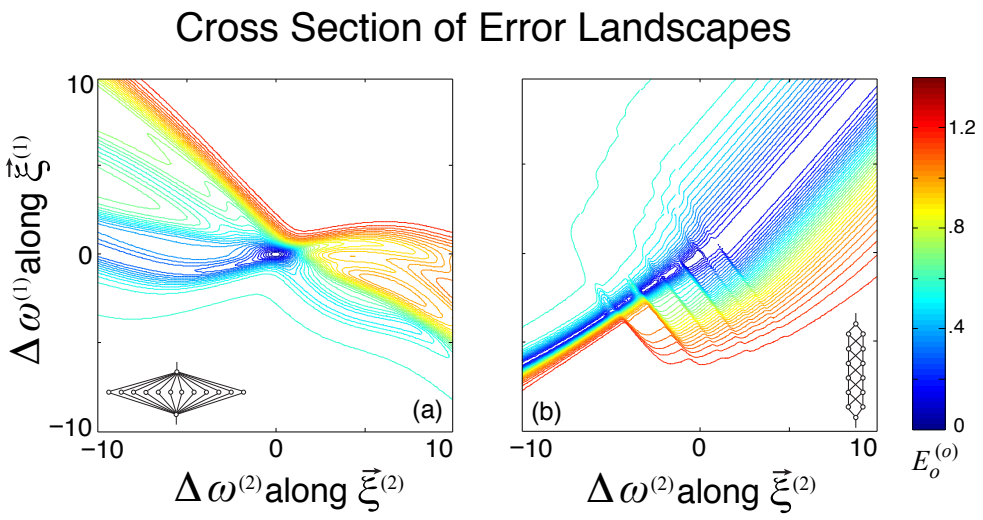

Figure 2.18: Cross Section of Network Error Landscapes. Error $E_o^{(o)}$ is projected onto the two stiffest eigenvector directions $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\xi}^{(2)}$ about minima produced by the (a) fan and (b) stacked network given unlimited training time. The two minima were chosen for comparison because they have the same number and similar magnitude of nonzero eigenvalues, although similar behavior was observed for alternative minima. The insets show zoomed in views of the contour plots about their central minima. (a) The projection of the fan landscape shows a single deep minimum surrounded by smooth peaks. (b) In contrast, the projection of the stacked landscape shows a long, deep valley of several local putative minima separated by low barriers. The surrounding landscape is much bumpier than that of the fan network.

2.7 Dependence on Error Landscape Structure

To better understand how the underlying error landscape constrains network performance, we assess the local properties of error landscape minima. As both the network structure and the external training information dictate error landscape features, this characterization provides a direct link between the structural variations in network topology (Figure 2.11) and the observed differences in network performance (Figures 2.14, 2.15, and 2.16).

2.7.1 Characterizing Landscape Minima

Given unlimited training time, the distributions in Figure 2.17(a) mark the error of local minima found within the error landscape of each network. Each minimum can be characterized by the degree of local landscape curvature, where directions of high curvature specify combinations of weight adjustments that produce large changes in error. We adopt the terminology used in previous studies and refer to directions with high and low curvature as stiff and sloppy, respectively [37, 38]. Stiff and sloppy directions are found by diagonalizing the error Hessian $H_{pq} = \partial^2 E / \partial \omega_p \partial \omega_q$ evaluated at the set of weights that produces the local error minimum. For computational efficiency, we use the approximate Levenberg-Marquardt (LM) Hessian [39], defined as:

$$\frac{\partial^2 E}{\partial \omega_p \partial \omega_q} \approx \sum_{d=1}^{N_D} \frac{\partial r_d^{(o)}}{\partial \omega_p} \frac{\partial r_d^{(o)}}{\partial \omega_q}, \tag{2.6}$$

where $r_d^{(o)} = (y_d(\boldsymbol{\omega}) - y_d^{(o)})$ is the residual of the $d\text{th}$ original point.

The LM Hessian is a good approximation to H when the error of local minima, and thus the residual $r_d^{(o)}$, is small and the additional Hessian term $r_d^{(o)} \partial^2 r_d^{(o)} / \partial \omega_p \partial \omega_q$ can be neglected. For a given model and data set, the LM Hessian agrees well with the stiffest eigenvectors of H and is equivalent to H when the model perfectly fits the data. In addition, it has a known number of exactly zero eigenvalues equal to the difference in the number of model parameters N_p and the number of data points N_d [37, 38].

We diagonalize the LM Hessian about each of the 500 minima with the error values $\{E_o^{(o)}\}$ shown in Figure 2.17(a). Each error minimum produces a set of N_p eigenvalues

λ and normalized eigenvectors $\boldsymbol{\xi}$, which give the degrees and directions of stiffness in weight space.

As an illustrative example of landscape features observed along these relevant directions, Figure 2.18(a-b) shows the projection of the error landscape onto the two stiffest eigenvector directions $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\xi}^{(2)}$ centered on zero error minima produced by the fan and stacked networks, respectively.

The fan landscape shows a single deep basin surrounded by smoothly varying peaks. In contrast, the stacked landscape is rugged, showing a deep valley with several minima separated by small barriers. While these minima appear to be distinct, they may be connected by higher dimensional pathways that cannot be seen in this reduced space.

2.7.2 Participation of Network Connections

The ability of a network to move along relevant eigenvector directions may depend on the number of weights that must be significantly adjusted, or equivalently the localization of eigenvector components. To quantify the degree of localization of the $p\mathrm{th}$ eigenvector $\boldsymbol{\xi}^{(p)}$, we calculate its participation ratio [40]:

$$\rho^{(p)} = \sum_{q} (\xi_q^{(p)})^4 \tag{2.7}$$

where individual eigenvector components $\xi_q^{(p)}$ correspond to specific weights ω_q in the network. $\rho^{(p)}$ is a dimensionless quantity that ranges between a completely delocalized minimum of $1/N_P$, for which all components have equal weight $1/\sqrt{N_P}$, and a completely

localized maximum of 1, for which a single component carries unit weight.

For the set of minima with error values $\{E_o^{(o)}\}$, we quantify $\{\rho^{(1)}\}$ and $\{\lambda^{(1)}\}$ of the stiffest eigenvectors $\{\boldsymbol{\xi}^{(1)}\}$, as combinations of weight changes specified by these eigenvector directions produce the largest changes in error. The covariances $\mathrm{C}_{E,\rho} = \mathrm{Cov}(E_o^{(o)}, \rho^{(1)})$ and $\mathrm{C}_{E,\lambda} = \mathrm{Cov}(E_o^{(o)}, \lambda^{(1)})$ in these quantities are shown by the ellipses centered about their average values in Figures 2.20(a) and 2.20(b), respectively.

Figure 2.20 highlights the variability in basin structure within and between the networks. As h/ℓ increases, both the average and variance in $\{E_o^{(o)}\}$, $\{\rho^{(1)}\}$, and $\{\lambda^{(1)}\}$ increase. Higher variance leads to lower confidence in predicting the success of the network, but it also suggests that the network has more options when exploring its error landscape.

The orientations of the covariance ellipses in Figures 2.20(a) and 2.20(b) provide information regarding the relationships between $E_o^{(o)}$, $\rho^{(1)}$, and $\lambda^{(1)}$. The semi-major axis of each $\mathrm{C}_{E,\rho}$ ellipse in Figure 2.20(a) lies along the trend swept out by the average values of $\{E_o^{(o)}\}$ and $\{\rho^{(1)}\}$, suggesting a general, positive correlation between $E_o^{(o)}$ and $\rho^{(1)}$. While the average values of $\{E_o^{(o)}\}$ and $\{\lambda^{(1)}\}$ would suggest that these quantities are also positively correlated, Figure 2.20(b) shows that for a given value of h/ℓ, larger values of $E_o^{(o)}$ correspond to smaller values of $\lambda^{(1)}$. These results reveal general characteristics of error landscape structure; higher error minima $(\mathrm{larger}\ E_o^{(o)})$ tend to be shallower $(\mathrm{smaller}\ \lambda^{(1)})$ and require the adjustment of fewer weights $(\mathrm{larger}\ \rho^{(1)})$.

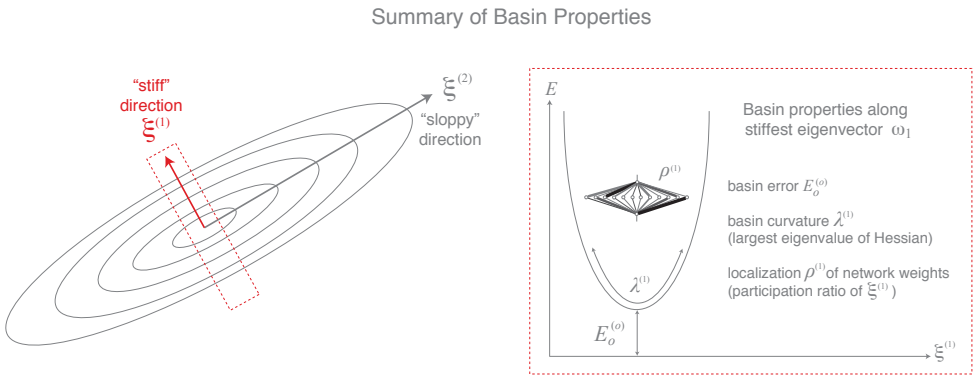

Figure 2.19: Basin Properties of Interest. Stiff and sloppy landscape directions. which correspond respectively to directions of high and low curvature about local landscape minima, are computed as the eigenvectors $\boldsymbol{\xi}$ of the LM Hessian. For each of 500 minima within network error landscapes, we assess landscape properties along the stiffest direction $\boldsymbol{\xi}^{(1)}$. We compare the error $E_o^{(o)}$ of each minimum, which is independent of the landscape direction in consideration, with two direction-specific properties: the curvature $\lambda^{(1)}$ (stiffest eigenvalue) and the localization $\rho^{(1)}$ of weights (participation ratio of eigenvector components) along the stiffest landscape direction.

2.7.3 Landscape Characteristics and Successful Learning

Variations in landscape structure provide insight into the way in which each network searches for solutions. In particular, fan solutions are characterized by low error and participation ratio, indicating that the fan network must adjust nearly all of its weights in order to navigate zero error basins. In contrast, stacked solutions span a range of error values. The corresponding basins are characterized by a variety of eigenvalues and participation ratios, indicating that the stacked network can navigate many types of basins by adjusting variable numbers of weights. Larger participation ratios correspond to higher errors and lower eigenvalues, suggesting that the stacked network can navigate shallow, high error basins by adjusting few connections. Narrow, low error basins, found by both the fan and stacked networks, require fine tuning of a large number of connections.

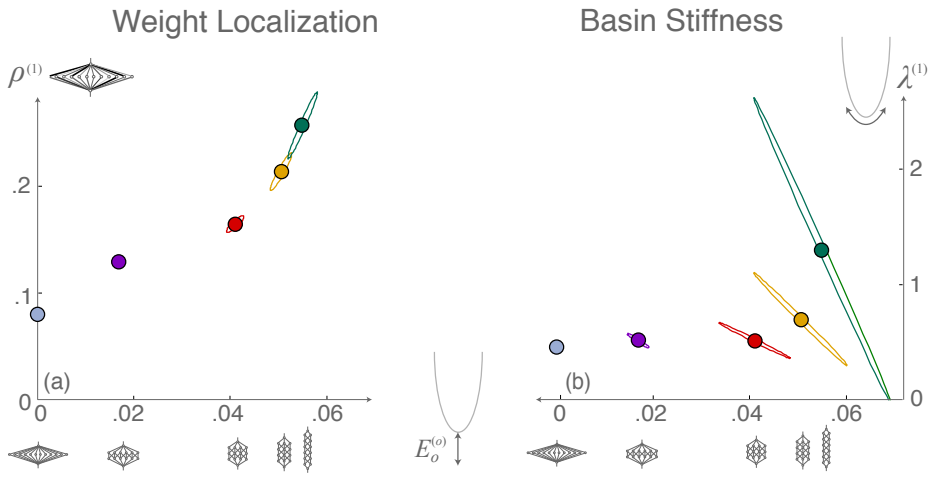

Figure 2.20: Covariance in Basin Properties. Covariances between (a) $\{\rho^{(1)}\}$ and $\{E_o^{(o)}\}$ and between (b) $\{\lambda^{(1)}\}$ and $\{E_o^{(o)}\}$ are shown for error landscape minima produced by the five networks shown in Figure 2.11. For each network, the values of $\{E_o^{(o)}\}$ are taken from the distributions shown in Figure 2.17(a). Covariances, indicated by ellipses, are centered about their average values, indicated by markers. The semimajor axis of each ellipse marks the direction of maximum covariance. Increasing h/ℓ increases both the average and variance in all three quantities. For a given network, larger values of $E_o^{(o)}$ generally correspond to smaller values of $\lambda^{(1)}$ and larger values of $\rho^{(1)}$.

In combination, landscape characteristics help explain the results shown in Figures 2.15 and 2.17. Given unlimited training time, landscape variability is disadvantageous and can prevent a network from finding a low error minimum. Once time is limited, landscape variability can be advantageous in preventing failure by providing the network with high error, shallow basins that can be navigated with the adjustment of relatively few connections. If limited training time is coupled with extremely noisy information, landscapes with high error basins can be advantageous by decreasing average error relative to landscapes with no easily reachable basins. Because our sequential sessions combined both limited and unlimited training time and both clean and noisy data, we see an additional tradeoff between the two sessions. Unlimited training time and well constrained data favor the fan over the stacked network in minimizing average error, while limited time and noisy data favor the stacked network over the fan.

2.8 Discussion

In investigating tradeoffs in learning and memory performance that arise from structural complexity, we found that none of the considered architectures simultaneously mastered both learning and memory tasks. This suggests that systems whose function depends on such simultaneous success might require architectures that are complex combinations of both parallel and serial structures. Indeed, this inherent sensitivity of function to underlying architecture may help to explain the high degree of variability evident in architectural motifs of large-scale biological and technical systems. For in-

stance, in natural neuronal networks, cortical connection patterns display a variety of architectural complexities at varying spatial scales. Examples of fan architectures are found in hub-and-spoke motifs, which form an important part of the small-world architecture [41, 42, 43], as well as in the decomposition of cortical network architectures into subnetworks or modules which may simultaneously process differential information [44, 45, 46, 12, 47]. Moreover, stacked architectures are evident within cortical lamina [3], within the hierarchical organization displayed in the sequential ordering of the visual system [48], and within the nested modularity of large-scale cortical connectivity [49, 12, 47]. Similarly, artificial neural networks display complex combinations of fan and stacked motifs including modularity [50], hierarchy [51], and small-worldness [52, 53].

2.8.1 Parallel versus Layered Architectures

Given the wealth of structural motifs present in real world systems, it is of interest to first isolate the tradeoffs in performance associated with small parallel and layered network structures which together form the complex architectural landscape of larger systems and thereby constrain their overall performance. Here we found that the deep, narrow basins within the error landscape enabled the fan network to produce very accurate solutions. However, the difficulty of simultaneously adjusting many network connections in order to escape deep basins may have hindered the ability of the fan network to adapt, a result that helps explain the susceptibility of parallel networks to the problems of overfitting and failure to generalize [27]. In contrast, higher variability in the width and depth

of local minima enabled the stacked network to quickly find coarse but generalizable solutions through the adjustment of a smaller fraction of weights. In combination, these results support the hypothesis that the number and width of local landscape minima may increase with increasing number of hidden layers [54], and we suggest that this variability helps explain why layered networks may require fewer computational units and may better generalize than parallel networks [55, 56]. However, the impact of structural variations on functional tradeoffs, for example between specificity and generalizability, extends beyond artificial network studies and is crucial for understanding the interaction of learning processes in large scale models of the brain [57]. While parallel architectures are often preferred in artificial network studies due to their consistency and accuracy [54, 56], our results highlight the advantages of layered architectures when performance criteria favor generalizability and minimization of failure.

2.8.2 Intermediate Architecture

Building on the intuition gained from the two benchmark extremes–fan and stacked–we further assessed the characteristics of intermediate networks, which can be used to more directly probe the expected behavior of structurally complex composite systems. In particular, our intermediate structures were composed of several adjacent stacked networks and therefore shared principal features of both parallel and layered systems. Additionally, these networks had slightly larger numbers of connections than the fan and stacked networks.

Due to these structural differences, the depth of local minima within the intermediate landscapes displayed more variation than fan minima but more continuity than stacked minima. As landscape variability was linked to improved generalization capabilities, a continuous range of basin depths may have enabled the more successful balance between flexible learning and stable memory observed in the intermediate networks. This performance supports the hypothesis that short path lengths (similar to the serialization h/ℓ [58]) and low connection densities may facilitate simultaneous performance of information segregation (memory retention) and integration (generalization) within natural neuronal systems [59]. These competing processes are also maintained in natural neuronal systems and neural circuit models through homeostatic plasticity mechanisms such as synaptic scaling [60, 61] and redistribution [62, 63], in addition to the rehearsal methods employed here [32, 33, 34, 35, 36]. Even in the absence of such homeostatic plasticity mechanisms, we found that the architectural combination of parallel and layered connectivity helped foster a balance between learning and memory.

2.8.3 Variable Learning Conditions and Network Efficiency

We extended our analysis from the case of unlimited training time, which revealed information about error landscape structure, to the biologically-motivated case of limited training time. Comparison of these two cases revealed a tradeoff in performance between training speed and solution accuracy. In the absence of temporal constraints, the production of highly accurate representations required longer training times. Sim-

ilarly, temporal constraints led to larger solution errors. This tradeoff between speed and accuracy has been observed in cortical networks, where emphasis on performance speed during perceptual learning tasks increased the baseline activity but decreased the transient task-related activity of neurons within the decision-making regions of the human brain [64, 65]. Here we found that network architecture played a significant role in the manifestation of this tradeoff, and the presence of additional hidden layers helped minimize network susceptibility to changes in training time. In particular, the fan network demonstrated the greatest change in performance under temporal constraints, showing a decrease in consistency coupled with occasional catastrophic error values. In contrast, the intermediate and stacked networks improved consistency and minimized inaccuracy once training time was limited.

Upon closer inspection, we found that the intermediate networks produced solutions with increased speed given unlimited time and with increased potential for accuracy when time was limited as compared to the fan and stacked extremes. The presence of additional connections may have influenced the number of iterations required to find a solution, or similarly the minimum error found with a fixed number of iterations. While the graph measure of path length is known to influence network efficiency [58], these results imply that the number of networks connections may additionally enable the network to quickly find an accurate solution.

In addition to static variations in connectivity, dynamic structural changes such as synapse formation [66] can facilitate learning and memory processes. The converse

case of network degradation, or disruptions to structural connectivity, is also known to have widespread consequences in functional properties of the brain [67, 68, 69]. A more detailed study of the relationship between connection number and robustness could provide additional insight into the effects of synapse formation and degradation on functional performance. Our analysis of error landscape features revealed that different architectures showed variable localization properties in the eigenvectors associated with local error minima, and we therefore expect robustness to depend on both the architecture and the location of growth or damage within the network.

2.8.4 Methodological Considerations

We found that parallel networks suffered from the creation of excessively detailed representations of information, an "overfitting" problem that is often addressed through the use of cross-validation [70] and weight regularization [71] techniques. As one goal of this study was to uncover the structural basis for differences in representational capabilities, it was crucial to understand network behavior in the absence of task-specific cross-validation schemes. Additionally, as the number of parameters was roughly constant across all network structures (and identical for the fan and stacked networks), we were able to draw comparisons across network architectures in the absence of additional weight regularization constraints.

While parallel network models have commonly been used in machine learning studies, multi-layer "deep" networks have recently gained interest due to their potential ability to

compactly represent (using fewer computational units and parameters) highly variable functions [55, 56]. The "deep belief" framework has been successful for training large, multi-layered networks, and training methods often couple unsupervised, layer-wise (greedy) training with supervised fine-tuning [72]. Recent studies of deep belief networks found that classification performance improved with the addition of layers [54]. In addition, it was suggested that a reduction in the number of hidden layers would require an exponential increase in the number of hidden units in order to achieve similar network performance [56]. These results emphasize the capabilities of layered networks and provide an additional framework in which to explore structure-function tradeoffs.

Although biologically-motivated, the FFBP framework includes several simplifying assumptions that could be modified to include additional, realistic complexity. First, we assumed that only the connection weights, analogous to synaptic strengths, were variable. Real neurons also exhibit changes in intrinsic dynamics [73] that interact with network architecture to constrain functionality in the brain [74]. Accounting for such relationships could be particularly relevant, for example, in the study of neuron response profiles within different cortical layers [15]. Second, we assumed that signals passed between nodes had no temporal structure, analogous to representing steady state neuron firing rates. Temporally varying signals could be included to study the dependence of dynamic properties, such as synchronization [75, 76, 74] and signal propagation [77], on structural organization [78]. Lastly, we assumed feedforward connectivity. The addition of recurrent connections could be used to study the relationship between recurrent

structure and oscillatory functions such as cortical sleep rhythms [79] and oscillation couplings relevant for associative learning and memory [80]. In each of these directions, we anticipate that underlying structural complexity will continue to impact performance through functional tradeoffs.

In summary, different network architectures produce error landscapes with distinguishable characteristics, such as the height and width of local minima, which in turn determine performance features such as speed, accuracy, and adaptability. Inherent tradeoffs, observed across a range of architectures, arise as a consequence of the underlying error landscape structure. The presence of local landscape minima enable greater speed, more generalizable solutions, and minimization of catastrophic failure. However, these successes come at the cost of decreased accuracy. Understanding how both the landscape characteristics and the resulting performance features vary across a range of architectures is crucial for both understanding and guiding the design of more complex biological and technical systems.

Chapter 3

Network Performance Constrained by Internal versus External Structure

"If you just have a single problem to solve, then fine, go ahead and use a neural network. But if you want to do science and understand how to choose architectures, or how to go to a new problem, you have to understand what different architectures can and cannot do."

–Marvin Minsky

3.1 Introduction

The results of the previous chapter raise several important questions about the capabilities and limitations imposed by different network architectures on the interaction between the network and the external environment. Given that both artificial and biolog-

ical learning systems continually interact with and adapt to their environment, it is of interest to investigate how internal network structure interacts with and takes advantage of structure in the external environment. This is particularly relevant in the context of evolution and development, where systems are able to maintain functionality in the presence of ongoing internal structural changes.[1]

In this chapter, we explore several extensions of the network studies introduced in Chapter 2, which identified tradeoffs in the ability of different network structures to robustly and efficiently construct representations of their external environment. We first assess the robustness of the results shown in Figures 2.13, 2.14, and 2.15 to variations in network size and input training pattern. We then explore two different directions that focus on the limitations imposed by network structure versus environmental structure. We first explore the extent to which small network modules, such as the single-layer two-node network that forms the basis of the stacked network, impose limitations on the performance of larger, composite networks. We then explore the extent to which network structure interacts with and takes advantage of structure in the external environment. Together, these directions highlight several avenues of work in progress, each of which may provide detailed insight into the behavior of larger network structures that must perform more complex learning and memory tasks.

[1] Portions of this chapter originally appeared in Hermundstad *et al.*, "Structural drivers of function in information processing networks," *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on* , 2011 [81],

3.2 Robustness of Functional Tradeoffs to Structural Variations

As an extension of the results presented in the previous chapter, we evaluate the performance of larger networks consisting of 18 nodes arranged into configurations with $h\text{x}\ell = (1\text{x}18, 2\text{x}9, 3\text{x}6, 6\text{x}3, 9\text{x}2)$. We additionally evaluate the representations produced by networks shown in Figure 2.11 using a permuted set of values for the original points $(\boldsymbol{x}^{(o)}, \boldsymbol{y}^{(o)})$. These variants in network architecture and training data are highlighted in Figure 3.1.

Figures 3.2, 3.3, and 3.4 show that the results observed in the previous chapter, namely the qualitative features of network solutions, error distributions, and the resulting learning and memory tradeoffs, are consistently observed using larger networks and permuted training points.

3.2.1 Larger Networks

Given a fixed set of training points, the addition of nodes and associated connections should, in principle, improve network performance because of the increased ratio of parameters to constraints. However, we find that the addition of nodes and connections does not guarantee an improvement in performance.

We find that larger networks produce qualitatively similar solutions to those observed in smaller networks. The 1×18 fan network fits points with high order polynomials,

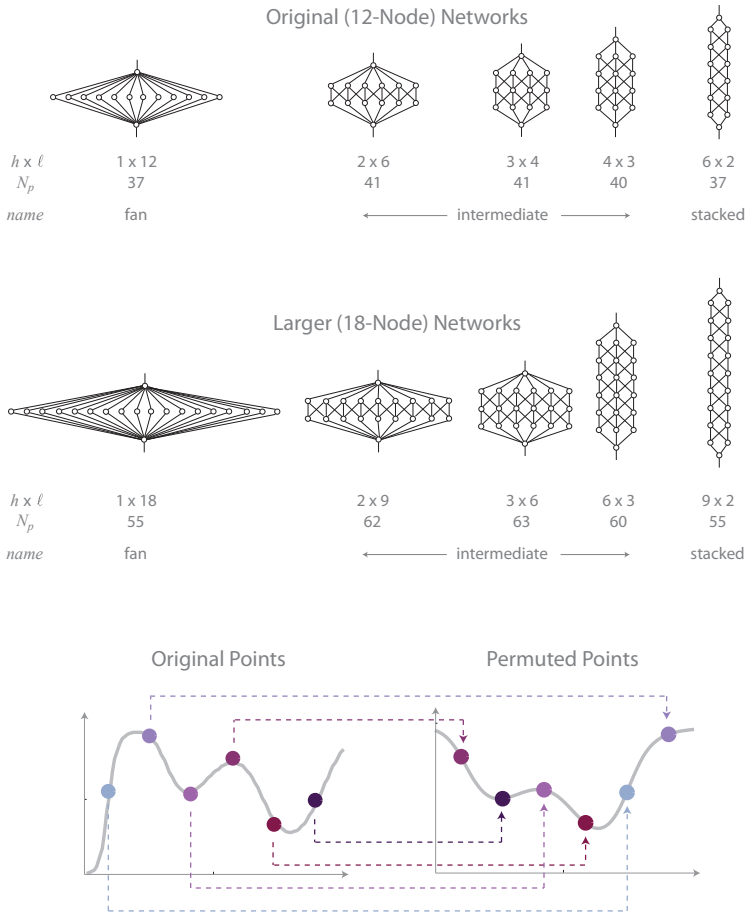

Figure 3.1: Illustration of Larger Networks and Permuted Training Points. The previous chapter considered the performance of five 12-node networks (top row). Here, we consider larger, 18-node networks (middle row) arranged into similar configurations as the 12-node networks. We further consider the performance of the 12-node networks trained on a randomly permuted set of the original training points (bottom row).

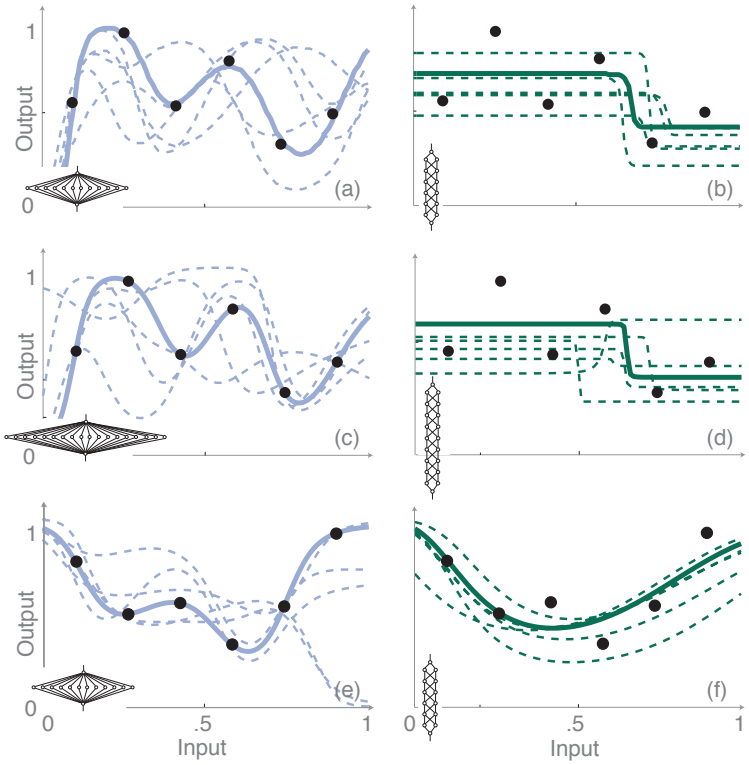

Figure 3.2: Impact of Network Size and Training Function on Network Solutions. The left and right columns show solutions produced respectively by the fan and stacked networks, indicating for each network the approximation f_o (solid curve) of the original points (point markers) and a subset of approximations $\{f_n\}$ (dashed curves) of the new and buffer points. Panels (a) and (b) are identical to those shown in Figure 2.13 of Chapter 2 for the original 12-node networks. Panels (c) and (d) confirm that the qualitative features of these solutions are maintained in 18-node networks. In particular, the 18-node stacked network produces largely linear solutions with sharper kinks than its 12-node counterpart. Panels (e) and (f) confirm that the qualitative features of these solutions are similarly maintained for permuted sets of training points. In this realization, the fan network again fits the permuted original points with a high-order polynomial. In comparison, the stacked network produces a lower order polynomial that averages over the variation in the permuted points.

while the 9×2 stacked network produces largely linear fits. However, the 18-node stacked network produces sharper nonlinearities than did its 12-node counterpart, an effect that arises from the successive application of the sigmoidal transfer function.

The distributions of error values computed from these solutions are qualitatively similar to those produced by smaller networks. The addition of nodes to the fan network decreases its maximum error in learning, while the addition of nodes to the stacked network decreases its maximum error in memory. This confirms that the extent to performance is impacted by the addition of nodes depends on where the nodes are added.

Evaluation of these distribution properties across intermediate networks shows that the observed tradeoffs between best success and worst failure in memory, and between the average errors achieved during the first and second training session, are consistently maintained across the full set of larger networks. Larger networks show higher errors during the first session but lower errors during the second session as compared to smaller networks.

Together, these results show that the addition of nodes produces different effects on network performance depending on whether nodes are added in parallel or serial. Furthermore, we do not see a uniform improvement in performance with the mere addition of nodes. Rather, the observed tradeoffs shift in favor of improvement in the second training session over improvements in the first. It would be interesting, in future work, to systematically investigate the change in these tradeoffs across larger variations in network sizes.

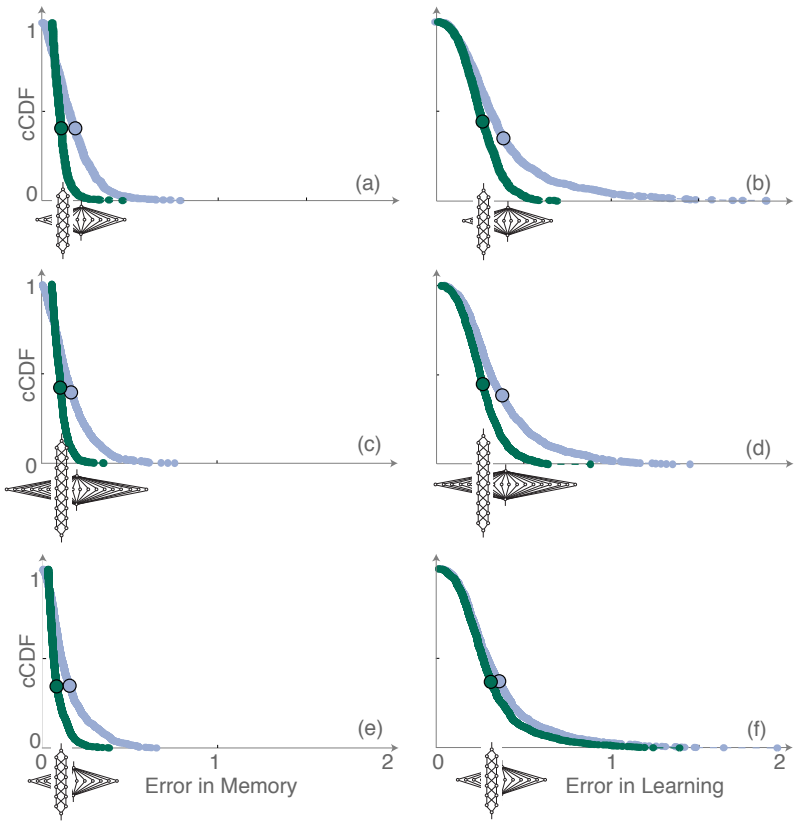

Figure 3.3: Impact of Network Size and Training Function on Error Distributions. The left and right columns respectively show the cCDFs of $\{E_n^{(o)}\}$ and $\{E_n^{(n)}\}$, with the average value of each distribution marked by a filled circle. Panels (a) and (b) are identical to those shown in Figure 2.14 of Chapter 2 for the original 12-node networks. Panels (c) and (d) confirm that the qualitative features of the cCDFs are maintained in 18-node networks, with the stacked network producing lower averages errors in learning and memory and lower maximum but higher minimum errors in memory as compared to the fan network. Panels (e) and (f) confirm that these qualitative relationships are similarly maintained using permuted sets of the original points. In particular, we see less separation in the minimum, maximum, and average errors produced by the fan and stacked networks using the permuted, as compared to the unpermuted, points.

3.2.2 Permuted Training Points

In training the fan and stacked networks on a permuted set of the original points, we find that both networks produce solutions that, although quantitatively different, show similar qualitative features to those produced with the unpermuted training points. The fan network fits the permuted points with a high-order polynomial, and subsequent solutions show similarly high variance. In comparison, the stacked network produces a solution to the permuted points that, although exhibiting less striking linearity than was observed with the unpermuted points, averages over the variation in the points to produce a low-order polynomial fit.

The distributions of error values computed from these solutions show similar, albeit less pronounced, separations in the minimum, maximum, and average error values produced by the fan versus stacked networks. Both networks produce lower maximum errors in memory but higher maximum errors in learning as compared to their performance when training with the unpermuted set of points. Furthermore, we observe a reduced separation in the average errors, measured with respect to both memory and learning, produced by the fan versus stacked networks.

By extending this assessment of distribution properties across the set of intermediate networks, we confirm that the tradeoffs in learning and memory performance observed in Figure 2.15 are maintained for permutations of the training points. In comparison to the behavior observed with the larger networks, we find that both sets of tradeoffs become less pronounced when training on the permuted points. Across networks, we observe a

decrease in both the minimum and maximum errors in memory, and we similarly see a decrease in the average errors produced during the first versus second training sessions.

Taken together, these results show that performance tradeoffs are observed within larger networks and with different sets of training points. However, variation in size versus training input have different quantitative impacts on the extent to which these tradeoffs are observed.

3.3 Dependence on Internal Structure: Characterization of Layered Networks

The results of the previous section suggest that the qualitative relationships between parallel and layered network performance remain similar across variations in the number of network nodes. It is unclear, however, as to whether the observed difference in performance across network structures is a consequence of the restricted width or the increased height of layered relative to parallel networks. In an effort to decouple these dependencies, we have begun analyses on the performance of the stacked network across variations in network height, keeping the width fixed at $\ell = 2$.

We train 7 stacked networks, with $h = 1, 2, 3, 6, 9, 12, 13$, on the original six points $\{x^{(o)}, y^{(o)}\}$ introduced in Chapter 2. Each network is given unlimited training time (10^5 iterations) to represent the original points, and each training session is repeated 1000 times to generate a distribution of error values measured from these representations.

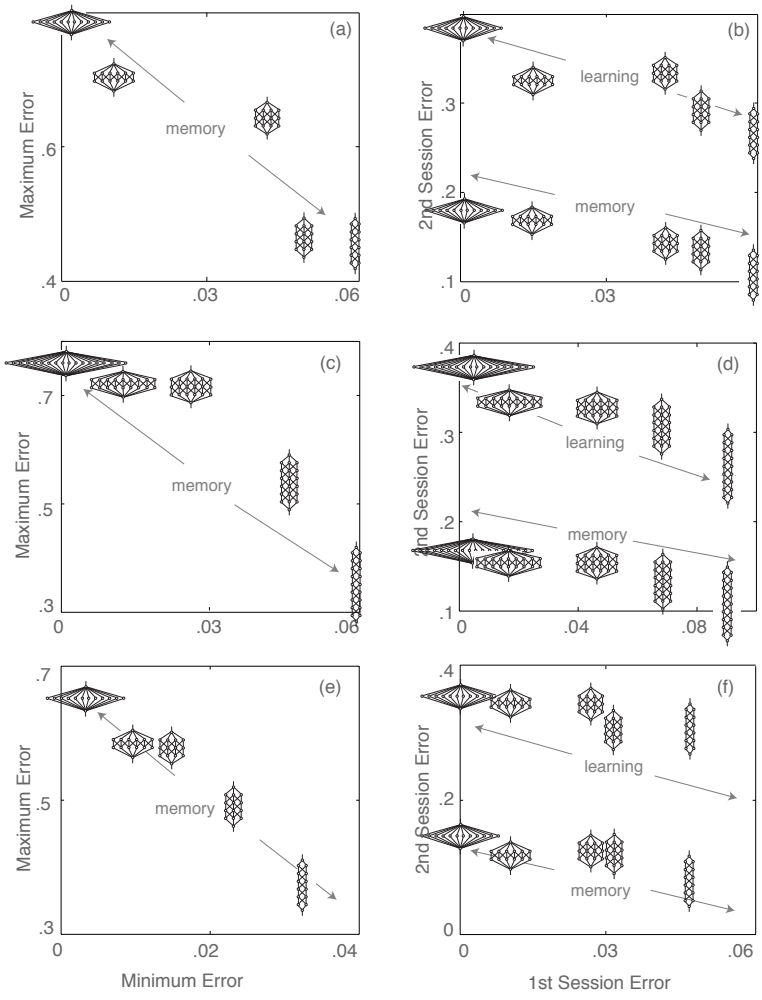

Figure 3.4: Impact of Network Size and Training Function on Performance Tradeoffs. Tradeoffs between best and worst memory performance (left column) and between first and second training sessions (right column), where network performance measured with respect to the errors $\{E_n^{(o)}\}$ and $\{E_n^{(n)}\}$. Panels (a) and (b) are identical to those shown in Figure 2.15 of Chapter 2 for the original 12-node networks. We see that the observed tradeoffs in learning and memory are consistent across (c,d) larger networks and with (e,f) permuted training points.

