Quantifying dynamical properties of brain activity using complex systems analysis

A thesis submitted by

Annie Gilmore Bryant

to fulfil the requirements of the degree of

Doctor of Philosophy

The University of Sydney
School of Physics

June, 2025

Acknowledgements

First and foremost, I wish to express my deepest gratitude to my PhD supervisor, Ben, who took a chance on a random email from a prospective student halfway across the world with next-to-no background in physics. For many reasons, this PhD opportunity seemed intimidating and near impossible at first, but your sincere enthusiasm, encouragement, and care for your students turned this into an extraordinarily rewarding and fulfilling experience. Words cannot express how thankful I am for your commitment to this group and to my success, both professionally and personally. These last 3.5 years have been the most transformative of my life, and I appreciate your compassion and patience while I navigated some difficult challenges. I feel extraordinarily lucky to have you as my mentor, and I aspire to emulate your passion, kindness, and care for both your work and those with whom you work. Thank you for always encouraging me to cultivate each newfound research interest throughout my PhD and for always being available to talk through ideas.

I would also like to extend my heartfelt thanks to Mac, my thesis co-supervisor, for 'adopting' me into your lab—from the moment I arrived in Sydney, you immediately made me feel welcome and at home with your group. Thank you for always being enthusiastic to discuss all things research, career growth, and GIFs (not necessarily in that order). Your support, guidance, and kindness extended toward me and everyone else who has the privilege of learning from you has been invaluable.

Thank you to my other co-supervisor, Joe, for your patience and willingness to help me navigate the world of information theory for the first time. I am deeply appreciative of your support and encouragement throughout my thesis journey.

To the American Australian Association, I am exceedingly grateful for the Graduate Education Fund scholarship that directly supported my move to Australia and enabled my research to flourish. I would also like to sincerely thank the Australian Government Department of Education Research Training Program (RTP) and the University of Sydney Physics Foundation for supporting my stipend and tuition fees as an international student, without which this PhD would not have been possible. I wish to also acknowledge the Postgraduate Research Support Scheme and the Paulette Isabel Jones Career Development Award from The University of Sydney for supporting international conference travel and other professional development opportunities during my candidature and examination period.

To the lab postdocs Oliver, Eli, Brandon, Chet, and Jayson: Thank you for being such outstanding mentors and friends throughout my PhD. A rising tide lifts all boats, and I so admire your dedication to uplifting the students across our groups.

A special thank you to my fellow PhD student crew (past and present!), including Gabe, Josh, Brendan, Aria, Jungwoo, Teresa, Kevin, Dan, Trent, Rishi, Kieran, Ruby, and Lachie. The best moments of the last 3.5 years have come from sharing in this experience together, especially outside of the 'lab', from trivia nights to weekend hikes, to birthday celebrations and conference shenanigans, and everything in between.

I cherish these lifelong friendships forged through this PhD program, and I look forward to many more adventures together in the years to come.

To the wonderful friends who came into my life outside of academia in Sydney, Sedem, Kaitlyn, Jenny, Michelle, Nat, Kate, Ellen, Ruairi, Emily, Paul, Fionnuala, Sam, Liam, and Cliodhna: Thank you for being a truly fantastic bunch of humans who have so enriched my life here, from brunches to ski trips to beach days, and a boogie or two along the way. I am also really grateful for your understanding and kindness as I've navigated the ups and downs of my PhD journey.

To my dear friends back home (and beyond), Morgan, Halle, Nicole, Francis, Valentina, Benno, and Brandon: Thank you all from the bottom of my heart for keeping in touch after I moved to Australia. The distance to home shrinks with each video call, voice note, and Insta reel we've exchanged over the years. While I so love the life I have cultivated here in Sydney, I miss you all very much, and I really appreciate feeling welcomed right back with open arms every time I visit home. Thank you especially to Morgan, Halle, and Nicole for making the trips to Australia to visit me during my PhD; it was so special to share my new life here with each of you, and I look forward to many more adventures together in the future.

To Ashna, who also came into my life in Sydney despite being a fellow Jersey gal: Sydney isn't quite the same since you moved back to the US, but I am so grateful for every meal we cooked together, wine cork we removed with a straightener (girls in STEM, right Sedem?), coastal road trip we spontaneously took, and Bad Bunny song we discreetly requested on the DF. Thank you so much for gracing my life with your friendship, warmth, and care, which got me through some of the most challenging moments of my PhD. I am so excited to see you stateside again soon, and in the meantime, thank goodness for voice notes!

To Christopher, thank you for being a sincere source of care and support through the final stretch of this thesis (and far beyond, too). Your calm presence and kind encouragement over the years has made more of a difference than I can put into words. I am very grateful for the friendship we have grown into, and look forward to our continued work together (and the fun conference adventures that await us in the future!).

To Bella and Giulia, thank you both for being such radiant beams of light in my life. I had no idea when moving to Sydney that I would have the great fortune of working (and becoming such great friends) with two of the most brilliant, hilarious, and kind-hearted people I have ever met. Thank you both so much for always being there for me, especially during the last stretch of my thesis, during which your encouragement, support, and reformer Pilates sessions kept me going. Despite all growing up in totally different parts of the world, I feel so loved, understood, and supported by you two.

To Angelica, mi hermana del otro lado del mundo: Words cannot express how deeply I value our friendship, from the moment you first arrived in Sydney to join Gabe (and I made myself right at home over-sharing, tee hee). Thank you for always making time and space for a couch snuggle with your kitties, a good yap, and a girl dinner with me, no matter how busy your schedule at the hospital may get. I am so grateful that our life

journeys coincided in Sydney, and I look forward to many years to come of beach days, coffee walks, and picnics together.

To Tash, thank you for being the black cat to my Golden retriever, the tee to my hee, and the best friend I could have ever asked for. I had no idea when Mac set us up on a 'lab playdate' over coffee that we would become such close friends, and boy am I glad we still went for that coffee even though it was bucketing out! From our hot girl walks to hiking adventures to concerts where I totally also knew the band playing, your friendship has been a constant source of joy and fulfillment during my PhD and beyond. While I miss you dearly since your move, I am so grateful we stay in touch just as much as ever and that we have a conference rotation that will always give us an excuse to plan another world tour together. Thank you for always being there for me and offering a supporting shoulder from the tiniest inconvenience to the biggest challenges.

Finally, I want to extend gratitude from the bottom of my heart to my family, whose unwavering support, love, and positivity buoyed me through the ups and downs of my candidature. I feel incredibly lucky to have such a wonderful family who has always encouraged me to pursue my dreams, no matter how far away they may take me. To my Mom and Dad: Thank you for always answering my video calls with a smile despite the late hour given timezone differences; for remaining my biggest lifelong cheerleaders; for setting incredible examples for how to lead a life marked by kindness, tenacity, and inspiration. To my Aunt Marlene and Uncle Gerry: Thank you for a lifetime of love, support, and encouragement; I am endlessly inspired by your compassion and brilliance that have shaped my life, and the lives of so many others. To my Nana, Nana Anna, Pop-Pop, and Papa Kenny: Thank you for joining me on this journey in spirit and for cultivating a legacy of love in our family.

Annie Gilmore Bryant
Ph.D. Student

Contents

Abstract

The human brain exhibits a complex and multiscale dynamical structure, from microscopic cellular crosstalk to macroscopic networks. While complex systems science offers a plethora of analytical tools, brain dynamics are typically explored with only a few hand-picked statistics. This thesis quantifies the brain's intricate dynamical structure by integrating methods from complex systems analysis into a systematic and interpretable framework for comparing across measures sensitive to different aspects of neural activity. Chapter 2 introduces time-series feature analysis through the lens of information theory, bridging distinct measures with common notation and terminology. Chapter 3 introduces and applies the Python library, *pyspi*, to time-series classification problems (including neuroimaging datasets) to uncover the most salient features for a given task. Building on this foundation, Chapters 4–7 apply these highly comparative methods to compelling questions in modern neuroscience. Chapter 4 comprehensively evaluates measures of inter-areal coupling to characterize its potential role in conscious visual perception. Chapter 5 expands the scope to integrate local activity and pairwise interactions in studying neuropsychiatric disorders, supporting the continued use of linear measures while underscoring the importance of multiscale approaches. Chapter 6 probes functional, structural, and molecular correlates of homotopic connectivity, a robust property of inter-hemispheric network architecture. Finally, Chapter 7 extends this framework to systematically compare algorithms that capture overlapping communities in the structural connectome to test new biological hypotheses. Collectively, this thesis presents a highly comparative framework for capturing the complex and multiscale dynamical structure of the human brain. This work points to exciting future directions in fields from lifespan development to personalized medicine in an era of expanding openly available data.

Statement of Originality

I hereby declare that the work presented in this thesis titled '*Quantifying dynamical properties of brain activity using complex systems analysis*', submitted to The University of Sydney in fulfillment of the requirements for the award of the degree of Doctor of Philosophy, is a bonafide record of the research work carried out under the supervision of Associate Professor Ben Fulcher. This is to certify that the content of this thesis is my own work. This thesis has not been submitted for any other degree or purpose. I certify that the intellectual content of this thesis is the product of my own work, and that all assistance received in preparing this thesis and all sources have been acknowledged.

Annie Gilmore Bryant

Certificate of Originality

As the primary supervisor for the candidature upon which this thesis is based, I hereby certify that this thesis is sufficiently well presented for examination and that it adheres to the prescribed word limit. I confirm that the Author's declaration above is correct and that the authorship attribution statements below are correct.

Associate Professor Ben Fulcher
PhD Thesis Primary Supervisor

Author attribution statement

We, the undersigned, acknowledge the following author attributions:

This thesis represents the work of Annie Gilmore Bryant. The primary supervisor Associate Professor Ben D. Fulcher, co-supervisor Professor James M. Shine, and auxiliary supervisors Associate Professor Joseph T. Lizier and Professor Michael Breakspear provided guidance in the studies and were involved in reviewing the introduction and discussion of this thesis. The work presented in Chapters 2 through 7 of this thesis was collectively performed as part of a collaborative research program, with contributions of all authors summarized below.