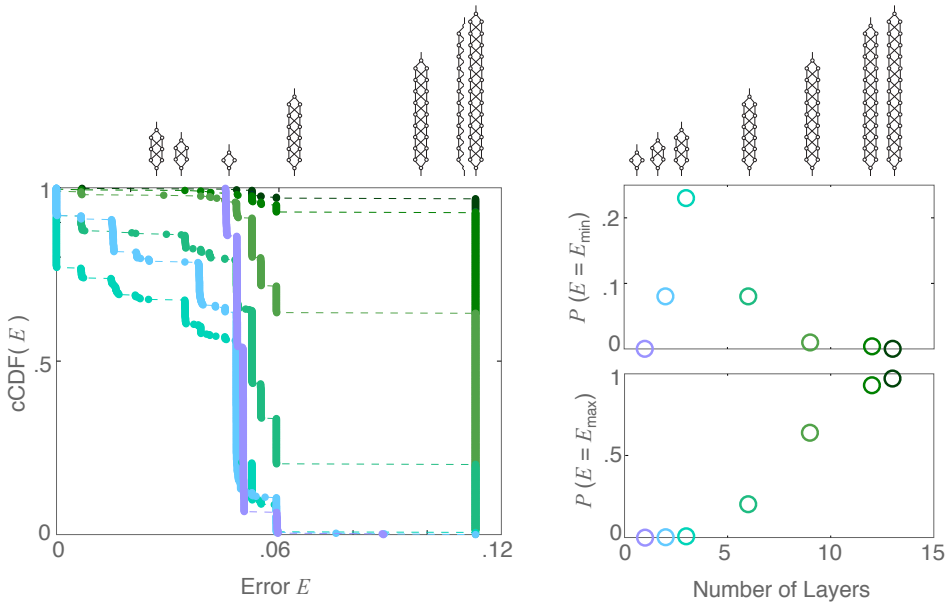

Figure 3.5: Dependence of Error Distributions on Stacked Network Size. Distributions and extremal values of error $\{E_o^{(o)}\}$ produced by stacked networks with a fixed width ℓ but with increasing height h. The single layer (2 node) network produces nearly constant error values, while all other networks produce a range of error values. None of the networks exceed the maximum error produced by fitting the points with a horizontal line. The probability of finding a minimum versus maximum error value, denoted respectively by the probabilities $P(E = E_{\min})$ and $P(E = E_{\max})$, are shown in the righthand panels and are calculated from the frequency of errors $E_{\min} = 0$ and $E_{\max} = .1131$. The probability of achieving a minimum error value increases for increasing $h \leq 3$ but decreases for increasing $h > 3$. In comparison, the probability of achieving a maximum error value steadily increases for increasing h.

Figure 3.5 shows the distribution of error values produced by each of the 7 networks. For all but the single-layer network, networks produce a wide range of error values that span from a minimum value $E_{\text{min}} = 0$ to a maximum value $E_{\text{max}} = E_{\text{line}} = .1131$ (minimum error achieved by fitting the points with a horizontal line). While networks with comparable numbers of nodes were shown in the previous chapter to produce error values exceeding E_{line}, the finding that none of the stacked networks exceeds this value suggests that the maximum error is fixed by the width, and not the height, of the network.

As all but the single-layer network span the full range of error values, we can compare the frequency with which each network produces the best- versus worst-case solutions by computing the probabilities $P_{\text{min}} = P(E = E_{\text{min}}$ and $P_{\text{max}} = P(E = E_{\text{max}})$ (right panels of Figure 3.5). We find that increasing the network height h improves performance for $h \leq 3$ by increasing the probability P_{min} of zero-error solutions while keeping P_{max} near zero. In contrast, increasing $h > 3$ worsens performance by both decreasing P_{min} and increasing P_{max}. Together, these results suggest that, for a given width, there is a network height $h \geq 1$ that maximizes performance. Furthermore, the observed monotonic increase in P_{max} for increasing h suggests that there is a fixed limit above which stacked networks can only produce linear solutions.

These results motivate the identification of scaling relationships across variations in the network height h. In an effort to isolate these relationships, we shift our focus to the single-layer (1×2) and double-layer (2×2) networks. Comparison of the distribution of error values produced by these two networks show that each network produces a relatively

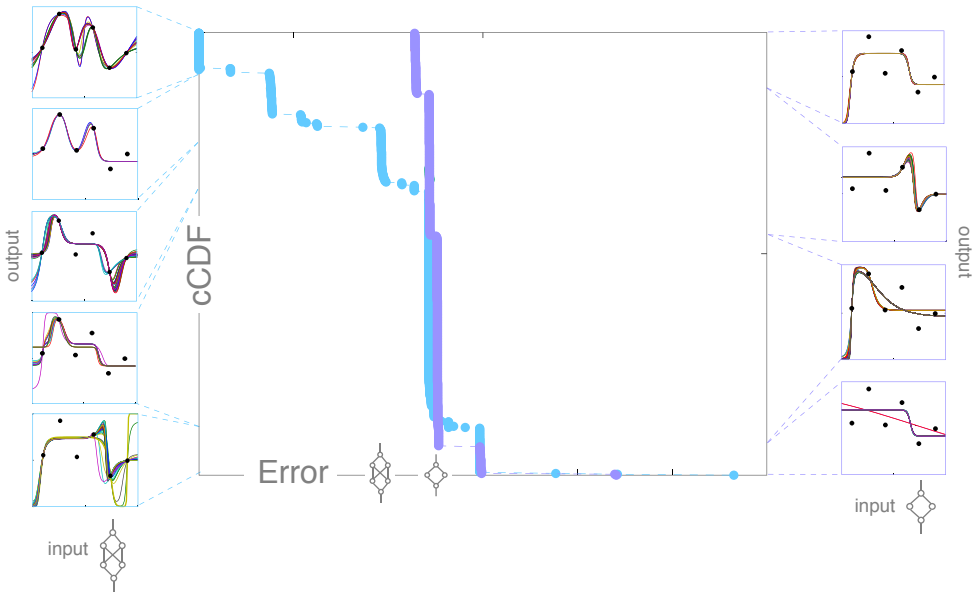

Figure 3.6: Restricted Solution Space of Small Stacked Networks. Comparison of solutions and error values produced by single-layer versus double-layer networks. Both networks produce discrete sets of error values, which correspond to distinct sets of solutions (indicated to the left and right of the main distributions). The single-layer network produces four sets of solutions (right panels), while the double-layer network produced five sets of solutions (left panels).

discrete set of errors, and these errors correspond to qualitatively distinct solutions (Figure 3.6). Furthermore, the solutions produced by the single- versus double-layer network show similarities in qualitative features. For example, the third set of solutions produced by the double-layer network shares the same qualitative features as the second and third solutions produced by the single-layer network. These similarities suggest that double-layer network solutions could be composed of linear combinations of single-layer network solutions. This is by no means an obvious statement, as the double layer network in an inherently non-linear extension of the single-layer network. Rather, this could only be achieved through constrained relationships in weight space.

To gain a better understanding of how single- versus double-layer network solutions could relate to each other through constrained weight space relationships, we examine the combinations of connection weights that produce the sets of solutions shown in Figure 3.6. These combinations are shown in Figures 3.7 and 3.8 for the single- and double-layer networks, respectively.

Inspection of Figure 3.7 reveals that the set of four single-layer solution types are produced by very constrained weight-space relationships. Furthermore, solutions with similar qualitative features are in closer proximity in weight space. The second and third solution sets, which both show tighter kinks than the first and fourth solution sets, are characterized by smaller input weights but larger output weights. In addition, the magnitudes of the two output weights tend to be proportional to one another. In comparison, the first and fourth solution sets show larger input but smaller output weights,

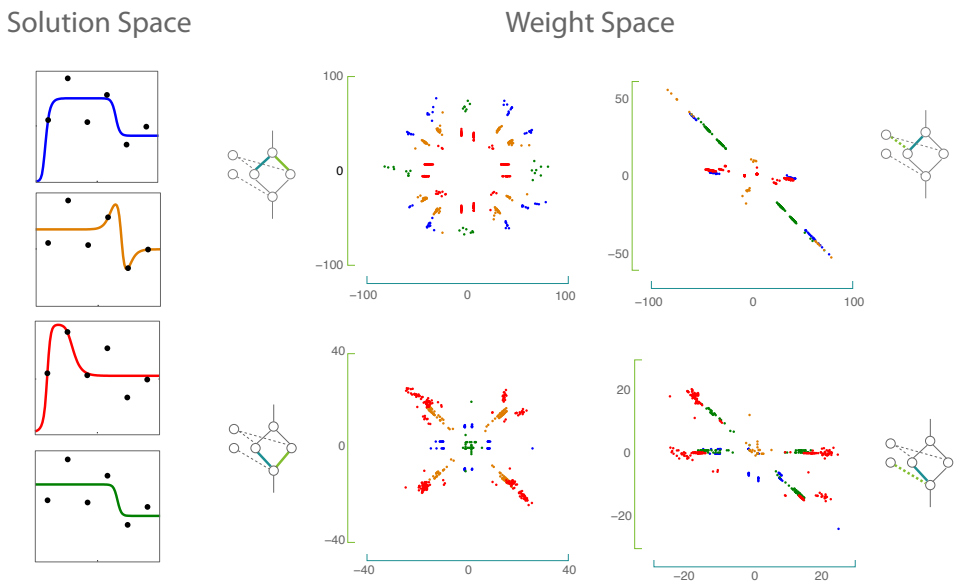

Figure 3.7: Restricted Weight Space of Single-Layer Network. Relationships between weights in the single-layer network that produce the sets of solutions shown in the right panels of Figure 3.6. Top row: projection of first input weight onto second input weight (left) and bias weight (right). Bottom row: projection of first output weight onto second output weight (left) and bias weight (right). For all weight combinations, the weight space is sparsely sampled. Solutions with similar qualitative features show similar relationships in weight space.

and weight strength tends to be localized across variations in one output weight rather than spanning linear combinations of both output weights.

Comparison of Figures 3.7 and 3.8 reveals that the double-layer network is much less constrained in the weight-space relationships that it exhibits. The input and output layers show more constrained localization of the weights than does the hidden layer, which shows a high degree of intermixing between weights. Comparison between the one-layer and two-layer sets of weight space relationships suggests the composition of two-layer solutions from one-layer solutions may be achieved by different combinations of one-layer weight-space relationships .

Together, these findings provide insight into the limitations of performance imposed by small network modules. As this work is still in progress, future work could provide a more quantitative assessment of the relationships between the solution space produced by single- and double-layer networks. While it may not be possible to predict the solution space of double-layer networks from the performance of single-layer networks, it may be possible to identify constraints on the size of solution space that could be achieved, and similarly the localization or delocalization of the corresponding occupation in weight space. The identification of such a recursion relationship between networks of increasing numbers of layers would be extremely useful in predicting the performance of larger network structures.

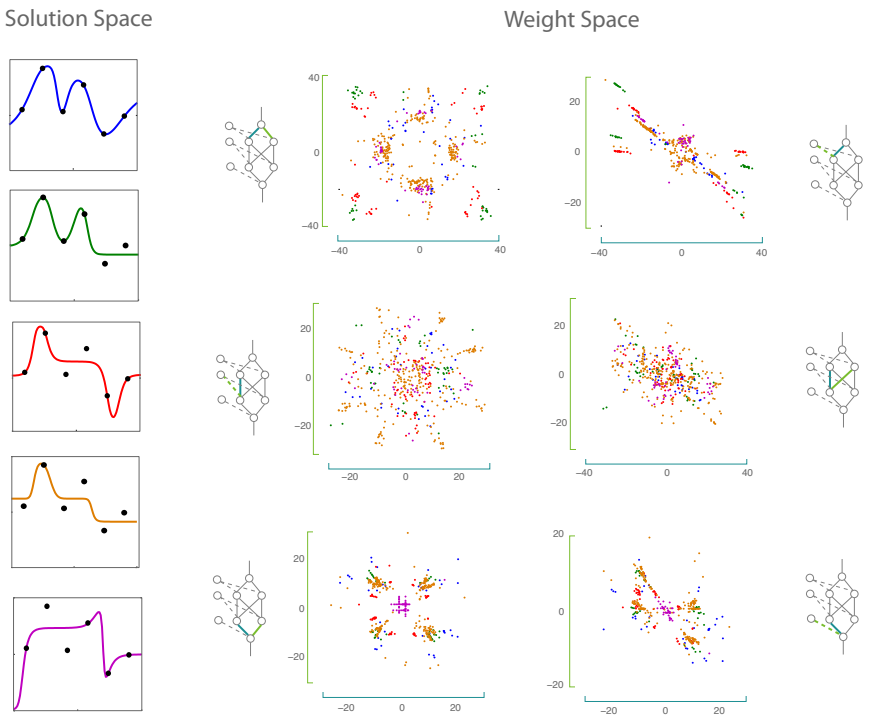

Figure 3.8: Restricted Weight Space of Double-Layer Network. Relationships between weights in the double-layer network that produce the sets of solutions shown in the left panels of Figure 3.6. Top row: projection of first input weight onto second input weight (left) and bias weight (right). Middle row: projection of first hidden weight onto second hidden weight (left) and bias weight (right). Bottom row: projection of first output weight onto second output weight (left) and bias weight (right). Comparison with Figure 3.7 shows that the double-layer network more densely samples weight space than does the single-layer network. Across solutions produced by the double-layered network, the input and output weight space is more sparsely sampled than is the hidden weight space.

3.4 Dependence on External Structure: Higher-Dimensional Training Functions

Both the previous sections and the previous chapter focused on the sequential learning of one-dimensional functional approximations. Increasing the complexity of the training pattern introduces new questions about the environmental features to which the network is sensitive. One can ask, for example, whether different features of an image are more or less relevant for successful learning versus successful memory, and whether these features are processed within different structural components of the network.

The direct extension of our analysis from one- to two-dimensional training functions would involve training networks on a string of raw pixel intensities (Figure 3.9), where pixel intensities are analogous to the y-values of the training points in Figure 2.13. However, this approach necessarily eliminates the spatial relationships between pixels that could in principle carry important information about the structure of the training function. Sensitivity to different spatial features, such as spatial correlations between pixels, spatial frequencies, and boundaries between groups of pixels, may differentially impact the ability of a network to learn and remember the training function. Furthermore, as suggested by our previous findings, different network structures may vary in their sensitivity to these different features. By simultaneously varying both internal network structure and external environmental structure, we can probe the extent to which internal network features take advantage of external environmental features during learning. The

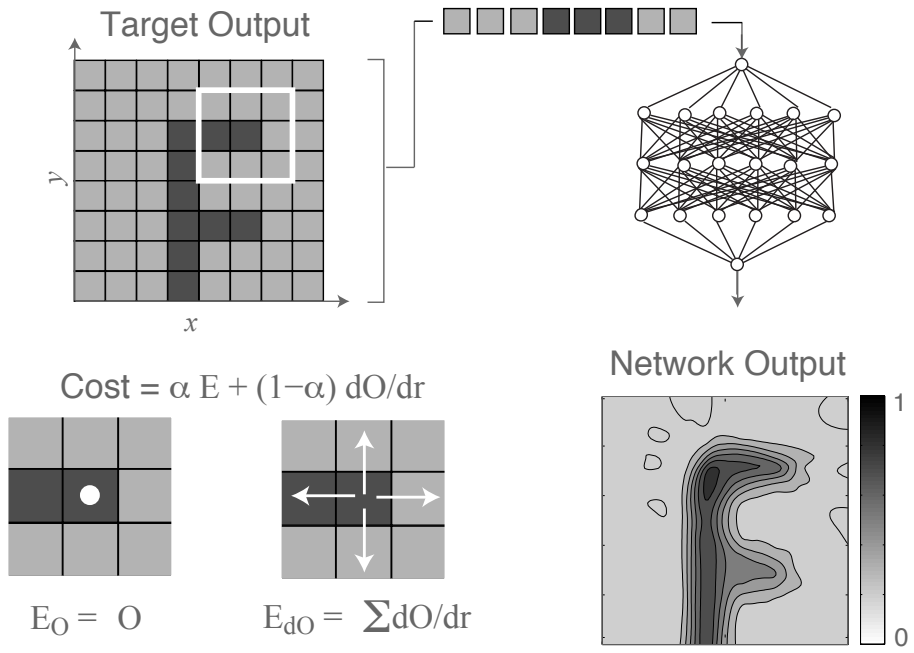

Figure 3.9: Illustration of Training Task. We choose to train networks on an 8×8 pixel greyscale image of the letter "F". As a basis network module, we consider a 18-node, 3-layer fully-connected network. Image pixels are fed sequentially into the network, with coordinates x_p and y_p fed into separate input nodes. The output of the network, when probed across continuous x and y values, produces a functional approximation of the training image. In addition to the standard cost E_O assigned to the raw pixel intensity values, and we assign an additional cost E_{dO} to spatial relationships between pixels. As the name would suggest, E_dO measures the error in the spatial derivative of pixel intensities measured from a given pixel across its four face-adjacent pixels. The total cost is a weighted combination of E_O and E_{dO}, where the weighting parameter α specifies the contribution from the raw pixel cost E_O.

extension of these ideas to competitive learning and memory processes, not discussed here, is a potential avenue for future research.

To this end, we extend our one-dimensional training function to a two-dimensional image. The spatial dimensions of this image are $x_p \times y_p$, which respectively give the number of pixels along the x- and y-dimensions and are analogous to the x_d values used in the one-dimensional case. The pixel intensity $z(x, y)$ is analogous to the previously-used values y_d and can be either binary (for a black and white image) or continuous (for a greyscale image). We consider an 8×8 grid of pixels, and we select pixel intensities to draw the letter "F", as shown in Figure 3.9. We choose to use the pixel values of 0.3 and 0.7, which lie well within the dynamic range of the sigmoid function, rather than the binary values of 0 and 1, which lie at the sigmoid extremes. Extension to more complicated images is the subject of future work.

Within our network model, the sensitivity to environmental features is controlled by the error E. The assignment of an error to spatial, rather than raw, input information alters the features of the error landscape such that navigation in the gradient direction moves the network toward regions of parameter space with lower error in the spatial properties of the external environment. Such an error then serves as a measure by which we can assess successful learning of spatial, rather than raw, information.

As a first step toward understanding the role of spatial information on learning and memory, we consider the case in which the total error, which we will now refer to as the cost C, is computed from a weighed combination of raw and spatial information

(illustrated schematically in Figure 3.9):

$$C(\boldsymbol{\omega}, \alpha) = \alpha E_O(\boldsymbol{\omega}) + (1 - \alpha) E_{dO}(\boldsymbol{\omega}), \tag{3.1}$$

where E_O is the error in the raw pixel intensities, analogous to the error previously assigned to our one-dimensional training functions:

$$E_O(\boldsymbol{\omega}) = \frac{1}{2} \sum_{d=1}^{N_d} \left(z_d(\boldsymbol{\omega}) - z_d \right)^2 . \tag{3.2}$$

The error E_{dO} could represent any spatial information of interest. We choose to assign to E_{dO} the error in the spatial derivative of face-adjacent adjacent pixel intensities:

$$E_{dO}(\boldsymbol{\omega}) = \frac{1}{2} \sum_{n.n.} \left(\frac{\partial z(\boldsymbol{\omega})}{\partial r} - \frac{\partial z}{\partial r} \right)^2 \tag{3.3}$$

The parameter α controls the relative weighting between spatial and non-spatial errors. In the case of $\alpha = 1$, the total cost C reduces to the least squares error in the output, as was used in Chapter 2.

The inclusion of a weighting parameter between different types of information introduces several questions. Should this weighting be fixed, and if so, at what value? If this weighting is variable, should it vary temporally during the learning process, spatially across the network structure, or evolve according to some internal or external rule? By assessing the performance of the network across variations in this weighting, we can explore whether performance is dependent not only upon sensitivity to different types of information, but also upon the time and location at which this sensitivity is introduced.

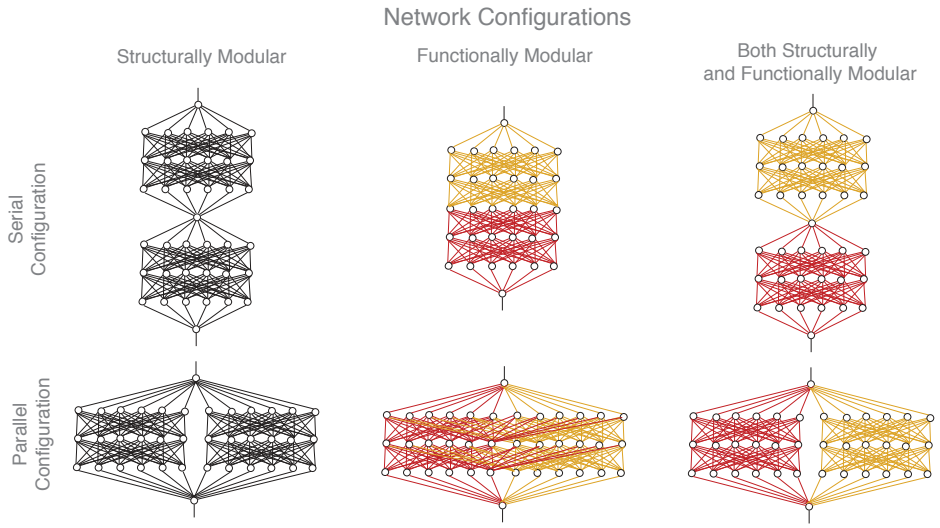

Figure 3.10: Structural versus Functional Modularity. Possible serial and parallel arrangements of two fully connected, 3x6 network modules. Networks can show structural modularity (left column), functional modularity (middle column), or both structural and functional modularity (right column). Structural modules are distinguished by a topological discontinuity in connectivity. Functional modules are distinguished by the implementation of a different cost function to the update of weights within each module.

To address these questions, we assess the performance of networks constrained to move within a fixed error landscape $C_L = .5 * E_O + .5 * E_{dO}$ that is equally weighted toward raw versus spatial information. The enables us to use the same measure of performance, namely the fixed external cost C_L, across all networks.

Internal sensitivity to raw versus spatial information can then be controlled by the internal cost C given by Equation 3.4, a quantity that determines how the network navigates the fixed landscape C_L. Variations in α alter the landscape direction along which network weights are updated. In this way, local sensitivity, as governed by $C(\boldsymbol{\omega}, \alpha)$, can differ from global sensitivity, as governed by $C_L(\boldsymbol{\omega})$.

We investigate network performance across different implementations of α by com-

bining variations in α with variations in structural and functional modularity. This approach allows us to vary both the location and degree of internal sensitivity to spatial information implemented within the network. We consider the simple case of two fully-connected network modules, each of size $h \times \ell = 3 \times 6$, that can be structurally or functionally distinct and can be arranged in parallel or in series. If arranged in parallel, structurally distinct modules are connected at the input and the output layer but are not connected within hidden layers. If arranged in series, structurally distinct modules are connected via one intermediate node that collects all output from module and relays it as input into the second module (Figure 3.10).

Functional modularity is defined such that distinct modules obey different cost weighting schemes and are therefore distinguished by their sensitivity to raw versus spatial information. When implemented within the network, functional modularity equates to the update of module-specific connection weights along different directions within the error landscape.

As mentioned previously, there are several ways in which we could choose the parameter α. Figure 3.11 highlights possible weighting schemes in α, along with possible configurations of structural and functional modularity in which these weighting schemes could be implemented. We distinguish between four separate cases in which two modules are (*i*) structurally and functionally integrated, (*ii*) structurally distinct but functionally integrated, (*iii*) structurally integrated but functionally distinct, and (*iv*) structurally and functional distinct. We further distinguish between cases in which the weighting is fixed

during the training process ($\alpha = [0, .5, 1]$) from those in which the weighting varies as a function of iteration number ($\alpha = \alpha(i)$). We first prescribe a specific function $\alpha(i)$ that is either monotonically increasing or decreasing in i, such that the network is smoothly biasing toward either raw or spatial information as training progresses. We later suggest a method by which α could evolve during the learning process.

Together, the number of possible arrangements of structural and functional modules, combined with the number of potential weighting schemes for α, gives a wide range of possible network configurations to assess. However, we find that two types of configurations are highly unsuccessful in learning external representations.

The first unsuccessful configuration is the set of structurally distinct modules arranged in series (regardless of α). The inability of this configuration to learn information likely arises as a result of combining and redistributing network outputs, which creates an information bottleneck that limits the impact of weight adjustments. As a test of the effect of this bottleneck, future work could incrementally decrease the number of nodes in a intermediate layer that lies between the two modules. This could be used to identify the minimum degree of structural separation that still allows for successful learning.

The second unsuccessful configuration is the set of modules that are initially (or indefinitely) sensitive solely to spatial information. Because E_{dO} contains no information about absolute pixel intensity, the navigation of the error landscape based solely on spatial information can lead the network to produce solutions that do not mirror absolute intensity values.

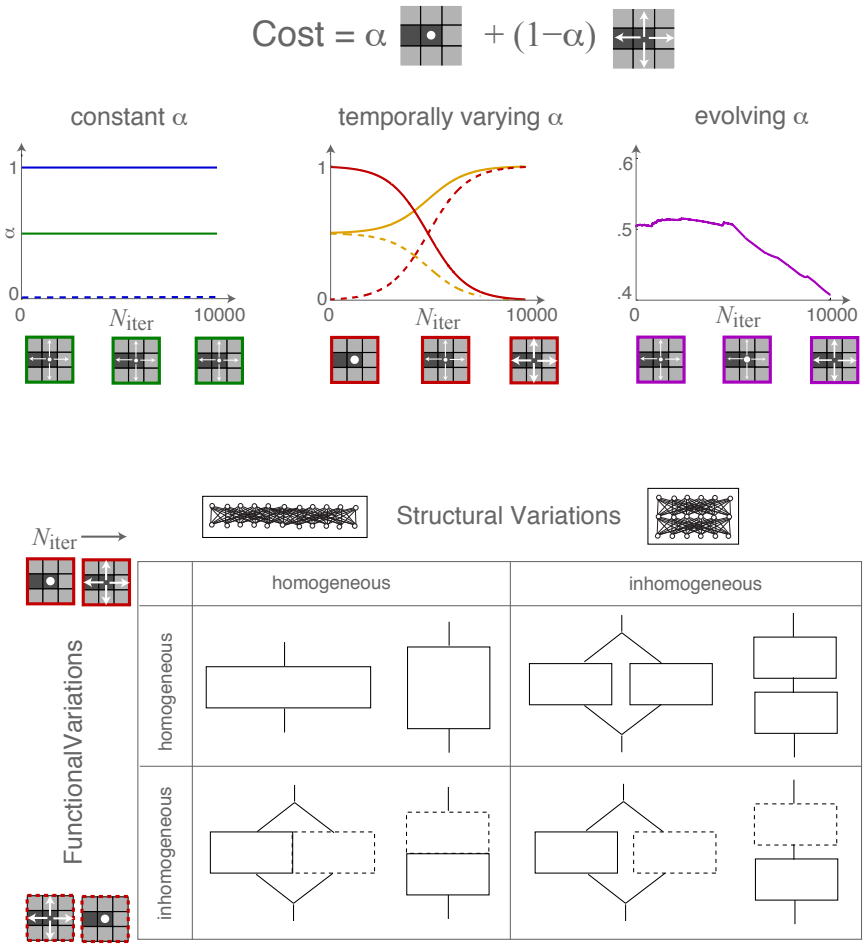

Figure 3.11: Possible Schemes for Assigning Variable Costs. Top row: we consider three possible schemes for varying the weighting parameter α as a function of iteration number N_{iter}: fixed, temporally varying, and evolving. Below each scheme is an illustration of the relative contribution to the cost from raw pixel intensity versus spatial derivatives in pixel intensity. Pairs of solid and dashed lines represent complementary schemes. Bottom row: we implement these schemes within different types of the structurally and functionally homogeneous and inhomogeneous networks shown in Figure 3.10. Functionally distinct modules, which need not be structurally distinct, implement different schemes for varying α.

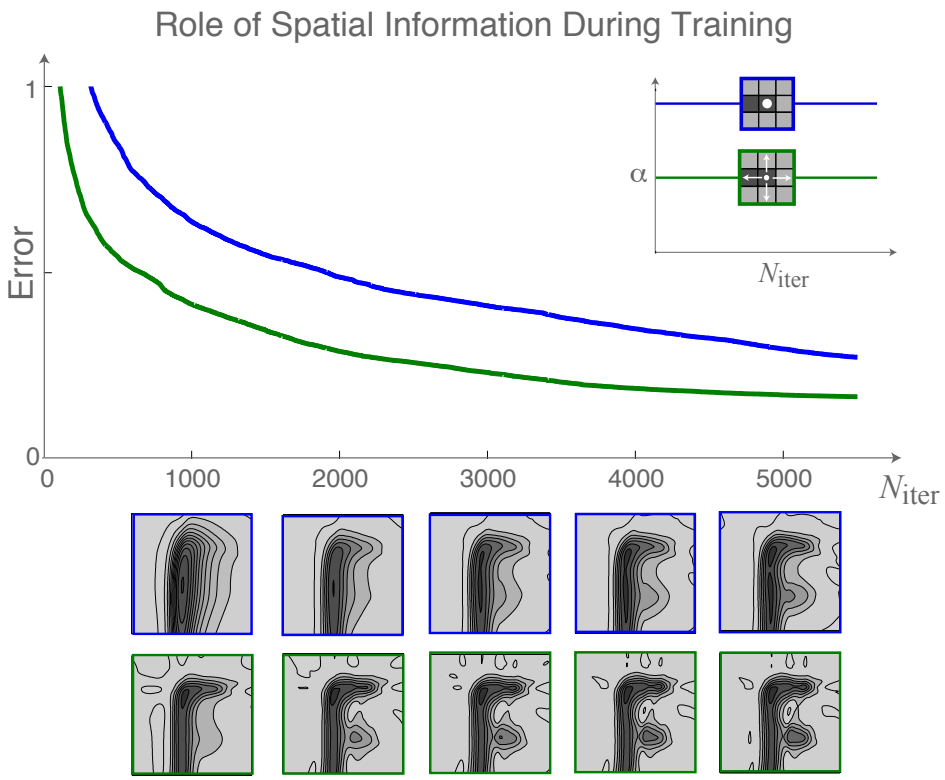

Figure 3.12: Impact on Learning of Sensitivity to Spatial Information. Representation of training function (letter "F") at different point in the training process for $\alpha = 1$ (blue) and $\alpha = 0.5$ (green). The inclusion of spatial information increases the speed with which the network learns the training information and results in a more accurate final representation.

Across the remaining configurations, we find that the functional consequence of including sensitivity to spatial information vary depending on how and where this sensitivity is implemented within the network.

As a preliminary assessment of the role of spatial sensitivity, we compare the performance of a single network module whose internal cost is uniformly sensitive ($\alpha = .5$) versus insensitive ($\alpha = 1$) to spatial information. Figure 3.12 compares, between these two cases, the cost C_L as a function of iteration number i averaged across 100 trials. We

see that sensitivity to spatial information improves the rate of learning (rate of decrease in cost) and results in a more accurate final representation of the training function.

Given that sensitivity to spatial versus raw information can improve performance when implemented within a structurally homogeneous network, we further explore whether this improved performance depends on the structural configuration of the network. Figure 3.13 compares three cases in which the functional modularity is held fixed, with each of two functional modules either increasing or decreasing in α from $\alpha = .5$. Across these three cases, we vary the configurational arrangement of functional modules between a serial arrangement (structurally homogeneous) and parallel arrangement (both structurally homogeneous and inhomogeneous).

We find that the network shows poor performance when it exhibits both structural and functional modularity. In comparison, both the serial and parallel arrangements of functional modules perform well when implemented within a structurally homogeneous network. However, the parallel configuration of functional modules has a higher rate of learning than the does serial configuration, and the final output more closely resembles, from visual inspection alone, the desired output "F".

Together, these results show that sensitivity to spatial information improves the rate of learning, and the extent to which sensitivity improves overall accuracy depends on how it is implemented within the network. This suggests that varied sensitivity to different types of external information is important for successful learning and may further impact ability of a network to retain information in subsequent training sessions. However,

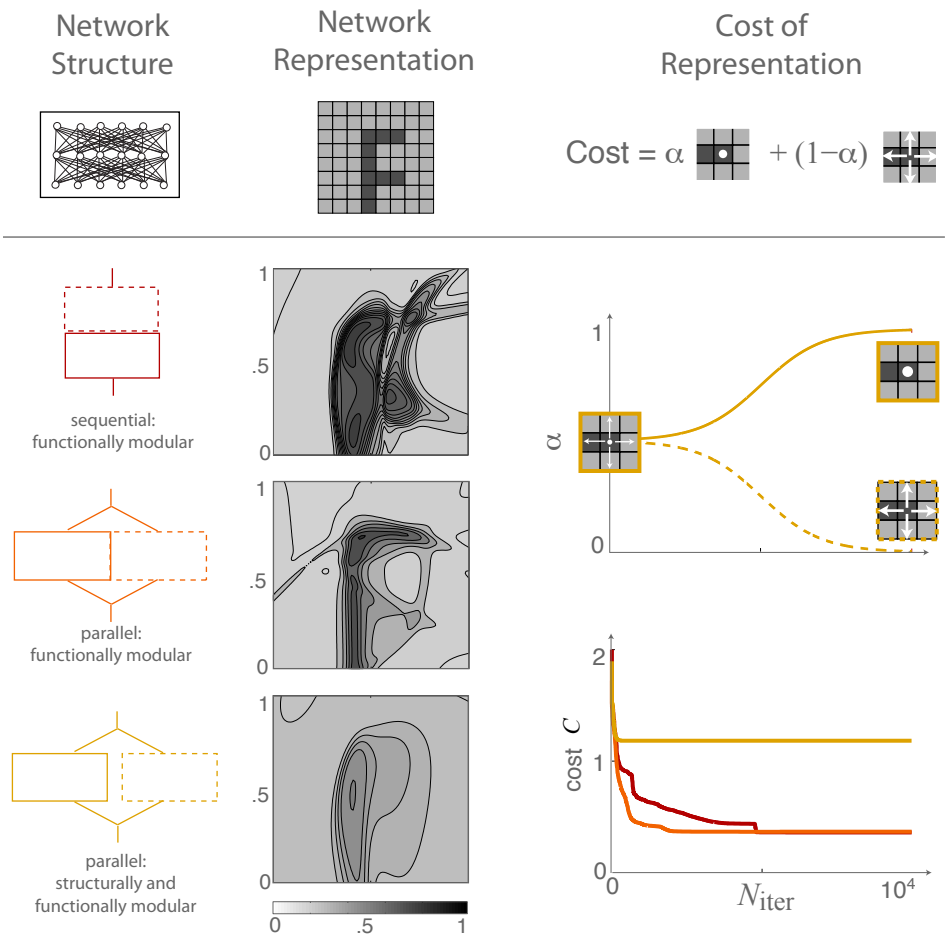

Figure 3.13: Performance of Structurally- versus Functionally-Distinct Modules. We compare three network configurations that exhibit varying degrees of structural and functional modularity (left column), with functionally distinct modules indicated by solid versus dashed lines. For all three network configurations, we implement the weighting scheme shown in the upper right. At the beginning of the training session, the cost of raw and spatial information are equally weighted within both modules ($\alpha = .5$). As training progresses, one module biases toward raw information (solid curve), while the second module biases toward spatial information (dotted curve). Comparison of network performance reveals that both structurally uniform networks achieve low final errors, while the structurally modular network plateaus at a high error value. Comparison of the two structurally uniform networks reveals that, although both networks achieve high accuracy, the parallel configuration achieves a representation that more closely resembles the letter "F".

biological systems are marked by their ability to dynamically interact with and adapt to their surroundings. This suggests that internal learning rules should be continually evolving during the learning process rather than following a fixed prescription.

One way in which this adaptation could be achieved is by allowing α to evolve during training, whereby the network is allowed to continually adjust its sensitivity to spatial versus nonspatial information. We propose one possible scheme for the evolution of α, in which $\alpha(i)$ is updated at each iteration in order to bias toward the larger of the two error components in our cost function:

$$\alpha(i+1) = \begin{cases} \alpha(i) + d\alpha(E_O - E_{dO}), \text{if } E_O > E_{dO} \\ \\ \alpha(i) - d\alpha(E_{dO} - E_O), \text{if } E_O < E_{dO} \end{cases} \tag{3.4}$$

where $d\alpha$ is a parameter that sets the rate of evolution. This evolution process enables α to act as a reinforcement plasticity mechanism that strengthens beneficial sensitivity and weakens detrimental sensitivity.

This work is still in its initial stages, and we have yet to understand how network performance differs between the cases in which $\alpha(i)$ is prescribed versus allowed to evolve dynamically. Importantly, this could help understand how sensitivity to different types of information changes (or should ideally change) at different points in the learning process. Furthermore, we are only beginning to understand interactions between structural and functional modularity, and further investigation could explore more complex interactions between different modules. Combined, these studies provide insight into the interaction

between internal and external structure, and they inform learning models in which systems are differentially sensitive to many types of information.

3.5 Discussion

The varied work presented in this chapter suggests several possible directions for future research. As adaptability and evolvability are inherently important biological processes, it is of significant importance to understand how these processes are encoded in network structure and function. Furthermore, as these processes are implemented within a given network architecture, it is of equal importance to understand the limitations imposed by this architecture. Future work could solidify the relationships between network modules of varying sizes, as these could be combined to created a diverse range of complex architectural motifs. Such studies could be complimented with a more thorough investigation of the different types of information to which a network is sensitive, such as spatial frequencies and higher order correlations, which would further inform our understanding of the external environmental features that shape and constrain adaptable system function.

Chapter 4

Structural and Functional Connectivity of the Human Brain

"The number of possible "on-of" patterns of neuronal firing is immense, estimated at a staggering ten to the millionth power. The brain is obviously capable of an imponderably huge variety of activity; the fact that it is often organized and functional is quite an accomplishment!"

–Daniel J. Siegel, from *The Developing Mind*, 1999

4.1 Introduction

Human cognitive function is supported by large-scale interactions between different regions of the brain. The anatomical scaffolding that mediates these interactions can be described by a structural connectome that maps the spatial layout of hard-wired white

matter tracts [82]. Structural connectivity, defined by the physical properties of these direct anatomical connections, supports the relay of electrical signals between brain regions. Functional interactions can similarly be described by a functional connectome that maps the strength of task-dependent communication between brain regions. Functional connectivity, defined by coordinated changes in energy consumption that support local neural activity [83], is thought to reflect the strength of anatomically-mediated signal transmission between brain regions. However, the indirect inference of structural and functional connectivity from different experimental techniques raises two complementary questions about the quantitative relationships between structural and functional connectomes: (*i*) to what extent can the task-dependent strength of functional interaction between brain regions be predicted from structural connectomes, and (*ii*) to what extent can the hard-wired properties of anatomical connections be inferred from functional connectomes?

Connectomes, whether examined at the neural or systems level, are networks whose structural properties, such as the length, number, and spatial location of connections, can differentially impact functional properties, such as the strength of local or global network activity. In the human brain, long-distance connections are thought to assist in the integration of multiple sensory modalities [84], while local dense connections are thought to support a balance between information integration and segregation [59]. However, the extent to which variations in structure and function are differentially related on a large scale in the human brain is not well understood. Previous studies of structure-

function relationships have described the properties of specific anatomical connections [43] that mediate interactions between select brain regions [85, 86, 87, 88]. Similar analyses of large-scale structural and functional networks have been conducted across parcellations of the cortical surface [89] but have been limited to small sample sizes and to resting-state neural activity.

To address this gap in understanding, we combine the specificity of detailed anatomical and functional analysis with the statistical power of 84 subjects measured noninvasively at rest and during the performance of attention- and memory-demanding tasks. Functional interactions are predicted from diffusion tensor imaging (DTI) measurements of white matter connectivity, while structural properties are inferred from functional magnetic resonance imaging (fMRI) measurements of changes in blood oxygenation level dependent (BOLD) signals [90]. This two-pronged approach isolates consistent and complimentary relationships between structural and functional connectivity across subjects and brain states.

In what follows, we provide an overview of the principles behind magnetic resonance imaging that enable the inference of structural and functional pathways in the brain. We then introduce a novel set of multimodal approaches for relating structural and functional connectivity in human brain networks. We compare the subject-dependent structural measures of tract length and tract number, thought to differentially impact sensory processing [84, 59], with the subject-dependent functional measures of resting-, attention-, and memory-state BOLD correlations, thought to reflect the coordinated and

task-dependent control of different brain regions. We further distinguish between the subject-independent measures of inter- versus intra-hemispheric connectivity, known to exhibit different structural properties that could impact functional interactions (e.g. [91]). Together, these delineations enable (*i*) the prediction of task-dependent functional connectivity from underlying anatomical properties and (*ii*) the inference of structural connectivity from functional interaction strength. Resting-state function is found to be supported by local dense intra-hemispheric connectivity and length-independent inter-hemispheric connectivity. In comparison, task-driven function is found to be supported by long-range intra-hemispheric connectivity. These relationships, consistently maintained across subjects, provide insight into the architectural design of the human brain and the functional capabilities and constraints imposed by its architecture.

4.2 Magnetic Resonance Imaging

The problem of mapping the structural and functional architecture of the human brain faces significant experimental limitations because, with the exception of postmortem studies or rare cases of medical monitoring, we do not have direct access to the brain (for good reasons–imagine the implications of such direct access). Instead, we rely on non-invasive measurements acquired via magnetic resonance techniques. Such noninvasive techniques are advantageous because they enable the acquisition of neural data in awake, behaving subjects without damage to brain tissue. This disadvantage, however, is that the properties of brain tissue must be inferred indirectly from these measurements. The ques-

tion then arises as to what these techniques are measuring, and how these measurements inform our understanding of human brain architecture.

To understand how magnetic resonance techniques provide estimates of structural and functional brain connectivity, we first review the physics behind nuclear magnetic resonance (NMR) on which magnetic resonance imaging (MRI) relies.

4.2.1 Overview of MRI Principles

MRI techniques are used to reconstruct the average magnetization of hydrogen atoms within localized regions of tissue. These techniques are particularly sensitive to variations in soft tissue found, for example, in the heart and brain. Measurements are acquired from localized volumes of tissue, or voxels (volume-pixel), within narrow tissue slices, and slices are later combined to reconstruct three-dimensional tissue properties.

These measurements rely on the principle of NMR, which describes the precession of nuclear spins in an external magnetic field about the field vector $\boldsymbol{B}$ at a frequency proportional to the field strength $|\boldsymbol{B}|$. MRI techniques measure nuclear magnetization by first aligning nuclear spins with an external magnetic field ($1.5 - 3T$) and then perturbing spin alignment through the controlled application of electromagnetic pulses. The perturbed spins create local variations in magnetic field strength that can be detected by external antennas.

To illustrate this process (see Figure 4.1), consider a spin aligned along the longitudinal direction defined by the external field $\boldsymbol{B}$. This spin can be excited by an electromagnetic

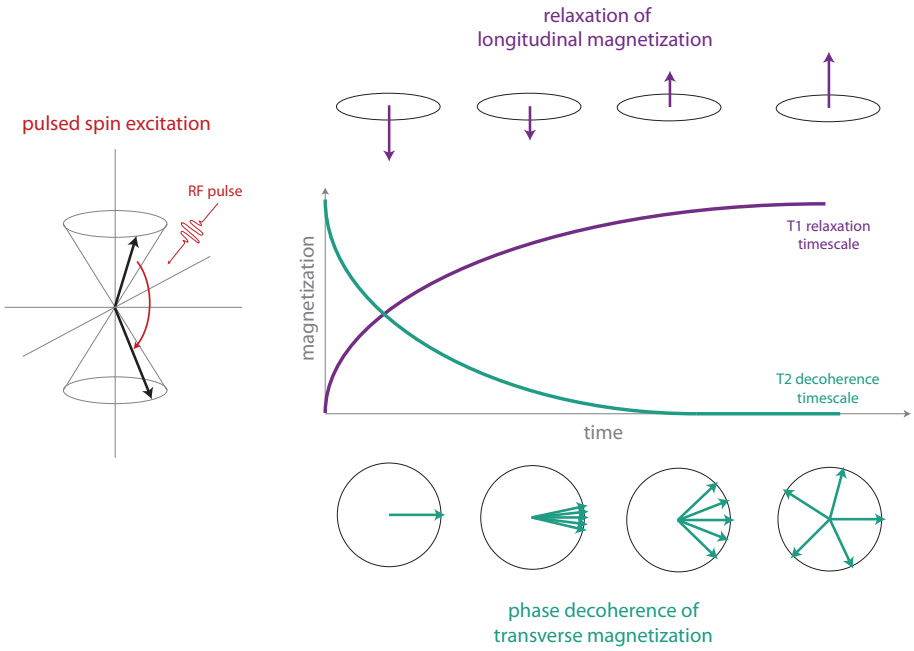

Figure 4.1: Relaxation of Longitudinal and Transverse Magnetization. A radio frequency (RF) pulse tuned to the appropriate frequency can be used to excite nuclear spins, thereby anti-aligning the spins with the external magnetic and inducing phase precession. Over time, individual spins will lose energy via spin-lattice interactions, and they will lose phase coherence via spin-spin interactions. Given a local spin density, energy loss results in a net increase in longitudinal magnetization, while phase decoherence results in a net decrease in transverse magnetization. The timescales over which the longitudinal and transverse magnetization relax are termed the T_1 and T_2 timescales, respectively.

pulse of radio frequency (RF) to a higher energy state. This pulse can act to flip the spin, such that it is antialigned with and precesses about the longitudinal direction. This precession introduces a new transverse component of magnetization. Once the RF pulse has passed, the spin will decay back to its original state through energy exchanges with the local surrounding lattice of molecules, and it will similarly lose phase coherence with other spins. Given a local spin density, this will result in a gradual increase in the net longitudinal magnetization, and a gradual decrease in the net transverse magnetization. The timescale over which the longitudinal magnetization returns to $1 - 1/e = 63\%$ of its original value is defined to be the T1 relaxation timescale, which reflects local spin-lattice interactions. Similarly, the timescale over which transverse magnetization decay to $1/e = 37\%$ of its original value is defined to be the T2 relaxation timescale, which reflects local spin-spin interactions. T1 and T2 timescales vary between different tissue types, (e.g. liquid has relatively long T1 and T2 timescales, while fat has short timescales), enabling the differentiation between different tissue types based on the relaxation properties of local magnetization.

The repeated application of RF pulses can be used to control local changes in magnetization. Consider, for example, two tissues (A and B) with different T1 relaxation timescales. Given a sufficient amount of time following the first RF pulse, both tissues will have fully returned to their original longitudinal magnetization. However, if a second RF pulse is applied after a repetition time $TR < T1$, the magnetization of the two tissues will have relaxed to different values, and they will therefore show different responses to

the application of the second pulse. In this way, the time TR between pulses can be used to control the relative magnetization within, and therefore the relative signal transmitted by, different tissues.

In one application of this technique (Figure 4.2), the initial RF pulse is set to align the spin with the transverse plane, such that there is no longitudinal magnetization (90° RF pulse) . After the the pulse is shut off, precessing spins will begin to lose phase coherence over a timescale T2* (T2 star). A second pulse set to 180°, applied after an echo time TE, will cause the spins to reverse the direction of their precession such that they regain phase coherence. This refocusing process, called a spin-echo, allows for the reapplication of RF pulses needed to amplify the transmitted signal. Although this spin-echo pulsing can be repeated many times, the net signal decays over a timescale $T2 > T2*$.

The controlled measurement of transverse magnetization via RF pulsing underlies a wide range of neuroimaging techniques. We focus on two techniques, diffusion tensor imaging (DTI) and functional magnetic resonance imaging (fMRI).

4.2.2 Diffusion Tensor Imaging

Diffusion tensor imaging (DTI) measures variations in the local microstructure of brain tissue. Microstructural variations alter the diffusion of water molecules in the brain, which can in turn be measured via MR techniques.

Water molecules in the brain diffuse via Brownian motion. In the absence of microstructural tissue variations, this diffusive process would be random and would not

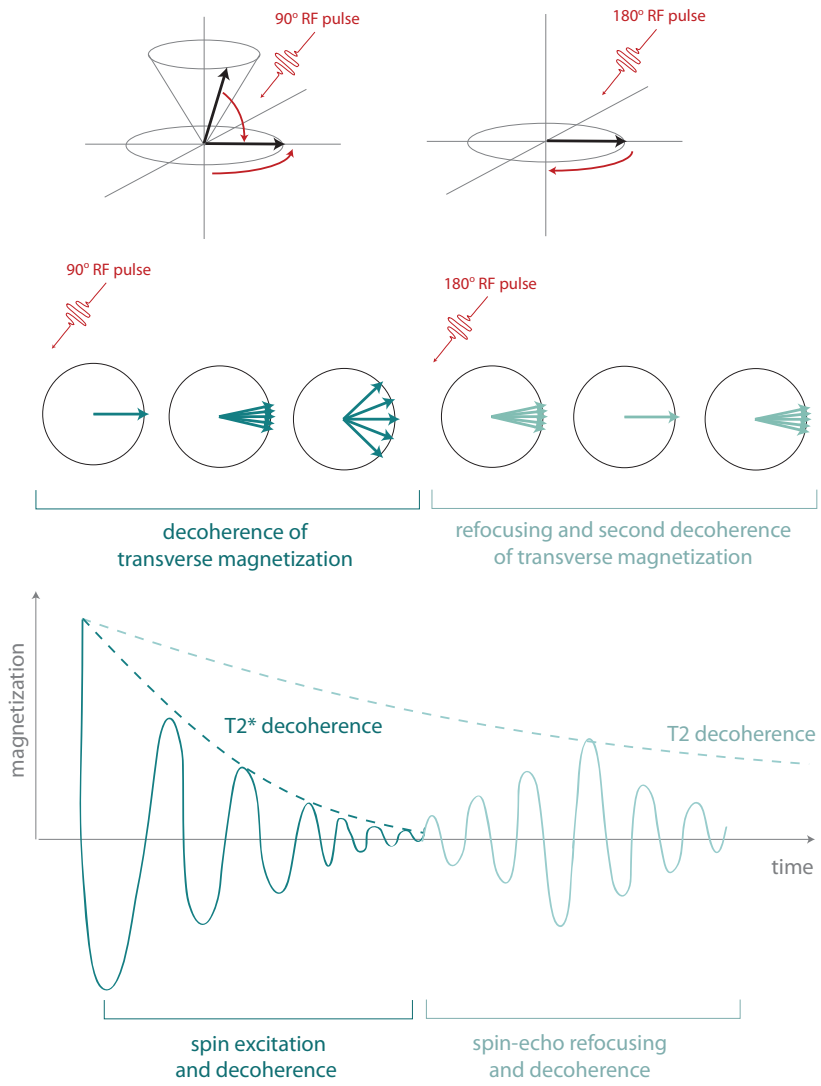

Figure 4.2: Spin-Echo Excitation. The application of a 90° RF pulse polarizes nuclear spins in the transverse plane and causes them to coherently precess in the counterclockwise direction. Phases begin to decohere over the T_2^* timescale, after which a second 180° RF pulse reverses the direction of precession and causes the phases to refocus in a "spin-echo" response. This is followed by a second decoherence, after which the process can be successively repeated. Each spin-echo serves to amplify the total signal transmitted by the transverse magnetization. However, the amplitude of each individual spin-echo decreases in magnitude over the T_2 decoherence timescale.

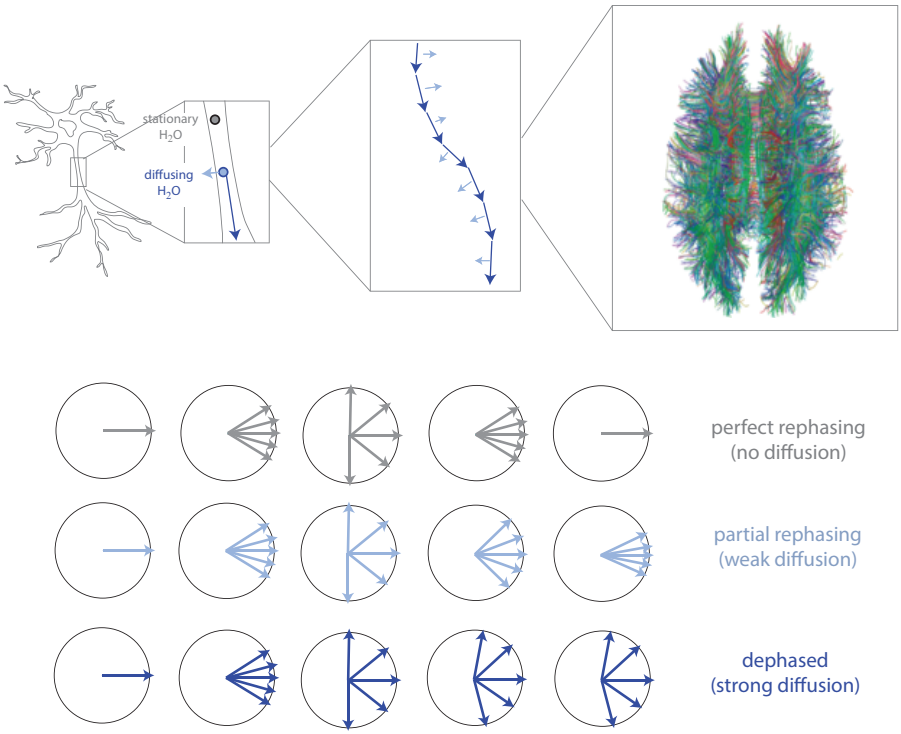

Figure 4.3: Diffusion Tensor Imaging. DTI measures the preferential diffusion of water molecules in different tissues. In a uniform tissue, water would show no preferential diffusion. However, when microstructural tissue variations are present, as occurs at the boundary of white matter tracts, water diffuses more rapidly along the tract versus perpendicular to the tract. Under the application of a pulsed magnetic field gradient, water molecules that have remained stationary will show perfect rephasing, while molecules that have traversed longer distances will show partial or no rephasing. The relative difference in phase coherence can then be detected by the MR scanner.

exhibit a preferential diffusion gradient. As the rate of water diffusion varies within different types of brain tissue, however, preferential diffusion gradients can be observed at tissue boundaries.

This diffusion anisotropy is particularly useful for identifying the boundaries of white matter tracts in the brain. These tracts consist of bundles of myelinated axons along which water diffuses more rapidly than it would in a orthogonal direction. DTI measurements infer the direction and degree of net diffusion within individual tissue voxels, which can then be used to reconstruct anatomical fiber pathways.

The diffusive properties of molecules can be measured by applying a 90° RF-pulsed magnetic field gradient, which aligns nuclear spins in the transverse plane and causes precession at different rates. The application of a second 180° RF-pulsed field gradient can then be used to realign the spins, as described previously. If water molecules have moved in time between the first and second pulses, this realignment will be imperfect, resulting in a detectable MR signal. Strongly preferential diffusive processes, such as those along while matter tracts, produce strong local signals that can be used to map tract boundaries (Figure 4.3).

4.2.3 Functional Magnetic Resonance Imaging

Functional MRI (fMRI) measures local changes in blood oxygenation, which are in turn thought to reflect local changes in neural activity.

The relationship between neural activity and blood oxygenation can be understood

in terms in local energetics. Active brain cells require more energy, in the form of glucose and oxygen, than inactive brain cells. This energy is delivered to cells in the form of oxygenated blood, which results in a local (2-3mm) increase in blood flow, a local expansion of blood vessels, and a local net increase in oxygenated hemoglobin Hb (the oxygen carrier in blood). Changes in neural activity can therefore by characterized by changes in oxygenated blood flow to localized brain regions.

Changes in blood oxygenation are detectable by MRI measurements due to the differences in magnetic properties between oxygenated and deoxygenated blood. Deoxygenated blood (deoxyhemoglobin, dHB) is paramagnetic, while oxygenated blood (hemoglobin, Hb) is diamagnetic. dHB therefore interacts with and distorts the local magnetic field such that nuclear spins decohere more quickly in the presence of dHB than in the presence of Hb. As a result, regions with higher Hb content produce stronger MR signals, enabling the mapping of local changes in blood oxygen content (Figure 4.4).

4.3 Network Approaches for Studying the Human Brain

The magnetic resonance imaging measurements discussed in the previous section are used to measure the local properties of brain tissue. Variations in these properties can be related to underlying structural inhomogeneities, such as the physical boundaries between axons and their surrounding tissue, that affect the translational motion of molecules. Similarly, local tissue properties can reflect functional responses, such as the degree of energy consumption, that affect the local density of different types of molecules. Our goal

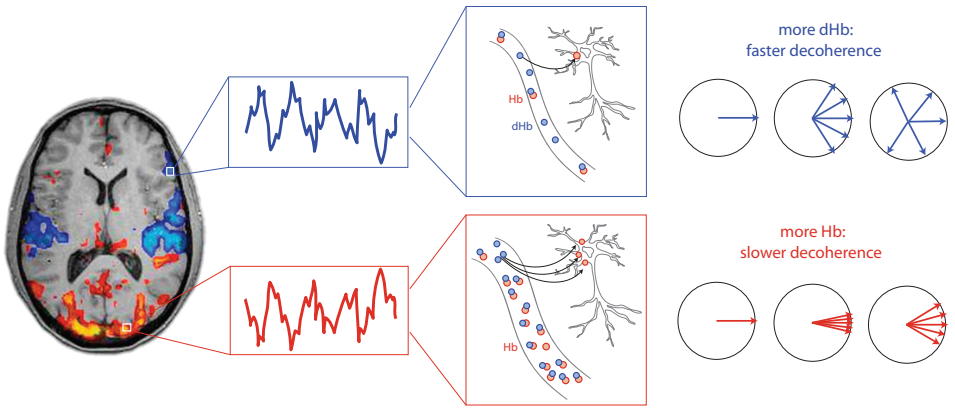

Figure 4.4: Functional MRI. fMRI measures fluctuations in blood oxygen content, which reflect local changes in neural activity. When neurons become more active, they require more oxygen, which is carried in Hemoglobin (Hb) in the bloodstream. This results in a local expansion of blood vessels with a higher content of Hb relative to dHb (deoxygenated hemoglobin). As Hb and dHb differ in their magnetic properties, the average magnetization in a region with Hb content shows slower phase decoherence than does a region with high dHb content. This relative difference in phase coherence can then be detected by the MR scanner.

is to use these local measurements to gain an understanding of the large scale structural and functional properties of the human brain.

Recall that DTI measurements record local diffusion gradients within a single tissue voxel, and these gradients reflect the boundary between white matter (composed of myelinated axons) and the surrounding tissue. If diffusion gradients in two adjacent voxels are largely overlapping, the assumption is made that the axon boundary extends between voxels along this gradient direction. Tractography, a computational method that searches for significant gradient overlap, can be used to reconstruct large-scale white matter pathways linking different regions of the brain from voxel-wise measurements of water diffusion.

Similarly, recall that fMRI measurements record local fluctuations blood oxygenation within single tissue voxels, quantified in terms of a blood oxygen level dependent (BOLD) signal. This signal varies over time and during the performance of different tasks (for example, moving an arm with elicit increases in the BOLD signal within the motor cortex). The activity of a local region of the brain can be described by the regional mean time series averaged over constituent voxels, and interactions between brain regions can be described by correlations between regional mean time series. In this way, the coordinated control of different brain regions can be constructed from voxel-wise measurements of fluctuations in blood oxygenation.

Given that communication between brain regions is mediated by physical connections, we expect functional interactions, as measured via fMRI, to reflect the existence of structural connections, as measured via DTI. While one would no doubt question that the function of the brain should be reflective of its underlying structure, neither the extent to which this relationship can be observed experimentally nor the specific features of these relationship are known. The identification of such large-scale structure function relationships is therefore important for both assessing the experimental techniques used to acquire these measurements and for understanding the specific constraints that anatomical architecture imposes on human cognitive function.

Importantly, the brain is neither fixed in structure nor static in function. Anatomical pathways can differ between individuals, as can the ways in which these pathways are utilized for different cognitive functions. Structure can be altered during development,

learning, and aging and can be disrupted by neurological disorders, each of which can have significant functional consequences. The identification of robust relationships between structural and functional brain connectivity is therefore made difficult both by the variability in these properties observed across individuals and in the experimental cost of neuroimaging measurements required to obtain a significant cross-section of the population. Recent analyses of large-scale connectivity have therefore been limited in statistical power and in scope to a small number of subjects, to specific brain regions, and to resting-state activity. As a result, our comparison of widespread structural and functional connectivity across 84 subjects across different task states significantly contributes to the current understanding of human brain architecture.

We rely on power tools drawn from network theory for the assessment of structure-function relationship within and across individuals. The nature of the experimental measurements at hand lends itself to network descriptions, for which localized brain regions are represented as nodes, and the strengths of interaction (structural or functional) between brain regions are represented by weighted, undirected connections between nodes. In the following subsections, we describe the specific methodological considerations needed for mapping the human brain onto structural and functional brain networks.

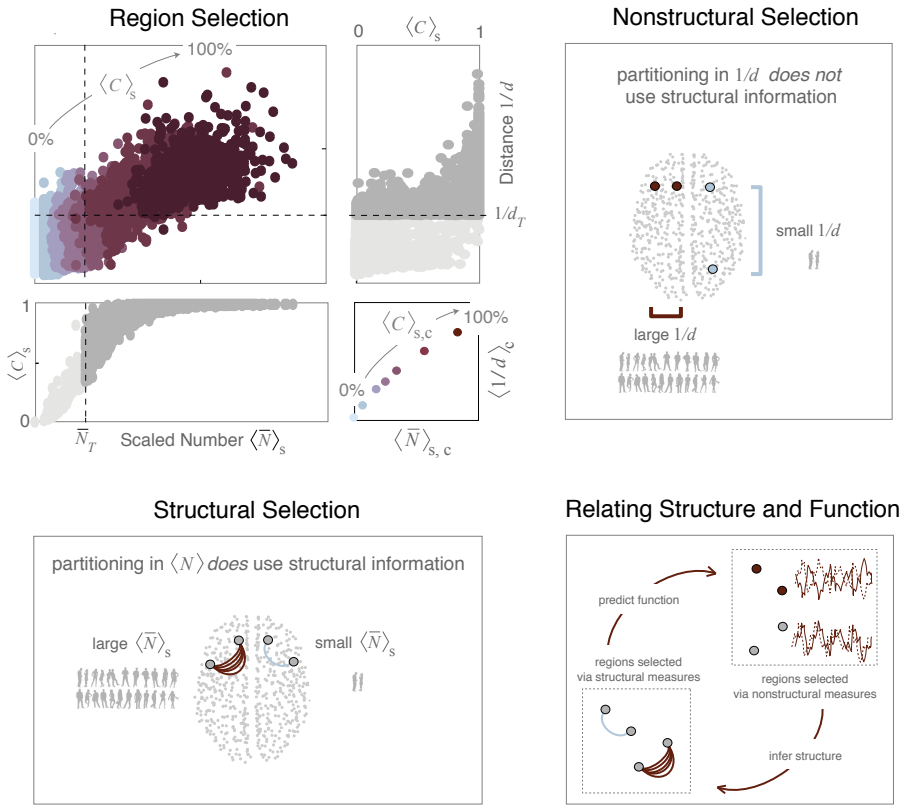

Figure 4.5: Selection of Robustly-Connected Region Pairs. Upper left: Consistency in connectivity $\langle C \rangle_s$ increases as a function of scaled number $\bar{N}$ and inverse interregional distance $1/d$, with average values $\langle C \rangle_{s,c}$ indicated in the lower right inset. Horizontal and vertical projections of $\langle C \rangle_s$ are shown in grey as a function of $\bar{N}$ and $1/d$, respectively. Note that variations in $\bar{N}$ more tightly constrain $\langle C \rangle_s$ than do variations in $1/d$. We impose thresholds $\bar{N}_T$ and $1/d_T$ (dashed lines) to select two largely overlapping subsets of region pairs with high $\langle C \rangle_s$, whereby regions linked by high rescaled numbers $\bar{N}$ (lower left) and within small interregional distances (upper right) are consistently present across a large number of subjects. Regions selected via $\bar{N}_T$ are used to predict FC from SC, and regions selected via $1/d_T$ are used to infer SC from FC (lower right).

4.3.1 Structural versus Functional Brain Networks

The construction of brain networks relies on the appropriate choice of brain regions to be used as network nodes, between which connectivity measures are then assessed. The Automated Anatomical Labeling (AAL) Atlas [92] describes a parcellation scheme in which regions are defined based on known anatomical boundaries. The 90 regions in this atlas consist of both cortical and subcortical structures, but they do not include the brainstem. Because they are derived from anatomical boundaries, the regions in this atlas vary significantly in size. As this variation has been shown to impact connectivity estimates, we choose to use an upsampled version of this atlas consisting of 600 regions that are similar in volume while remaining anatomically constrained. From this set of regions, we compute structural and functional networks for 84 individual subjects.

Structural brain networks are obtained from DTI measurements via a tractography algorithm used to identify white matter tracts linking brain regions. For each subject, we compute two measures of structural connectivity (SC): the total number N and average length L of tracts linking two regions. Qualitatively similar results were obtained by rescaling $N \rightarrow N/L$, a method that was previously suggested to account for bias in the tractography algorithm [43], and we therefore report results using the direct measure N. We additionally define a binary number C that specifies the presence or absence of SC between two regions, such that $C_{i,j} = 1$ if regions i and j are linked by one or more white matter tract, and $C_{i,j} = 0$ otherwise.

Functional brain networks are obtained from fMRI measurements of BOLD time

series. Pearson's correlations are computed between scale 2 wavelet coefficients of voxel-averaged time series measured within two regions. For each subject, we compute three measures of functional connectivity (FC): the correlation between two time series measured at rest (resting state), during the performance of an attention task (attention state), and during the performance of a memory task (memory state). Given that task-driven changes in FC are small relative to resting-state values [93], we compare the strength of FC measured at rest (rsFC) to that measured in deviations $\Delta\mathrm{asFC} = \mathrm{asFC} - \mathrm{rsFC}$ of the attention state (asFC) from rest and in deviations $\Delta\mathrm{msFC} = \mathrm{msFC} - \mathrm{rsFC}$ of the memory state (msFC) from rest.

In the following analysis, we identify structural properties that are predictive of function $(\mathrm{SC} \rightarrow \mathrm{FC})$ and functional properties that are indicative of structure $(\mathrm{FC} \rightarrow \mathrm{SC})$, assessing these relationships on both a representative and a subject-specific level.

Notation: In comparing different quantities, we compute the average $\langle O \rangle$ and standard deviation $\sigma(O)$ of a given quantity O. When computed across subjects, we reference the quantity with the subscript s (e.g. $\langle O \rangle_s$), and when computed across connections within a single subject, we reference the quantity with the subscript c (e.g. $\langle O \rangle_c$).

4.3.2 Selecting Robustly-Connected Region Pairs

The brain exhibits both sparse and variable connectivity, with far fewer anatomical connections than would be expected at random [47] and with patterns of connectivity that vary between individuals [82]. By isolating regions that are consistently connected across

many subjects, we can reliably compare representative and subject-specific relationships in SC and FC.

Of the possible 179700 pairings between 600 regions, less than 2% are structurally connected within a given subject. Even fewer region pairs are consistently connected across subjects, a discrepancy that may be mitigated by the use of Freesurfer surface-based region parcellations. To robustly relate SC and FC across subjects, we isolate region pairs that consistently show nonzero SC. The selection of region pairs need not be the same for the analyses of $\mathrm{SC} \rightarrow \mathrm{FC}$ and $\mathrm{FC} \rightarrow \mathrm{SC}$ so long as the former does not use information about FC and the latter does not use information about SC.

As a measure of inter-subject consistency in SC, we compute the fraction of subjects $\langle C \rangle_s$ that show nonzero SC between a given region pair. We find that this consistency increases with both the normalized number of tracts $\bar{N} = \langle N \rangle_s / \sigma_s(N)$ (a purely structural measure) and the inverse interregional distance $1/d$ (a purely metric measure that has been linked to single-subject SC [89]). We therefore impose the thresholds $\bar{N}_T = 0.6$ and $1/d_T = 0.1$ to select two largely overlapping subsets of region pairs for the separate analyses of $\mathrm{SC} \rightarrow \mathrm{FC}$ and $\mathrm{FC} \rightarrow \mathrm{SC}$ (upper left of Figure 4.5). Both subsets are similar in size (3085 versus 3079 region pairs, respectively) and average consistency $\langle C \rangle_{s,c}$ (86% versus 79%, respectively). Furthermore, because $\bar{N}$ and $1/d$ scale roughly linearly with one another, it is straightforward to appropriately tune $\bar{N}_T$ and $1/d_T$ to achieve a desired consistency and subset size while maintaining similar results (see Chapter 5 for further discussion).

Of the two measures, variations in $1/d$ less tightly constrain $\langle C \rangle_s$ than do variations in $\bar{N}$ (horizontal projections in Figures 4.6 and 4.8). d must therefore be restricted to relatively small values in order to achieve high consistency, thereby disfavoring the selection of long connections. Even so, a small fraction of the selected region pairs are not structurally connected within any subject, introducing a small amount of noise into the analysis of $\mathrm{FC} \rightarrow \mathrm{SC}$.

Note that qualitatively similar results can be achieved by selecting region pairs via the direct measure $\langle C \rangle_s$ (see Chapter 5). However, we choose instead to select region pairs via $\bar{N}$ and $1/d$ because this selection avoids two drawbacks of using $\langle C \rangle_s$ directly: (*i*) $\langle C \rangle_s$ inherently requires information about SC and is therefore less optimal than $1/d$ for the assessment of $\mathrm{FC} \rightarrow \mathrm{SC}$, and (*ii*) $\langle C \rangle_s$ lacks a single-subject correlate that would enable the direct extension of these methods to single-subject brain networks and is therefore less optimal than $\bar{N}$ for the assessment of $\mathrm{SC} \rightarrow \mathrm{FC}$.

4.3.3 Representative and Subject-Specific Brain Networks

We construct representative and subject-specific brain networks by respectively measuring subject-averaged and subject-specific strengths of SC and FC between consistently-connected region pairs. To assess $\mathrm{SC} \rightarrow \mathrm{FC}$ (or analogously $\mathrm{FC} \rightarrow \mathrm{SC}$), we partition each network into subgroups of connections within similar structural (functional) properties, and we compare the functional (structural) connectivity within partitioned subgroups. We apply the same partitions to both representative and subject-specific networks,

and we confirm that our analysis is robust to partition choices (see Chapter 5).

To compare the properties of partitioned subgroups of connections, we evaluate shifts in the cumulative distribution functions (CDFs) of a given quantity O, where O takes on representative or subject-specific values of SC and FC. The CDF(O), which measures the probability of finding $O > O^*$ for every value of O^*, enables the simultaneous comparison of different instantiations of the quantity O. In assessing representative networks, we report the full CDFs of $\langle O \rangle_s$. In assessing subject-specific networks, however, we report subject-specific distribution averages $\langle O \rangle_c$ to enable the comparison of distribution properties across subjects.

4.4 Predicting Function from Structure

4.4.1 Structural Partitions

Structural resources are unevenly distributed between different regions of the brain [94]. For example, a significant amount of material is dedicated to myelination in the corpus callosum. Similarly, certain brain regions are more densely or distantly connected than other regions [95]. We investigate the extent to which variations in these structural properties are reflected in the strength of communication between brain regions.

We find that a majority of consistently-connected brain regions are linked by few numbers of short connections. However, there exist regions that are densely linked by short connections or sparsely linked by long connections. Here, we use "dense" and

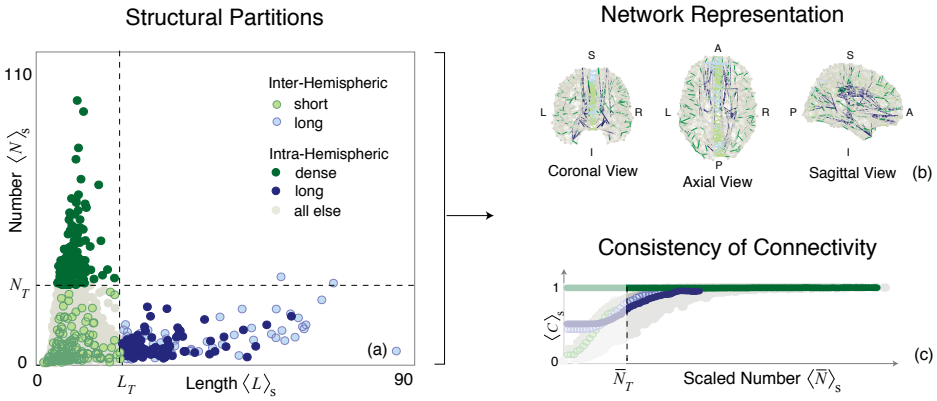

Figure 4.6: Structural Partitions in the Representative Brain. Number $\langle N \rangle_s$ versus length $\langle L \rangle_s$ of connections between region pairs selected via $\bar{N}_T$. We apply a length threshold $L_T = 20$ and a number threshold $N_T = 30$ (dashed lines) to separate long from short and dense from sparse connections, and we further distinguish inter-hemispheric connections (outlined markers). In combination, these partitions separate four non-overlapping subgroups of connections, short (light green) and long (light blue) inter-hemispheric connections and dense (dark green) and long (dark blue) intra-hemispheric connections, from the remaining bulk of short, sparse intra-hemispheric connections (tan). The average consistency of these connections (superimposed onto the horizontal projection in the upper left) increases with $\bar{N}_T$ and is largest within the subgroup of densely-connected regions. Coronal, axial, and sagittal views of structural and functional subgroups of connections. Grey nodes mark region centers, and lines mark region pairs that are linked by one or more tract in the representative brain. Note that curvilinear tracts are represented as straight lines.