Chapter 2 of this thesis is the first half of a manuscript that is now in submission, as **Bryant, A. G.**, Cliff, O. M., Fulcher, B. D., Shine, J. M., & Lizier, J. T. (2025). Unifying concepts in information-theoretic time-series analysis. arXiv preprint: arXiv:2505.13080
Contributions: Annie G. Bryant conceptualized the review, performed the empirical analyses, created visual schematics, and wrote the original draft of the chapter. Joseph T. Lizier contributed to the conceptualization, design, methodology, interpretation of results, and critical review of the chapter; Oliver M. Cliff, James M. Shine, and Ben D. Fulcher contributed to the design and critical review of the chapter. The second half of this work is included as an Appendix item, and is not included in the thesis body.

Chapter 3 of this thesis is comprised of excerpts from work that is published as Cliff, O. M., **Bryant, A. G.**, Lizier, J. T., Tsuchiya, N., & Fulcher, B. D. (2023). Unifying pairwise interactions in complex dynamics. *Nature Computational Science*, *3*(10), 883-893. https://doi.org/10.1038/s43588-023-00519-x.
Contributions: This chapter comprises two components: a theoretical review of pairwise interaction statistics in complex systems (along with their algorithmic implementation), and a series of case studies systematically applying these measures to diverse real-world problems to demonstrate their utility. Oliver Cliff led the conceptualization, design, and first half of analyses, corresponding to the theoretical comparison and algorithmic implementation of all statistics for pairwise interactions. **Annie G. Bryant** led the section on case studies, which serves as the central empirical counterpart to the theoretical review and quantitative framework developed in the first half of this chapter. Specifically, these contributions included identifying and preprocessing neuroimaging datasets, performing statistical analysis in the application of highly comparative pairwise analysis in neuroimaging, visualizing and interpreting results, and writing the components of the draft corresponding to this case study series. Joseph T. Lizier and Naotsugo Tsuchiya contributed to critical review of the manuscript; and Ben D. Fulcher contributed to the conceptualization, design, and critical review of the manuscript. All excerpts in this chapter correspond to work led directly by Annie G. Bryant; the full manuscript is available as an Appendix item.

Chapter 4 of this thesis comprises a manuscript that is now in submission, as **Bryant, A. G.** & Whyte, C. J. (2025). Highly comparative analysis of inter-areal coupling from MEG data to quantify neural correlates of visual experience. bioRxiv preprint: 10.1101/2025.04.06.646695
Contributions: Annie G. Bryant conceptualized the study design, preprocessed and analyzed the neuroimaging data, performed classification analyses, visualized results, and wrote the drafts of the chap-

ter. Christopher J. Whyte developed and implemented the modeling pipeline, interpreted results, and contributed to the design and critical review of the chapter.

Chapter 5 of this thesis is published as **Bryant, A. G.**, Aquino, K., Parkes, L., Fornito, A., & Fulcher, B. D. (2024). Extracting interpretable signatures of whole-brain dynamics through systematic comparison. *PLoS Computational Biology*, *20*(12), e1012692. https://doi.org/10.1371/journal.pcbi.1012692.
Contributions: Annie G. Bryant consolidated and analyzed the data, visualized results, and wrote the manuscript drafts. Kevin Aquino and Linden Parkes contributed to data analysis and critical review; Alex Fornito and Ben D. Fulcher contributed to the conceptualization, design, and critical review of the manuscript.

Chapter 6 of this thesis is closely related to a manuscript that is now in submission, as **Bryant, A. G.**, Shine, J. M., & Fulcher, B. D. Mapping functional, structural, and molecular correlates of homotopic connectivity. bioRxiv preprint: 10.1101/2025.06.10.657943
Contributions: Annie G. Bryant conceptualized the study design, analyzed the data, visualized results, and wrote the drafts of the manuscript. James M. Shine and Ben D. Fulcher contributed to the design and critical review of the chapter.

Chapter 7 of this thesis is published as **Bryant, A. G.***, Jha, A.*, Agarwal, S., Cahill, P., Lam, B., Oldham, S., Arnatkevičiūtė, A., Fornito, A., & Fulcher, B. D. (2025). Tailoring overlapping community detection methods to the brain's structural connectome. *Network Neuroscience*, 1-58. https://doi.org/10.1162/netn.a.39. *, Co-first authors.
Contributions: Annie G. Bryant analyzed data, visualized results, and wrote drafts of the chapter—with primary contributions to the empirical brain connectivity analysis component of the chapter. Aditi Jha (co-first author) led conceptualization, implementation, and interpretation for the first half of the chapter, and wrote the original draft of the first half of the chapter; Sumeet Agarwal, Patrick Cahill, Brandon Lam, Stuart Oldham, and Aurina Arnatkevičiūtė contributed to data curation and analysis, as well as critical review of the chapter; and Alex Fornito and Ben D. Fulcher contributed to the conceptualization, design, and critical review of the chapter.

Annie Gilmore Bryant
Candidate
27 August 2025

Associate Professor Ben D Fulcher
Primary Supervisor
27 August 2025

Professor James M Shine
Co-Supervisor
27 August 2025

Other publications co-authored during candidature

2025

Bryant, A. G.*, Malla, S.*, Jayakumar, R., Woost, B., Wolf, N., Li, A., Das, S., Van Veluw, S. J., & Bennett, R. E. (2025). Molecular profiling of frontal and occipital subcortical white matter hyperintensities in Alzheimer's disease. *Frontiers in neurology*, 15. *Co-first authors.

Tan, J. B., Orlando, I., Whyte, C. J., **Bryant, A. G.**, Munn, B. R., Baracchini, G., King, M., O'Callaghan, C., Müller, E. J., & Shine, J. M. (2025). Cerebellar and subcortical contributions to working memory manipulation. *Nature communications biology*, 8.

Munn, B. R., Tan, J., Müller, E. J., **Bryant, A. G.**, Orlando, I. F., Grimm, C., Zerbi, V., & Shine, J. M. Neuromodulation controls adaptive multiscale brain dynamics for cognition. *In submission*.

2024

Serrano-Pozo, A., Li, H., Li, Z., Muñoz-Castro, C., Jaisa-Aad, M., Healey, M. A., Welikovitch, L. A., Jayakumar, R., **Bryant, A. G.**, Noori, A., Connors T. R., Hu, M., Zhao, K., Liao, F., Gen, L., Pastika, T., Tamm, J., Abdourahman, A., Kwon, T., Bennett, R. E., Woodbury, M. E., Wachter, A., Talanian, R. V., Biber, K., Karran, E. H., Hyman, B. T., & Das, S. (2024). Astrocyte transcriptomic changes along the spatiotemporal progression of Alzheimer's disease. *Nature neuroscience*, 1-17.

Lee, H., Fu, J. F., Gaudet, K., **Bryant, A. G.**, Price, J. C., Bennett, R. E., Johnson, K. A., Hyman, B. T., Hedden, T., Salat, D. H., Yen, Y-F., & Huang, S. Y. (2024). Aberrant vascular architecture in the hippocampus correlates with tau burden in mild cognitive impairment and Alzheimer's disease. *Journal of cerebral blood flow & metabolism*, 44(5), 787-800.

2023

Bryant, A. G., Li, Z., Jayakumar, R., Serrano-Pozo, A., Woost, B., Hu, M., Woodbury, M. E., Wachter, A., Lin, G., Kwon, T., Talanian, R. V., Biber, K., Karran, E. H., Hyman, B. T., Das, S., & Bennett, R. E. (2023) Endothelial cells are heterogeneous in different brain regions and are dramatically altered in Alzheimer's disease. *Journal of neuroscience*, 43(24), 4541-4557.

Henderson, T., **Bryant, A. G.**, & Fulcher, B. D. (2023). Never a dull moment: Distributional properties as a baseline for time-series classification. *arXiv preprint*.

Links to the above publications or preprints are included in the Appendix, where available.

Artificial Intelligence Statement

During the preparation of the thesis, the Author (Annie Gilmore Bryant) used Grammarly and ChatGPT for the purposes of grammar and text enhancement to improve clarity of the text. The Author confirms that where such modifications were suggested by these tools, the content was manually reviewed for accuracy and appropriateness before inclusion in the thesis.

During the preparation of the thesis, the Author used Github Copilot embedded within Visual Studio Code to assist with Python, R, and Matlab code development in support of analysis and figure generation.

The author takes full responsibility for the submitted thesis and ensures the work is their own and has used generative AI within the parameters of use.

Australian Government Support

This research was supported by an Australian Government Research Training Program (RTP) Scholarship and Stipend.

1
Introduction

Sometimes I'll start a [thesis] and I don't even know where it's going, I just hope I find it along the way.
Michael Scott, The Office, Season 5 Episode 12

This chapter serves as an introduction to this thesis, beginning with an overview of a standard example neuroimaging analysis pipeline and an orientation to the spatial and temporal scales of brain structure and function. I then introduce a more general view of the brain as a complex system, which provides a framework for understanding the brain's dynamical structure and how it can be quantified using a variety of statistical measures. This motivates the development and application of a highly comparative approach that enables the systematic evaluation of different representations of brain structure and function, which is the central theme of this thesis. To methodologically anchor this approach, I briefly review the key neuroimaging modalities used in this thesis and how each can be used to capture brain structure and function at different spatial and temporal scales. In parallel, I also review the importance of methodological parsimony and accessible scientific communication throughout the thesis, which comprise key components of the highly comparative approach. At the conclusion of this chapter, I outline the aims of this thesis and the rationale for its organization, which is primarily structured around a series of manuscripts that each contribute to the overarching theme of quantifying dynamical properties of brain activity using complex systems analysis.