"sparse" to refer to connections with numbers greater and less than a threshold value N_T, and we use "long" and "short" to refer to connections whose lengths are greater and less than a threshold value L_T. In combination with the delineation between inter- and intra-hemispheric connections, these partitions define four non-overlapping structural subgroups, long and short inter-hemispheric connections and long and dense intra-hemispheric connections, whose properties we compare to those of the remaining bulk of short, sparse intra-hemispheric connections (Figure 4.6).

Connections within each structural subgroup differ in their degree of consistency $\langle C \rangle_s$. Regions that exhibit dense connectivity are consistently connected within nearly all subjects, regardless of the selection threshold $\bar{N}_T$ (horizontal projection in Figure 4.6). The observed properties of this subgroup are therefore less susceptible to noise introduced by inconsistent SC. The remaining subgroups are consistently-connected within 85% of subjects, but $\bar{N}_T$ can be tuned to larger values in order to increase this consistency and reduce associated noise (see Chapter 5 for further discussion).

4.4.2 Functional Connectivity of the Representative Brain

In the resting state, we find striking differences in the strength of FC between regions linked by different types of structural connections (Figure 4.7a,d). All inter-hemispheric connections, regardless of length, show strong rsFC. The reduced sensitivity of inter-hemispheric correlations to variations in tract length may arise due to the insulating properties of heavy myelination along inter-hemispheric tracts. Dense intra-hemispheric

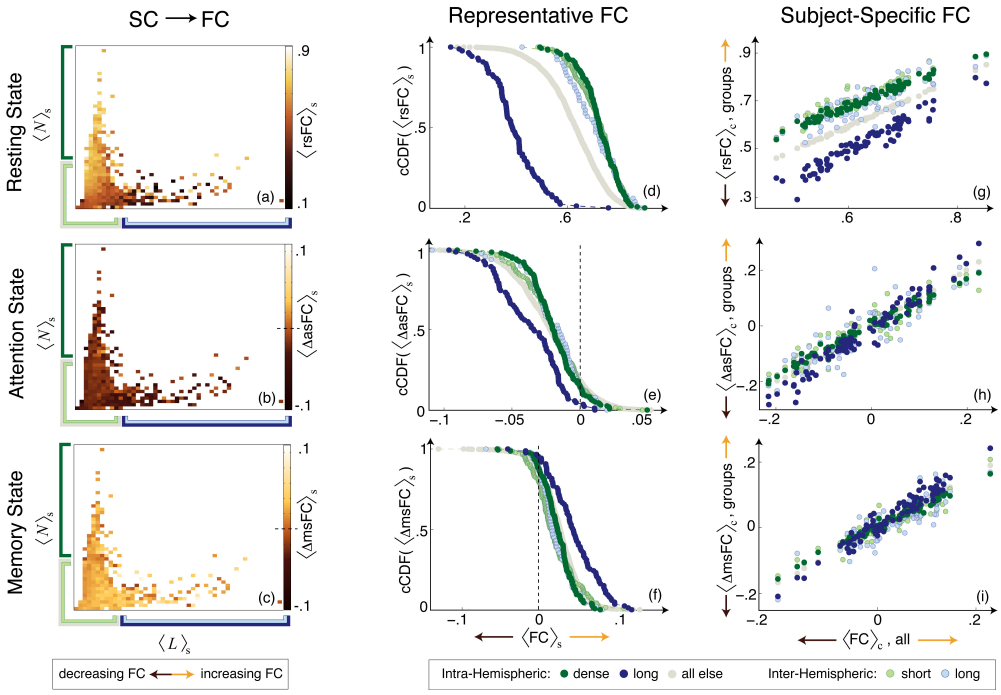

Figure 4.7: Functional Connectivity of Structural Subgroups. Functional measures $\langle \mathrm{rsFC} \rangle$ (top row), $\langle \Delta \mathrm{asFC} \rangle$ (middle row), and $\langle \Delta \mathrm{msFC} \rangle$ (bottom row) produced by structural subgroups of connections in representative and subject-specific brain networks. (a-c) Density maps of $\langle \mathrm{FC} \rangle_s$ versus $\langle N \rangle_s$ and $\langle L \rangle_s$ in the representative brain, with the partitions used to define structural subgroups indicated below and to the left of each subfigure. (a) In the resting state, we see significant variation in FC across variations in SC. During task performance, we see overall (b) suppression during attention and (c) activation during memory, with less variation in FC across variations in SC. (d-f) CDFs of $\langle \mathrm{FC} \rangle_s$ produced by structural subgroups of connections in the representative brain. (d) In the resting state, inter-hemispheric connections and dense intra-hemispheric connections show strong FC, while long intra-hemispheric connections show weak FC. During task performance, long intra-hemispheric connections show larger (e) decreases in $\langle \mathrm{FC} \rangle_s$ during attention and (f) increases in $\langle \mathrm{FC} \rangle_s$ during memory as compared to the remaining subgroups of connections. (g-i) Subject-specific values of $\langle \mathrm{FC} \rangle_c$ produced by structural subgroups of connections, where subjects are ordered by overall FC such that each vertical cross-section represents a single subject. (g) In the resting state, all subjects show strong $\langle \mathrm{FC} \rangle_c$ between inter-hemispheric and densely-connected intra-hemispheric regions, and all subjects show weak $\langle \mathrm{FC} \rangle_c$ between distant inter-hemispheric regions. (h-i) During task performance, all subjects show similar changes in the strength of $\langle \mathrm{FC} \rangle_c$ measured within different structural subgroups.

connections show similarly strong rsFC, a property that may reflect signal amplification from high numbers of connections (Figure 4.7e,f). Long intra-hemispheric connections, however, show notably weak rsFC despite being of similar length and number to the set of long inter-hemispheric connections. These observations extend beyond previous findings of increasing rsFC with decreasing interregional distance [89] to identify structural mechanisms that support strong rsFC between nearby inter- versus intra-hemispheric region pairs.

During task performance, we find that a majority of brain regions decrease in FC during attention tasks (Figure 4.7b,e) but increase in FC during memory tasks (Figure 4.7c,f) relative to their behavior at rest. Inter-hemispheric and dense intra-hemispheric connections, which displayed relatively strong rsFC, show similar changes in both asFC and msFC to the remaining bulk of connections. Long intra-hemispheric connections, however, show significant changes in FC between the two tasks, exhibiting stronger suppression in the attention state (Figure 4.7e) and stronger activation in the memory state (Figure 4.7f) than the remaining bulk of connections. The magnitude of these changes, which distinguishes between attention and memory tasks, is consistent across selection thresholds and becomes more pronounced when biasing toward long connections (see Chapter 5).

4.4.3 Individual Variability in Functional Connectivity.

The overall strength of FC varies significantly across subjects. Within subject-specific brain networks, however, we find that structural subgroups of connections show qualitatively similar *shifts* in FC to those observed in the subject-averaged representative brain network. In the resting state, all subjects exhibit strong correlations between inter-hemispheric and densely-connected intra-hemispheric regions, and they similarly exhibit weak correlations between distantly-linked intra-hemispheric regions (Figure 4.7g). In the attention and memory states, all subjects exhibit similar changes in FC produced by different structural subgroups (Figure 4.7h-i). Notably, the subgroup of long intra-hemispheric connections shows higher variability both across subjects (Figure 4.7g-i) and across connections within a given subject (see Chapter 5) as compared to the remaining structural subgroups.

4.4.4 Conclusions from SC to FC Analysis

The strong values of rsFC but consistent changes in asFC and msFC exhibited by inter-hemispheric and dense intra-hemispheric connections suggest that these connections support strong resting-state function. In contrast, the weak values of rsFC but large changes in asFC and msFC exhibited by long intra-hemispheric connections suggest that these connections support task-dependent changes in attention and memory function.

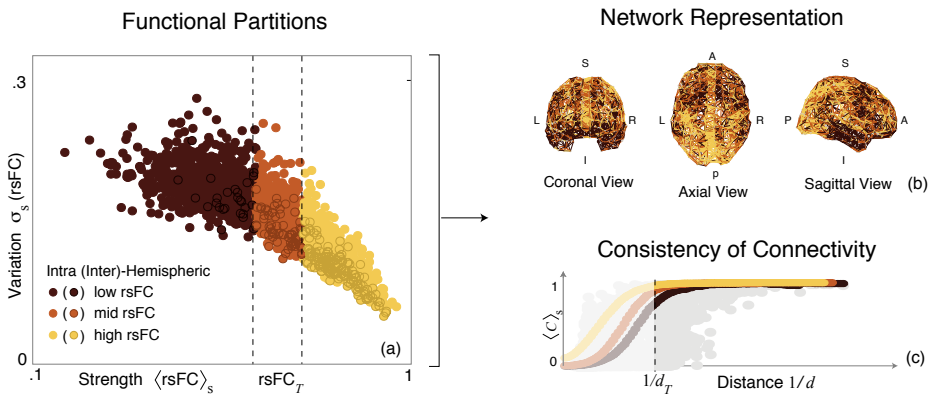

Figure 4.8: Functional Partitions in the Representative Brain. Inter-subject variance $\sigma_s(\text{rsFC})$ decreases for increasing FC strength $\langle \text{rsFC} \rangle_\text{s}$ between region pairs selected via $1/d$. We apply functional thresholds rsFC_T (dashed lines) to separate low (bottom 33% in brown), intermediate (middle 33% in orange), and high (top 33% in yellow) rsFC, and we further distinguish inter-hemispheric connections (outlined markers). The average consistency of these connections (superimposed onto the vertical projection in the upper left) increases with $1/d_T$ and is largest within the strongly correlated subgroup. Coronal, axial, and sagittal views of structural and functional subgroups of connections. Grey nodes mark region centers, and lines mark region pairs that are linked by one or more tract in the representative brain. Note that curvilinear tracts are represented as straight lines.

4.5 Inferring Structure from Function

The results of the previous section revealed that the structural features of anatomical connections differentially impact functional interactions between brain regions. As a stronger test of the relationship between SC and FC, we investigate whether the functional interactions can be used to infer underlying structural properties.

4.5.1 Functional Partitions

Given the pronounced separation in the resting-state properties of structural subgroups, we use rsFC to infer underlying SC. We apply fixed thresholds rsFC_T to separate

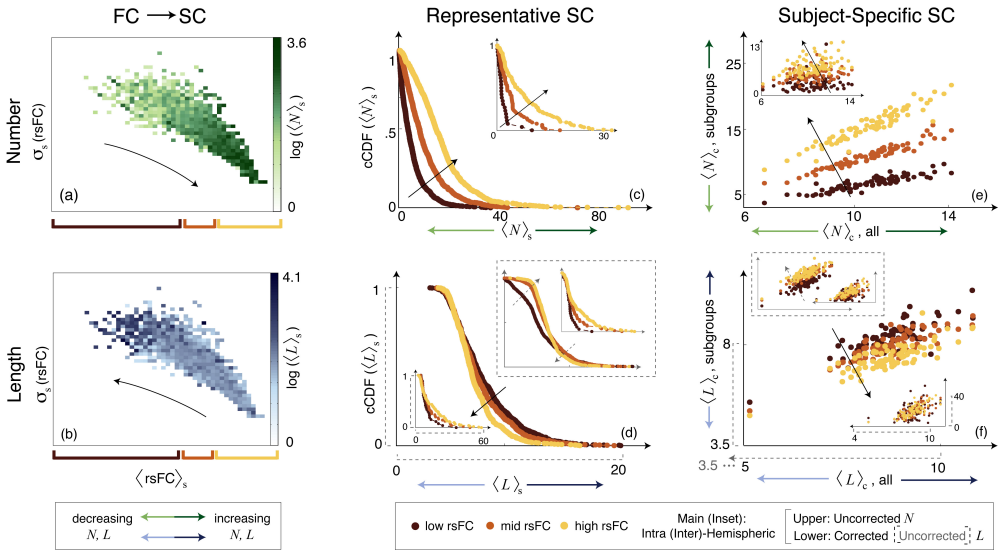

Figure 4.9: Structural Connectivity of Functional Subgroups. Structural measures $\langle N \rangle$ (top row) and $\langle L \rangle$ (bottom row) produced by functional subgroups of connections within representative and subject-specific brain networks, with the properties of inter-hemispheric connections shown as insets in subfigures c-f. All lengths are given in mm. (a-b) Density maps of $\log(\langle \mathrm{SC} \rangle_\mathrm{s})$ versus $\langle \mathrm{rsFC} \rangle_\mathrm{s}$ and $\sigma_s(\mathrm{rsFC})$, with functional partitions indicated below each subfigure. We see significant variation in SC across variations in rsFC, with rsFC tending to increase for (a) increasing $\langle N \rangle_s$ and (b) decreasing $\langle L \rangle_s$. (c-d) cCDFs of $\langle \mathrm{SC} \rangle_\mathrm{s}$ produced by structural subgroups of connections in the representative brain. (c) Increasingly large numbers of both inter- and intra-hemispheric connections support increasingly strong rsFC, and intra-hemispheric connections are more numerous, on average, than inter-hemispheric connections. (d) Corrected distributions of $\langle L \rangle_s$ show that increasingly strong correlations are supported by increasingly short intra-hemispheric connections and increasingly long inter-hemispheric connections. Uncorrected distributions are shown in the dotted inset (with axis scales indicated by dotted lines along the main axes), reveal that the inclusion of inconsistent connectivity alters the distribution of short lengths. (e-f) Subject-specific values of $\langle \mathrm{SC} \rangle_\mathrm{c}$ produced by functional subgroups of connections, where subjects are ordered by overall SC strength such that each vertical cross-section represents a single subject. (e) Across subjects, strong rsFC is supported by large numbers of inter- and intra-hemispheric connections. (f) Corrected values of $\langle L \rangle_c$ show that, across subjects, strong rsFC is consistently supported by short intra-hemispheric connections. Uncorrected values (dotted inset), reveal that the inclusion of absent connections reverses the relationships between intra-hemispheric connection length and rsFC strength.

weak (bottom 33%), intermediate (middle 33%), and strong (top 33%) rsFC, and we further distinguish inter- from intra-hemispheric interactions (Figure 4.8). In agreement with the results of the previous section, we find that nearly two-thirds of all inter-hemispheric interactions fall into the strongly-correlated subgroup.

We find that the functional properties of strongly-correlated regions are less variable across subjects, showing low inter-subject variance $\sigma_s(\mathrm{rsFC})$ (Figure 4.8). Strongly-correlated region pairs also tend to be structurally connected across a larger percentage of subjects (vertical projection in Figure 4.8). Together, these findings suggest that strong rsFC is supported by consistency in both structural and functional connectivity.

4.5.2 Structural Connectivity of the Representative Brain

Consistent with the results found in the previous section, we find striking differences in the structural properties of connections that support strong versus weak correlations (Figure 4.9a,b). Both inter- and intra-hemispheric regions that show increasingly strong rsFC are linked by an increasingly high number of connections (Figure 4.9c). As previously observed, intra-hemispheric regions are more densely connected and tend to be linked by shorter connections, on average, than inter-hemispheric regions.

Inspection of the distribution of connection lengths reveals that increasingly long inter-hemispheric connections consistently support increasingly strong rsFC (insets of Figure 4.9d). The separation in the average length of connections linking strongly- versus weakly-correlated regions is small and can change across variations in the selection and

partitioning of region pairs (see Chapter 5). This suggests, in agreement with the results of the previous section, that the strength of inter-hemispheric rsFC is less sensitive to variations in connection length.

While increasingly long intra-hemispheric connections appear to support increasingly strong rsFC over short distances (dotted inset of Figure 4.9d), this relationship is an artifact of averaging over inconsistently-connected region pairs. Inconsistent connectivity can produce apparent connection lengths $\langle L \rangle_s$ that are shorter than the interregional distance d between the interconnected regions. Removing this artifact (a process that requires knowledge of SC) reveals that strong rsFC is consistently supported by short intra-hemispheric connections (main portion of Figure 4.9d). This artifact does not qualitatively impact the previously-discussed relationships between SC and FC observed in Figures 4.7 and 4.9 (see Chapter 5 for further discussion).

4.5.3 Individual Variability in Structural Connectivity

Just as the overall strength of FC varied across subjects, we find significant inter-subject variation in the overall number and length of connections. Within subject-specific brain networks, however, strong and weak correlations are supported by connections with similar structural properties to those observed in the subject-averaged representative brain network.

Individual subjects show strikingly consistent separation in the number of intra- and inter-hemispheric connections that link strongly- versus weakly-correlated regions,

with higher numbers of connections consistently supporting stronger rsFC. The observed length of these connections, however, is sensitive to the presence of inconsistent connectivity, as was observed in the representative brain network.

In the absence of accounting for inconsistent connectivity, strong intra-hemispheric correlations appear to be supported by long connections (dotted inset of Figure 4.9f). If we remove absent connections from subject-specific brain networks (a process analogous to that used in the representative brain network to remove unphysical lengths), we recover the relationship that strong correlations are consistently supported by short intra-hemispheric connections (main portion of Figure 4.9f). In both cases, inter-hemispheric regions show varied separation in the length of connections that support strong versus weak correlations (insets in Figure 4.9f).

4.5.4 Conclusions from FC to SC Analysis

The consistent link between connection number and rsFC strength suggests that high connection number is a broadly conserved property for supporting strong FC. In comparison, the observed variation between inter- versus intra-hemispheric connection length and rsFC strength suggests that connection length is not a broadly conserved property but rather differentially relates to the anatomical properties of both the connections and the specific regions linked by these connections.

4.6 Methodological Considerations

While the analyses of $\mathrm{FC} \rightarrow \mathrm{SC}$ and $\mathrm{SC} \rightarrow \mathrm{FC}$ produce consistent and complimentary results, the inference of SC from FC is the more difficult of the two analyses. The difficulty arises in selecting, without the knowledge of SC, region pairs that show consistent SC across subjects. The noise introduced by selecting region pairs with inconsistent SC can be partially, but not completely, eliminated by restricting the analysis to small interregional distances. The disadvantage, however, is that this approach biases against the selection of the task-relevant set of long intra-hemispheric connections, making the inference of SC from task-driven FC more difficult. Furthermore, the distribution of connection lengths inferred from this approach is sensitive to artifacts that arise from averaging over inconsistently-connected region pairs. Robust relationships between connection length and rsFC strength can be achieved by removing from the representative brain network connections whose average length $\langle L \rangle_s$ is less than the minimum possible length $d - d_{\min}$, or analogously by removing absent connections from subject-specific brain networks.

The application of indirect or higher-order connectivity measures, such as modularity, clustering, and path length, to the present analysis may help bridge the findings of previous studies that have separately assessed structural [43, 47] and functional [96, 97, 46] human brain networks (see [98] for a comprehensive review). Together, these approaches may inform and constrain the connectivity used in computational models of network dynamics.

In concluding, we can speculate as to why the brain might be structured in this manner, with many short and few long connections that differentially impact resting versus task-driven activity. Connections are energetically expensive to both maintain and use [99, 100, 101, 102], favoring short and sparse over long and dense connectivity. However, few long connections more efficiently transmit information between distant regions than do many short connections, supporting the role of long connections in task-directed activity. Conversely, dense connectivity may enhance the robust properties of default mode function by reducing the potential impact of local structural disruptions, supporting the role of high numbers of connections in strong resting-state activity.

Together, these findings provide insight into the principles that may have constrained the evolution and development of the anatomical architecture of the human brain, and they make specific predictions about the functional implications of degradations to this architecture. Identifying such links between anatomical and functional connectivity patterns is crucial for understanding both the capabilities and constraints on human cognitive function.

4.7 Discussion

Identifying relationships between structural and functional human brain networks is crucial for understanding the large-scale organization of the human brain. Previous studies of structure-function relationships have been limited to specific brain regions, small sample sizes, and resting-state activity [43, 89, 88] for which is it difficult to assess

widespread structure-function interactions. In the present study, we introduce novel methods for both predicting and inferring consistent relationships between structural and functional connectivity across subjects and cognitive states.

The complimentary analyses of $\mathrm{FC} \rightarrow \mathrm{SC}$ and $\mathrm{SC} \rightarrow \mathrm{FC}$ identify conserved structural properties that support strong resting-state function. We find that high numbers of connections consistently underlie strong rsFC, a result that supports both empirical and computational studies of resting-state activity [89, 103]. The lengths of these connections are differentially important for distinguishing inter- from intra-hemispheric function, with local intra-hemispheric and nonlocal inter-hemispheric connections consistently supporting strong rsFC.

Analysis of $\mathrm{SC} \rightarrow \mathrm{FC}$ further identified structural properties that support task-dependent changes in function. Long-range intra-hemispheric connections, which link brain regions important for attention [104, 105, 106, 107] and memory [108, 109, 110] (see Chapter 5), are found to both support and distinguish between attention and memory tasks. The observed variability in task-driven FC, a property that has been linked to behavioral measures [111], suggests that the strength of connectivity between distant intra-hemispheric regions may be predictive of attention and memory performance.

The finding that state-dependent shifts in FC reflect known properties of both anatomical connections and the regions linked via these connections provides support for the utility of correlations in BOLD signals as a measure of functional interaction between brain regions. Recent results have indicated that fiber pathways can exhibit abrupt turns

that would not be identified by the tractography algorithm used here [95]. The inclusion of such pathways in the analysis could strengthen the observed relationship between structural and functional connectivity. The observation of structurally-mediated FC, however, does not discount previous findings that functional correlations can persist in the absence of SC, as is observed in persons with agenesis of the corpus callosum [112], and may be mediated by indirect SC [89].

Changes in both resting and task-driven FC have been linked to development [113, 114], aging [115, 116], and neurological disease [117, 118]. The results presented here provide insight into the structural mechanisms that could contribute to such altered functional states. Disruptions to dense connections could alter the topological nodal properties of network hubs, a consequence that has been linked to altered rsFC in diseases such as epilepsy [119]. Disruptions to inter-hemispheric connections could similarly reduce rsFC, as is observed in patients with axonal injury in the corpus callosum [120]. As changes in FC have additionally been linked to variability in task performance [121, 111], structural disruptions are expected to alter behavior across a range of cognitive tasks.

While the analyses of $\mathrm{FC} \rightarrow \mathrm{SC}$ and $\mathrm{SC} \rightarrow \mathrm{FC}$ produce consistent and complimentary results, the inference of SC from FC is the more difficult of the two analyses. The difficulty arises in selecting, without the knowledge of SC, region pairs that show consistent SC across subjects. The noise introduced by selecting region pairs with inconsistent SC can be partially, but not completely, eliminated by restricting the analysis to small interregional distances. One disadvantage, however, is that this approach biases against

the selection of the task-relevant set of long intra-hemispheric connections, making the inference of SC from task-driven FC more difficult. Furthermore, the distribution of connection lengths inferred from this approach is sensitive to artifacts that arise from averaging over inconsistently-connected region pairs. Robust relationships between connection length and rsFC strength can be achieved by removing inconsistent connections from the representative brain network, or analogously by removing absent connections from subject-specific brain networks.

The application of indirect or higher-order connectivity measures, such as modularity, clustering, and path length, to the present analysis may help bridge the findings of previous studies that have separately assessed structural [43, 47] and functional [96, 97, 46] human brain networks (see [98] for a comprehensive review). Together, these approaches may inform and constrain the connectivity used in computational models of network dynamics.

In concluding, we can speculate as to why the brain might be structured in this manner, with many short and few long connections that differentially impact resting versus task-driven activity. Connections are energetically expensive to both maintain and use [99, 100, 101, 102], favoring short and sparse over long and dense connectivity. However, few long connections might more efficiently transmit information between distant regions, as is needed during task performance, than do many short connections. Conversely, dense connectivity could enhance the robust properties of default mode function by reducing the potential impact of local structural disruptions. Finally, an

insensitivity of resting-state correlations to variations in connection length may be crucial for functionally binding the two hemispheres which, although structurally segregated, must support a single cognitive identity.

Together, these findings provide insight into the principles that may have constrained the evolution and development of the anatomical architecture of the human brain, and they make specific predictions about the functional implications of degradations to this architecture. Identifying such links between anatomical and functional connectivity patterns is crucial for understanding both the capabilities and constraints on human cognitive function.

Chapter 5

Impact of Network Construction on Brain Connectivity

"There are billions of neurons in our brains, but what are neurons? Just cells. The brain has no knowledge until connections are made between neurons. All that we know, all that we are, comes from the way our neurons are connected."

–Tim Berners-Lee, from *from Weaving The Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*, 1999

5.1 Introduction

The following chapter provides a detailed analysis of relationships between structural and functional connectivity in the human brain across variations in the techniques used

to assess these relationships. In the previous chapter, we used structural and nonstructural measures to identify pairs of brain regions that were consistently structurally connected within a large percentage of subjects. Within the subset of regions selected via these measures, we constructed subject-averaged ("representative") and subject-specific brain networks from measures of structural and functional connectivity. We examined the extent to which functional connectivity could be predicted from structural connectivity, and to what extent structural connectivity could be inferred from functional connectivity. Subgroups of connections partitioned based on structural and functional measures (Figures 4.6 and 4.8) were shown to differ in their strength of functional and structural connectivity (Figures 4.7 and 4.9). Here, we highlight the specific anatomical regions of the brain involved in these structural and functional interactions. We then show that the observed relationships between structural and functional connectivity are robust to the specific choices made in selecting and partitioning subgroups of connections. By comparing two memory tasks, one for faces (assessed in 4) and one for words (assessed in this chapter), we confirm that the observed properties of memory-state functional connectivity are consistent across different memory conditions. Finally, we discuss the sensitivity of these results to inconsistent structural connectivity within representative and subject-specific brain networks.

5.2 Review of Methodology

In the previous chapter, we assessed the strength of structural and functional connectivity (SC and FC, respectively) between the 600 brain regions within the uniform-600 atlas. The strength of SC between regions was defined by the average length and total number of white matter tracts. The strength of FC was defined by the strength of BOLD correlations and was separately assessed at rest (rsFC), in deviations $\Delta\mathrm{asFC} = \mathrm{asFC} - \mathrm{rsFC}$ of the attention state (asFC) from rest, and in deviations $\Delta\mathrm{msFC} = \mathrm{msFC} - \mathrm{rsFC}$ of the memory state (msFC) from rest. The memory state was defined in the text to be strength of deviations $\Delta\mathrm{msFC}$ measured during the faces memory task, although we later compare within this supplement the strength of $\Delta\mathrm{msFC}$ measured during the faces memory task versus the words memory task.

By relating these structural and functional connectivity measures, we examined the extent to which (*i*) functional connectivity could be predicted from structural connectivity ($\mathrm{SC} \rightarrow \mathrm{FC}$) and (*ii*) structural connectivity could in turn be inferred from functional connectivity ($\mathrm{FC} \rightarrow \mathrm{SC}$). We assessed these relationships within subject-averaged, or "representative", brain networks, for which the average strength of connectivity $\langle \mathrm{O_{ij}} \rangle_{\mathrm{s}}$ between regions i and j is given by:

$$\langle \mathrm{O_{ij}} \rangle_{\mathrm{s}} = \frac{1}{\mathrm{N_s}} \sum_{\mathrm{s}=1}^{\mathrm{N_s}} \mathrm{O_{ij}^{s}}, \tag{5.1}$$

where $N_s = 84$ is the number of subjects and O takes on values of SC or FC. The construction of representative brain networks from subject-averaged measures of SC and

FC is advantageous for identifying structure-function relationships that are common to many subjects. We then compared representative structure-function relationships to those observed within subject-specific brain networks. In comparing connectivity measures across subjects, we computed subject-specific connectivity strengths $\langle \mathrm{O} \rangle_{\mathrm{c}}$ as averages across connections within a single subject:

$$\langle \mathrm{O} \rangle_{\mathrm{c}} = \frac{1}{\mathrm{N_c}} \sum_{\mathrm{i,j}}^{\mathrm{N_c}} \mathrm{O_{ij}}, \tag{5.2}$$

where N_c gives the number of region pairs used in the analysis, with $N_c = 3079$ for the analysis of $\mathrm{SC} \rightarrow \mathrm{FC}$, and $N_c = 3079$ for the analysis of $\mathrm{FC} \rightarrow \mathrm{SC}$.

The regions pairs selected for each analysis were chosen based on their likelihood of being consistently structurally connected across a large percentage of subjects. To measure this consistency, we defined the subject-specific binary quantity C, such that $C_{ij} = 1$ if regions i and j are connected by one or more white matter tract within a single subject, and $C_{ij} = 0$ otherwise. We defined the consistency of connectivity to be the subject-averaged quantity $\langle C_{ij} \rangle_s$, which specifies the fraction of subjects that show one or more white matter tract linking regions i and j (e.g. a value of $\langle C_{ij} \rangle_s = .75$ indicates that regions i and j are linked by one or more structural connection within 75% of subjects).

Because the consistency $\langle C_{ij} \rangle_s$ does not have an analogously continuous subject-specific correlate (but rather has only the binary correlate C), and because the computation of $\langle C_{ij} \rangle_s$ inherently requires the use of structural information, we identified two additional measures that relate to this consistency. Importantly, these measures (one

structural and one nonstructural) can be defined as subject-averaged or subject-specific quantities, enabling the comparison between representative and subject-specific brain networks. Furthermore, the use of a structural measure to select regions for the prediction of function does not make any *a priori* assumptions of functional information, nor does the use of a nonstructural measure to select regions for the inference of structure make any *a priori* assumptions of structural information. These measures are defined as follows:

The rescaled number of connections $\bar{N}_{ij}$ between regions i and j, given by:

$$\bar{N}_{ij} = \frac{\langle N_{ij} \rangle_s}{\sigma_s(N_{ij})}, \tag{5.3}$$

is a purely structural measure that assesses the reliability in connection number across subjects. The denominator of this quantity gives the standard deviation in number $\sigma_s(N_{ij})$ computed across subjects:

$$\sigma_s(N_{ij}) = \sqrt{\frac{1}{N_s} \sum_{s}^{N_s} \left(N_{ij}^s - \langle N_{ij} \rangle_s \right)^2}. \tag{5.4}$$

We additionally specify the purely nonstructural measure of inverse interregional distance $1/d_{ij}$, defined by the metric distance d_{ij} between the center of mass positions (x, y, z) of regions i and j:

$$\frac{1}{d_{ij}} = \frac{1}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}} \tag{5.5}$$

In the previous chapter, we showed that increasing values of both $\bar{N}$ and $1/d$ relate to increasing values of the consistency $\langle C \rangle_s$. We therefore say that $\bar{N}$ and $1/d$ are *indirect*

measures of consistency, while $\langle C \rangle_s$ is, by our definition, the *direct* measure of consistency. The dependence of $\langle C \rangle_s$ on $\bar{N}$ and $1/d$ enables us to isolate region pairs that are consistently structurally-connected (high values of $\langle C \rangle_s$) by selecting region pairs above the threshold values $\bar{N}_T = 0.6$ and $1/d_T = 0.1$. From the set of region pairs selected via $\bar{N}_T$ and $1/d_T$, we constructed two representative brain networks for the separate assessment of (*i*) $\mathrm{SC} \rightarrow \mathrm{FC}$ and (*ii*) $\mathrm{FC} \rightarrow \mathrm{SC}$.

In (*i*), we imposed a length threshold $L_T = 20$ to separate short ($L \leq L_T$) from long ($L > L_T$) connections, and we imposed a number threshold $N_T = 30$ to separate dense ($N > N_T$) from sparse ($N \leq N_T$) connections. Note that our use of the terms "dense" and "sparse" to refer to high and low connection numbers differs from definitions of density in which the number of tracts is scaled by the total cross sectional tract area. In combination with the separate consideration of inter- and intra-hemispheric connections, these partitions in length and number define four non-overlapping structural subgroups: short and long inter-hemispheric connections, and short but dense and long but sparse intra-hemispheric connections, whose functional properties we compared to the remaining bulk of short, sparse intra-hemispheric connections. In Figure 4.7, we showed that these structural subgroups were predictive of shifts in the strength of rsFC, $\Delta\mathrm{asFC}$, and $\Delta\mathrm{msFC}$ observed within representative and subject-specific brain networks.

As the resting state was shown to exhibit the most pronounced separation in the functional properties of structural subgroups, we used resting state FC, rather than task-driven FC, to infer underlying structural properties in (*ii*). We imposed two functional

thresholds $\mathrm{rmFC_T}$ to partition region pairs into three equal-sized functional subgroups that show weak (bottom 33%), intermediate (mid 33%), and strong (top 33%) rsFC. We showed that these functional subgroups were supported by structural connections that differed in both length and number.

In what follows, we address four specific points of this analysis. (1) We identify the anatomical regions that participate in the structural subgroups whose regions were selected via $\bar{N}_T$ and partitioned via L_T and N_T, and similarly those regions that participate in the functional subgroups whose regions were selected via $1/d_T$ and partitioned via $\mathrm{rsFC_T}$. (2) We show that our results are robust to variations in the selection of regions via $\bar{N}_T$ and $1/d_T$ (versus the selection via thresholds $\langle C \rangle_T$ in the direct consistency $\langle C \rangle_s$) and to variations in the structural and function partitions imposed by L_T, N_T, and $\mathrm{rsFC_T}$. (3) We compare the results of using memory-state connectivity msFC measured from the faces memory task versus the words memory task. (4) We discuss the extent to which variability in the consistency of structural connectivity, introduced by the selection of region pairs via $\bar{N}$ and $1/d$, impacts the results of the previous chapter.

5.3 Robustness to Thresholding

In selecting and partitioning groups of region pairs for the construction of brain networks, we made specific choices about both the number of retained region pairs and the divisions used to group connections between these region pairs. In what follows, we evaluate the structural and functional connectivity of the representative brain networks

across variations in the thresholds used to construct ($\bar{N}_T$ and $1/d_T$) and partition (L_T, N_T, and $\mathrm{rsFC_T}$) these networks. We assess this robustness to thresholding within our two analyses, $\mathrm{SC} \rightarrow \mathrm{FC}$ and $\mathrm{FC} \rightarrow \mathrm{SC}$, and we show that the correlations between structure and function observed in each analysis are robust to thresholding.

5.3.1 Predicting Function from Structure

The analysis of $\mathrm{SC} \rightarrow \mathrm{FC}$ shown in the previous chapter was performed on a representative brain network constructed from region pairs linked by high rescaled numbers of structural connections ($\bar{N} > \bar{N}_T$). Connections between these region pairs were partitioned based on their length and number via the structural thresholds N_T and L_T. In combination with the delineation between inter- and intra-hemispheric connectivity, these partitions defined four structural subgroups, (1) short and (2) long inter-hemispheric connections, and (3) dense and (4) long intra-hemispheric connections, whose properties were compared to the remaining bulk of short, sparse intra-hemispheric connections. The subject-averaged strengths $\langle \mathrm{rsFC} \rangle_\mathrm{s}$, $\langle \Delta \mathrm{asFC} \rangle_\mathrm{s}$, and $\langle \Delta \mathrm{msFC} \rangle_\mathrm{s}$ were then compared across these partitioned subgroups.