1.1 STANDARD FUNCTIONAL IMAGING ANALYSIS PIPELINE

Consider a researcher interested in studying the brain's structure and function using neuroimaging data. This researcher has access to a neuroimaging dataset, which contains time series data from multiple brain regions across many participants. Perhaps the researcher seeks to investigate how inter-hemispheric synchrony develops throughout the lifespan, or how protein aggregation in neurodegenerative disease disrupts network function, or how a particular neurotrophic factor facilitates multiscale changes to brain connectivity. Regardless of the specific topic, the researcher may seek a way to analyze the neuroimaging data in a way that captures the brain's 'dynamical structure'—that is, the activity of the brain evolving over time—and how it relates to the research question at hand. Typically, one might begin by discretizing the brain into 'regions' according to an atlas based on function, anatomy, and/or cytoarchitecture [1–3]. In this case, for whichever imaging modality is used, the methodological process will probably look something like this (after performing the requisite preprocessing steps to spatially normalize and remove artifacts and noise, such as head motion, physiological noise, and scanner drift):

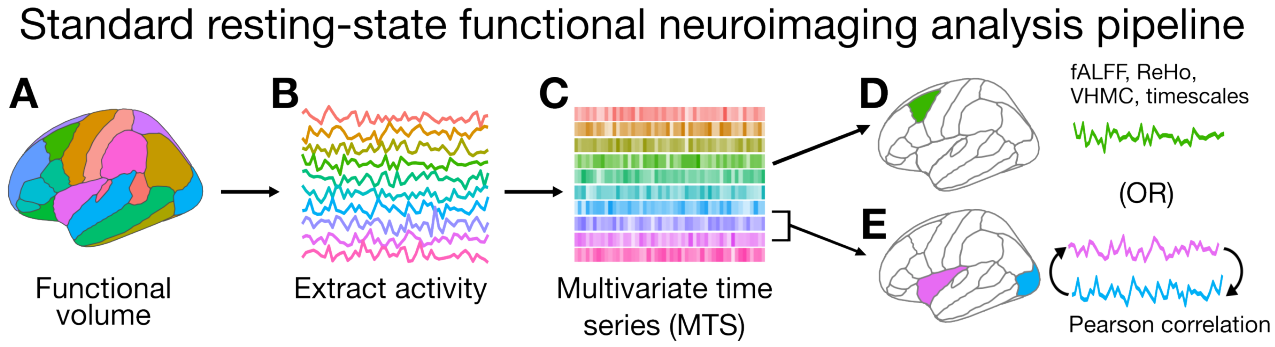


Figure 1.1.1: The general form for traditional approaches to addressing a research question of interest, depicted schematically for illustrative purposes. **A.** After preprocessing, the first step is to delineate regions of interest in the brain, which can be done using a parcellation or segmentation atlas. **B.** For each parcel/region, the time series of activity is extracted, typically by averaging the signal across all voxels or vertices in that region. **C.** The set of regional time series forms a multivariate time series (MTS), which can be represented as a tabular matrix with regions as rows and time points as columns. **D.** Local dynamics may be quantified by focusing on the activity of each region individually, using measures like the fractional amplitude of low-frequency fluctuations (fALFF) [4, 5], regional homogeneity (ReHo) [6], voxel-mirrored homotopic connectivity (VMHC) [7], or intrinsic timescale (typically, the decay constant of the linear autocorrelation function) [8, 9]. **E.** Alternatively, the functional connectivity between regions can be assessed, usually by computing the Pearson correlation coefficient between the time series of each pair of regions [10, 11].

First, the researcher may parcellate the cortex and/or segment subcortical structures into regions of interest according to a given atlas, as depicted in Fig. 1.1.1A. From these regions, the researcher then extracts the voxel- (or vertex-) averaged time series, which yields a univariate time series for each region, as shown in Fig. 1.1.1B. The set of time series for all regions then comprises a multivariate time series (MTS) matrix, as shown in Fig. 1.1.1C, with each row corresponding to a region and each column corresponding to a time point. There are many different ways to represent and analyze the brain's dynamical structure from this resulting MTS, and the choice of which to use is often determined by the specific research question at hand. For example, as shown in Fig. 1.1.1D the researcher might partition the MTS into single-region time

series and measure the spontaneous low-frequency fluctuations in each region using a standard measure like the fractional amplitude of low-frequency fluctuations (fALFF) [4, 5], regional homogeneity (ReHo) [6], voxel-mirrored homotopic connectivity (VMHC) [7], or intrinsic timescale (typically, the decay constant of the linear autocorrelation function) [8, 9]. Alternatively, as shown in Fig. 1.1.1E, the researcher might compute the pairwise functional connectivity (FC) between regions, which is typically done by computing the Pearson correlation coefficient between each pair of time series [10, 11].

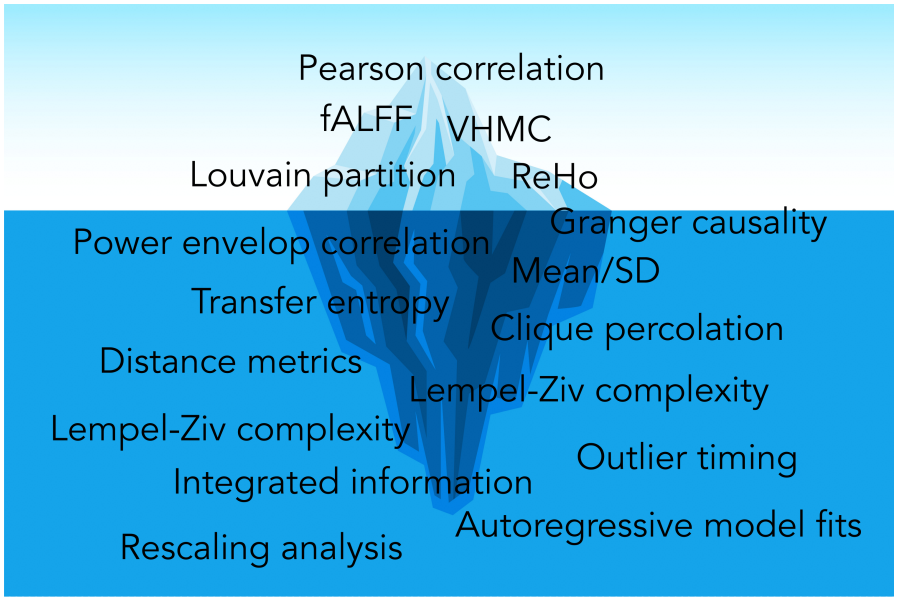


Figure 1.1.2: Statistical analysis of brain-imaging data typically only explores the tip of the iceberg. Measures that are often used in structural and functional neuroimaging analysis form important components in the repertoire of tools for quantifying brain activity and connectivity. However, they are only a small subset of the many possible measures that can capture unique aspects of the brain's dynamics—analogous to the tip of an iceberg floating above the waterline, with the bulk of the iceberg remaining below the surface. *Iceberg image adapted from Adobe Stock Image #1256896566 by user* ***vesvocrea*** *(licensed under Adobe Education License).*

Often, functional neuroimaging studies will only explore the tip of the metaphorical 'iceberg' of methods with which to capture the complex dynamical structure of the brain, as depicted in Fig. 1.1.2. There are thousands of statistical measures that can capture unique properties of activity within a brain area (i.e., local dynamics), coupling between pairs of brain areas (i.e., pairwise interactions), or network-level community dynamics from distributed (pairwise interaction-based) graphs [12]. However, historically, the combination of disciplinary silos and logistical challenges with algorithmic implementation meant that only a small handful of measures were used in practice, and seldom with any systematic comparison. As a result, the vast majority of measures that could be used to capture the brain's dynamical structure have not been explored in standard neuroimaging analyses, and the choice of which measure to use is usually made manually based on heuristics or literature conventions. Until very recently, only a handful of studies directly compared across measures for functional connectivity—or, less commonly, local regional dynamics—and the set of statistics examined was generally limited in scope [13–15]. Many applications, from basic neuroscience to translational research, could greatly benefit from a more systematic and comprehensive approach to quantifying the brain's dynamical structure. In this thesis, I aim to address this gap with two key approaches:

(1) zooming out and viewing the brain as a complex system; and (2) parsing currently disparate literature to streamline information about what measures are out there and why/when/how to use them. In order to appreciate the rich tapestry of neural structures and interactions that give rise to the brain's complex dynamics, I will begin by briefly reviewing the relevant scales of brain structure and function, as well as the key neuroimaging modalities used in this thesis.

1.2 THE BRAIN AS A COMPLEX SYSTEM: SCALES AND HIERARCHIES

1.2.1 SPATIAL SCALES IN THE BRAIN

The brain exhibits multiple complex hierarchies, with linear and nonlinear interactions that are orchestrated across spatial and temporal scales [16–18]. We begin by first considering the spatial continuum along which we can describe and analyze the brain's structure and function. As depicted in Fig. 1.2.1, the brain's intricate spatial architecture can be decomposed into 'microscale' components, such as individual neurons and synapses, which are organized into 'mesoscale' local circuits and networks, and finally into 'macroscale' large-scale brain systems. While the work presented in this thesis focuses primarily on the macroscopic level (i.e., brain regions and their interactions within macroscale networks), it is important to acknowledge the contributions of the microscopic and mesoscopic levels to our understanding of brain function. Moreover, the methods and principles introduced in this thesis can be applied to data at all spatial scales, from the microscopic level of single-neuron activity to the macroscopic level of whole-brain networks.

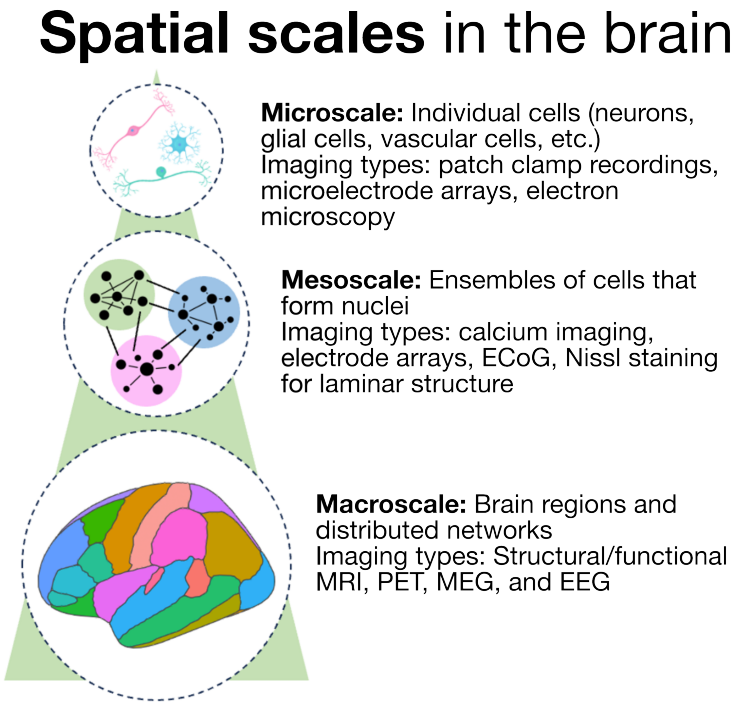


Figure 1.2.1: The brain exhibits a range of spatial scales, from single cells up to distributed networks. The brain's spatial scales can be broadly organized into three levels: Microscale, which encompasses individual cells (including neurons, glial cells, and vascular cells) and can be probed using techniques like microelectrode arrays or two-photon microscopy; Mesoscale, which includes ensembles of neurons that form nuclei or local circuits, and can be studied using techniques like calcium imaging, electrode arrays, electrocorticography (ECoG), or Nissl staining for laminar structure; Macroscale, which involves larger brain regions and distributed networks, and can be investigated using techniques like structural/functional magnetic resonance imaging (MRI), positron emission tomography (PET), magnetoencephalography (MEG), or electroencephalography (EEG). *Microscale cell types image adapted from Adobe Stock Image #189041780 by user* ***achiichiii*** *(licensed under Adobe Education License).*

1.2.2 ZOOMING OUT TO VIEW THE BRAIN AS A COMPLEX SYSTEM

The plethora of ways to quantify and analyze the brain's dynamical structure speaks to its membership in the constellation of complex systems, meaning a system comprised of individual components, each exhibiting its own distinctive activity patterns, with dynamic and nonlinear interactions among components [19]. In the context of the brain's spatial scales, the individual neurons (microscale) organize into local circuits (mesoscale) that exhibit localized firing dynamics, which collectively shape macroscale network interactions that contribute to behaviors like perception or emotion [20]. From this perspective, the brain can be conceptually mapped in the same way as systems across diverse domains, as depicted schematically in Fig. 1.2.2A(i). Other such complex systems include economics, where individual consumers, government regulations, and international trading yield broader trends like growth or recession (Fig. 1.2.2A[ii]); social networks, where friends, family, and coworkers form community groups (Fig. 1.2.2A[iii]); and ecological food webs, where individual species interact with each other and the environment to form complex ecosystems (Fig. 1.2.2A[iv]). As with other types of complex systems, the brain's activity evolves over time in a structured and non-random way, even without any externally evoked dynamics.

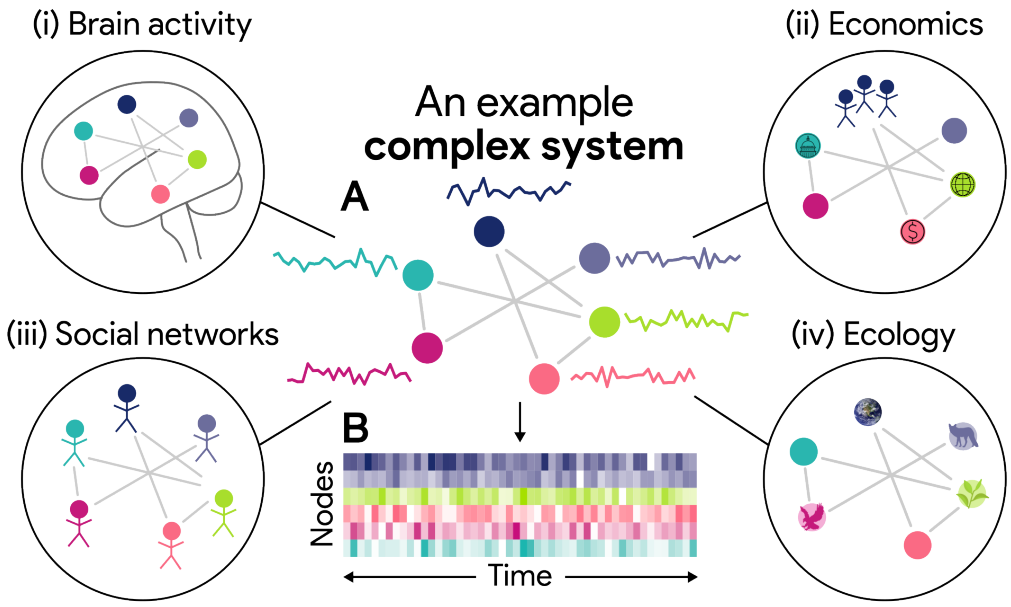


Figure 1.2.2: The brain is a specific type of a more general class of complex systems. A. The general form of a complex system comprises a set of nodes, each with its own local dynamics, that collectively exhibit nonlinear interactions (represented as edges) which give rise to emergent global behaviors. (i) The brain is a complex system with nodes representing brain regions, with multiple types of interactions between them spanning spatial and temporal scales. Other examples of complex systems include (ii) economics, (iii) social networks, and (iv) an ecological food web. Importantly, in each system, the components may represent concrete physical entities (e.g., neurons, people, or members of different species) or abstract processes (e.g., international trade, environmental fluctuations, or governmental regulations). **B.** In each case, when the activity of the constituent components is measured over time, the system dynamics can be represented in the form of a multivariate time series (MTS) matrix.

By zooming out from neural-specific analytical conventions, we can abstract the brain's dynamical structure to a more generalized representation, examining how the internal trajectory evolves in the brain as a complex system. The activity of each component, measured over time, can be represented as a multivariate time series (MTS) matrix, as introduced above in Fig. 1.1.1C (Sec. 1.1). Here, each row corresponds to an individual component (be it a region in a brain, a person in a social network, an economic indicator, etc.) and each column corresponds to a time point, as depicted schematically in Fig. 1.2.2B. Measuring the interactions within a distributed system presents a fundamental question in science, with a long history of developing methods for quantifying such relationships—albeit, often isolated within disciplinary silos. For example, economists capitalize on (no pun intended) the 'cointegration' of paired non-stationary[1] time series, like financial indices and their associated future contracts in the stock market, to build econometric models with significant coupling [21]. Over in the geosciences, pairwise causal models have been refined for inferring underlying mechanisms of natural processes, like atmospheric factors driving air circulation [22] and the influence of sea-surface temperature on anchovy and sardine populations [23]. Collectively, this brings us to the crux of the thesis: **How can we invoke the wealth of statistical methods developed across an interdisciplinary landscape to better understand the brain as a complex system?** And how can we do so in a way that prioritizes methodological parsimony and biological interpretability?

[1] A *non-stationary* process exhibits statistical properties that change over time, such as mean or variance.

As with any complex system, activity within and between components can be meaningfully analyzed at different levels of granularity. We can concretely anchor the local dynamics and pairwise coupling between regions along a graded spectrum that describes the number of components considered, as depicted in Fig. 1.2.3—from one in the case of local dynamics, to two in the case of pairwise coupling, and many in the case of distributed interactions. In this thesis, for distributed interactions, I focus on networks constructed from pairwise adjacency matrices—though I note there are many other ways to represent distributed dynamics beyond those discussed here, such as examining higher-order interactions among three or more brain areas at once [24–26]. Moreover, while this thesis focuses on contiguous brain regions with discrete boundaries, these principles of topological interaction can also flexibly extend to spatially distributed modes, extracted as components based on geometric embedding [27, 28], structural connectivity [29], or dimensionality reduction techniques [30, 31].

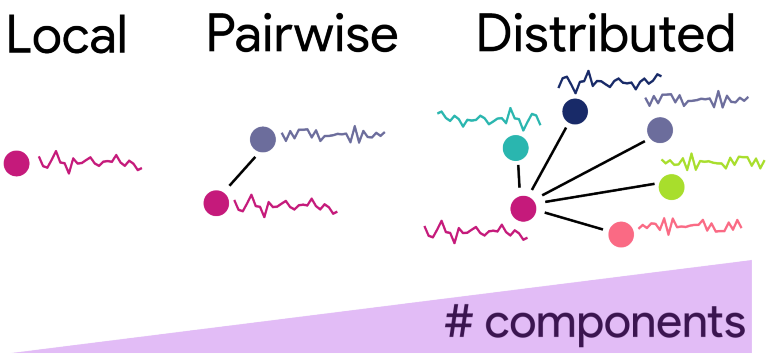


Figure 1.2.3: The brain's dynamical structure unfolds in a multiscale framework of interactions spanning different levels of component granularity. We can organize the brain's dynamical structure based on the granularity of the activity within and between its components, roughly into three scales: Local, which focuses on the intrinsic dynamics of an individual region or cell, such as the firing rate of a neuron or the activity of a brain region; Pairwise, which considers the interaction between a pair of regions or cells, such as the correlation between the activity of two neurons or the functional connectivity between two brain regions; and Distributed, which encompasses the overall structure emerging from the pairwise interactions across many regions or cells, like global (i.e., brain-wide) functional connectivity or collective dynamics across a network of neurons.