We now evaluate the robustness of structurally-dependent shifts in $\langle \mathrm{rsFC} \rangle_\mathrm{s}$, $\langle \Delta \mathrm{asFC} \rangle_\mathrm{s}$, and $\langle \Delta \mathrm{msFC} \rangle_\mathrm{s}$ to variations in $\bar{N}_T$, L_T, and N_T. As a measure of comparison across thresholds, we compute the averages $\langle \mathrm{rsFC} \rangle_\mathrm{s,c}$, $\langle \Delta \mathrm{asFC} \rangle_\mathrm{s,c}$, and $\langle \Delta \mathrm{msFC} \rangle_\mathrm{s,c}$ of the complementary cumulative distribution functions (cCDFs) produced by connections within the representative brain network. These cCDFs were shown in Figure 4.7d-f for the

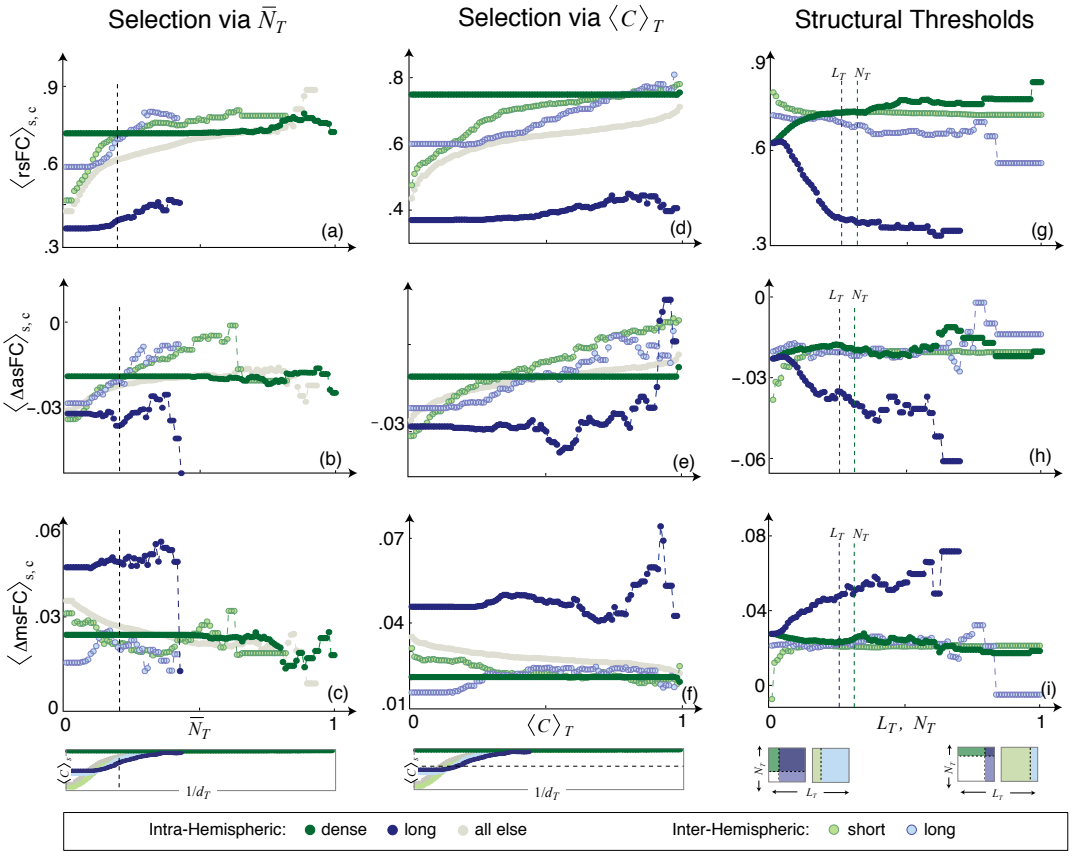

Figure 5.1: Predicting FC from SC: Robustness to Thresholding. Changes in average functional connectivity $\langle \text{rsFC} \rangle_{\text{s,c}}$, $\langle \Delta\text{asFC} \rangle_{\text{s,c}}$, and $\langle \Delta\text{msFC} \rangle_{\text{s,c}}$ produced by structural subgroups of connections across variations in the selection of regions via (a-c) $\bar{N}_T$ and (d-f) $\langle C \rangle_T$ and across variations in the (g-i) structural thresholds L_T and N_T. Thresholds are scaled by their maximum values, with the values used in the previous chapter indicated by dashed lines. Variations in the selection of region pairs via the indirect ($\bar{N}_T$) and direct ($\langle C \rangle_T$) measures of consistency produce similar shifts in $\langle \text{FC} \rangle_{\text{s,c}}$ that collectively increase with increasing threshold values. In comparison, variations in structural thresholds alter the separation in $\langle \text{FC} \rangle_{\text{s,c}}$ across structural subgroups. Structural subgroups show the most pronounced separation in $\langle \text{FC} \rangle_{\text{s,c}}$ in the resting state, with inter-hemispheric and dense intra-hemispheric connections producing consistently strong $\langle \text{rsFC} \rangle_{\text{s,c}}$ and long intra-hemispheric connections producing consistently weak $\langle \text{rsFC} \rangle_{\text{s,c}}$ (top row). During task performance, inter-hemispheric and dense intra-hemispheric connections produce similar changes in $\langle \text{FC} \rangle_{\text{s,c}}$ to the bulk of remaining connections, while long intra-hemispheric connections producing consistently weak changes in $\langle \text{asFC} \rangle_{\text{s,c}}$ (middle row) and strong changes in $\langle \text{msFC} \rangle_{\text{s,c}}$ (bottom row).

threshold values $\bar{N}_T = 0.6$, $N_T = 30$, and $L_T = 20$. The notation $\langle O \rangle_{s,c}$ indicates that the average of the quantity O was computed first across subjects, in constructing the representative brain network, and then across connections within the representative network. Variations in $\langle O \rangle_{s,c}$ measure relative shifts in the average values of the cCDFs across variations in thresholding.

We find that the observed shifts in FC shown in Figure 4.7d-f in Chapter 4 are robust to variations in thresholding, as shown here in Figure 5.1 and as discussed in detail below. Across all threshold values, inter-hemispheric FC shows minimal dependence on connection length, with both short and long inter-hemispheric connections supporting strong rsFC and similar changes in asFC and msFC to the remaining bulk of connections. Dense intra-hemispheric connections show nearly consistent FC strength across thresholding scenarios, exhibiting notably strong rsFC. Long intra-hemispheric connections consistently show task-dependent changes in FC, exhibiting low values of rsFC, decreases in asFC from rest, and increases in msFC from rest.

Variations in Selection Thresholds. Increasing the selection threshold $\bar{N}_T$ corresponds to retaining fewer region pairs that are more densely connected across subjects (left column of Figure 5.1). As mentioned in Section 5.2, $\bar{N}$ is an indirect measure of the consistency in connectivity $\langle C \rangle_s$, and therefore the densely-connected region pairs selected via $\bar{N}_T$ also tend to be consistently connected within a large fraction of subjects. The selection of region pairs via $\bar{N}$, rather than via $\langle C \rangle_s$, is advantageous because $\bar{N} = \langle N \rangle_s / \sigma_s(N)$ has an analogous single-subject correlate given by the number of connections N scaled

by the uncertainty in the measurement of N. In comparison, the single-subject correlate of the measure $\langle C \rangle_s$ is the subject-specific binary number C, which cannot be used as a continuous threshold for region pair selection.

To confirm that region pair selection via $\bar{N}_T$ produces similar results to selection via $\langle C \rangle_s$, we compare shifts in $\langle \mathrm{rsFC} \rangle_{\mathrm{s,c}}$, $\langle \Delta \mathrm{asFC} \rangle_{\mathrm{s,c}}$, and $\langle \Delta \mathrm{msFC} \rangle_{\mathrm{s,c}}$ across variations in $\bar{N}_T$ and $\langle C \rangle_T$, shown respectively in the left and middle columns of Figure 5.1.

Because $\bar{N}$ scales with the consistency $\langle C \rangle_s$, and because the distributions of $\langle \mathrm{FC} \rangle_{\mathrm{s}}$ are insensitive to inconsistent connectivity (to be discussed in Section 5.5), variations across $\bar{N}_T$ and $\langle C \rangle_T$ produce similar results, with the latter producing stronger separations in FC across structural groups (middle column of Figure 5.1). For both choices of selection threshold, the overall strength $\langle \mathrm{FC} \rangle_{\mathrm{s,c}}$ increases with increasing threshold.

Variations in Structural Thresholds. Variations in the structural thresholds L_T and N_T alter the partitioning of connections into long ($L > L_T$) versus short ($L \leq L_T$) and dense ($N > N_T$) versus sparse ($N \leq N_T$) subgroups, which in turn alter the degree of separation in FC observed across structural subgroups.

Biasing toward higher numbers of connections (by increasing N_T) increases the strength of $\langle \mathrm{FC} \rangle_{\mathrm{s,c}}$ in the resting state but has limited effect on task-driven states, producing a slight increase in $\langle \Delta \mathrm{asFC} \rangle_{\mathrm{s,c}}$ and a slight decrease in $\langle \Delta \mathrm{msFC} \rangle_{\mathrm{s,c}}$. Biasing toward longer connections (by increasing L_T) increases the separation in $\langle \mathrm{FC} \rangle_{\mathrm{s,c}}$ between short and long inter- and intra-hemispheric connections. The separation in inter-hemispheric $\langle \mathrm{FC} \rangle_{\mathrm{s,c}}$, however, is consistently smaller than the separation in intra-hemispheric $\langle \mathrm{FC} \rangle_{\mathrm{s,c}}$.

Most notably, biasing toward longer connections reveals that long intra-hemispheric connections show increasingly weak $\langle \mathrm{FC} \rangle_{\mathrm{s,c}}$ in the resting and attention states and increasingly strong $\langle \mathrm{FC} \rangle_{\mathrm{s,c}}$ in the memory state, thereby confirming that the task-dependent shifts observed in Figure 4.7d-f are not artifacts of thresholding.

5.3.2 Inferring Structure from Function

The analysis of $\mathrm{FC} \rightarrow \mathrm{SC}$ shown in Chapter 4 was performed on a representative brain network constructed from region pairs within close interregional distances ($1/d > 1/d_T$). Region pairs were partitioned based on their their strength of rsFC via the functional thresholds $\mathrm{rsFC_T}$ into three functional groups exhibiting (1) weak (bottom 33%), (2) intermediate (middle 33%), and (3) strong (top 33%) rsFC. These subgroups were further delineated into inter- and intra-hemispheric functional groups. The subject-averaged structural properties $\langle L \rangle_s$ and $\langle N \rangle_s$ were then compared across these partitioned subgroups.

We now evaluate the robustness of functionally-dependent shifts in $\langle L \rangle_s$ and $\langle N \rangle_s$ to variations in $1/d_T$ and $\mathrm{rsFC_T}$. As a measure of comparison across thresholds, we compute the averages $\langle L \rangle_{s,c}$ and $\langle N \rangle_{s,c}$ of the cCDFs produced by connections within the representative brain network. These cCDFs were shown in Figure 4.9c-d for the threshold values $1/d_T = 0.1$and $\mathrm{rsFC_T} = 1/3$.

We find that the observed shifts in SC shown in Figure 4.9c-d are robust to variations in thresholding, as shown here in Figure 5.2 and as discussed in detail below. Across all

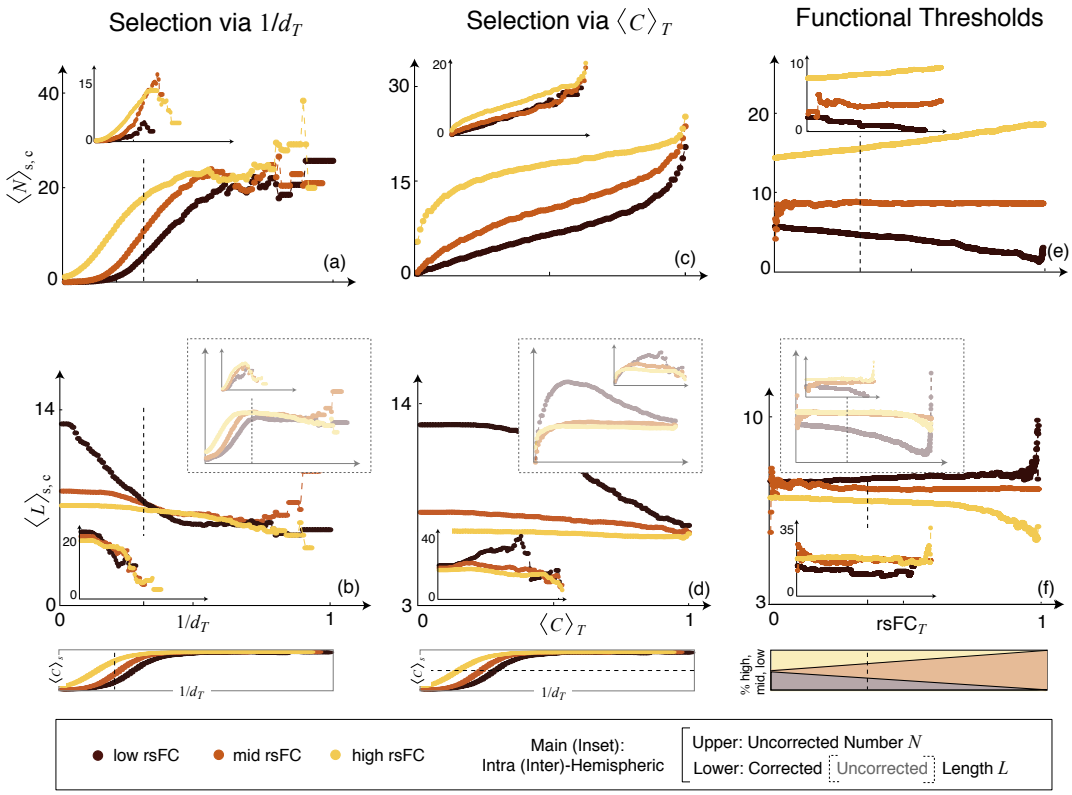

Figure 5.2: Inferring SC from FC: Robustness to Thresholding Changes in average structural measures $\langle N \rangle_{s,c}$ and $\langle L \rangle_{s,c}$ produced by functional subgroups of connections across variations in the selection of regions via (a-c) $1/d_T$ and (d-f) $\langle C \rangle_T$ and across variations in the (g-i) functional thresholds $\mathrm{rsFC_T}$, with inter-hemispheric values shown in insets. Thresholds are scaled by their maximum values, and the specific values used in the previous chapter are indicated by dashed lines. Corrected length distributions are shown in the lower row, with uncorrected distributions shown in dotted insets. Variations in the selection of regions pairs via the indirect ($1/d_T$) and direct ($\langle C \rangle_T$) measures of consistency produce similar shifts in $\langle \mathrm{SC} \rangle_{\mathrm{s,c}}$, with the overall values of $\langle N \rangle_{s,c}$ and $\langle L \rangle_{s,c}$ tending to respectively increase and decrease with increasing threshold values. In comparison, variations in the functional thresholds alter the separation in $\langle N \rangle_{s,c}$ and $\langle L \rangle_{s,c}$ across functional subgroups. All threshold variations maintain the finding that strong rsFC is supported by high numbers of connections and short intra-hemispheric connection lengths. Inspection of uncorrected $\langle L \rangle_{s,c}$ reveals that the relationship between increasingly short intra-hemispheric connections and increasingly strong rsFC can be recovered if (b) unphysical connection lengths selected via $1/d_T$ are separately removed, if (d) regions are selected via $\langle C \rangle_T$, or if (f) the functional thresholds $\mathrm{rsFC_T}$ are tuned to select very strongly- and very weakly-correlated region pairs.

threshold variations, high numbers of inter- and intra-hemispheric connections consistently support strong rsFC. In assessing the distributions of connection lengths, which are shown in a later section to be sensitive to the presence of connections with unphysical connection lengths, we separately consider the uncorrected (dotted insets) and corrected (main figure) distributions for which connections with unphysical lengths are respectively included or excluded. The corrected distributions show that short intra-hemispheric connection lengths consistently support strong rsFC. A thorough discussion of the identification and implications of unphysical connections is presented in Section 5.5.

Variations in Selection Thresholds. Increasing the selection threshold $1/d_T$ corresponds to retaining region pairs that are closer in physical proximity (left column of Figure 5.2). As mentioned previously, $1/d$ is an indirect measure of the consistency in connectivity $\langle C \rangle_s$, and therefore region pairs in close physical proximity also tend to be consistently connected within a large fraction of subjects. The selection of region pairs via $1/d$, rather than via $\langle C \rangle_s$, is advantageous because $1/d$ does not rely on knowledge of anatomical connectivity and therefore does not suffer from the circular use of structural information for the inference of structural information. In comparison, the measure $\langle C \rangle_s$ relies on the knowledge of structural connectivity between region pairs.

To confirm that region pair selection via $1/d_T$ produces similar results to region pair selection via $\langle C \rangle_s$, we compare shifts in $\langle N \rangle_{s,c}$ and $\langle L \rangle_{s,c}$ produced by functional subgroups across variations in both $1/d_T$ and $\langle C \rangle_T$, shown respectively in the left and middle columns of Figure 5.2.

Connection Numbers. Because $1/d$ scales with the consistency $\langle C \rangle_s$, and because the distribution of $\langle N \rangle_s$ is insensitive to inconsistent connectivity (discussed in the following section), variations across $1/d_T$ and $\langle C \rangle_T$ produce similar distributions of connection numbers, with the latter producing a more pronounced separation in the large number of inter- and intra-hemispheric connections that support strong versus weak correlations (middle column of Figure 5.1). Increases in both threshold values result in an overall increase in connection number.

Connection Lengths. In comparison to the distribution of $\langle N \rangle_s$, the distribution of $\langle L \rangle_s$ varies depending on whether regions are selected via $1/d_T$ or $\langle C \rangle_T$, and whether connections with unphysical lengths are included in or excluded from in the analysis. When unphysical connections are excluded (Figure 5.2b,d), variations in both $1/d_T$ and $\langle C \rangle_T$ produce consistent results, with increasingly short intra- and inter-hemispheric connections supporting increasingly strong rsFC. Region selection via $\langle C \rangle_T$ produces a stronger separation in the length of inter and intra-hemispheric connections that support strong versus weak rsFC, as was similarly observed in the distributions of connection numbers. Increases in both threshold values result in an overall decrease in the average connection length.

When unphysical connections are included, however, variations in $1/d_T$ and $\langle C \rangle_T$ produce different results (dotted insets in Figure 5.2b,d). Region selection via $1/d$ shows that strongly-correlated regions are linked by longer, rather than shorter, intra- and inter-hemispheric connections. Region selection via $\langle C \rangle_s$ recovers the relationship between

short intra- and inter-hemispheric connection length and strong rsFC, confirming that inconsistency in connectivity (quantified by low values of $\langle C \rangle_s$) are responsible for the altered relationship between connection length and rsFC strength.

Regardless of the selection of regions via $1/d$ versus $\langle C \rangle_s$, and regardless of whether unphysical connections are included in or excluded from the analysis, the distributions of inter-hemispheric connection lengths show much less separation across functional subgroups than do the distributions of intra-hemispheric connection lengths.

5.3.3 Variations in Functional Thresholds

Variations in the functional threshold $\mathrm{rsFC_T}$ shift the delineations between region pairs that exhibit weak, intermediate, and strong rsFC (right column of Figure 5.2). We vary this threshold such that the strongly- and weakly-correlated groups consist of the same number of region pairs but can differ in number from the intermediate group, such that the strongly- and weakly-correlated groups are highly populated for small values of $\mathrm{rsFC_T}$, while the intermediate group is highly populated for large values of $\mathrm{rsFC_T}$. Variations in functional threshold values alter the observed separation in SC across functional subgroups.

Connection Numbers. Biasing toward very strong and very weak rsFC by increasing $\mathrm{rsFC_T}$ in turn increases the separation in the number of connections that support strong versus weak rsFC, with high numbers of inter- and intra-hemispheric connections consistently supporting strong rsFC (Figure 5.2e).

Connection Lengths. The impact of variations in $\mathrm{rsFC_T}$ on the distribution of connection lengths depends on the inclusion versus exclusion of connections with unphysical lengths (main portion versus dotted inset of Figure 5.2f). When unphysical connections are excluded from the analysis, variations in $\mathrm{rsFC_T}$ increase the separation in the length of connections that support strong versus weak correlations, with short intra-hemispheric connections consistently supporting strong rsFC. When unphysical connections are included in the analysis, however, strong rsFC appears to be supported by longer, rather than shorter, intra-hemispheric connections across variations in $\mathrm{rsFC_T}$. In both cases, stronger inter-hemispheric correlations are supported by slightly longer, rather than shorter, connections. The observed separation in length across functional subgroups, however, is much less pronounced for inter- versus intra-hemispheric connections.

5.3.4 Considerations for Inter- versus Intra-Hemispheric Connections

Because inter- and intra-hemispheric connections differ in their structural properties, their distributions of SC show different sensitivities to variations in threshold values and to the inclusion versus exclusion of connections with unphysical lengths.

Inter-Hemispheric Connections. Relatively few inter-hemispheric regions are linked by connections with unphysical lengths (see later sections on 'Inconsistent Connectivity in the Representative Brain Network'), and therefore the distributions of inter-hemispheric number and length are less sensitive to the exclusion of unphysical connections than are intra-hemispheric distributions. Regardless of whether unphysical connections are in-

cluded in or excluded from the analysis, inter-hemispheric regions are consistently linked by fewer connection numbers and greater connection lengths, on average, than intra-hemispheric regions, a trend that is maintained across variations in all threshold values. In contrast to consistently large separation in SC observed between intra-hemispheric functional subgroups, inter-hemispheric subgroups show reduced separation in the structural properties that support strong versus weak correlations. Furthermore, the relative contribution from long versus short inter-hemispheric connections is variable across thresholding scenarios. Together, these results provide additional support for our previous finding that the functional interactions between inter-hemispheric regions are less sensitive to underlying structure than are interactions between intra-hemispheric regions.

Intra-Hemispheric Connections. The majority of connections with unphysical lengths link intra-hemispheric regions, and therefore the distributions of intra-hemispheric connection length (but not number; see Section 5.5) are sensitive to the effects of connections with unphysical length. When included in the analysis, these unphysical connections alter the apparent relationship between connection length and rsFC, such that increasingly long, rather than short, connections appear to support increasingly strong rsFC. However, large increases in functional threshold values, which bias toward the very strongest and very weakest correlations, can be used to recover the trend that short intra-hemispheric connections support strong rsFC (far right of dotted inset in Figure 5.2f). Importantly, this technique requires no knowledge of SC. This trend can additionally be recovered by selecting regions pairs via the consistency $\langle C \rangle_s$ (dotted inset in Figure 5.2d). The

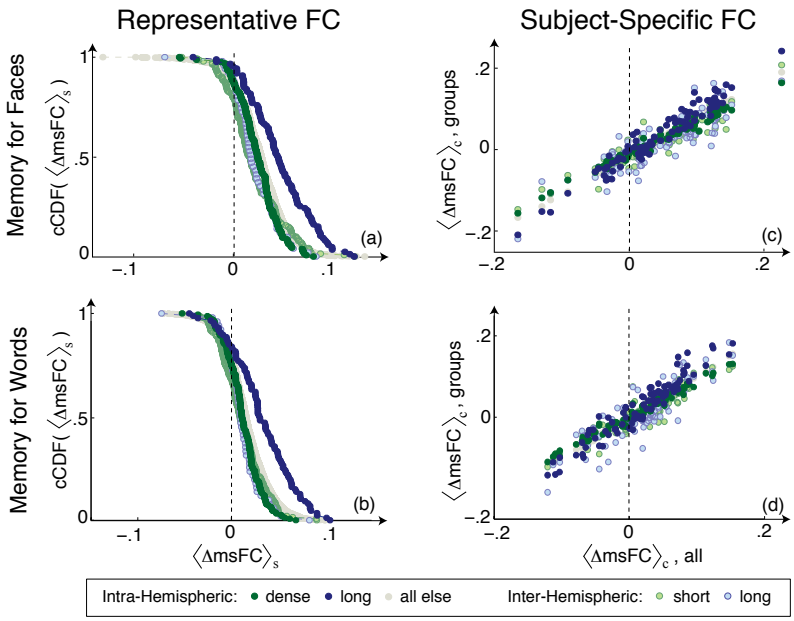

Figure 5.3: Comparison of Representative and Subject-Specific FC Between Memory for Words versus Faces. The representative brain network shows similar shifts in memory state functional connectivity ($\langle \Delta \mathrm{msFC} \rangle_\mathrm{s}$) between the memory tasks for (a) faces versus (b) words, with the distribution of $\langle \Delta \mathrm{msFC} \rangle_\mathrm{s}$ showing less variance across connections for words versus faces. For both words and faces, long intra-hemispheric connections showing pronounced shifts toward higher values of $\langle \Delta \mathrm{msFC} \rangle_\mathrm{s}$, while inter-hemispheric and dense intra-hemispheric connections show similar shifts in $\langle \Delta \mathrm{msFC} \rangle_\mathrm{s}$ to the remaining bulk of short, sparse intra-hemispheric connections. Individual subjects show similar variability in $\langle \Delta \mathrm{msFC} \rangle_\mathrm{c}$ across structural subgroups for (c) faces and (d) words, with the words task showing less variance in $\langle \Delta \mathrm{msFC} \rangle_\mathrm{c}$ across subjects as compared to the faces task.

different results obtained from the direct (via $\langle C \rangle_T$) versus indirect (via $1/d_T$) selection of consistently-connected region pairs suggests that increasing the minimum consistency, rather than the average consistency, is important for the robust inference of structure from function.

5.4 Comparison of Memory for Words versus Faces

The comparison of functional connectivity across structural subgroups of connections, as assessed in the previous chapter, compared the strength of FC measured in the resting state, during the performance of the attention task, and during the performance of the memory task for faces. Subjects also performed a memory task for words that was not assessed in Chapter 4. Here, we show that the two memory tasks, for words and for faces, show similar shifts in the functional connectivity produced by different structural subgroups of connections. The details of the two memory tasks are discussed in [122].

Long intra-hemispheric connections were shown in the previous chapter to support pronounced shifts in the memory-state functional connectivity ($\langle \Delta \mathrm{msFC} \rangle_\mathrm{s}$) of the representative brain network associated with memory for faces. Here, we find that the representative brain network shows similar shifts in $\langle \Delta \mathrm{msFC} \rangle_\mathrm{s}$ associated with memory for words, with long intra-hemispheric connections supporting the largest increase in msFC among structural subgroups. Furthermore, we find that memory for words shows less variation in $\langle \Delta \mathrm{msFC} \rangle_\mathrm{s}$ across connections in the representative brain network than does memory for faces (Figure 5.3a,b). Individual subjects show similar variation in the functional connectivity of structural subgroups between memory for words versus faces, with memory for words showing less variation in $\langle \Delta \mathrm{msFC} \rangle_\mathrm{c}$ across subjects than memory for faces (Figure 5.3c,d).

In a manner identical to that described in the previous section, we compare the functional connectivity of memory for words versus memory for faces across variations

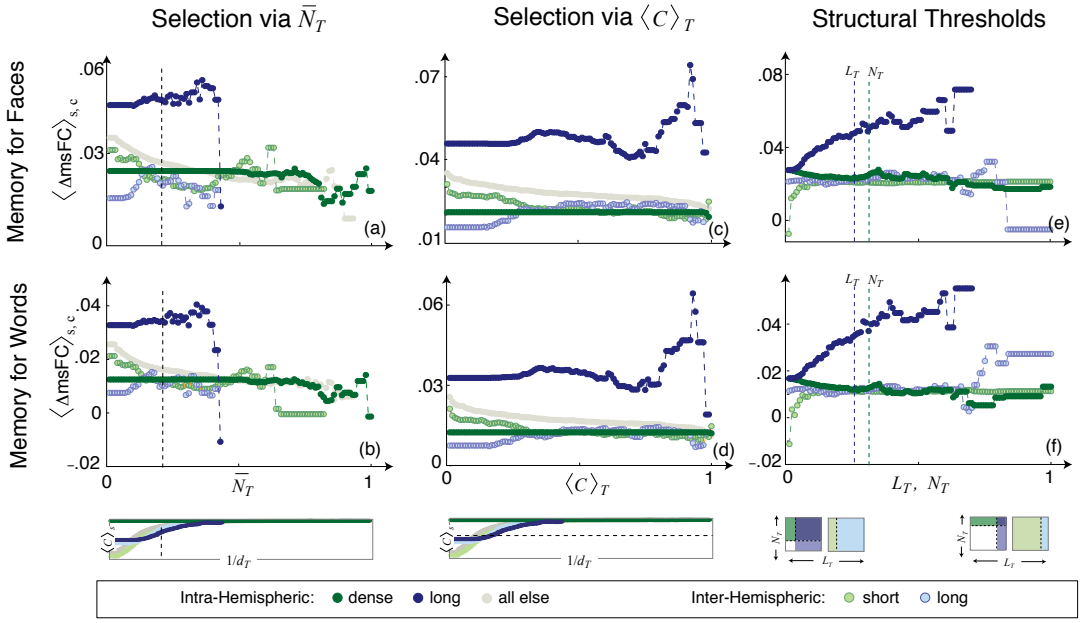

Figure 5.4: Comparison of FC Robustness toThresholding Between Memory for Words versus Faces. The memory state functional connectivity $\langle \Delta \mathrm{msFC} \rangle_{\mathrm{s,c}}$ of the representative brain network shows similar robustness to region selection via $\bar{N}_T$ (left column), region selection via $\langle C \rangle_T$ (middle columns) and the structural partitions N_T and L_T (right column) between memory for faces (top row) versus memory for words (bottom row), where the trends for the faces tasks are described in detail in Figure 5.1 above. Across all partitions, long intra-hemispheric connections consistently show strong $\langle \Delta \mathrm{msFC} \rangle_{\mathrm{s,c}}$ for both words and faces, while inter- and dense intra-hemispheric connections show similar values of $\langle \Delta \mathrm{msFC} \rangle_{\mathrm{s,c}}$ to one another and to the remaining bulk of short, sparse intra-hemispheric connections. Interestingly, in biasing toward long connections, long inter-hemispheric connections decrease in $\langle \Delta \mathrm{msFC} \rangle_{\mathrm{s,c}}$ in the faces task but increase in $\langle \Delta \mathrm{msFC} \rangle_{\mathrm{s,c}}$ in the memory task, suggesting that these connections may be important for distinguishing between different types of memory tasks.

in the selection and partitioning of region pairs in the representative brain network. Across variations in the selection thresholds $\bar{N}_T$ and $\langle C \rangle_T$, we find that memory for words shows similar qualitative shifts in $\langle \Delta \mathrm{msFC} \rangle_{\mathrm{s,c}}$ to those exhibited by memory for faces. Across variations in both selection thresholds, long intra-hemispheric connections consistently support large changes in $\langle \Delta \mathrm{msFC} \rangle_{\mathrm{s,c}}$ in both memory tasks (left and middle columns of Figure 5.4). These relationships are similarly maintained across variations in the structural thresholds N_T and L_T (right column of Figure 5.4). However, in biasing toward long connections, we find that very long inter-hemispheric connections show a decrease in $\langle \Delta \mathrm{msFC} \rangle_{\mathrm{s,c}}$ in the memory task for faces but an increase in $\langle \Delta \mathrm{msFC} \rangle_{\mathrm{s,c}}$ in the memory task for words. This suggests that both long inter- and long intra-hemispheric connections are important for supporting strong memory function, and long inter-hemispheric connections may be particularly important for distinguishing between different memory tasks.

Together, these results confirm that the memory state, whether defined by a memory task for words or faces, exhibits structure-function relationships that are distinct from those observed in the resting and attention states. Furthermore, these results suggest that long connections both support and distinguish between different types of memory tasks.

5.5 Inconsistent Connectivity in the Representative Brain

The two representative brain networks used for the separate assessments of $\mathrm{SC} \rightarrow \mathrm{FC}$ and $\mathrm{FC} \rightarrow \mathrm{SC}$ were constructed using two different, but largely overlapping, subsets

of region pairs. As discussed in Section 5.2, regions were selected via the structural and nonstructural measures $\bar{N}$ and $1/d$. Both measures relate to but are distinct from the consistency in connectivity $\langle C \rangle_s$, where our terminology is defined such that "high consistency" and "inconsistency" respectively refer to large and small values of $\langle C \rangle_s$. As a result of employing two different selection methods, the two representative networks vary in the degree to which regions are consistently connected across subjects, as measured by the distributions of $\langle C \rangle_s$ across connections within the representative brain networks.

5.5.1 Identifying Inconsistent Connectivity

In the previous section, we compared relationships between SC and FC across variations in the methods used to select of region pairs, separately considering region pair selection via indirect ($\bar{N}$ and $1/d$, as used in the previous chapter) versus direct ($\langle C \rangle_s$) measures of consistency. Figure 5.1 confirmed that the functional properties of structural subgroups are robust to the variations in consistency that arise from selecting region pairs indirectly via $\bar{N}_T$ versus directly via $\langle C \rangle_T$. Figure 5.2 further confirmed that the observed relationship between large connection numbers and strong rsFC is robust to similar variations in consistency arising from region pair selection via $1/d_T$ versus $\langle C \rangle_T$. However, Figure 5.2 showed that the apparent length of these connections varies significantly depending on region pair selection via $1/d_T$ versus $\langle C \rangle_T$. Given that $1/d$ and $\langle C \rangle_s$ are related, the question arises as to why the region pairs selected via each measure show qualitatively different relationships between $\langle L \rangle_s$ and $\langle \mathrm{rsFC} \rangle_\mathrm{s}$.

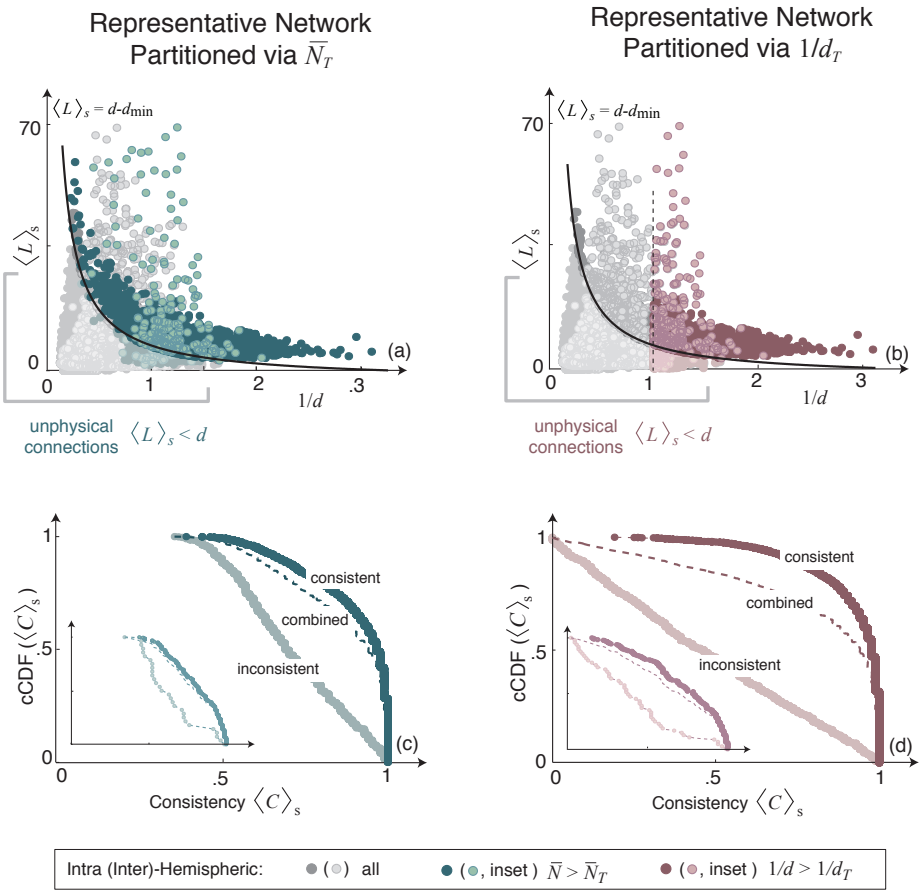

Figure 5.5: Inconsistent Connectivity in Representative Brain Networks. (a,c) Average length $\langle L \rangle_s$ versus inverse interregional distance $1/d$ of connections between regions within the two representative brain networks assessed in Chapter 4. Regions selected via (a) $\bar{N}$ and (c) $1/d$ are shown respectively in turquoise and maroon, with the minimum length $L_{\min} = d - d_{\min}$ superimposed in black. Both methods of region pair selection reduce the number of, but do not eliminate, connections with unphysical lengths $\langle L \rangle_s < L_{\min}$ (opaque). The presence of unphysical connections is an artifact of averaging over connections that are inconsistently present across subjects, as evidenced in the shifted distributions of $\langle C \rangle_s$ produced by these connections (b,d). The removal of these connections shifts the average consistency toward higher values.

To address this question, we note that the process of averaging structural properties across subjects can produce connections that disobey physical constraints. One such physical constraint requires that the subject-averaged length $\langle L \rangle_s$ of connections linking two regions not be less than the physical separation d between those regions. A low value of consistency $\langle C_{ij} \rangle_s$ between two regions i and j, for which a large number of subjects show no structural connectivity ($C_{ij} = 0$), can therefore be manifested in an apparent connection length $\langle L_{ij} \rangle_s$ less than the minimum physical length $L_{ij,\min} = d_{ij} - d_{\min}$, where the subtraction of the minimum interregional distance $d_{\min}$ allows for the possibility of adjacent regions to be linked via connections with near-zero length. We refer to connections that exhibit $\langle L \rangle_s < L_{\min}$ as "unphysical" connections.

We find that, within the sets of region pairs used to construct each representative brain network, a subset of region pairs exhibits this unphysical property arising from inconsistent connectivity (low values of $\langle C \rangle_s$). Figure 5.5a-b shows the average length $\langle L \rangle_s$ versus the inverse interregional distance $1/d$ of connections within the two representative brain networks whose region pairs were separately selected via $\bar{N}_T$ and $1/d_T$. By comparing the lengths $\langle L \rangle_s$ to the minimum physical length $L_{\min}$ (black line in Figure 5.5a-b), it can be seen that a fraction of connections within each network have unphysical lengths $\langle L \rangle_s < L_{\min}$. Furthermore, by comparing the distributions of $\langle C \rangle_s$ produced by connections whose lengths are greater and less than $L_{\min}$ (Figure 5.5c-d), we can see that this effect arises from inconsistency in connectivity. In both representative brain networks, connections with unphysical lengths less that $L_{\min}$ show significantly lower

inter-subject consistency than do connections with physical lengths greater than $L_{\min}$. Because $\bar{N}$ is more strongly related to $\langle C \rangle_s$ than is $1/d$, the region pairs selected via $\bar{N}_T$ show less separation in the distributions of $\langle C \rangle_s$ than do region pairs selected via $1/d_T$.

5.5.2 Impact of Inconsistent Connectivity on SC-FC Relationships

To assess the consequences of including unphysical connections in the representative brain networks, we repeat the analyses of $\mathrm{SC} \rightarrow \mathrm{FC}$ and $\mathrm{FC} \rightarrow \mathrm{SC}$ performed in Chapter 4, now selecting regions subject to the constraints $\bar{N} > \bar{N}_T$ and $\langle L \rangle_s > L_{\min}$ (for the analysis of $\mathrm{SC} \rightarrow \mathrm{FC}$) and $1/d > 1/d_T$ and $\langle L \rangle_s > L_{\min}$ (for the analysis of $\mathrm{FC} \rightarrow \mathrm{SC}$).

The comparisons between this selection method and the selection via $\bar{N}_T$ and $1/d_T$ alone (as was used in the previous chapter) are shown respectively in Figures 5.6 and 5.7. We find that, with the exception of the distribution of intra-hemispheric connection lengths inferred from rsFC (first shown in Figure 5.6c-d), all results remain qualitatively similar with the inclusion versus exclusion of these unphysical connections.