Each of the aforementioned levels of component granularity can be analyzed independently to yield meaningful insights into the brain's structure and function. For each brain region, its local dynamics are collectively shaped by (and in turn shape) its bidirectional interactions with the functional network in which it is embedded [2, 32–34]. Emerging evidence suggests that combining properties of resting-state activity that span both local dynamics *and* pairwise coupling can synergistically provide more information about the brain's structure and function than either representation on its own [35–39]. Methods for combining properties of structure and/or function across scales range from simple feature concatenation to more nuanced data fusion techniques [40, 41]. In the comparative spirit of this thesis, I will demonstrate how bridging across these levels of component granularity provides unique and informative insights into brain function in both health and disease. I will begin with the local dynamics of each region (i.e., the activity of a single region in isolation) and then move on to pairwise interactions between regions, which encompasses the coupling between two regions. Finally, I will discuss distributed interactions, which captures the collective activity

of many regions across a network. In each case, I will briefly summarize the typical measure(s) used to capture the dynamics at that scale to provide context for the highly comparative approach taken in this thesis.

1.2.3 LOCAL DYNAMICS IN THE BRAIN

'Local dynamics' refers to the activitzy of a single brain unit (region, voxel, cortical column, etc.) sampled over time, such as the resting-state activity of an individual's hippocampus or oscillatory activity of the primary visual cortex. In this thesis, I focus on the macroscale end of local dynamics, corresponding to contiguous brain regions—which can be defined practically as a set of voxels (or vertices) that are spatially adjacent to each other and share a common boundary based on a given parcellation scheme. There are a variety of physiological mechanisms governing the local dynamics of a brain region, including synaptic activity, neuromodulatory signaling, and intrinsic oscillatory activity of neuronal populations within that region [32]. Focusing on local dynamics can yield brain-wide maps of activity properties in health and disease, which offers a spatially interpretable way to quantify and visualize regional dynamics in contexts ranging from targeted brain stimulation [42, 43] to localized disruptions in diverse neurological disorders [44–46]. Depending on the imaging modality and question at hand, a researcher could focus on a single property of the activity in a single region, or scale up to multiple properties of activity in a single region, or even multiple properties of activity across multiple regions. In any of these cases, of the plethora of measures that could capture local dynamics, the choice of which to use is often made manually based on heuristics and/or literature conventions—typically constrained to measures like the fractional amplitude of low-frequency fluctuations (fALFF) [4, 5], regional homogeneity (ReHo) [6], or voxel-mirrored homotopic connectivity (VMHC) [7].

1.2.4 PAIRWISE CONNECTIVITY IN THE BRAIN

In addition to the local dynamics of a single brain region, we can also consider pairwise functional connectivity, meaning the statistical dependency between activity time series measured from two brain regions [47]. We can also represent the brain's physical architecture through pairwise *structural* connectivity, estimated using tractography applied to diffusion-weighted imaging (DWI, discussed in more detail in Sec. 1.3.1). While not all structural and/or functional connections reflect direct monosynaptic axonal projections, there is an overall positive association between structural and functional connectivity, meaning that regions with more axonal connections are likely to exhibit stronger functional coupling [48, 49]. Functional connectivity between two regions can arise from multiple physiological mechanisms, including shared afferent inputs or efferent targets, as well as bidirectional anatomical connections [50–52]. Most resting-state functional neuroimaging studies have examined functional connectivity between pairs of brain regions by computing the Pearson correlation coefficient [10, 11, 53], a zero-lag cross-correlation that is exclusively sensitive to linear interactions between two time series. In contrast to functional connectivity, the ways to capture structural connectivity are constrained to (probabilistically) estimating the number of streamlines physically connecting the two regions, which in turn represents the number of axonal fiber bundles connecting the two regions.

1.2.5 DISTRIBUTED NETWORKS FORMED FROM PAIRWISE INTERACTIONS

Structural and functional connectivity can both be investigated in a single region–region pair, which is particularly useful in the context of a specific hypothesis about a given biological pathway. However, it is also often informative to consider the interactions between many region–region pairs altogether, which can be represented as a distributed network of pairwise interactions [54]. This representation opens the door to a variety of sophisticated yet interpretable analytical techniques deriving from network science and graph theory that capture unique aspects of both local and global brain organization. While structural and functional connectivity arise from different physiological mechanisms—anatomical connections versus correlated activity, respectively—they can both be represented as networks, in which brain regions are the nodes and the connections between them (structural or functional) form the edges [55, 56]. These connectomic networks are often directly compared to study the associations between brain structure and function, and how such associations can be disrupted across disease settings [57–60]. While functional connectivity is generally computed as a measure of pairwise dependence between two regions, one could extend this to encapsulate higher-order dependencies, such as triplet or quartet interactions [24–26]. By contrast, since a white-matter fiber bundle intrinsically innervates two regions, structural connectivity is almost exclusively represented through pairwise adjacency matrices that may be weighted by the number of fibers connecting each region–region pair.

For both structural and functional connectivity, the resulting distributed network can be represented as an adjacency matrix, which can be too high-dimensional to meaningfully analyze directly. In practice, to distill biological inferences about structural or functional connectivity in a given participant (or group of participants), researchers often apply a community detection algorithm to the pairwise connectivity matrix—which identifies groups of nodes that are more densely connected to each other than to the rest of the network (i.e., forming 'communities'). Such community detection analysis offers insights into the underlying organization of the brain's network structure and enables statistical analysis of nodal embedding within the topological network, using measures like 'participation coefficient' or 'within-module degree' [57, 61, 62]. There are many algorithms conducive to this type of analysis, such as the Louvain method [63], which is the most widely used community-detection method in network neuroscience [64]. The Louvain algorithm seeks the optimal non-overlapping division of the network into mutually exclusive communities, with 'non-overlapping' meaning that each node (region) is assigned to a single community. This method is consistent with evidence for structural [65] and functional [66] specialization within different corners of the brain (e.g., the visual cortex is principally involved in processing visual information [67], while the motor cortex is geared toward motor control [68]). However, a non-overlapping partition of a given structural or functional network may not accurately reflect the underlying connectivity structure [69, 70], particularly in the case of transmodal cortical association areas that are flexibly involved in multifaceted information processing [71, 72].

1.2.6 THIS GENERALIZED REPRESENTATION OPENS THE DOOR TO HIGHLY-COMPARATIVE ANALYSIS

Even within each level of component granularity (e.g., local activity or pairwise coupling), many different statistical representations have been developed to capture some aspect of dynamical structure, each with

its own strengths and weaknesses. Prior work indicates that resting-state functional magnetic resonance imaging in particular (fMRI, introduced in more detail in Sec. 1.3.2 of this chapter) is primarily sensitive to macroscopically linear interactions owing to a combination of factors, including low temporal resolution and minimal dynamical perturbations from a fixed point equilibrium [73–75]. Similarly, it has been posited that resting-state fMRI is too noisy and poorly sampled in time to justify more sophisticated time-series analysis methods (e.g., those applied to imaging modalities with higher temporal resolution) [76–78]. While this suggests that the Pearson correlation coefficient is generally well-suited to capture interactions from resting-state fMRI [79, 80], the statistic also makes strong assumptions about the data—including that the joint bivariate distribution is Gaussian [81]—which may not always hold across imaging modalities [11, 82].

Without systematic comparison to alternative methods, it is difficult to determine whether the Pearson correlation coefficient is the most appropriate method for a given dataset and research question; indeed, recent data-driven analyses of resting-state fMRI have highlighted the utility of exploring beyond standard linear time-series analysis for both intra-regional [83] and inter-regional dynamics [13, 84]. Prior to the work presented in this thesis, recent work has suggested that time-series measures developed outside of neuroscience can yield meaningful insights into brain function—such as dynamic time warping, which is increasingly applied to capture patterns in resting-state fMRI activity [85–87] despite its origins in spoken word recognition [88]. While these analyses investigated a small curated subset of time-series measures, they underscore the potential benefits of venturing beyond the conventional measures used in neuroimaging to explore a wider range of time-series statistics.

In recent years, hundreds to thousands of time-series measures have been organized and implemented in cohesive code-based libraries, such as the *hctsa* library [89, 90] and related *catch22* library [91] for local activity and the *pyspi* library for pairwise interactions [92]. The goal with each of these libraries is to encapsulate a very large range of statistical approaches to summarize dynamical patterns of activity in a given multivariate time series, in the form of comprehensive feature vectors. Crucially, the unified algorithmic implementations in these libraries enable the systematic comparison of different representations of dynamical structure, which would otherwise be impractical to apply piecemeal—thus opening up a large repertoire of methods with which we can quantify brain activity across different levels of component granularity. For example, the *hctsa* library collates thousands of univariate time-series features from across the time-series literature, including those that capture aspects of predictability, stationarity, variability, and the scaling of fluctuations across different timescales [89, 90]. Similarly, the *pyspi* library unifies hundreds of statistics of pairwise interactions, such as those which capture instantaneous or time-delayed responses, linear or nonlinear interactions, and condition on the past or allow for non-constant delays [92]; the development and first application of this library comprises Chapter 3 of this thesis. Mounting evidence collectively suggests that such highly comparative methods can yield meaningful insights into the brain's dynamical structure, both in terms of local activity and pairwise interactions, as well as distributed networks formed from pairwise interactions. For example, novel methods have elucidated important principles of macroscopic brain organization, in applications from accurately tracking independent biological gradients [83] to distinguishing experienced meditators from controls [30] to classifying stimulation experiments in mice [42, 43].

In addition to functional neuroimaging, highly comparative methods can also yield meaningful insights into macroscale properties of structural connectivity when applied to diffusion-weighted imaging data (DWI, as

introduced in Sec. 1.3.1 of this introductory chapter). Community detection algorithms (e.g., the Louvain method [63]) are often applied to structural connectivity networks to infer the modular organization of the brain's physical wiring. While these methods constrain each node (e.g., region) to participate in a singular community, recent years have seen the development of overlapping community detection algorithms (OCDAs) that allow nodes to belong to multiple communities—which can more organically capture the hierarchical and overlapping nature of brain organization [69, 70]. Amidst this growing variety of algorithmic choices, it is often unclear which OCDA is most appropriate for a given dataset and research question, particularly since each algorithm makes different assumptions about how communities and overlapping nodes are defined. The 'no free lunch' theorem [93] states that no single algorithm can perform optimally on all possible data, and community detection methods are no exception [94, 95].

In the present thesis, I will implement a highly comparative approach with both structural and functional neuroimaging modalities to systematically compare a wide range of measures that capture different aspects of the brain's dynamical structure, with applications to both basic and translational neuroscience. This comprehensive methodological approach stands in contrast to the more subjective (or, perhaps, hypothesis-driven) conventional approach of selecting a single method for analyzing a given neuroimaging dataset, as introduced at the start of this introductory chapter. In the chapters of this thesis, we build upon emerging work to develop and/or apply highly comparative methods to a wide range of neuroimaging datasets, including both structural and functional data, and across multiple levels of component granularity, from local (one component) to pairwise (two components) to distributed networks (constructed from pairwise interactions across many components). This approach directly enables the systematic comparison of many different representations of the brain's dynamical structure to identify the most informative and interpretable measures for a given research question, with applications ranging from consciousness to inter-hemispheric synchronization to neuropsychiatric disorders. Collectively, these steps start to bridge the critical gap between interdisciplinary theory in complex systems and practical applications for human neuroimaging analysis, which forms the central theme of this thesis.

1.3 AN OVERVIEW OF THE NEUROIMAGING MODALITIES EXAMINED IN THIS THESIS

Neuroimaging is a powerful tool for visualizing and quantifying the brain's structure and function, offering a window into the nuanced activity within and communication between different brain areas. While there are invasive modalities that require physical tissue resection and/or implantation of recording devices, in this thesis, I focus on non-invasive imaging techniques to capture distinct aspects of macroscale brain structure and function. Each neuroimaging modality involves different acquisition parameters and timescales that render them suitable for different aspects of visualizing and analyzing brain structure and function. In this section, I will briefly review the key neuroimaging modalities used in this thesis, focusing on their spatial and temporal scales, and how they can be used to capture different aspects of brain structure and function.

1.3.1 KEY STRUCTURAL NEUROIMAGING MODALITIES

Magnetic resonance imaging (MRI) forms the basis for most of the work presented in this thesis, so I will begin with a brief review of different types of MRI, before introducing other relevant imaging modalities. MRI is a non-invasive imaging technique that applies a strong magnetic field together with radiofrequency pulses to reconstruct detailed images of the brain's structure. There are also functional variants of MRI, though due to the trade-off between spatial and temporal resolution, functional MRI generally exhibits lower spatial resolution than structural MRI [96]. MRI exploits the intrinsic magnetic properties of protons in different biological tissues, each of which exhibits a unique 'signature' of relaxation times—thus enabling the differentiation of tissue types based on proton relaxation [97, 98]. Importantly, MRI is a cornerstone of both basic and translational neuroscience research, offering multifaceted insights into brain structure and function without any radiation exposure or injection of contrast agents (unless specifically required for a given perfusion or angiography sequence).

T1-weighted MRI

A typical research neuroimaging study will include a T1-weighted sequence, which derives its contrast (i.e., the distinction between different tissue types) based on the spin angle of the hydrogen protons in water molecules, which at baseline are randomly oriented [97, 98]. Once a patient or participant enters the scanner, the scanner applies a magnetic field (called the 'B0' field) and a subsequent radiofrequency pulse that causes the protons to align and spin along an axis parallel to the magnetic field. After a brief period, the scanner measures the amount of energy the protons release as they relax, which is directly proportional to the time it takes for them to return to their original orientation—known as the T1 relaxation time. The anatomical volume is then reconstructed from the T1 relaxation time in each voxel, which will then undergo several phases of preprocessing, including field map correction, motion adjustment, tissue segmentation, and alignment to a standard template space. Subsequent functional imaging acquisitions are aligned to the individual's 'native-space' T1-weighted MRI as an intermediate point of reference, which can then be parcellated in native-space or transformed to a standard-space (e.g., MNI-152 [99]) before parcellation.

Diffusion-weighted imaging (DWI)

Diffusion-weighted imaging (DWI) is a specialized MRI technique that measures how freely water can move in different areas of the brain, providing insights into microstructural properties of the brain parenchyma [100, 101]. Specifically, DWI derives its contrast from the Brownian motion of water molecules, which varies dramatically across different tissue types—for example, water can move more freely in CSF than in gray matter, particularly compared to tumors or infarcts. DWI is particularly useful for mapping white matter architecture in the brain, which is comprised of myelinated axons that innervate different cortical and subcortical regions. In the related diffusion tractography imaging (DTI), the diffusion of water molecules is measured in multiple directions across the three-dimensional volume of the brain, enabling reconstruction of the orientation and integrity of white-matter tracts. This modality is often used to study the brain's structural connectivity—which refers to the physical connections between different brain regions—in tandem with a preprocessed T1-weighted volume, enabling the estimation of fibers connecting specific pairs of regions [102]. In chapter 7, we use DWI data to estimate the right-hemisphere structural connectome across a large cohort of

healthy participants ($N = 973$), which we then use to compare the ability of different overlapping community detection algorithms (OCDAs) to recover the modular organization of anatomical connections between brain regions.

1.3.2 KEY FUNCTIONAL NEUROIMAGING MODALITIES

Functional magnetic resonance imaging (fMRI)

Functional magnetic resonance imaging (fMRI) measures the blood oxygen level-dependent (BOLD) signal, which measures the ratio of deoxygenated to oxygenated hemoglobin driven by localized changes to brain oxygenation and blood flow [103, 104]. These processes are spatially and temporally coupled to neuronal activity via a phenomenon called neurovascular coupling, in which the onset of neural activity in a given region increases the metabolic demand for oxygen and glucose, such that more oxygen is shuttled from the blood vessels out to the parenchyma [105]. This causes an initial increase in the relative amount of deoxygenated hemoglobin remaining in the blood that induces a change in magnetic properties of the blood that can be detected by the scanner. With sustained neural activity, there is increased blood flow through vasodilation (i.e., increased diameter of blood vessels), which brings more oxygenated hemoglobin to the region, such that the ratio of deoxygenated hemoglobin decreases. This change in the ratio of deoxygenated to oxygenated hemoglobin is referred to as the BOLD signal, which the scanner measures over time to yield a time series of BOLD activity for each voxel in the brain. Importantly, the BOLD signal is not a direct measure of neuronal activity, but rather a proxy that is separated from the underlying neural activity by several steps that span multiple seconds due to the hemodynamic response function (HRF) [106]. BOLD fMRI can be used to investigate both resting-state brain dynamics and task-evoked responses; this thesis focuses almost exclusively on the resting-state condition for fMRI (in Chapters 2, 5, and 6), with the exception of one naturalistic movie-viewing setting in Chapter 3.

Magnetoencephalography (MEG)

Magnetoencephalography (MEG) is a functional neuroimaging modality that uses magnetometers to detect the magnetic field generated by neural activity, providing millisecond-level temporal resolution of activity (aggregated across tens of thousands of neurons per channel) [107]. While fMRI provides a proxy of neural activity via changes in blood flow, MEG directly measures oscillatory electrical activity in the brain with improved temporal resolution (at the expense of spatial resolution) [108]. Physiologically, the primary source of MEG signal is slow, synchronized excitatory postsynaptic potentials (EPSPs) in large populations of synchronously active pyramidal neurons, mainly in cortical layer 5 [109, 110]. These pyramidal neurons are long and aligned in parallel along cortical columns, and the EPSPs generate current flow along their dendrites. MEG is sensitive to the tangential component of these currents, making it maximally sensitive at sulcal walls (i.e., where pyramidal neurons are aligned perpendicular to the skull surface). While a signal can be obtained from anywhere in the cerebrum, physics inherently limits the depth from which *reliable* signals may be captured. In other words, the signal-to-noise ratio (SNR) decreases exponentially with depth into the brain due to the increasing distance between the neural magnetic current generators and the external MEG sensors [107], owing to factors including signal attenuation and field spread. As a result, MEG is most sensitive to superficial cortical structures (e.g., upper parts of gyri and sulci), whereas signal

from deeper areas (e.g., subcortical gray matter structures and the brainstem) is more difficult to resolve—though still possible with advancements in preprocessing techniques [111]. Given the relative advantages and limitations of temporal and spatial resolution, respectively, MEG is particularly well-suited to examine task-evoked responses in larger (and more superficial) cortical areas.

In Chapter 4 of this thesis, we use MEG data to compare the performance of a wide range of functional connectivity measures (from the *pyspi* library, introduced in Chapter 3) in distinguishing between consciously perceived visual stimuli, with the goal of identifying the most informative and interpretable measures that can quantitatively evaluate theoretical models of consciousness. The high temporal resolution of MEG is vital for this analysis in order to capture the rapid dynamics of neural activity that underlie conscious perception of short (1-second) visual stimuli, unfolding on the scale of tens to hundreds of milliseconds.

Electroencephalography (EEG)

Electroencephalography (EEG) is a functional neuroimaging modality that uses electrodes placed at specific intervals along the scalp's surface to measure electrical activity within the brain. As with MEG, EEG measures millisecond-resolution neural activity generated by EPSPs in large populations of synchronously active pyramidal neurons in the cerebral cortex [112, 113]. However, while MEG measures the magnetic field generated by these EPSPs, EEG measures the difference in electrical potential detected by each active electrode and a common 'reference' signal, either a physically placed electrode or a virtual reference based on the average across all physical electrodes. EEG offers comparable temporal resolution to that of MEG, but with lower spatial resolution owing to the reliance on physical electrodes, which in turn yield greater distortion of the signal due to the skull and skin compartments [114]. However, EEG offers logistical advantages over MEG, as it is considerably less expensive to set up, simpler to use, and more portable, vastly improving its accessibility for both research and clinical applications. Both EEG and MEG suffer from the 'inverse problem', which refers to the challenge of inferring the underlying neural activity from the measured signals (either electric or magnetic, respectively) on the scalp surface due to the superposition of signals from multiple sources [115]. However, several advanced source reconstruction techniques have been developed to address this challenge, called 'forward solutions' [116], many of which use biophysically informed models that incorporate measurements of the participant's head position. These techniques include beamforming methods [117], which estimate the strength and location of the underlying neural sources based on the empirically measured signal at the scalp surface.