In the assessment of $\mathrm{SC} \rightarrow \mathrm{FC}$, unphysical connections predominantly occupy the bulk of short, sparse intra-hemispheric connections (Figure 5.6b). Due to the large number of region pairs that fall within this structural subgroup, the relative number of region pairs linked by unphysical connection lengths is small in comparison, and therefore their inclusion does not significantly alter the observed shifts in resting-state (Figure 5.6c-d) or task-driven (Figure 5.6e-h) functional connectivity produced by different structural subgroups of connections.

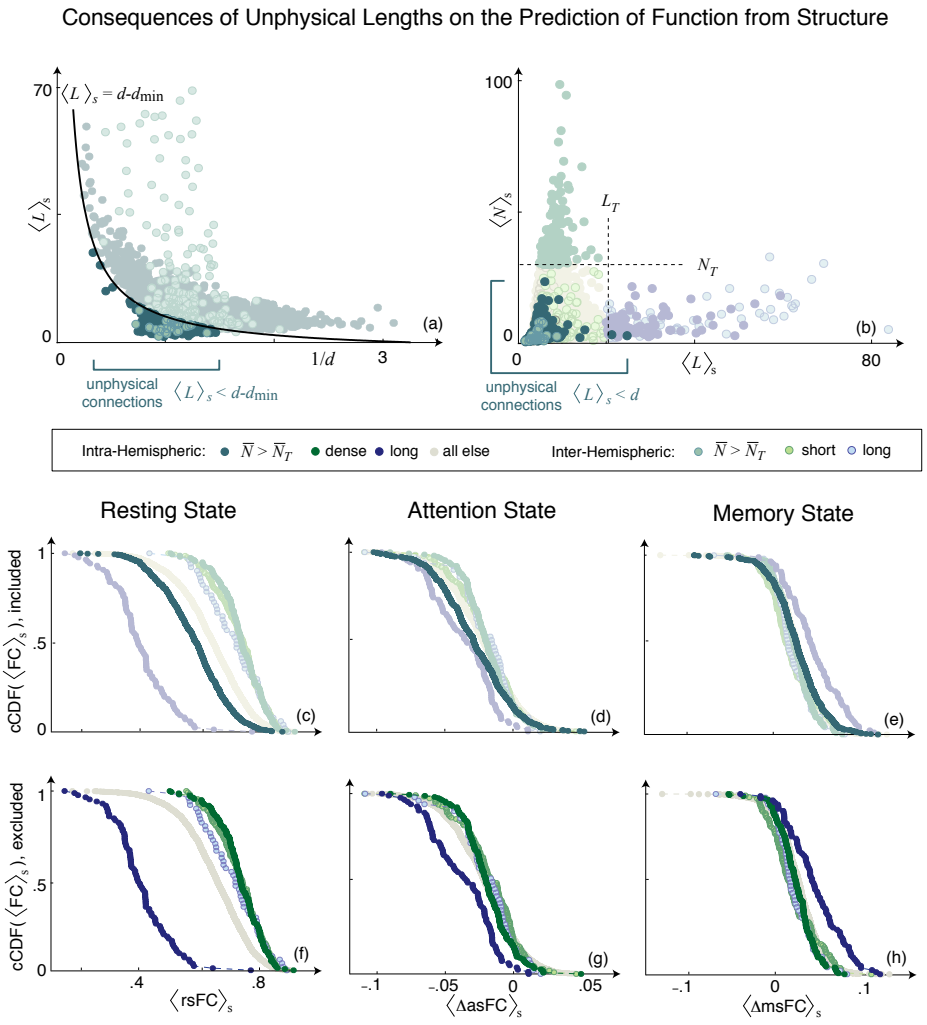

Figure 5.6: Consequences of Inconsistent Connectivity on SC → FC. (a) The presence of unphysical average connection lengths $\langle L \rangle_s < L_{\min}$ arises from inconsistently connected region pairs selected via $\bar{N}$. (b) Connections with unphysical lengths are predominantly low in number and short in length. (c-h) Distributions of average functional measures (c,f) $\langle \mathrm{rsFC} \rangle_\mathrm{s}$, (d,g) $\langle \Delta \mathrm{asFC} \rangle_\mathrm{s}$ and (e,h) $\langle \Delta \mathrm{msFC} \rangle_\mathrm{s}$ produced by structural subgroups of connections in the (c-e) presence and (f-h) absence of unphysical connections. As these unphysical connections predominantly fall into the large group of short, sparse intra-hemispheric connections, their presence does not qualitatively affect the observed shifts in the functional connectivity produced by inter-hemispheric, dense intra-hemispheric, and long intra-hemispheric connections.

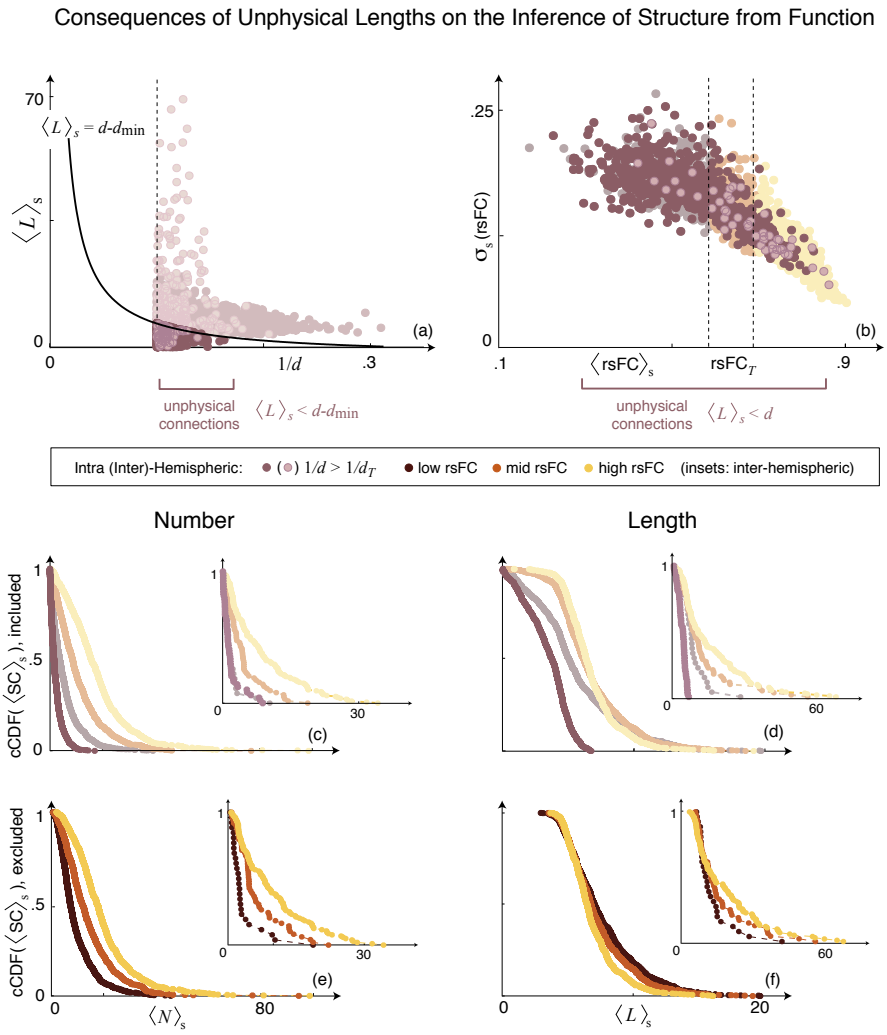

Figure 5.7: Consequences of Inconsistent Connectivity on FC $\rightarrow$ SC. (a) Unphysical average connection lengths $\langle L \rangle_s < L_{\min}$ arising from inconsistently connected region pairs selected via $1/d_T$. (b) Unphysical connections predominantly link weakly-correlated intra-hemispheric regions, and therefore the removal of these connections from the analysis does not qualitatively alter inter-hemispheric distributions. (c-f) Distribution of average structural measures (c,e) $\langle N \rangle_s$ and (d,f) $\langle L \rangle_s$ that support weak, intermediate, and strong correlations, shown in the (c-d) presence and (e-f) absence of unphysical connections. Inconsistent connections are (c) low in number (d) and short in length. Inconsistent connectivity produces a disproportionately high density of short intra-hemispheric connections (d). The removal of these regions from the analysis recovers a consistent relationship between increasingly short intra-hemispheric connections and increasingly strong rsFC (f).

In the assessment of $\mathrm{FC} \rightarrow \mathrm{SC}$, unphysical connections predominantly occupy the weakly-correlated subgroup, and they show noticeably low connection numbers and lengths (Figure 5.7). The presence of unphysical connections increases the observed separation in connection number across functional subgroups by biasing the weakly-correlated subgroup toward lower connection numbers. The exclusion of unphysical connections from the analysis only slightly decreases, but does not remove, this separation. In comparison, the presence of unphysical connections significantly affects the distribution of connection lengths. Inspection of the length distributions shown in Figure 5.7d,f reveals that the inclusion of unphysical connections produces a disproportionately high density of connections with very short lengths, such that longer local connections appear to show stronger rsFC. The exclusion of unphysical connections eliminates the excess of short connection lengths, thereby altering the distributions such that increasingly strong intra-hemispheric correlations are consistently supported by increasingly short connections. As was shown above (Figure 5.2), this finding is robust to variations in the thresholds used to both select and partition functional subgroups.

Note that a similar argument can be made to exclude connections with unphysical average numbers $\langle N \rangle_s < 1$. This constraint excludes a smaller subset of connections that is nearly fully contained within the subset excluded via $\langle L \rangle_s < L_{\min}$. This constraint therefore reduces, but does not eliminate, the altered distribution of short connection lengths. All other results remain qualitatively similar with the removal of connections with unphysical numbers.

5.5.3 Extension to Subject-Specific Networks

Analogous techniques can be applied to single-subject networks, for which the two possible constraints imposed on representative brain networks (implemented via the exclusion of connections with $\langle L \rangle_s < L_{\min}$ or $\langle N \rangle_s < 1$) reduce to the single constraint of excluding absent connections within subject-specific networks.

Consistent with the analysis of representative brain networks, the removal of absent connections does not alter the subject-specific values of resting-state $\langle \mathrm{rsFC} \rangle_\mathrm{c}$, attention-state $\langle \Delta \mathrm{asFC} \rangle_\mathrm{c}$, or memory-state $\langle \Delta \mathrm{msFC} \rangle_\mathrm{c}$ functional connectivity observed across structural subgroups (Figure 5.8), nor does it alter the subject-specific values of inter- and intra-hemispheric connection number $\langle N \rangle_c$ observed across functional subgroups (left column of Figure 5.8). However, as was observed in the representative brain network, the removal of absent connections alters the subject-specific values of inter- and intra-hemispheric connection length $\langle L \rangle_c$ (right column of Figure 5.8 and Figure 4.9f). Whereas with the inclusion of absent connections, strongly-correlated inter- and intra-hemispheric regions appear to be linked by longer connections, the removal of absent connections reveals that increasing intra-hemispheric rsFC is supported by decreasing connection lengths, while increasing inter-hemispheric rsFC shows minimal variations in connection length. These results confirm that the properties of representative brain networks are consistently observed within subject-specific networks for both the inclusion and exclusion of unphysical connections.

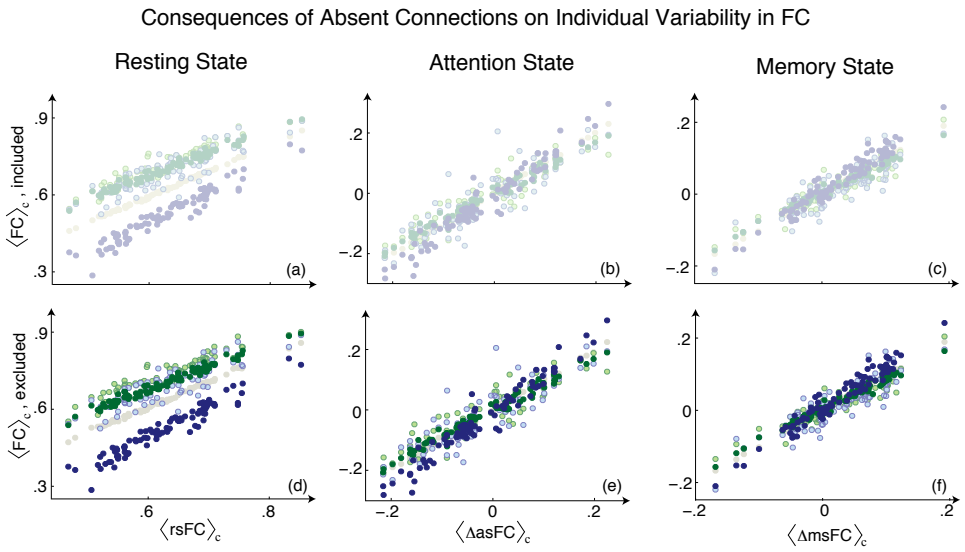

Figure 5.8: Consequences of Absent Connectivity on Individual Variability in Functional Connectivity. Absent connections (with $L = N = 0$) do not affect the average resting-state ($\langle \mathrm{rsFC} \rangle_c$), attention-state ($\langle \Delta\mathrm{asFC} \rangle_c$), or memory-state ($\langle \Delta\mathrm{msFC} \rangle_c$) functional connectivity across subjects. The inclusion (top row) versus exclusion (bottom row) of absent connections produces qualitatively similar shifts in FC across subjects.

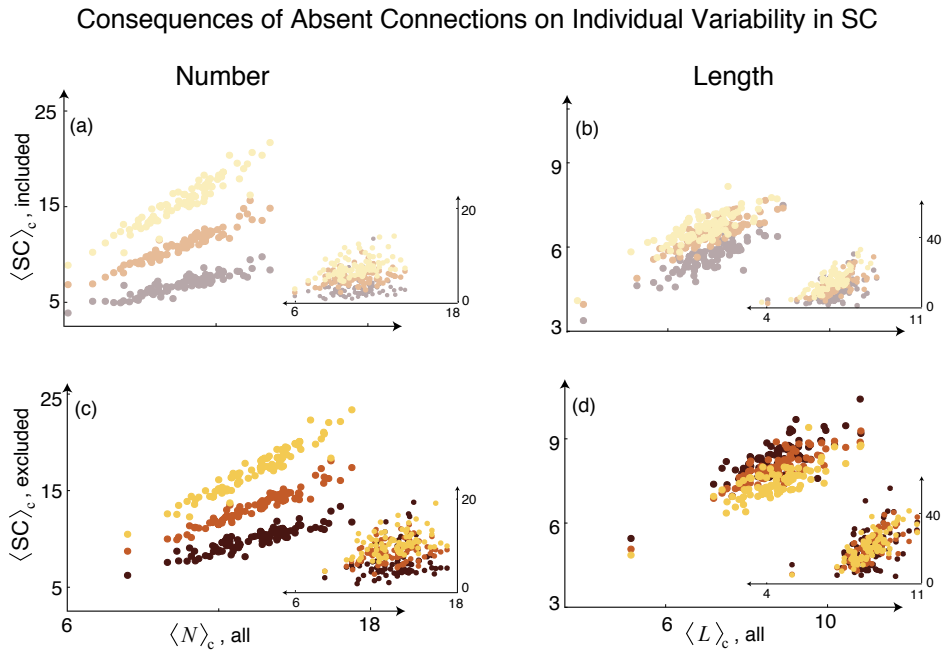

Figure 5.9: Consequences of Absent Connectivity on Individual Variability in Structural Connectivity. Absent connections (with $L = N = 0$) are shown to reduce, across subjects, the (a) average number $\langle N \rangle_c$ and (b) average length $\langle L \rangle_c$ of both inter- and intra-hemispheric connections. (c) Removal of absent connections from the analysis increases $\langle N \rangle_c$ across subjects while preserving the qualitative relationship between increasing $\langle N \rangle_c$ and increasing rsFC. Removal of absent connections similarly increases $\langle L \rangle_c$, but it also affects the qualitative relationship between $\langle N \rangle_c$ and rsFC. With the inclusion of absent connections, strongly-correlated intra- and inter-hemispheric regions appear to be linked by longer connections; however, the removal of absent connections shows that strongly-correlated intra-hemispheric regions are linked by shorter, rather than longer, connections, while strongly- and weakly-correlated inter-hemispheric regions show minimal differences in length. These results are consistent with those results obtained from the removal of connections with unphysical lengths in the representative brain.

5.6 Discussion

Together, this analysis confirms that the relationships between structural and functional connectivity, detailed in Chapter 4, are consistently observed across variations in the selection and thresholding of region pairs and across different choices of cognitive (memory) tasks. However, care must be taken in the selection of region pairs for the inference of structural from functional properties. The introduction of inconsistently-connected region pairs can alter the distributions of apparent connection lengths, an artifact that can be removed through appropriate thresholding or through the inclusion of structural information during the region selection process.

Chapter 6

Structurally-Constrained Relationships between Functional Brain States

"The modern geography of the brain has a deliciously antiquated feel to it–rather like a medieval map with the known world encircled by terra incognito where monsters roam."

–David Bainbridge, from *The Strange Anatomy of the Brain, New Scientist*, 2008

6.1 Introduction

The brain is inherently dynamic, whether observed at rest or during the performance of complex tasks. Despite predictions that resting-state neural activity would be noisy and unconstrained, the human brain has been shown to exhibit correlated patterns of neural activity, even in the absence of any goal-directed behavior [123]. In particular, a diffuse

set of brain regions has been shown to exhibit consistently strong correlations both across individuals and across cognitive states [124]. This default mode network (DMN) is widely considered to serve as a baseline of large-scale neural activity, from which task-driven fluctuations above (activations) and below (deactivations) can be assessed.

The robust features of the default mode are thought to reflect the underlying functional organization of the brain [125], such that disruptions to this organization may be indicative of altered functional states [126]. As a result, there is much interest in describing the structural and functional properties of the DMN. Furthermore, given that the brain employs the same structural architecture for both default mode and task-driven function, the question has been raised as to whether the propensity for task-driven activity is encoded in the properties of such resting-state functional networks. However, the extent to which widespread correlations in neural activity are robustly related across cognitive states is still unclear, as are the structural features that could support such state relationships.

Recent studies of resting-state neural activity have identified two functional networks, denoted task-positive (TP) and task-negative (TN) networks, composed of regions known to activate and deactivate during the performance of attention tasks relative to their behavior at rest [127]. The resting-state correlation strength between a subset of these TP and TN regions has further been shown to predict individual variability in task-induced correlation strength during a goal-directed attention task [128]. While these findings suggest that TP and TN regions encode both resting-state and task-driven

function, it is unclear as to whether widespread interactions between these networks can distinguish between different tasks. Furthermore, the extent to which structural or functional disruptions could impact these interactions is not presently known.

Motivated by these open questions, we build upon the work presented in Chapters 4 and 5 to quantify the structural features that constrain TP and TN interactions across resting, attention, and memory states. In comparison to the work presented in the previous chapters, which assumed no knowledge of the specific anatomical regions involved in structure-function interactions, we now ask whether the inclusion of region-specific information impacts the observed relationships between structure and function both across individuals and across cognitive states.

We group anatomical brain regions based on their involvement in the TP and TN networks defined in [127]. We show that the numbers of connections that mediate interactions within versus between these task-related networks differentially support strong resting-state and task-directed functional correlations. We further show that a mapping onto the space of interactions within versus between TP and TN networks can be used to distinguish between resting, attention, and memory activity. Comparison across subjects reveals that this space shows striking order, enabling the identification of groups of subjects that exhibit similar state-space relationships. Importantly, we show that subjects exhibiting improbable state-space properties show abnormal behavioral performance during both attention and memory tasks. This suggests that further characterization of state-space groupings may enable the prediction of task-directed function from resting-

state activity and may additionally help identify signatures of altered functional states.

6.2 Constructing Brain Networks

Evidence suggests that spontaneous neural activity can be separated into two functional networks, denoted task-positive (TP) and task-negative (TN), based on resting-state correlations within and anti-correlations between networks [127]. Although defined based on resting-state neural activity, these networks are composed of regions known to routinely increase and decrease in activation during attention-demanding cognitive tasks relative to their activity at rest [127].

To assess interactions within and between these networks, we identify regions within our 600 ROI atlas (introduced in Chapter 4) that overlap wholly or partially with the TP and TN regions defined in [127]. We define the remaining regions as “other" (OTH) regions. We evaluate structural and functional connectivity between three types of region pairs composed of (*i*) two task-positive regions (PP), (*ii*) two task-negative regions (NN), and (*iii*) one task-positive and one task-negative region (PN). These interactions can then be compared to the remaining set of interactions between task-positive, task-negative, and other regions (PO, NO, and OO; see Figure 6.1 for a schematic of possible interactions).

6.2.1 Structural versus Functional Brain Networks

As described in Chapter 4, structural connectivity (SC) is obtained from DTI measurements via a tractography algorithm used to identify white matter tracts linking ROI pairs.

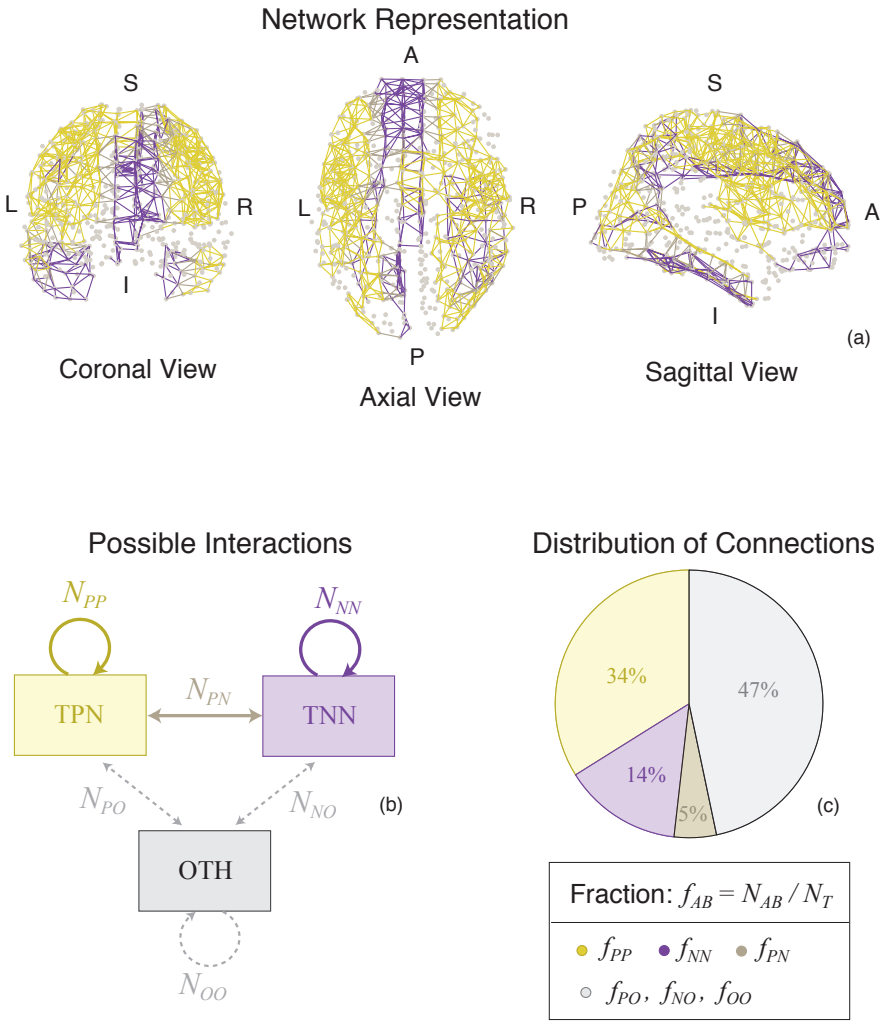

Figure 6.1: Interactions within and between Task-Positive and Task-Negative Networks. We identify regions in our atlas that belong to the task-positive (TP) and task-negative (TN) brain regions described in [127], and we label all remaining regions as "other" (OTH) regions. There are then six possible types of interactions between these three types of regions. We focus on three of these interactions: those between two task-negative regions (NN), between two task-positive regions (PP), and between a task-positive and a task-negative region (PN). These connections are highlighted in the coronal, axial, and sagittal views of the representative brain network. Grey nodes mark region centers, and lines mark region pairs that are linked by one or more tract in the representative brain. Note that curvilinear tracts are represented as straight lines.

We isolate one measure of SC, namely the total number N of white matter tracts linking two regions. We can then partition N into three groups based on the previously-defined interactions: N_{PP}, N_{NN}, and N_{PN}. We can compare these connection numbers to those of the remaining interactions, $N_{\text{bulk}} = [N_{PO}, N_{NO}, N_{OO}]$

Functional connectivity (FC) is obtained from fMRI measurements by computing Pearson's correlations between scale 2 wavelet coefficients of BOLD time series measured in ROI pairs, where time series are averaged across voxels within each ROI. We define the strength of FC to be the correlation between two ROI time series, where FC is measured at rest (rsFC), in deviations $\Delta\text{asFC} = \text{asFC} - \text{rsFC}$ of the attention state (asFC) from rest, and in deviations $\Delta\text{msFC} = \text{msFC} - \text{rsFC}$ of the memory state (msFC) from rest. As with groupings in N, we can similarly partition state-specific FC into three groups: FC_{PP}, FC_{NN}, and FC_{PN}.

As described in Chapter 4, we will again refer to the average $\langle O \rangle$ of a given measure O, where O takes on values of SC or FC. When computed across subjects, we reference the quantity with the subscript s (e.g. $\langle O \rangle_s$), and when computed across connections within a single subject, we reference the quantity with the subscript c (e.g. $\langle O \rangle_c$).

6.2.2 Selection of Robustly-Connected Region Pairs

To identify robust relationships between SC and FC, we select the subset of PP, NN, and PN region pairs that are linked by one or more white matter tract in at least 80% of subjects. This equates to selecting region pairs with a consistency $\langle C \rangle_s$ greater than a

threshold value $\langle C \rangle_T = 0.8$, where C is a subject-specific binary number that quantifies the presence of absence of one of more white matter tract linking two brain regions (as initially defined in Chapter 4). For the subset of task-related region pairs above this consistency threshold, we construct representative and subject-specific brain networks by respectively measuring subject-averaged and subject-specific strengths of SC and FC between consistently-connected region pairs.

This selection method is a variation of the methods used in Chapter 4, where region pairs were selected via the structural and nonstructural measures $\bar{N}$ and $1/d$. These indirect measures of consistency were shown to be related to but distinct from the direct consistency $\langle C \rangle_s$. The selection of region pairs via purely structural and nonstructural measures was important for ensuring that no functional or structural bias was introduced into the respective analyses of $\mathrm{SC} \rightarrow \mathrm{FC}$ and $\mathrm{FC} \rightarrow \mathrm{SC}$. Here, we seek to understand how SC constrains relationships between FC across brain states, but we do not assess the reverse process of identifying constraints on SC imposed by FC. We can therefore use a purely structural measure to select region pairs, and we choose to use the direct measure of consistency $\langle C \rangle_s$.

The observed relationships between SC and FC presented in the remainder of the chapter are robust to the specific choice of $\langle C \rangle_T$ used to select region pairs (see Section 6.6). Furthermore, although the 2132 connections selected via this thresholding process comprise a small fraction of the possible 179700 connections between 600 ROIs, this number accounts for roughly half of all connections measured within any given subject.

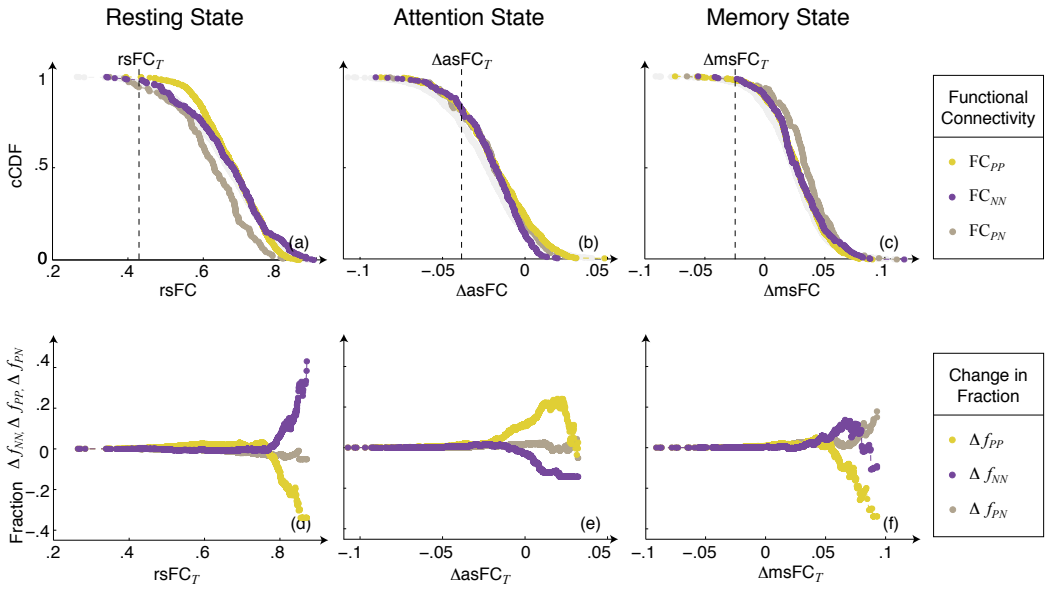

Figure 6.2: Contributions of Network Interactions to Strong Functional Connectivity. Top row: cCDFs of functional measures $\langle \mathrm{rsFC} \rangle$, $\langle \Delta\mathrm{asFC} \rangle$, and $\langle \Delta\mathrm{msFC} \rangle$ produced by PP, NN, and PN connections in the representative brain network. In the resting state, PP and NN connections show slightly stronger correlations than do PN connections. In the attention state, all distributions are strongly overlapping, and in the memory state, PN connections show slightly stronger correlations than do PP and NN connections. Bottom row: we compute the total number of tracts linking two regions whose correlations lie above a sliding threshold value, and we assess deviations Δf in this number from the value computed using the entire set of connections. In the resting state, we see an increased contribution from NN connections and decreased contributions from PP and PN connections in supporting strong correlations. In the attention state, we an increased contribution from PP connections and a decreased contribution from NN connections. In the memory state, we see an increased contribution from PN and NN connections and a decreased contribution from PP connections in supporting strong correlations.

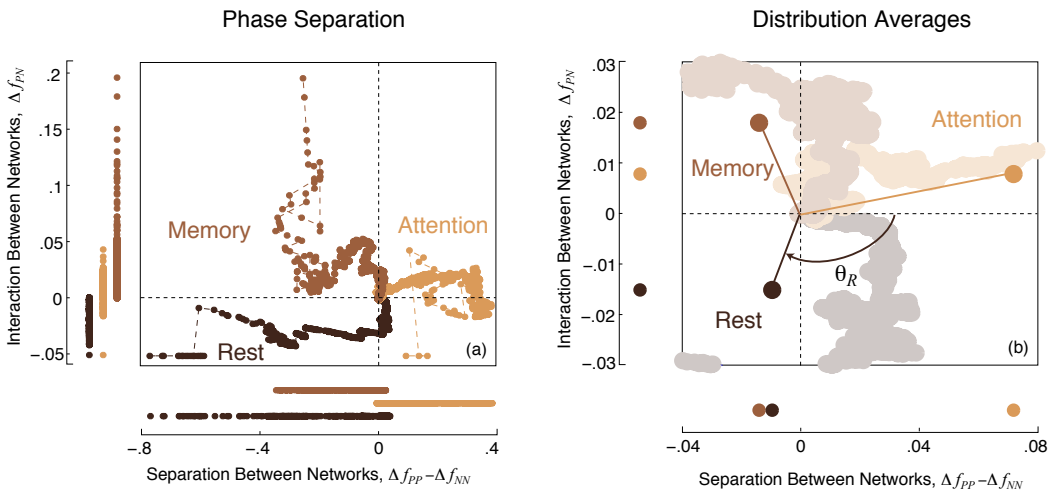

Figure 6.3: Structurally-Constrained Relationships Between Cognitive States. Deviations $\langle \Delta f_{PP} - \Delta f_{NN} \rangle_s$ and $\langle \Delta f_{PN} \rangle_s$ distinguish between rest, attention, and memory states. The averages of these distributions can be characterized by an angular component θ and a radial component r, which mark the positions of each distribution in the phase-like space defined by $\langle \Delta f_{PP} - \Delta f_{NN} \rangle_s$ and $\langle \Delta f_{PN} \rangle_s$.

6.3 Functional Connectivity of Task-Based Networks

In assessing the strength of functional interaction between TP and TN regions, we expect to observe differences in the strength of functional correlation measured within versus between networks. As TP and TN networks were defined based on strong resting-state correlations within but anti-correlations between regions within each network, we predict that PP and NN connections will exhibit stronger rsFC than PN connections. Furthermore, given that TP and TN regions are also known to respectively activate and deactivate during attention tasks, we predict that PP connections will show more pronounced changes $\langle \Delta \mathrm{asFC} \rangle_\mathrm{s}$ than NN and PN connections. To evaluate these predicted relationships, we compare shifts in the complementary cumulative distribution functions

(cCDFs) of $\langle \mathrm{rsFC} \rangle_\mathrm{s}$, $\langle \Delta \mathrm{asFC} \rangle_\mathrm{s}$, and $\Delta \mathrm{msFC} \rangle_\mathrm{s}$ produced by PP, NN, and PN connections relative to the distribution produced by the remaining bulk of PO, NO, and OO connections.

We find that PP and NN connections produce similar distributions of functional connectivity to one another, but their relationship to the distribution of PN connections varies between resting, attention, and memory states (Figure 6.2a-c). In the resting state, correlations within networks are stronger than correlations between networks, a result that is consistent with the definition of TP and TN networks based on strong resting-state correlations within each network and anti-correlations between networks (Figure 6.2a). In the attention state, we see similar changes in the strength of correlations within and between networks (Figure 6.2b), while the memory state is marked by an increase in the strength of correlation between relative to within networks (Figure 6.2c).

In comparing the leading edges of each distribution, we see task-dependent differences in the types of connections that support the strongest correlations. To assess differences in the relative contribution of PP, NN, and PN connections to strong functional correlations, we apply sliding thresholds to the resting-state $(\mathrm{rsFC_T})$, attention-state $(\Delta \mathrm{asFC_T})$, and memory-state $(\Delta \mathrm{msFC_T})$ distributions of functional connectivity. For the set of connections above these sliding thresholds, we compute the change in the fraction $\langle \Delta f \rangle_s$ of connections that link a task-positive with a task-negative region $(\langle \Delta f_{PN} \rangle)_s$ versus the fraction that link two task-positive $(\langle \Delta f_{PP} \rangle_s)$ and two task-negative $(\langle \Delta f_{NN} \rangle_s)$ regions, where Δf is given by:

$$\Delta f_{ab} = \frac{N_{ab}\left(\mathrm{FC} > \mathrm{FC_T}\right)}{N_{\mathrm{all}}} - \frac{N_{ab}}{N_{\mathrm{all}}}, \tag{6.1}$$

and $ab = \{PP, NN, PN\}$. The quantity N_{ab}/N_{all} specifies the baseline values $f_{PP} = .31$, $f_{NN} = .13$, and $f_{PN} = .05$ computed from the full set N_{ab} of unthresholded connections.

We find that increasingly strong rsFC is supported by an increasingly large contribution from NN connections and a decreasingly small contribution from PP connections (Figure 6.2d), and we find that the reverse is true in the attention state (Figure 6.2e). Furthermore, we see an increase in contribution of NN connections to $\Delta\mathrm{msFC}$ (Figure 6.2f), suggesting an overlap between the resting and memory states. We additionally find task-dependent changes in the contribution of connections that couple task-positive and task-negative regions. In particular, increasingly strong rsFC is marked by a consistent decrease in contribution from PN connections, while increasingly strong $\Delta\mathrm{msFC}$ is marked by a consistent increase in their contribution.

Comparison of the relative contributions from PP, NN, and PN connections reveals that two quantities, $\langle \Delta f_{PN} \rangle_s$ and $\langle \Delta f_{PP} - \Delta f_{NN} \rangle_s$, vary in a state-dependent manner. The first quantity measures the relative contribution of PN connections to supporting strong correlations and can be understood as the strength of coupling between TP and TN networks. The second quantity measures the difference in contribution from PP versus NN connections to supporting strong correlations and can be understood as the degree of separation between TP and TN networks. In this manner, $\langle \Delta f_{PN} \rangle_s$ and

$\langle \Delta f_{PP} - \Delta f_{NN} \rangle_s$ represent the functional integration versus segregation between TP and TN networks.

Evaluation of the space define by $\langle \Delta f_{PN} \rangle_s$ and $\langle \Delta f_{PP} - \Delta f_{NN} \rangle_s$ reveals a striking separation between brain states. This is shown in Figure 6.3a, where the overlap of all distributions at the center point marks the bulk properties, and deviations from the center point mark changes in $\langle \Delta f \rangle_s$ as a function of increasing $\mathrm{FC_T}$. Biasing toward stronger correlations by increasing $\mathrm{FC_T}$ incrementally decreases the number of regions in consideration, resulting in the observed jitter at the edges of each distribution.

Importantly, this representation highlights the specific TP and TN interactions that support and distinguish each state; namely, (*i*) the resting state shows increased TN activity and decreased TP-TN coupling, (*ii*) the attention state shows increased TP activity and increased TP-TN coupling, and (*iii*) the memory state shows increased TN activity and increased TP-TN coupling.

The quantities $\langle \Delta f_{PN} \rangle_s$ and $\langle \Delta f_{PP} - \Delta f_{NN} \rangle_s$ can be viewed as a mapping of a "phase space," whereby the distribution averages $\langle \Delta f_{PN} \rangle_{s,c}$ and $\langle \Delta f_{PP} - \Delta f_{NN} \rangle_{s,c}$ can be characterized by a radial component r:

$$r = \sqrt{\langle \Delta f_{PN} \rangle_{s,c}^2 + \langle \Delta f_{PP} - \Delta f_{NN} \rangle_{s,c}^2} \tag{6.2}$$

and an angular phase θ:

$$\theta = \arctan\left(\frac{\langle \Delta f_{PN} \rangle_{s,c}}{\langle \Delta f_{PP} - \Delta f_{NN} \rangle_{s,c}} \right), \tag{6.3}$$

where θ is measured from the horizontal axis counterclockwise to $+180^\circ$ and clockwise to -180°, as shown in Figure 6.3b. In this representation, the resting state is characterized by a negative phase angle $\theta_R < -90^\circ$, for which the projections $\langle \Delta f_{PP} - \Delta f_{NN} \rangle_{s,c}$ and $\langle \Delta f_{PN} \rangle_{s,c}$ onto the x and y axes are both negative (strong NN interactions and reduced TP-TN coupling). The attention state is characterized by a positive phase angle $0 < \theta_A < 90^\circ$, for which the projections onto the x and y axes are both positive (strong PP interactions and increased TP-TN coupling). Finally, the memory state is characterized by a positive phase angle $\theta_M > 90^\circ$, for which the projection onto the x axis is negative (strong NN interaction) but onto the y axis is positive (increased TP-TN coupling). The angular component therefore characterizes the relative change in interactions within versus between networks, while the radial component characterizes the absolute strength of these interactions.

In addition to providing an objective measure of phase-space relationships, the variables (r_R, θ_R), (r_A, θ_A), and (r_M, θ_M) enable us to compare the state-dependent properties of the representative brain network, assessed here, with subject-specific brain networks, to be assessed in the following section.

6.4 Individual Variability in Phase Relationships

Given the pronounced separation between states observed in the representative brain network, we investigate the extent to which subject-specific brain networks show similar features of "phase separation."

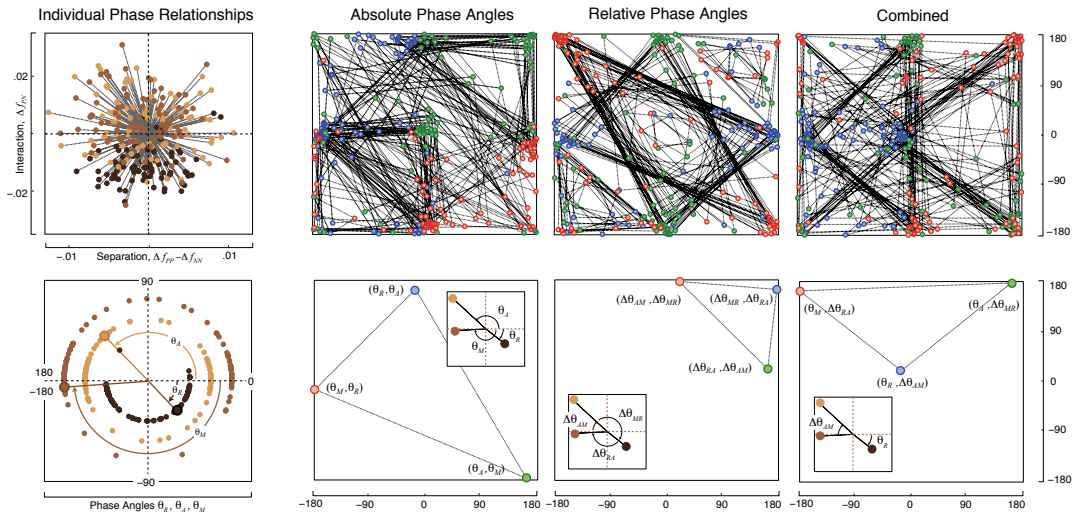

Figure 6.4: Individual Variability in Phase Relationships. Far upper left: subject-specific relationships between resting (r_R, θ_R), attention (r_A, θ_A), and memory (r_M, θ_M) states, shown respectively in black, tan, and brown. Individual subjects, marked by sets of three phases, show significant variability in phase relationships. Far lower left: angular components θ_R, θ_A, and θ_M of subject-specific phases. A majority of subjects show that the resting state is localized within the lower half plane($\theta < 0$), while attention and memory are localized about the horizontal axis ($\theta \sim 0°, \pm 180°$). Middle left, middle right, far right: remapping of subject-specific phase relationships by projecting θ_a onto θ_b (middle left), $\Delta\theta_{ab} = \theta_a - \theta_b$ onto $\Delta\theta_{bc} = \theta_b - \theta_c$ (middle right) and θ_a onto $\Delta\theta_{bc}$ (far right), where $\{a, b, c\}$ are cyclical permutations of R, A, and M. Below each remapping is a subject-specific example of the phase remapping, where each subject is represented by a triangle whose vertices mark the remapped phases. Subjects show a striking degree of order in the relationships between absolute phases (middle left), between relative phases (middle right), and between absolute and relative phases (far right).

To compare phase-space relationships across subjects, we compute subject-specific distributions of $\langle \Delta f_{PP} \rangle$, $\langle \Delta f_{NN} \rangle$, and $\langle \Delta f_{PN} \rangle$ across increasing values of $\mathrm{rsFC_T}$. For each subject, we compute the distribution averages $\langle \Delta f_{PN} \rangle_c$ and $\langle \Delta f_{PP} - \Delta f_{NN} \rangle_c$, from which we can define subject-specific values of (r_R, θ_R), (r_A, θ_A), and (r_M, θ_M).

Comparison of phase relationships across subjects reveals significant variability in both the angular and radial components of resting-, attention-, and memory-state phases

(figure 6.4). Isolating state-specific angular components shows that the resting state tends to be localized in the lower half plane, for which the coupling between TP and TN networks is reduced. This behavior agrees with the definition of TP and TN networks based on anticorrelations between networks observed in the resting state ([127]). In comparison, the angular components of both the attention and memory states tend to be localized around $\theta = 0°, \pm 180°$, for which there is a large functional separation between TP and TN networks.

To isolate similarities in phase relationships across subjects, we perform three remappings of the original phase diagram:

1. $[(r_R, \theta_R),\ (r_A, \theta_A),\ (r_M, \theta_M)] \longrightarrow [(\theta_R, \theta_A),\ (\theta_A, \theta_M),\ (\theta_M, \theta_R)]$

2. $[(\theta_R, \theta_A),\ (\theta_A, \theta_M),\ (\theta_M, \theta_R)] \longrightarrow [(\Delta\theta_{RA}, \Delta\theta_{AM}),\ (\Delta\theta_{AM}, \Delta\theta_{MR}),\ (\Delta\theta_{MR}, \Delta\theta_{RA})]$

3. $[(\theta_R, \theta_A),\ (\theta_A, \theta_M),\ (\theta_M, \theta_R)] \longrightarrow [(\theta_R, \Delta\theta_{AM}),\ (\theta_A, \Delta\theta_{MR}),\ (\theta_M, \Delta\theta_{RA})]$

where $\Delta\theta_{ab} = \theta_a - \theta_b$, and $\Delta\theta \rightarrow \Delta\theta \mp 360°$ if $\Delta\theta \gtrless \pm 180°$ The first remapping isolates absolute phase angles, the second remapping isolates relative differences in phase angles, and the third remapping compares absolute and relative phase angles. All three remappings are shown in Figure 6.4, where each individual is represented by a triangle whose vertices are given by the coordinates in (1)-(3) above.

Despite the significant variability observed in Figure 6.4a, we see that there is significant order in the relationships between absolute and relative phase angles exhibited across subjects. This order is marked in Figure 6.4 (middle left, middle right, and far right)

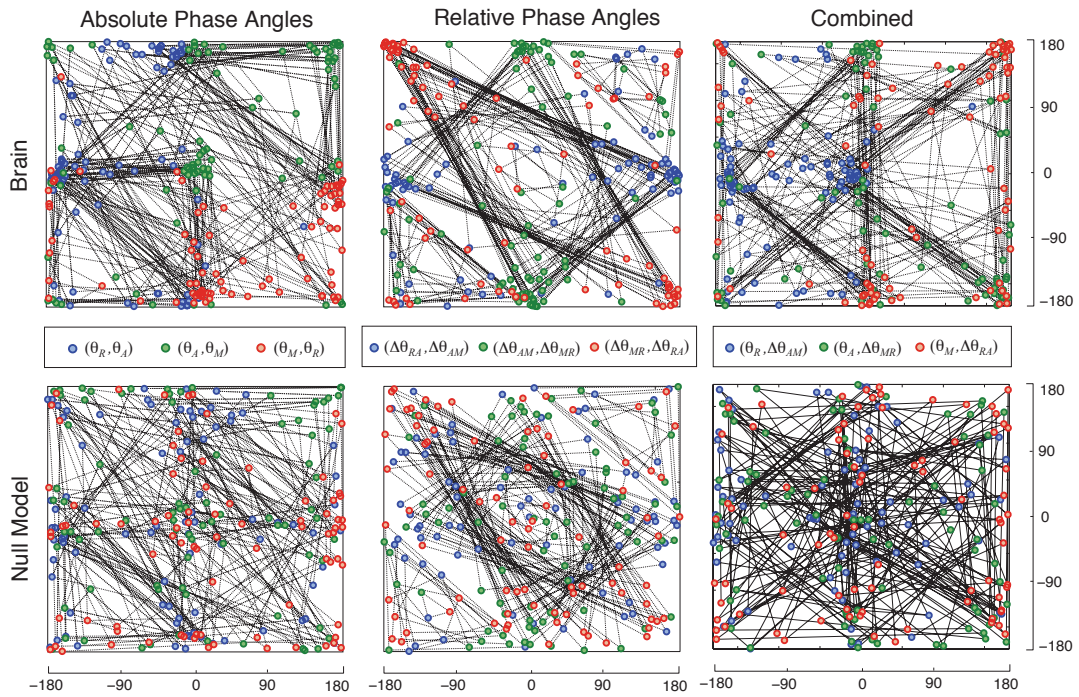

Figure 6.5: Comparison of Individual Phase Relationships to Null Model. Comparison between remappings of phase space relationships shown in Figure 6.4 (top row) and null model in which TP, TN, and OTH region labels are randomly reassigned. In all cases, the phase space relationships produced by the real brain show significantly more organization that do null model phase space relationships. The real brain shows groupings of many individuals that show similar relationships between absolute and relative phases, marked by the clustering of points with the same color. In comparison, the null model shows significant noise in the relationships between relative and absolute phases, and there is no clustering of points of the same color.

by the large degree of clustering in points of the same color, and similarly in the largely overlapping sets of triangles that link these clusters. This order is particularly striking when comparing relative phase differences between brain states, as shown in the middle right of Figure 6.4. Together, these results show that there are probable configurations of absolute phase angles, relative phase angles, and mappings between the two.

To confirm that the observed order is not an artifact of our analysis techniques, we

compare these remappings to those obtained from a null model in which TP, TN, and OTH region labels are randomly reassigned to regions within our 600-ROI atlas. These comparisons, shown in Figure 6.5, reveal that random reassignment of region labels removes the observed structure in phase relationships. Under this reassignment, we no longer see clustering of points of the same color, nor do we see as strong a degree of overlapping sets of triangles linking these points. Similar results were achieved using a null model in which the numbers of connections linking brain regions, rather than region labels, were randomly reassigned.

6.5 Links between Phase-Space Groupings and Behavior

The existence of structured relationships between resting, attention, and memory states suggests that the shared features of these relationships may reflect underlying organizational principles that constrain both structural and functional connectivity. Furthermore, the observation of multiple clusters of subjects, each exhibiting different phase-space relationships, raises the question as to whether these clusters relate to external measures of behavioral performance. Given that the existence of phase space groupings can only be uncovered by assessing large numbers of subjects, it may additionally be possible to identify regions of phase space that, being unoccupied by healthy individuals, are predictive of altered functional states.

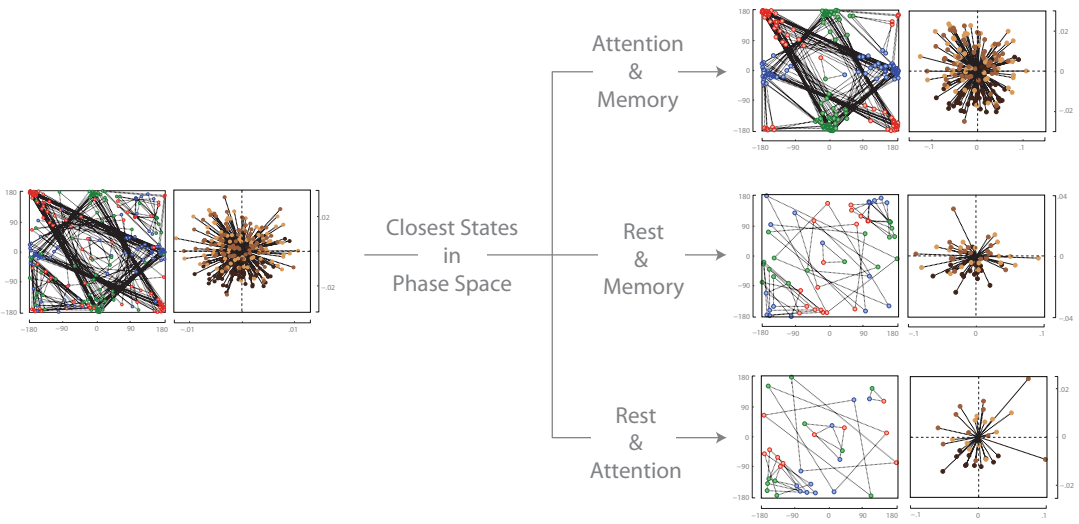

Figure 6.6: Phase Space Groupings of Individuals. Phase relationships for all subjects are shown on the left. Each of the three subgroups, separated based on the relative proximity of brain states, are further partitioned in Figures 6.8, 6.9, and 6.10.

6.5.1 Characterization of Phase-Space Groupings

In what follows, we detail the method by which we separate groups of subjects with similar phase relationships. As we see the most striking order in the relative phases between brain states (Figure 6.4), we first group subjects based on the relative proximity of different brain states to one another:

Primary Grouping: Relative Proximity of Brain State Phases

(1) $|\Delta\theta_{AM}| < \{\Delta\theta_{RA}, \Delta\theta_{MR}\}$

(2) $|\Delta\theta_{MR}| < \{\Delta\theta_{RA}, \Delta\theta_{AM}\}$

(3) $|\Delta\theta_{RA}| < \{\Delta\theta_{AM}, \Delta\theta_{MR}\}$

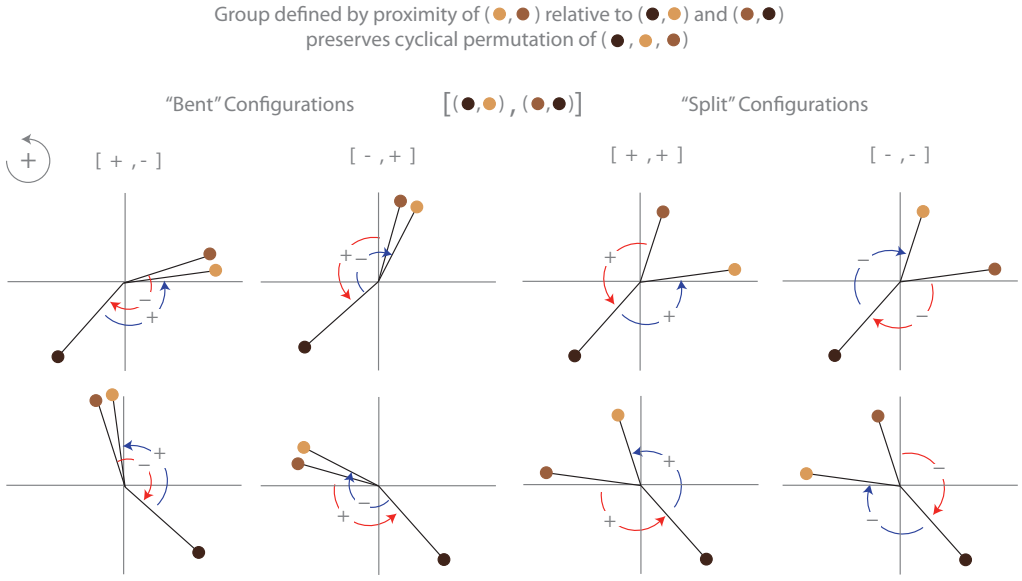

Figure 6.7: Schematic of Phase Relationships used to Group Subjects. Examples of the "bent" and "split" configurations defined by the Second Partition Groupings in the text.

These groups are shown in the right column of Figure 6.6. We see that a majority (55 of 84) of individuals fall into the first grouping, for which attention and memory states are closest in phase space. The next most populous group is that for which rest and memory are closest (18 of 84), and the least populous is that in which rest and attention are closest (11 of 84). In a broad sense, this suggests that, within this representation of TP and TN interactions, the two task states are most similar, while attention and rest (the two states used in [127] to define TP and TN networks) are the most dissimilar.

Groupings (1), (2), and (3) listed above are characterized by the relative phase angle $\Delta\theta_{AM}$, $\Delta\theta_{MR}$, and $\Delta\theta_{RA}$, respectively. We further partition each group based on relationships between the remaining two relative phases ($[\Delta\theta_{RA}, \Delta\theta_{MR}]$, $[\Delta\theta_{RA}, \Delta\theta_{AM}]$, and $[\Delta\theta_{AM}, \Delta\theta_{MR}]$ for groups (1), (2), and (3), respectively). We distinguish groups by the sign $[+, -]$ of these relative phases. Because relative phases are defined on the interval $[-180°, 180°]$, the sign of each relative phase characterizes the arrangement of the two closest phases relative to the third phase (see Figure 6.7). These groups are given by the following relationships for a pair of relative angles $[\Delta\theta_{\alpha\beta}, \Delta\theta_{\beta\gamma}]$:

Secondary Grouping: Configuration of Brain State Phases

(a) $[\Delta\theta_{\alpha\beta}, \Delta\theta_{\beta\gamma}] = [+, -]$, $[\Delta\theta_{\alpha\beta} > 0, \Delta\theta_{\beta\gamma} < 0]$

(b) $[\Delta\theta_{\alpha\beta}, \Delta\theta_{\beta\gamma}] = [-, +]$, $[\Delta\theta_{\alpha\beta} < 0, \Delta\theta_{\beta\gamma} > 0]$

(c) $[\Delta\theta_{\alpha\beta}, \Delta\theta_{\beta\gamma}] = [+, +]$, $[\Delta\theta_{\alpha\beta} > 0, \Delta\theta_{\beta\gamma} > 0]$

(d) $[\Delta\theta_{\alpha\beta}, \Delta\theta_{\beta\gamma}] = [-, -]$, $[\Delta\theta_{\alpha\beta} < 0, \Delta\theta_{\beta\gamma} < 0]$

where (α, β, γ) is any permutation of (R, A, M). Figure 6.7 shows a schematic of these definitions. We refer to the groups (a) and (b) as "bent" configurations because all three phases lie within one half plane, and we refer to groups (c) and (d) as "split" configurations because there is no half plane that contains all three phases. These secondary groups (a,b,c,d) are shown in Figures 6.8, 6.9, and 6.10 for the primary groups (1), (2), and (3).

Despite the wide range of phase relationships that could be realized under these broad groupings (see, e.g., Figure 6.7), subjects grouped based on relative phase relationships tend to also show similarities in absolute phases. This is a reflection of the structure observed in Figure 6.4e. For example, the most populated grouping G1a (Figure 6.8) could be achieved by the two different configurations shown in the left column of Figure 6.7. These two configurations would be characterized by the localization of rest in the lower left versus lower right quadrants. However, we see that subjects in grouping G1a primarily exhibit one type of configuration in which rest is localized in the lower right quadrant and attention and memory are bent toward the horizontal axis in the upper left quadrant (analogous to the configuration in the lower left panel of Figure 6.7).

6.5.2 Comparison of Behavioral Measures Across Groups

We can use the groupings enumerated above to assess whether subjects with similar phase-space relationships show similar performance characteristics during behavioral tasks. We focus on three performance measures taken for both the attention and memory tasks: average reaction time (RT), d' ("d-prime"), and a criterion switch score (CS).

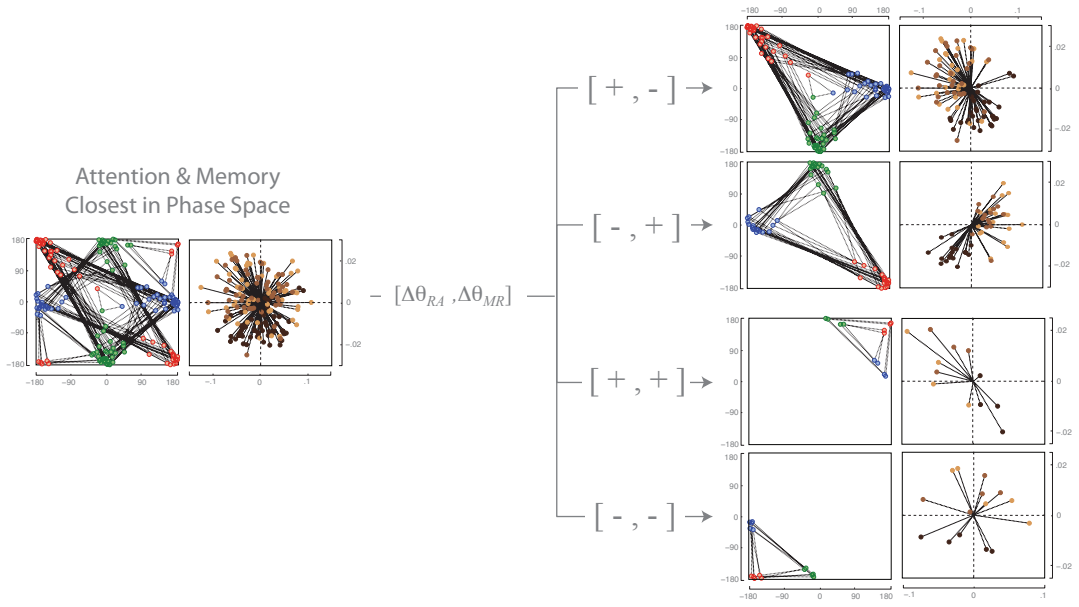

Figure 6.8: Group 1: Attention and Memory are Closest in Phase Space. Series of left and right panels indicate relative phase relationships and absolute phase diagrams, respectively. Group 1 separates very cleanly into four subgroups, each with very similar phase relationships. The two most populated groups are "bent" configurations and show rest and attention/memory in the (G1a) lower right and upper left quadrants and in the (G1b) lower left and upper right quadrants, respectively.

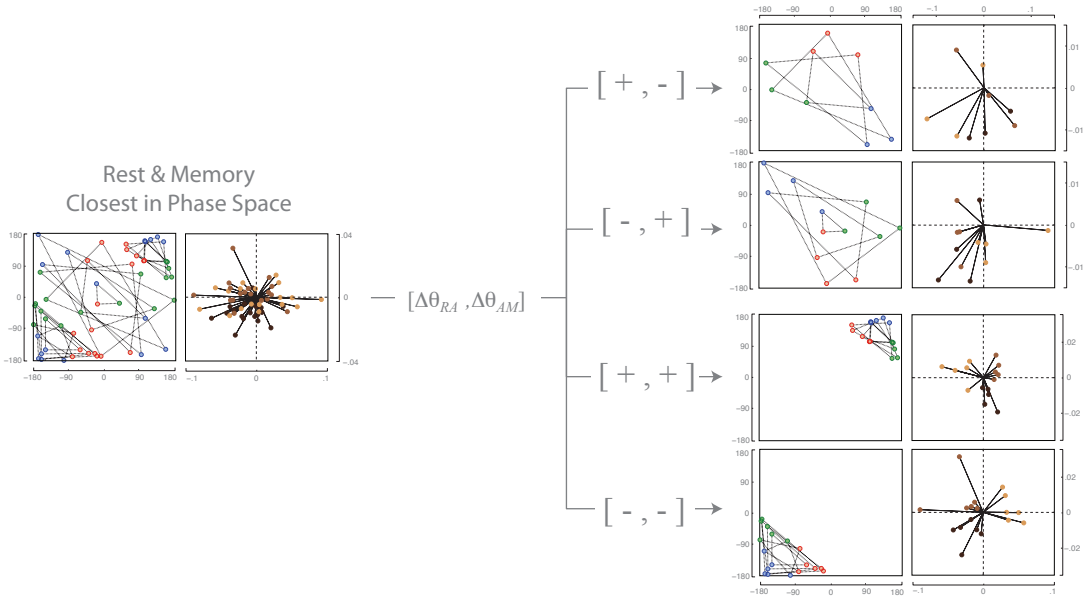

Figure 6.9: Group 2: Rest and Memory are Closest in Phase Space. Series of left and right panels indicate relative phase relationships and absolute phase diagrams, respectively. The two most populated groups, G2c and G2d, are "split" configurations, each of which cleanly separates all three brain states across all subjects within the group.

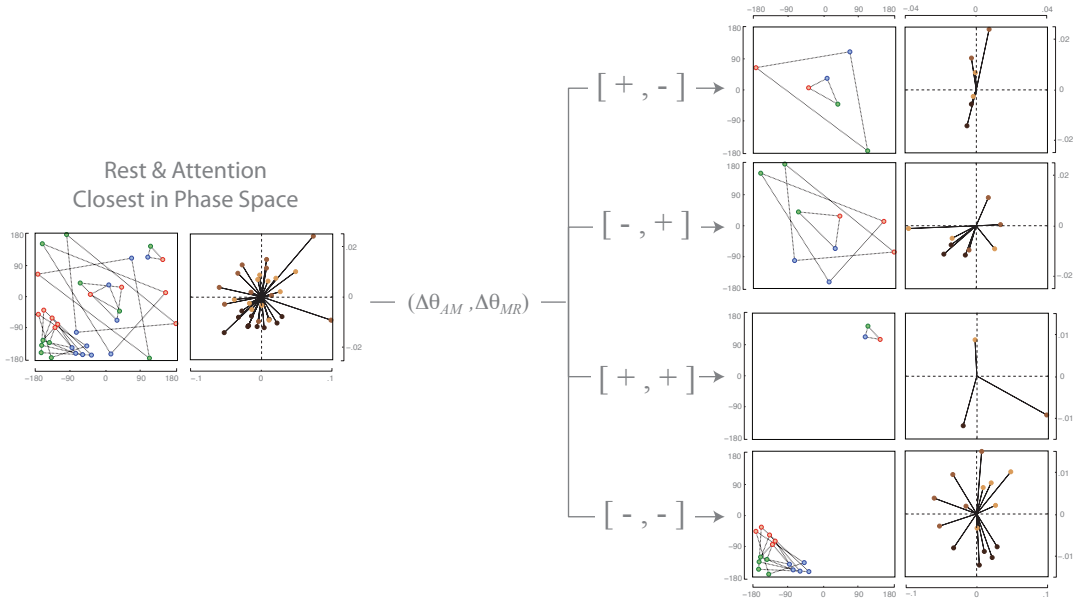

Figure 6.10: Group 3: Rest and Attention are Closest in Phase Space. Series of left and right panels indicate relative phase relationships and absolute phase diagrams, respectively. The most populated group, G3d, is a "split" configuration. Group G3a, which shows strong interaction and weak separation between TP and TN networks (vertical versus horizontal projection of phases in absolute phase diagram), also deviates significantly in performance from the other subjects (see Figure 6.11)

Average RT measures the average time between the appearance of a stimulus and a subject's response (via a button press) to that stimulus. The measure d' assesses perceptual sensitivity, with higher values indicating higher sensitivity. Finally, CS measures strategic flexibility in switching between decision making strategies, with higher values indicating the ability to more readily switch strategies. A detailed discussion of these performance measures can be found in [122].

We find that subjects exhibit significant variability in both attention and memory performance. Although it is possible to distinguish trends in the average performance of subjects within different groupings, the high variance in performance across subjects undermines the significance of such trends. This may be a result of the limited number of subjects in this study, and comparisons across larger sample sizes may reveal more reliable relationships between phase space groupings and behavioral performance. However, one notable group, G3a, deviates significantly from the performance exhibited by the remaining subjects. Figure 6.11 shows the values of RT, d', and CS for group G3a relative to the remaining subjects. This group is characterized by slow reaction times, low perceptual sensitivity, and high strategic flexibility. Inspection of the raw phase relationships (lower right of Figure 6.11) reveals that this group shows strong interactions between TP and TN networks (vertical projections) and small separations between networks (horizontal projections), suggesting that this combination of TP and TN interactions is simultaneously detrimental to reaction time and perceptual sensitivity but beneficial for strategic flexibility.

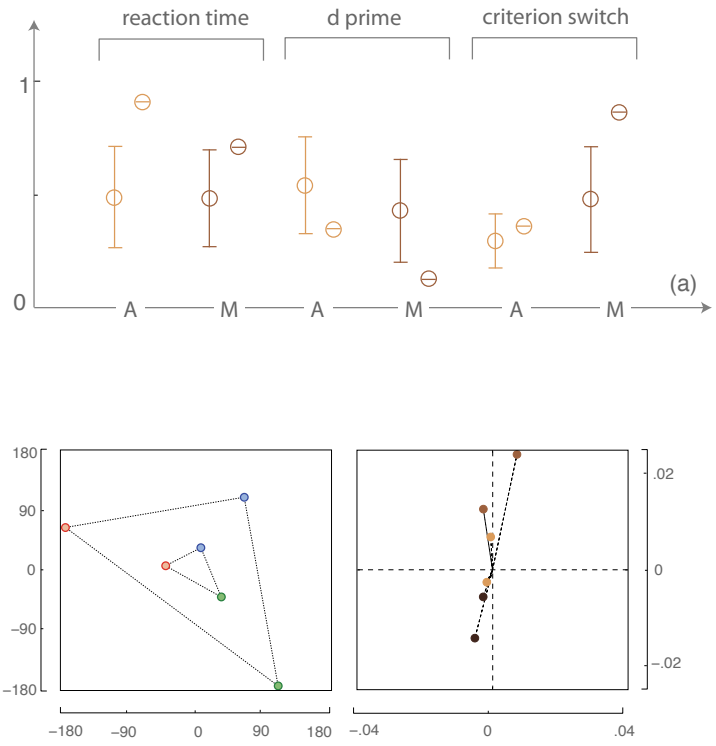

Figure 6.11: Phase Space Grouping Exhibiting Abnormal Behavioral Performance. Trop row: Average values of reaction time RT, perceptual sensitivity d', and criterion switch CS scores of strategic flexibility for attention and memory tasks produced by subjects in group G3a relative to the same average measures produced by the full group of individuals. Performance measures are rescaled between their minimum and maximum values. Subjects in G3a exhibit abnormally high reaction times, low values of d', and high criterion switch scores in both the attention and memory tasks. Bottom row: Representation of relative phase relationships (left) and raw phase space values (right) of the two members of group G3a.

The finding that outliers in phase space also deviate significantly in their behavioral performance suggests that there are regions of phase space, or relationships between phases, that are linked to altered functional states. Further characterization of these relationships across a larger number of subjects, more performance measures, and genetic or cognitive traits may better isolate these outlier regions.

6.6 Robustness to Thresholding

In constructing the representative brain network for the comparison of TP and TN networks, we first isolated region pairs that were consistently structurally connected within 80% ($\langle C \rangle_T = .8$) of subjects. By varying this consistency threshold, we can explore how phase relationships may change with more or less robust connectivity.

In a manner similar to that used in the previous chapter, we repeat our analysis of the representative brain network across variations in the consistency threshold $\langle C \rangle_T$. As described in Section 6.3, we compute $\langle \Delta f_{PP} - \Delta f_{NN} \rangle_{s,c}$, and $\langle \Delta f_{PN} \rangle_{s,c}$ from $\langle \mathrm{rsFC} \rangle_\mathrm{s}$, $\langle \Delta \mathrm{asFC} \rangle_\mathrm{s}$, and $\langle \Delta \mathrm{msFC} \rangle_\mathrm{s}$, which mark the phase-space positions of the resting, attention and memory states.

Figure 6.12 shows these phase-space positions across variations in $\langle C \rangle_T$, where arrows indicate the direction of increasing $\langle C \rangle_T$. We find that the observed separation between brain states in the phase space defined by $\langle \Delta f_{PN} \rangle$ and $\langle \Delta f_{PP} - \Delta f_{NN} \rangle$ is consistently observed across all threshold values. The averages of both the resting state and attention state distributions move further from the origin with increasing $\langle C \rangle_T$, while the average

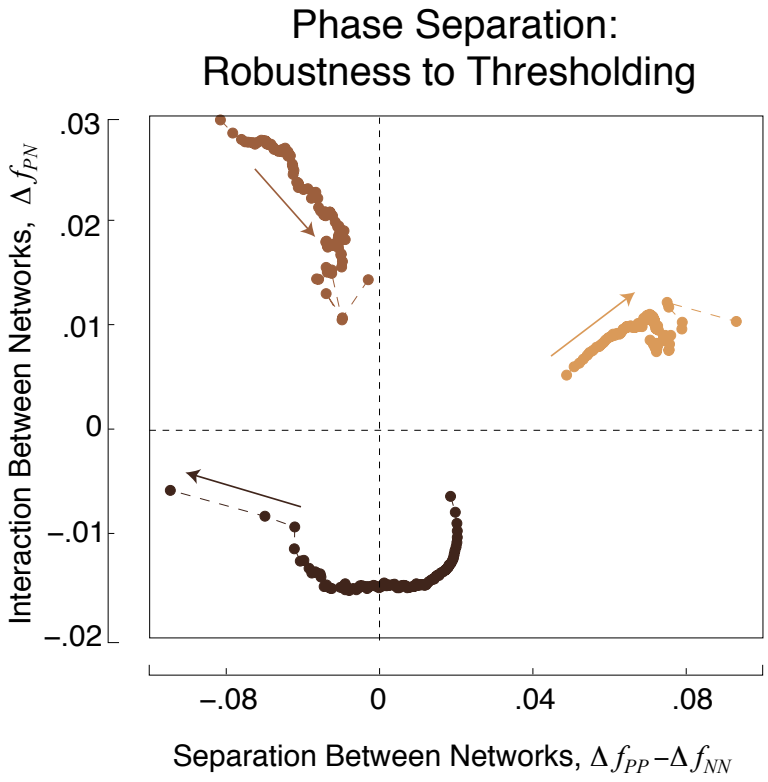

Figure 6.12: Robustness of Phase Relationships to Thresholding. Averages $\langle \Delta f_{PP} - \Delta f_{NN} \rangle_{s,c}$ and $\langle \Delta f_{PN} \rangle_{s,c}$ of the distributions produced by the representative brain network, shown in Figure 6.3, across variations in the consistency threshold $\langle C \rangle_T$. Arrows indicate the direction of increasing $\langle C \rangle_T$. Across all variations in threshold, distributions show a strong degree of separation. In biasing toward high consistency, the resting and attention states show stronger deviation from one another and from their expected values. In comparison, biasing toward higher consistency reduces the separation between the memory state and its expected value.

of the memory state distribution move closer to the origin. This suggests that robust connectivity facilitates a strong separation between rest and attention states, with memory sharing features of both states.

6.7 Discussion

Relating structural connectivity and functional activity across brain states is crucial for understanding how the brain utilizes the same architecture for the performance of different cognitive functions. Previous studies have been limited to resting-state or single-task neural activity without knowledge of the underlying structural architecture that supports this activity. Here, we developed a novel approach for relating both functional activity and underlying structural architecture across cognitive states. We showed that the relative numbers of connections mediating interactions within versus between task-related networks differentially support strong state-dependent correlations.

This study built upon the work presented in Chapters 4 and 5 to assess how the inclusion of region-specific information impacted the analysis of structural and functional connectivity. In addition to the structural measure of total tract number and the functional measures of resting-, attention-, and memory-state FC, we identified regions in our atlas belonging to putative task-positive (TP) and task-negative (TN) networks identified in [127]. Inclusion of this information revealed that brain states differ in the degree of interaction within and between TP and TN networks, as defined by the relative number of structural connections that support strong TP versus TN correlations.

When probed across subjects, the phase space defined by these interactions shows constrained relationships between cognitive states, suggesting that this space is reflective of underlying structural and functional organizational principles in the brain. By isolating groups of individuals that share similar phase-space relationships, we can distinguish common from rarely-achieved organizational principles, and we can ask whether these principles relate to external measures of performance.

The comparison of behavioral measures across groupings enabled the isolation of one group of outliers whose phase-space relationships and behavioral performance deviated significantly from the remaining subjects. This suggests that rarely-occupied regions of phase space may be indicative of altered structural or functional organization that can impact behavioral performance. A further characterization of these relationships across different behavioral and genetic measures could potentially be used to develop objective diagnostic measures for altered functionality.

The observed relationship between relative and absolute phases suggests that it may additionally be possible to predict relationships between cognitive states using only resting-state data. In combination with the observed relationship to behavioral measures, this would have significant implications for diagnosing conditions in which there is no access to attention and memory data, such as in cases where subjects cannot adequately perform directed tasks.

In summary, the observation of a constrained space of relationships between cognitive states suggests that the bounds of this space are reflective of common organizational

principles in the brain. The methods developed here for assessing this space provide an objective approach for relating neural activity across cognitive states both within and across individuals. Further application of these methods to the analysis of more extensive behavioral and genetic traits may help identify signatures of altered structural or functional brain organization.