In Chapter 3 of this thesis, we evaluate the full *pyspi* library on an EEG dataset acquired from a volunteer participant attempting to move a cursor on a screen by modulating their brain activity toward cortical 'up' or 'down' states. This dataset provides an important point of comparison as the other neuroimaging case study in this chapter examines BOLD fMRI, which exhibits a sampling frequency several orders of magnitude lower than that of EEG. As a result, EEG time series may capture rapid fluctuations in neural activity (that are not observable in fMRI) which may give rise to nonlinear and/or time-lagged pairwise interactions, such that different statistical measures may be more or less appropriate for capturing the underlying dynamics.

1.4 THE IMPORTANCE OF INTERPRETABILITY AND ACCESSIBLE COMMUNICATION

Having introduced the potential for tailoring methods from complex systems to compelling problems in modern neuroscience, I now focus on the importance of interpretability and accessible communication. This is motivated by the need to make complex methods more approachable for a wider audience, including researchers from other fields and those without formal training in statistics, mathematics, or physics. In the following sections, I will outline how this thesis aims to achieve this through the combination of methodological parsimony, visual interpretability, and interdisciplinary dialogue.

1.4.1 METHODOLOGICAL PARSIMONY AND INCREMENTAL COMPLEXITY ONLY AS NEEDED

In the context of this thesis, 'methodological parsimony' refers to the principle of using the simplest possible representation that sufficiently captures the underlying dynamical structure of the system being studied [118]—in this case, the brain. This principle is grounded in the notion that simpler models tend to be more interpretable and easier to understand, which is especially important in the context of neuroscience, where the goal is often to elucidate the mechanisms giving rise to both function and dysfunction [119–124]. With the growing sophistication of machine-learning algorithms in recent decades, there has been a tendency to default to more complex statistical methods, many of which employ deep learning techniques that can obscure their clear interpretation [124]. Such model complexity can lead to overfitting and poor generalizability on out-of-sample data [125], particularly in the context of neuroimaging, where the number of features (e.g., voxels or regions) can far exceed the number of samples (e.g., participants) [126]. For example, within the last five years, some studies have reported >90% accuracy in distinguishing neuropsychiatric disorder patients from clinically healthy controls, largely by combining deep neural network approaches with smaller sample sizes [127–131]. By contrast, two recent studies rigorously evaluated multiple different types of classifiers for hundreds [132] or thousands [133] of participants with major depressive disorder or healthy controls, finding in both cases that the maximum case–control classification accuracy did not surpass 65%.

Small sample sizes and complex classifiers can therefore lead to overestimated predictive performance, which is problematic in scientific settings (such as neuropsychiatric disorders) in which understanding and generalizability are paramount [124, 134]. Without any structured comparison to simpler methods, it is difficult to determine if and when such complexity is needed to achieve strong performance on a given problem (e.g., case–control classification). Recent work has directly compared the performance of complex classifiers to simpler ones, finding that the latter can often match (or even outperform) the former in diverse settings, including earthquake detection [135, 136], sleep-stage classification [137], and epilepsy identification [138], while also being more interpretable and computationally efficient. This highlights the importance of starting with a simple model and increasing complexity only as demonstrably needed. In the context of this thesis, I have prioritized methodological parsimony by using simple linear classifiers and minimal parameter tuning, with a focus on interpretability in both the methods and results.

Importantly, this does not mean that I have avoided complexity altogether; rather, I have sought to use the simplest possible representation that sufficiently captures the underlying structure and function of the system being studied. For example, in Chapter 5, we develop and implement a modular approach that intro-

duces incremental complexity from individual brain regions up to integrated brain-wide maps that combine local dynamics and pairwise interactions. Then, in Chapter 7, we develop a systematic approach to comprehensively compare different overlapping community detection algorithms (OCDAs) for structural connectivity networks, which allows us to refine the most appropriate method for a given dataset and research question. In both cases, there are simple methods that are typically used by default—for example, fALFF for local dynamics, Pearson correlation for pairwise interactions, and Louvain partition for distributed networks—which form a baseline against which we can compare more complex methods. This leaves open the possibility that the simplest approach may in fact be the most appropriate for a given dataset and research question, while also allowing us to systematically evaluate the performance of more complex methods in a principled way.

1.4.2 DATA VISUALIZATION: A HISTORICAL PERSPECTIVE

Throughout the work presented in this thesis, I have focused on data visualization as a means to clearly unpack and interpret methods and results from analyzing the brain's complex dynamical structure. Broadly speaking, effective visualization is a powerful tool that facilitates data exploration, hypothesis testing, and communication of results. An oft-quoted example in data visualization courses is that of Dr John Snow's pioneering work in 1854, overlaying cholera deaths atop a map of water pumps in central London, driving the discovery that cholera is water-borne and prompting the swift removal of the contaminated pump—contributing to the end of the cholera epidemic in the Broad Street neighborhood. In his seminal works, 'The Visual Display of Quantitative Information' [139] and 'Visual Explanations' [140], Professor Edward Tufte discusses how Dr Snow's work is one of the earliest examples of effective data visualization for both hypothesis testing and perceptual inference:

> [D]escriptive narration is not causal explanation . . . Instead of plotting a time-series, which would simply report each day's bad news, Snow constructed a graphical display that provided direct and powerful testimony about a possible cause-effect relationship. Recasting the original data from their one-dimensional temporal ordering into a two-dimensional spatial comparison . . . reveals a strong association between cholera and proximity to the Broad Street pump, in a context of simultaneous comparison with other local water sources and the surrounding neighborhoods without cholera. [140]

Prior research demonstrates that combining visual depictions with written text bolsters readers' abilities to construct mental models with less cognitive burden than pure 'sentential' explanations [141, 142]. To quote Professor Julie Jebeile, a philosopher of science at the University of Bern:

> [A] visual representation can exhibit a great amount of data in a structured way. Unlike a messy mass of data, from which it would be hard to draw relevant pieces of information, appropriate visual representations enable one to infer specific pieces of information about the target system. This is due not only to the fact that they can convey a great amount of data but also to the fact that they are . . . presentations of relations of dependence. They arrange within the space of the visual support the many data so that these data can be in relation with each other. [143]

In other words, visual representations can help us see and appreciate relationships between different parts

of the data, which is particularly important in the context of complex systems like the brain. Indeed, Bobek and Tversky [144] contend that visual explanations are critical to understanding the individual components of a complex system as interactions between these structural components. Many of the analyses presented in this thesis are the direct result of intermediate data visualizations, which I have used to explore and identify patterns that are not immediately evident from the raw data itself. As part of the presented thesis, I focus heavily on visual explanations to communicate the algorithmic implementation of time-series measures for local dynamics and pairwise interactions (specifically those derived from information theory; Chapter 2). Whenever possible, I depict empirical results directly in a spatial map of the brain, which is an intuitive way to communicate the results of a given analysis.

1.4.3 USING SCIENTIFIC COMMUNICATION TO FACILITATE INTER-DISCIPLINARY DIALOGUE

Having established the importance of methodological parsimony and data visualization, I finally turn to the importance of accessible communication in scientific discourse. Fostering interdisciplinary dialogue is vital for advancing our understanding of the brain as a complex system, as it helps to bridge gaps between different fields that offer unique and informative perspectives. This is particularly important in the context of statistical analysis for neuroimaging, in which the breadth of available methods can be daunting and present barriers to such fluid interdisciplinary exchange, especially in the context of greater technical complexity and associated jargon.

Accessibility and clarity are priorities throughout the present thesis, forming direct motivations for the work presented in Chapter 2, in which I provide a unified overview of measures derived from information theory to capture dynamical properties from a multivariate time series. Information theory was originally developed to characterize signal flow in communication channels [145], though it generalizes across domains as a model-free statistical framework with which researchers can analyze variance within and relationships between random variables [146, 147]. Recent years have seen increasing application of information-theoretic measures to computational neuroscience settings across spatial and temporal scales [148, 149]. However, amidst the statistical smorgasbord of potential measures at a researcher's disposal, inconsistent terminology and mathematical notation spread across a disjoint literature present a considerable hurdle in application, even for seasoned experts. This motivated the development of a cohesive review of information-theoretic measures for time series, with unified terminology, visual elements, and mathematical notation, in Chapter 2. More broadly, I have sought to make the methods and results presented in this thesis accessible to a wide audience, including those without any formal training in statistics, mathematics, or physics.

1.5 ORGANIZATION OF THIS THESIS

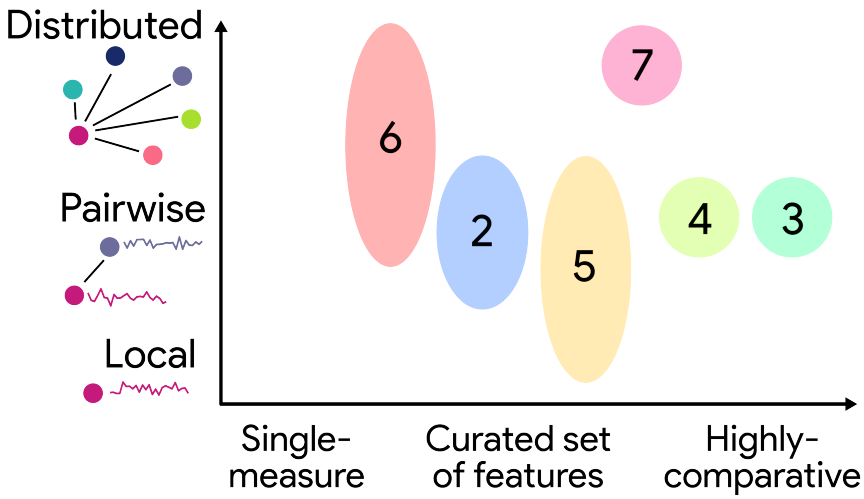


Figure 1.5.1: The chapters of this thesis are organized according to both component granularity and breadth of statistical measures. The six main chapters of this thesis are organized according to two distinct axes of scope: the level(s) of component granularity considered in the work (y-axis) and the breadth of statistical measures examined (x-axis).