Chapter 7

Conclusion

"The brain is the organ of destiny. It holds within its humming mechanism secrets that will determine the future of the human race."

–Wilder Penfield, from *The Second Career*, 1963

7.1 Overview

In concluding, we can return to the questions that motivated this study: what is the brain doing, how is it doing it, and why? Given that the brain supports functionality ranging from the blink of an eye to the debate of philosophical ideas, and given that it does so in the presence of ongoing structural changes that occur during development, learning, and aging, it is remarkable that such robust and coherent behavior is collectively produced by billions of individual neurons. In an effort to understand this incredible system, we have combined experimental, computational, and theoretical techniques to construct

mappings between the structural organization of neural systems and their functional capabilities. We explore these mappings across different spatial and temporal scales, from the small-scale interactions within networks of neurons to the large-scale interactions between whole brain regions. This integrative approach is advantageous for understanding how the multiscale design of neural systems can support varied functionality and adapt to an ever-changing environment.

Given the wide range of architectural and functional properties exhibited by neural systems, however, it remains a significant challenge to identify the structural features that constrain functional performance across spatial and temporal scales. This is a challenge not only in studies of the brain, but also in a variety of complex interconnected systems in which a balance between competing functions must be achieved within the constraints of a single architecture. Probing neural structure and function, specifically within the human brain, proves to be particularly difficult given the inherent complexity of neural systems and the experimental and computational limitations in probing this complexity. It is therefore crucial to develop multimodal approaches for identifying relevant structural and dynamical features that constrain neural system behavior across different levels of resolution.

7.2 Synopsis

This work developed novel approaches for relating structural system architecture and functional activity that can be broadly applied to a range of artificial and biological

systems. Traditional studies of system design often focus on understanding, and when applicable, optimizing the behavior of specific structural motifs for the performance of single functions. Such approaches have been very useful for identifying the "what" of neural processing, but they have provided limited insight into the "how" and "why" of neural circuit design. By comparing the balance of different functionalities supported by a range of different structural designs, we have identified functional advantages and disadvantages of different architectures that provide insight into underlying design principles that govern the complex behavior of neural systems.

7.2.1 Forward Modeling of Small-Scale Artificial Networks.

Our studies of small-scale neural networks identified tradeoffs in performance that arise from variations in network topology. We explored the ability of parallel and layered network architectures to produce adaptable representations of external information during sequential learning and memory tasks. We found that layered networks produce coarse representations of information that are less susceptible to interference from new information. In comparison, parallel networks produce highly accurate representations of information that can be more easily disrupted. We linked these differences in performance to network topology by characterizing the underlying error landscapes that govern the search for solutions. The observed differences in local landscape properties suggest that further landscape characterization, via methods such as nudged elastic band theory, could provide insight into the ability of a network to dynamically adapt its representation of

external information.

In extensions to this work, we explored the degree to which our results were limited by the properties of small networks. Our preliminary findings suggest that the reduced subspace of solutions produced by one- and two-layer networks may be related through linear approximations of nonlinear sigmoidal transfer functions. The construction of a recursion relation, from which the behavior of an ℓ-layer network could be predicted from the behavior of an $(\ell - 1)$-layer network, would significantly impact our understanding of the capabilities and limitations of larger, composite networks.

We further explored the extension of these computational models to more complex training functions. The extension from one-dimensional to two-dimensional training functions significantly increases the complexity of the information that can be presented to the network and allows for the comparison of different methods of information encoding. In exploring how performance is shaped by network sensitivity to different types of information, we found that sensitivity to spatial information improves the speed with which networks learn new information. When implemented nonuniformly throughout modular and nonmodular network structures, we additionally found that performance was sensitive to the arrangement of modules relative to the arrangement of spatial versus nonspatial node sensitivities. These results suggest several directions for future research. Further characterization of the specific spatial features to which networks are sensitive will provide a better understanding of the external environmental factors that constrain network performance. Exploring internal versus external structure-function relationships

could be particularly informative in the context of evolution and development, during which the network structure dynamically adapts during learning to take advantage of specific environmental features.

7.2.2 Inverse Analysis of Large-Scale Human Brain Networks

We complemented our studies of small scale computational networks with large-scale analysis of human brain networks, which provided direct evidence of constrained relationships between structural architecture and functional performance in biological systems. We related structural connectivity, defined based on the physical properties of white matter tracts as inferred from DTI measurements of water diffusion, to functional connectivity, defined by the strength of correlated fluctuations in energy consumption as inferred from fMRI measurements of BOLD activity. By comparing these relationships across 84 subjects and across three cognitive states, we identified robust relationships between the task-dependent strength of functional correlation between brain regions and the length, number, and spatial location of white matter tracts linking these regions. We further showed that functional correlation strength could in turn be used to infer the structural properties of the underlying anatomical architecture. The observed relationship between long intra-hemispheric connectivity and task-dependent functional activity suggests that the properties of these connections may be predictive of task-performance.

In extensions to this study, we confirmed that the observed relationships between structural and functional connectivity are robust to variations in our analysis techniques.

We discussed the sensitivity of our results to the inclusion of inconsistent connectivity, which altered the apparent length of connections inferred from functional correlations. We suggested methods for correcting for this sensitivity both in a representative case and a subject-specific case. Beyond isolating robust relationships between structural and functional connectivity in the human brain, this work highlighted several analysis tools that can be applied more broadly to network systems. The construction of representative brain networks from consistently-connected regions pairs within subject-specific brain networks allows for the comparison of properties that are representative of the entire group of subjects to those properties that are specific to individual subjects. The application of structural and functional thresholds further enables the isolation and comparison of different groups of connections.

Our final study, which extended the observed structure-function relationships in human brain networks, assessed constrained relationships between cognitive states. By including information about the anatomical regions involved in structurally-mediated functional interactions, we showed that interactions within and between putative task-related networks differentially support and distinguish between strong resting-, attention-, and memory-state correlations. By mapping structure-function relationships onto a phase-like space defined by these interactions, we identified phase variables that characterize different cognitive states. Remarkably, comparison across individuals reveals constrained relationships between phase variables, thereby enabling us to group individuals based on these relationships. We related these groupings to behavioral performance measures, and

we were able to identify one set of subjects whose abnormal phase-space relationships was indicative of abnormal behavioral performance. In future work, we will probe more extensively the relationship between phase-space variables and behavioral and genetic measures. Identifying mappings between internal brain-state properties and external behavioral measures may help predict task-driven brain activity or diagnose individuals with altered brain functionality.

7.2.3 Conclusions

Together, the combined study of small-scale artificial networks and large-scale biological networks enables the identification of structural properties that constrain and facilitate functionality across a wide range of spatial and temporal scales. Forward modeling approaches, such as those used to study feedforward neural networks, enable us to systematically control structural features in order to isolate functional capabilities and limitations of computational networks. By probing these relationships across a range of small network topologies, we gain insight into the performance of larger composite structures in which statistical studies of performance would be intractable. Conversely, inverse approaches, such at the MR imaging used to infer structural and functional connectivity in the human brain, enable us to identify the organizational principles that govern the functionality of large-scale biological systems. By assessing relationships between anatomical architecture and functional interactions in large-scale neural systems, we gain insight into the biological principles that constrain small-scale neural systems.

7.3 Future Directions

Beyond the ability to connect relevant neural phenomena across scales, we are faced with several fundamental challenges to understanding the structure and function of neural systems. In particular, one striking feature of neural systems is their ability to simultaneously process spatial and temporal features of sensory stimuli, such as the spectral and temporal components of speech or the position and velocity of moving objects in dynamic visual scenes. Although neuronal processing relies on both dimensions of sensory information, relatively little is known about the interaction between space and time within neural circuits. Furthermore, it is well understood that neural systems must be both robust and adaptive to growth, degradation, and noise. However, the mechanisms that enable neural systems to maintain functionality despite ongoing structural and dynamical changes are not well understood.

While the approaches discussed here can be appropriately modified to address these questions, more sophisticated modeling frameworks may help bridge the gap between small-scale artificial networks and large-scale biological networks. Models such as liquid-state machines and state-dependent networks rely on time-dependent neuronal properties, such as short term synaptic plasticity, to provide transient memory traces that interact with external stimuli to enable high-dimensional spatiotemporal representations [129]. Within these frameworks, structural properties of anatomical fiber pathways (as inferred from DTI) could be used to inform artificial network connectivity, which could in turn be systematically altered to predict large-scale functional activity (as measured via

fMRI). Together, this would enable the study of interactions between system architecture, adaptability, and robustness to noise.

7.3.1 Structural Regularity, Growth, and Degradation

Structural regularities, such as hierarchical [13] and modular [12] motifs, have been identified across a range of spatial scales and have been linked to dynamical properties such as the generation of nested rhythms [130] and a reduction in synchronization time [131]. Dynamic changes in connectivity, such as the growth observed during neurogenesis and dendritic spine formation [66] and the degradation associated with aging and lesion formation [69], can similarly have profound affects on functionality. However, the role of static and dynamic connectivity in facilitating spatiotemporal processing is not well understood.

Future work could explore the extent to which network topology alters the interaction between external stimulus and internal network dynamics, thereby enabling flexible representations of spatiotemporal stimuli. Consideration of the temporal dynamics produced by different static topologies could provide insight into the effects of structural growth and degradation. Statistical mechanical techniques, such as the sloppy model framework used to extract information from systems with many poorly-known parameters [1], could help identify the internal degrees of freedom most important for controlling dynamical features such as oscillation frequency or time to synchronization. Studying the topological localization of these degrees of freedom may reveal specific structural motifs that

maintain higher sensitivity to spatial or temporal features of the external stimuli. This could provide insight into the structural underpinnings of functional specialization, such as the specialization observed in orientational selectivity domains in the visual cortex.

Dynamic changes in structural (e.g. adding or removing connections) and temporal (e.g. locally degrading neuronal output) network features could further impact the representational capabilities of neuronal systems. If different topological motifs reveal distinct response characteristics, then localized perturbations may have drastically different effects depending on the location or timing of perturbations. In particular, different network topologies may differ in how they dynamically compensate for local degradation or incorporate new growth. Information theoretic approaches could be used, for example, to determine how the location and timing of local perturbations alter nonlocal information exchange. Identifying relationships between local fragilities and global robustness is crucial for understanding development and evolution in neuronal systems.

7.3.2 Coevolution of Plasticity and Topology

Synaptic plasticity mechanisms, such as synaptic scaling, spike timing dependent synaptic plasticity, and synaptic redistribution, facilitate adaptive responses to external stimuli [63]. These mechanisms are incorporated into network models via learning rules that modify synaptic strengths as a function of dynamic activity within the network. Recent theoretical studies have begun to investigate learning rules that may facilitate the creation of spatiotemporal representations of external stimuli (e.g. [132]). However, the

relative impact of different learning rules is not understood, particularly when embedded in networks with complex and dynamic connectivity. Given that plasticity mechanisms evolved alongside the structural regularities observed in the brain, plasticity is likely dependent on the structure of the network in which it is embedded.

Future work could investigate the extent to which synaptic plasticity rules take advantage of specific structural motifs in order to appropriately shape the interaction between internal and external dynamics. By considering learning rules that are sensitive to both the underlying structure and dynamic changes in connectivity, such as activity-dependent rewiring rules [78], future work could explore the interaction between structurally-sensitive plasticity, topology, and dynamics. In particular, plasticity may be differentially sensitive to local versus global structural motifs and activity patterns. Given that the role of long-term plasticity in spatiotemporal processing is not well understood, future work could additionally consider learning rules that vary in temporal sensitivity.

7.3.3 Physical Embedding and Scalability

The study of topological interactions purposefully neglects the constraints imposed by physical embedding. Given that neuronal systems evolved within these constraints, however, physical embedding plays an important role in the cost, efficiency, and scalability of neural architectures [12]. Importantly, ideal topologies may not be realizable when physical constraints are taken into account. Furthermore, even if these topologies are achievable in small networks, scaling in size may provide additional constraints on the

space of possible network motifs and the resulting network dynamics. In addition to illuminating design constraints, the processes of physically embedding and scaling enable more direct comparison of these models with experimental data such as the diffusion tensor imaging (DTI) and functional magnetic resonance imaging (fMRI) of large-scale structural and dynamical features of the brain.

Multi-scale models of emergent dynamic phenomena in neural systems could be used to probe the effects of physical embedding and scaling. Local variations in length and density of connections, for example, could disrupt synchronization patterns and alter transmission of information through the network. Physical constraints will also impact plasticity-dependent growth processes and rewiring schemes. Developing methods for appropriately coarse-graining structural and dynamical features is crucial for understanding behavior across the huge range of spatial and temporal scales observed within neuronal systems.

7.4 Outlook

The combination of forward modeling approaches and inverse data-driven techniques is powerful for studying a wide range of systems in which experimental access to mechanistic interactions is limited. This is particularly true of the human brain, where direct access to neural activity is only possible in severe medical cases. Inverse imaging techniques enable the reconstruction of structural and functional interactions from spatiotemporal images of the brain. Similarly, *in vitro* or *in vivo* measurements of neuronal

activity, largely limited to animal studies, can be used to develop mechanistic models of neuron-level interactions that can in turn be used to predict neural activity over coarser spatial and temporal scales. The interaction between these two approaches, forward modeling and inverse analysis, is crucial for understanding how neurons, as the building blocks of the brain, can support such a vast array of functionalities across scales ranging from individual cells to whole brain behavior.

Beyond understanding the relevant structural mechanisms that support and constrain functionality, the study of structure-function relationships in neural systems has important applications in the fields of bio-engineering and medicine. Artificial neural networks lie at the interface between biological and computational systems and can therefore inform the design of machines that maintain aspects of biological functionality, including robotic systems that must perform tasks in dangerous environments, prosthetic human-machine interfaces, and medical tools for invasively probe living systems with minimal biological interference.

In addition to the technological applications, knowledge of structure-function relationships has important medical applications. As changes in both small-scale and large-scale anatomical properties have been linked to neurological disorders, the ability to predict the functional consequences of structural disruptions is crucial for developing a mechanistic understanding of altered cognitive states. Furthermore, due to an inability to objectively characterize neurological disorders, diagnoses of such disorders are often made via personal interview. A characterization of structural and functional brain orga-

nization is therefore crucial for reliably identifying and distinguishing between different neurological disorders. Importantly, the work highlighted here has the potential to help distinguish the structural or functional signatures of healthy from unhealthy individuals and predict the functional impact of changes in structural organization.

With the rapid development of new experimental techniques for imaging the structure and function of neural systems, this is an exciting time for neuroscience research (especially for rogue physicists entering into the field). Given the high degree of detail achieved by many of these new measurements, it will be important to develop the appropriate quantitative tools for describing and coarse-graining relevant structural and dynamical features of neural systems across varying scales of resolution. The complementary use of theoretical modeling and data-driven analysis provides one approach for developing, modifying, and evaluating relationships between structural and functional interactions in both artificial and biological network systems. Together, such approaches will inform our understanding of the structural mechanisms that constrain functionality and may further enable the prediction of functional consequences arising from structural disruptions.

In the years to come, it will remain a significant challenge to connect the function of the neuron to the function of the brain, let alone to describe the mechanisms underlying conscious behavior. Yet, the mere ability to ask these questions is a testament to the utility of neuroscience research, to the unparalleled capabilities of the brain itself.

Bibliography

[1] K.S. Brown and J.P. Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E*, 68:021904, 2003.

[2] A. M. Hermundstad, K. S. Brown, D. S. Bassett, and J. M. Carlson. Learning, memory, and the role of neural network architecture. *PLoS Comput. Biol.*, 7:e1002063, 2011.

[3] V. B. Mountcastle. The columnar organization of the neocortex. *Brain*, 120:701–722, 1997.

[4] A K Jain, M N Murty, and P J Flynn. Data clustering: a review. *ACM Comput. Surv.*, 31:264–323, 1999.

[5] M Egmont-Petersen, D de Ridder, and H Handels. Image processing with neural networks–a review. *Pattern Recogn.*, 35:2279–2301, 2002.

[6] G Zhang, B E Patuwo, and M Y Hu. Forecasting with artificial neural networks: the state of the art. *Int. J. Forecasting*, 14:35–62, 1998.

[7] L Chittka and J J Niven. Are bigger brains better? *Curr. Biol.*, 19:R99535–R1008, 2009.

[8] C J Honey. Predicting human resting-state functional connectivity from structual connectivity. *Proc. Natl. Acad. Sci. USA*, 106:2035–2040, 2009.

[9] T Kenet, D Bibitchkov, M Tsodyks, A Grinvald, and A Arieli. Spontaneously emerging cortical representations of visual attributes. *Nature*, 425:954–956, 2003.

[10] A R McIntosh, M N Rajah, and N J Lobaugh. Functional connectivity of the medial temporal lobe relates to learning and awareness. *J. Neurosci.*, 23:6520–6528, 2003.

[11] J Scholz, M C Klein, T E J Behrens, and H Johansen-Berg. Training induces changes in white-matter architecture. *Nat. Neurosci.*, 12(11):1370–1371, 2009.

[12] Danielle S. Bassett, Daniel L. Greenfield, Andreas Meyer-Lindenberg, Daniel R. Weinberger, Simon W. Moore, and Edward T. Bullmore. Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Comput. Biol.*, 6:e1000748, 2010.

[13] M Kaiser and C C Hilgetag. Optimal hierarchical modular topologies for producing limited sustained activation of neural networks. *Front. Neuroinformatics*, 4:1–14, 2010.

[14] A T Reid, A Krumnack, E Wanke, and R Kotter. Optimization of cortical hierarchies with continuous scales and ranges. *NeuroImage*, 47:611–617, 2009.

[15] D Ress, G H Glover, J Liu, and B Wandell. Laminar profiles of functional activity in the human brain. *NeuroImage*, 34:74–84, 2007.

[16] C A Atencio and C E Schreiner. Columnar connectivity and laminar processing in cat primary auditory cortex. *PLoS ONE*, 5:e9521, 2007.

[17] H B Bakoglu. *Circuits, Interconnections, and Packaging for VLSI*. Addison-Wesley, Boston, 1990.

[18] A I Galushkin. *Neural Networks Theory*. Springer-Verlag, Berlin, 2007.

[19] K Fukushima. Neocognitron: a hierarchical neural network capable of visual pattern recognition. *Neural Netw.*, 1:119–130, 1988.

[20] A J Robinson. An application of recurrent nets to phone probability estimation. *IEEE Trans. Neural Netw.*, 5:298–305, 1994.

[21] W. S. McCulloch and W. H. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5:115–133, 1943.

[22] D. E. Rumelhart and J. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, 1986.

[23] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psych. Rev.*, 65:386–408, 1958.

[24] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Contr. Signals Syst.*, 2:303–314, 1989.

[25] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Netw.*, 4:251–257, 1991.

[26] P. Dayan and L. F. Abbott. *Theoretical Neuroscience Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.

[27] R Rojas. *Neural Networks - A Systematic Introduction*. Springer-Verlag, Berlin, 1996.

[28] Y. Chauvin and D. E. Rumelhart. *Back Propagation: Theory, Architectures, and Applications*. Erlbaum, 1995.

[29] R Fletcher and C M Reeves. Function minimization by conjugate gradients. *Comput. J.*, 7:149–154, 1964.

[30] E Polak and G Ribiere. Note sur la convergence de methodes de directions conjugees. *Rev. Franc. Inform. Rech. Oper.*, 16:35–43, 1969.

[31] M J D Powell. Convergence properties of algorithms for nonlinear optimization. *SIAM Rev.*, 28:487–500, 1986.

[32] M McCloskey and N J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower, editor, *The Psychology of Learning and Motivation*, volume 24, pages 109–159. Academic Press, New York, 1989.

[33] R Ratcliff. Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychol. Rev.*, 97:285–308, 1990.

[34] N E Sharkey and A J C Sharkey. An analysis of catastrophic interference. *Connect. Sci.*, 7(3):301–329, 1995.

[35] A Robins. Catastrophic forgetting, rehearsal, and pseudorehearsal. *Connect. Sci.*, 7:123–146, 1995.

[36] A Robins and S McCallum. Catastrophic forgetting and the pseudorehearsal solution in hopfield-type networks. *Connect. Sci.*, 10:121–135, 1998.

[37] K S Brown and J P Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E*, 68:021904, 2003.

[38] K S Brown, C C Hill, G A Calero, C R Myers, K H Lee, J P Sethna, and R A Cerione. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys. Biol.*, 1:184–195, 2004.

[39] R Fletcher. *Practical Methods of Optimization*. Wiley-Interscience, New York, 2nd edition, 1987.

[40] P A Mello and N Kuma. *Quantum transport in mesoscopic systems: complexity and statistical fluctuations*. Oxford University Press, New York, 2004.

[41] L M Bettencourt, G J Stephens, M I Ham, and G W Gross. Functional structure of cortical neuronal networks grown in vitro. *Phys. Rev. E*, 75(2):021915, 2007.

[42] S Achard, R Salvador, B Whitcher, J Suckling, and E Bullmore. A resilient, low-

frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.*, 26(1):63–72, 2006.

[43] P Hagmann, L Cammoun, X Gigandet, R Meuli, C J Honey, V J Wedeen, and O Sporns. Mapping the structural core of human cerebral cortex. *PLoS Biol.*, 6(7):e159, 2008.

[44] H Kim. Dissociating the roles of the default-mode, dorsal, and ventral networks in episodic memory retrieval. *NeuroImage*, 50(4):1648–1657, 2010.

[45] Z J Chen, Y He, P Rosa-Neto, J Germann, and A C Evans. Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cereb. Cortex*, 18(10):2374–2381, 2008.

[46] D Meunier, S Achard, A Morcom, and E Bullmore. Age-related changes in modular organization of human brain functional networks. *NeuroImage*, 44(3):715–723, 2009.

[47] D S Bassett, J A Brown, V Deshpande, J M Carlson, and S T Grafton. Conserved and variable architecture of human white matter connectivity. *NeuroImage*, 54(2):1262–1279, 2011.

[48] D J Felleman and D C van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex*, 1(1):1–47, 1991.

[49] D Meunier, R Lambiotte, and E T Bullmore. Modular and hierarchically modular organization of brain networks. *Front. Neurosci.*, 4(200):1–11, 2010.

[50] H-C Fu, Y-P Lee, C-C Chiang, and H-T Pao. Divide-and-conquer learning and modular perceptron networks. *IEEE Trans. Neural Netw.*, 12:250–263, 2001.

[51] O K Ersoy and S W Deng. Parallel, self-organizing, hierarchical neural networks with continuous inputs and outputs. *IEEE Trans. Neural Netw.*, 6(5):1037–1044, 1995.

[52] H Oshima and T Odagaki. Storage capacity and retrieval time of small-world neural networks. *Phys. Rev. E*, 76(3):036114, 2007.

[53] D Dominguez, M González, E Serrano, and F B Rodríguez. Structured information in small-world neural networks. *Phys. Rev. E*, 79(2):021909, 2009.

[54] H Larochelle, Y Bengio, J Louradour, and P Lamblin. Exploring strategies for training deep neural networks. *J. Mach. Learn. Res.*, 10:1–40, 2009.

[55] Y Bengio and Y LeCun. Scaling learning algorithms toward AI. In *Large Scale Kernel Machines*, pages 321–360. MIT Press, Cambridge, 2007.

[56] Y Bengio. Learning deep architectures for AI. *Found. Trends Mach. Learn.*, 2:1–127, 2009.

[57] H E Atallah, M J Frank, and R C O'Reilly. Hippocampus, cortex, and basal

ganglia: Insights from computational models of complementary learning systems. *Neurobiol. Learn. Mem.*, 82:253–267, 2004.

[58] E Bullmore and O Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.*, 10(3):186–198, 2009.

[59] G Tononi, O Sporns, and G M Edelman. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. USA*, 91:5033–5037, 1994.

[60] G G Turrigiano, K R Leslie, N S Desai, L C Rutherford, and S B Nelson. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391(6670):892–896, 1998.

[61] G G Turrigiano and S B Nelson. Homeostatic plasticity in the developing nervous system. *Nat. Rev. Neurosci.*, 5:97–107, 2004.

[62] H Markram and M Tsodyks. Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382(6594):807–810, 1996.

[63] L F Abbott and S B Nelson. Synaptic plasticity: Taming the beast. *Nat. Neurosci.*, 3:1178–1183, 2000.

[64] R Bogacz, E-J Wagenmakers, B U Forstmann, and S Nieuwenhuis. The neural basis of the speed-accuracy tradeoff. *Trends Neurosci.*, 33:10–16, 2009.

[65] V van Veen, M K Krug, and C S Carter. The neural and computational basis of controlled speed-accuracy tradeoff during task performance. *J. Cog. Neurosci.*, 20:1952–1965, 2008.

[66] T Xu, X Yu, A J Perlik, W F Tobin, J A Zweig, K Tennant, T Jones, and Y Zuo. Rapid formation and selective stabilization of synapses for enduring motor memories. *Nature*, 462:915–919, 2009.

[67] J Alstott, M Breakspear, P Hagmann, L Cammoun, and O Sporns. Modeling the impact of lesions in the human brain. *PLoS Comput. Biol.*, 5:e1000408, 2009.

[68] C J Honey and O Sporns. Dynamical consequences of lesions in cortical networks. *Hum. Brain Mapp.*, 29:802–809, 2008.

[69] R P Allred, D Adkins, M T Woodlee, L C Husbands, M A Maldonado, J R Kane, T Schallert, and T A Jones. The vermicelli handling test: A simple quantitative measure of dexterous forepaw function in rats. *J. Neurosci. Meth.*, 170:229–244, 2008.

[70] F Cucker and S Smale. On the mathematical foundations of learning. *Bull. Amer. Math. Soc.*, 39:1–49, 2001.

[71] O Bousquet, S Boucheron, and G Lugosi. Introduction to statistical learning theory. In *Advanced Lectures on Machine Learning*, volume 3176, pages 169–207. Springer-Verlag, Berlin, 2004.

[72] G E Hinton, S Osindero, and Y-W Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18:1527–1554, 2006.

[73] E Marder, L F Abbott, G G Turrigiano, Z Liu, and J Golowasch. Memory from the dynamics of intrinsic membrane currents. *Proc. Natl. Acad. Sci. USA*, 93:13481–13486, 1996.

[74] C Gaiteri and J E Rubin. The interaction of intrinsic brain dynamics and network topology in determining network burst synchrony. *Front. Comput. Neurosci.*, 5:1–14, 2011.

[75] P Bush and T Sejnowski. Inhibition synchronizes sparsely connected cortical neurons within and between columns in realistic network models. *J. Comput. Neurosci.*, 3:91–110, 1996.

[76] P R Roelfsema, A K Engel, P Konig, and W Singer. Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature*, 385:157–161, 1997.

[77] T P Vogels and L F Abbott. Signal propagation and logic gating in networks of integrate-and-fire neurons. *J. Neurosci.*, 25:10786–10795, 2005.

[78] M Rubinov, O Sporns, C van Leeuwen, and M Breakspear. Symbiotic relationship between brain structure and dynamics. *BMC Neurosci.*, 10:1–18, 2009.

[79] M V Sanchez-Vives and D A McCormick. Cellular and network mechanisms of rythmic recurrent activity in the neocortex. *Nat. Neurosci.*, 3:1027–1034, 2000.

[80] A B L Tort, R W Komorowski, J R Manns, N J Kopell, and H Eichenbaum. Theta-gamma coupling increases during the learning of item-content associations. *Proc. Natl. Acad. Sci. USA*, 106:20942–20947, 2009.

[81] A. M. Hermundstad, K. S. Brown, D. S. Bassett, and J. M. Carlson. Structural drivers of function in information processing networks. *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, pages 837–841, 2011.

[82] Kštter R Sporns O, Tononi G. The human connectome: A structural description of the human brain. *PLoS Comput. Biol.*, 1:e42, 2005.

[83] J. H. Lee, R. Durand, V. Gradinaru, F. Zhang, I. Goshen, D. S. Kim, L. E. Fenno, C. Ramakrishnan, and K. Deisseroth. Global and local fmri signals driven by neurons defined optogenetically by type and wiring. *Nature*, 465:788–792, 2010.

[84] Simon Clavagnier, Arnaud Falchier, and Henry Kennedy. Long-distance feedback projections to area v1: Implications for multisensory integration, spatial awareness, and visual consciousness. *Cog. Aff. Behav. Neurosci.*, 4:117–126, 2004.

[85] Dongyang Zhang, Abraham Z. Snyder, Joshua S. Shimony, Michael D. Fox, and Marcus E. Raichle. Noninvasive functional and structural connectivity mapping of the human thalamocortical system. *Cereb. Cortex*, 20:1187–1194, 2010.

[86] Lucina Q. Uddin, Kaustubh Supekar, Hitha Amin, Elena Rykhlevskaia, Daniel A. Nguyen, Michael D. Greicius, and Vinod Menon. Dissociable connectivity within human angular gyrus and intraparietal sulcus: Evidence from functional and structural connectivity. *Cereb. Cortex*, 20:2636–2646, 2010.

[87] M. P. van den Heuvel, R. C. W. Mandl, R. S. Kahn, and H. E. Hulsoff. Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human Brain Mapping*, 30:3127–3141, 2009.

[88] Z. P. Saygin, D. E. Osher, K. Koldewyn, G. REnolds, J. D. E. Gabrieli, and R. R. Saxe. Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nat. Neurosci.*, 15:321–327, 2012.

[89] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.*, 106(6):2035–2040, 2009.

[90] K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Hum. Brain Mapp.*, 2:56–78, 1994.

[91] R. Westerhausen, F. Kreuder, W. Woerner, R. J. Huster, C. M. Smit, E. Schweiger, and W. Wittling. Interhemispheric transfer time and structural properties of the corpus callosum. *Neurosci. Lett.*, 409:140–145, 2006.

[92] Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm

using a macroscopic anatomical parcellation of the mni mri single-subject brain. *NeuroImage*, 15:273–289, 2002.

[93] M. E. Raichle and M. A. Minturn. Brain work and brain imaging. *Annu. Rev. Neurosci.*, 29:449–476, 2006.

[94] S. N. Vaishnavi, A. G. Vlassenko, M. M. Rundle, A. Z. Snyder, M. A. Minturn, and M. E. Raichle. Regional aerobic glycolosis in the human brain. *Proc. Natl. Acad. Sci. U. S. A.*, 107:17757–62, 2010.

[95] V. J. Wedeen, D. L. Rosene, R. Wang, G. Dai, F. Mortazavi, P. Hagmann, and J. H. Kaas ansW Y I. Tseng. The geometric structure of the brain fiber pathways. *Science*, 335:1628–1634, 2012.

[96] A. Meyer-Lindenberg D. S. Bassett, S. Achard, T. Duke, and E. Bullmore. Adaptive reconfiguration of fractal small-world human brain functional networks. *Proc. Natl. Acad. Sci. U.S.A.*, 103:19518–19523, 2006.

[97] M. Valencia, J. Martinerie, S. Dupont, and M. Chavez. Dynamic small-world behavior in functional brain networks unveiled by an event-related networks approach. *Phys. Rev. E.*, 77:050905, 2008.

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

[99] D. Attwell and S. B. Laughlin. An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab*, 21:1133–1145, 2001.

[100] D S Bassett and E Bullmore. *Brain anatomy and small-world networks*. Betham, 2010.

[101] M. Kaiser and C. C. Hilgetag. Non-optimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput Biol*, 2:e95, 2006.

[102] Beth L Chen, David H Hall, and Dmitri B Chklovskii. Wiring optimization can relate neuronal structure and function. *Proc Natl Acad Sci U S A*, 103(12):4723–4728, 2006.

[103] Mario Senden, Rainer Goebel, and Gustavo Deco. Structural connectivity allows for multi-threading during rest: The structure of the cortex leads to efficient alternation between resting-state exploratory behavior and default mode processing. *NeuroImage*, 60:2274–2284, 2012.

[104] H. J. Heinze, G. R. Mangun, W. Burchert, H. Hinrichs, M. Scholz, T. F. Munte, A. Gos, M. Scherg, S. Johannes, H. Hundeschagen, M. S. Gazzaniga, and S. A. Hillyard. Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature*, 372:543–546, 1994.

[105] Shimin Fu, Pamela M. Greenwood, and Raja Parasurama. Brain mechanisms of

involuntary visuospatial attention: An event-related potential study. *Hum. Brain Mapp.*, 25:378–390, 2005.

[106] M. Corbetta, J. M. Kincade, J. M. Ollinger, M. P. McAvoy, and G. L. Shulman. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.*, 3:292–297, 2000.

[107] R. S. Rosenbaum, M. Ziegler, G. Winocur, C. L. Grady, and M. Moscovitch. Òi have often walked down this street beforeÓ: Fmri studies on the hippocampus and other structures during mental navigation of an old environment. *Hippocampus*, 14:826–835, 2004.

[108] H. Burianova and C. L. Grady. Common and unique activations in autobiographical, episodic, and semantic retrieval. *J. Cog. Neurosci.*, 19:1520–1534, 2007.

[109] A. Martin and L. L. Chao. Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.*, 11:194–201, 2001.

[110] D. A. Gusnard, E. Akbudak, G. L. Shulman, and M. E. Raichle. Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.*, 98:4259–4264, 2001.

[111] D. H. Weissman J. Prado, J. Carp. Variations of response time in a selective attention task are linked to variations of functional connectivity in the attentional network. *NeuroImage*, 54:541–549, 2011.

[112] J. M. Tyszka, D. P. Kennedy, R. Adolphs, and L. K. Paul. Intact bilateral resting-state networks in the absence of a corpus callosum. *J. Neurosci.*, 31:15154–15162, 2011.

[113] L. Q. Udin, K. S. Supekar, S. Ryali, and V. Menon. Dynamic reconfiguration of functional connectivity across core neurocognitive brain networks with development. *J. Neurosci.*, 31:18578–18589, 2011.

[114] K. Supekar, M. Musen, and V. Menon. Development of large-scale functional brain networks in children. *PloS Biol.*, 7:e1000157, 2009.

[115] L. Wang, Y. Li, P. Metzak, Y. He, and T. S. Woodward. Age-related changes in topological patters of large-scale brain functional networks during memory encoding and recognition. *NeuroImage*, 50:862–872, 2010.

[116] N. Chen, Y. Chou, A. W. Song, and D. J. Madden. Measurement of spontaneous signal fluctuations in fmri: adult age differences in intrinsic functional connectivity. *Brain Struct. Funct.*, 213:571–585, 2009.

[117] A. B. Waites, R. S. Biellmann, M. S. Saling, D. F. Abbott, and G. D. Jackson. Functional connectivity networks are disrupted in left temporal lobe epilepsy. *Ann. Neurol.*, 59:335–343, 2006.

[118] C. S. Monk, S. J. Peltier, J. L. Wiggins, S.-J. Weng, M. Carrasco, S. Risi, and C. Lord. Abnormalities of intrinsic functional connectivity in autism spectrum disorders. *NeuroImage*, 47:764–772, 2009.

[119] Z. Zhang, W. Liao, H. Chen, D. Mantini, J.-R. Ding, Q. Xu, Z. Wang, C. Yuan, G. Chen, and G. Lu. Altered functional-structural coupling of large-scale brain networks in idiopathic generalized epilepsy. *Brain*, 134:2921–2928, 2011.

[120] D. J. Sharp, C. F. Beckmann, R. Greenwood, K. M. Kinnunen, V. Bonnelle, X. De Boissezon, J. H. Powell, S. J. Counsell, M. C. Patel, and R. Leech. Default mode network functional and structural connectivity after traumatic brain injury. *Brain*, 134:2233–2247, 2011.

[121] Qi Zhu, Jiedong Zhang, Yu L. L. Luo, Daniel D. Dilks, and Jia Liu. Resting-state neural activity across face-selective cortical regions is behaviorally relevant. *J. Neurosci.*, 31:10323Ð10330, 2011.

[122] E. M. Aminoff, D. Clewett, S. Freeman, A. Frithsen, C. Tipper, A. Johnson, S. T. Grafton, and M. B. Miller. Individual differences in shifting decision criterion: A recognition memory study. *Mem. Cogn.*, 40:127–134, 2012.

[123] M E Raichle, A M MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman. A default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.*, 98:676–682, 2001.

[124] M. D. Greicius, B. Krasnow, A. L. Weiss, and V. Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U. S. A.*, 100:253–258, 2003.

[125] M. D. Fox and M. E. Raichle. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.*, 8:700–711, 2007.

[126] M. D. Greicius. Resting-state functional connectivity in neuropsychiatric disorders. *Curr. Opin. Neurol.*, 24:424–430, 2007.

[127] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.*, 102(27):9673–9678, 2005.

[128] M. Mennes, C. Kelly, X. N. Zuo, A. Di Martino, B. B. Biswal, F. X. Castellanos, and M. P. Milham. Inter-individual differences in resting-state functional connectivity predict task-induced bold activity. *NeuroImage*, 50:1690–1701, 2011.

[129] DV Buonomano and W Maass. State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.*, 10:113–125, 2009.

[130] T. Binzegger, R.J. Douglas, and K.A. Martin. Topology and dynamics of the canonical circuit of cat v1. *Neural Netw.*, 22:1071–1078, 2009.

[131] R. Vicente, L. L. Gollo, C.R. Mirasso, I. Fischer, and G. Pipa. Dynamical relaying can yield zero time lag neuronal synchrony despite long conduction delays. *Proc. Natl. Acad. Sci. U. S. A.*, 105:17157–17162, 2008.

[132] D V Buonomano. A learning rule for the emegence of stable dynamics and timing in recurrent networks. *J. Neurophysiol.*, 94:2275–2283, 2005.