The work presented in this thesis is organized as a series of six chapters, each of which builds upon previous ones to develop and refine a highly comparative framework for capturing the brain's complex dynamical structure. As depicted schematically in Fig. 1.5.1, I have mapped each of the six main chapters onto a specific level of component granularity (y-axis) and breadth of compared features (x-axis) to provide an overview of the scope encapsulated in this thesis. The chapters of this thesis are thus presented as follows. Chapter 2 provides a unified overview of measures derived from information theory to capture dynamical properties from a multivariate time series. This chapter includes important definitions and concepts that are used throughout the thesis, including realizations of a stochastic process, the concept of quantifying local dynamics and pairwise interactions in the brain, and the importance of effective data visualization. In subsequent chapters, ideas introduced in this chapter are algorithmically implemented and applied to compelling and diverse applications in human neuroscience.

Chapter 3 then introduces the 'Python toolkit of statistics for pairwise interactions' (*pyspi*), a Python library that implements over 200 measures of interaction between two time series—including all of the pairwise information-theoretic time-series measures discussed in Chapter 2. To demonstrate the utility of this highly comparative approach, we systematically compare the performance of all *pyspi* measures in three real-world classification tasks: (1) activity prediction from smartwatch actigraphy data; (2) classification of cortical 'up' or 'down' states from EEG data; and (3) distinction of brain dynamics during rest versus viewing a film from fMRI data. All three cases converge on the conclusion that conventional measures for a given domain generally do quite well, but that there are novel alternatives that surpass them in some cases.

In the remaining chapters of this thesis, we delve into biologically compelling questions in modern neuroscience through the lens of the highly-comparative framework introduced in Chapters 2 and 3. In Chapter 4,

we tackle an open challenge in consciousness research to quantify properties of inter-areal coupling that may encode the contents of (supra-threshold) conscious visual experience. Using data openly shared from the international COGITATE Consortium from an adversarial collaboration [150], we demonstrate and apply a comprehensive data-driven framework to evaluate measures of functional connectivity from MEG data—which then support the development and interpretation of simple neuro-dynamical models tailored to different theoretical predictions.

To venture beyond considering one level of component granularity in isolation, in Chapter 5, we ask if (and indeed, how) these scales can be combined to synergistically capture more information about the underlying dynamics of a system. We systematically compare across a range of local dynamics and pairwise interaction measures to characterize resting-state brain dynamics in the context of four neuropsychiatric disorders: schizophrenia, bipolar I disorder, attention-deficit hyperactivity disorder, and autism spectrum disorder. First, we demonstrate that linear properties of individual regions—and the Pearson correlation coefficient, which captures contemporaneous linear interactions between regions—rank among the most informative features for distinguishing cases from controls. However, we also show that all evaluated pairwise interaction measures are bolstered by the inclusion of local dynamical features, emphasizing the importance of capturing both local dynamics and pairwise coupling in tandem.

The subsequent chapters continue to expand the scope and scale of interactions analyzed. In Chapter 6, we characterize pairwise interactions between homotopic regions (i.e., regions that are spatially symmetric across the midline of the brain), examining how interactions relate to distributed functional organization within and across hemispheres. The comparative lens is also applied to quantify structural, functional, and molecular correlates of the hierarchical spatial variation in homotopic functional connectivity. Finally, in Chapter 7, we develop and implement a comprehensive framework for applying overlapping community detection algorithms (OCDAs) to characterize overlapping hierarchical structure in distributed structural networks in the right-hemisphere cortex. After identifying the optimal algorithm and parameter configuration, we apply this method to the empirically observed right-hemisphere cortical connectome, identifying anatomically and functionally meaningful communities bridged by fifteen ‘overlapping’ nodes.

To conclude this thesis, in Chapter 8, I summarize the key findings and contributions of this work, and discuss future directions for areas in which the highly comparative approach may offer major inroads for understanding the brain as a complex system.

2
Unifying concepts in information-theoretic time-series analysis

Information is power.
Dwight K. Schrute, The Office, Season 1 Episode 1

Information theory provides a powerful statistical repertoire for capturing complexity and dynamical richness across diverse domains in the life and physical sciences, from computational neuroscience to statistical mechanics. Despite the wide array of informative measures, a fragmented literature—including inconsistent mathematical notation, domain-specific terminology, and disparate visualization conventions—has hindered its interdisciplinary integration for time-series analysis. In this chapter, we address this challenge by presenting a novel and cohesive framework that unifies key information-theoretic time series measures through shared semantic definitions, mathematical notation, and visual representations. We compare eleven distinct information-theoretic measures in terms of theoretical foundations, computational formulations, and practical interpretability, mapping them into a common conceptual space through an illustrative case study based on resting-state macroscale brain activity. For brevity, we present excerpts from the associated manuscript that introduce the unified visual schematic and tabular representation of the information-theoretic measures, the brain activity case study, and the distinction between stochastic processes versus realizations. The second half of the manuscript is included as Appendix 1 of this thesis, providing a detailed breakdown of each individual measure, including its theoretical foundations, computational implementation, and interpretation in empirical data. My goal with this review is to provide a visually and semantically intuitive overview of the information-theoretic measures that we will use in subsequent chapters.

This chapter includes excerpts from work currently in submission as: Bryant, A. G., Cliff, O. M., Shine, J. M., Fulcher, B. D., & Lizier, J. T. Unifying concepts in information-theoretic time-series analysis. *arXiv* preprint (2025), doi: arXiv:2505.13080.

Abstract

Information theory is a powerful framework for quantifying complexity, uncertainty, and dynamical structure in time-series data, with widespread applicability across disciplines such as physics, finance, and neuroscience. However, the literature on these measures remains fragmented, with domain-specific terminologies, inconsistent mathematical notation, and disparate visualization conventions that hinder interdisciplinary integration. This work addresses these challenges by unifying key information-theoretic time-series measures through shared semantic definitions, standardized mathematical notation, and cohesive visual representations. We compare these measures in terms of their theoretical foundations, computational formulations, and practical interpretability—mapping them onto a common conceptual space through an illustrative case study with functional magnetic resonance imaging time series in the brain. This case study exemplifies the complementary insights these measures offer in characterizing the dynamics of complex neural systems, such as signal complexity and information flow. By providing a structured synthesis, our work aims to enhance interdisciplinary dialogue and methodological adoption, which is particularly critical for reproducibility and interoperability in computational neuroscience. More broadly, our framework serves as a resource for researchers seeking to navigate and apply information-theoretic time-series measures to diverse complex systems.

CODE AVAILABILITY

All code used for empirical analysis and visualization in this work is freely available at https://github.com/anniegbryant/info_theory_visuals.

THE NEED FOR A COHESIVE, ACCESSIBLE SUMMARY OF INFORMATION-THEORETIC TIME-SERIES MEASURES

Information theory, originally developed for studying communications channels, provides a model-free statistical framework for analyzing variance within and relationships between random variables [145–147]. Measures derived from information theory, like mutual information and transfer entropy, present powerful and domain-general tools with which researchers can quantify distinctive patterns of information flow and dependence, with particular utility in applications to multivariate time series samples of random variables. Information-theoretic measures are increasingly developed and applied across diverse domains, from economics [151] to epidemiology [152] to neuroscience [148, 149]. Despite this smorgasbord of different measures at a researcher's disposal—including different forms of entropy, directed information flow, and stochastic interactions—we are generally limited to a disjoint literature that employs inconsistent terminology and mathematical notations. These differences in nomenclature create a sizable learning curve for a researcher to understand the nuances of a single measure, which becomes even steeper when trying to understand intricate similarities and differences across multiple measures. As a result, one may select a measure based on subjective criteria and/or domain-specific conventions, potentially missing more appropriate information-theoretic measure(s) for the problem at hand. Prior work has endeavored to provide practical guides for applying and interpreting general information-theoretic measures (including joint/conditional entropy, mutual information, transfer entropy, and partial information decomposition) across disciplinary do-

mains [149, 153–157], but a unifying treatment of how information-theoretic methods can be adapted to time-series analysis remains lacking.

To our knowledge, no prior work has attempted to unify the disjoint literature on information theoretic measures for time series as a coherent presentation using consistent terminology and visual elements. To address this gap, here we conceptually unify information-theoretic time-series measures with common semantic nomenclature and mathematical notation in a cohesive guide that offers accessible categorization and explanations. Our aim with this paper is to unify information-theoretic measures into a common schematic and tabular representation to provide a high-level summary that is broadly accessible to interdisciplinary readers interested in applying information theory to time series data. As such, our approach is to actively incorporate the mathematical formulations of each measure alongside plain-language summaries and interpretable visuals to improve the accessibility of these powerful measures across disciplines. Visual schematics form the basis for much of this paper, motivated by work demonstrating how visual explanations facilitate conceptual links between the structural components that collectively interact to form a complex time-varying system [144].

2.1 COMPILING A TAXONOMY OF INFORMATION-THEORETIC TIME-SERIES MEASURES

Before delving into the information-theoretic measure taxonomy, it will be helpful to first outline a few key definitions. We define a 'time series' as a time-indexed stochastic process, comprised of a sequence of random variables $\mathtt{X} = \{X_1, X_2, X_3, \ldots, X_T\}$ (where T is the length of the process) as depicted schematically in Fig. 2.1.1. Each X_t is a random variable that can take the form of a single dimension (univariate), two dimensions (bivariate), or multiple dimensions (multivariate).[1] Here, we primarily consider the case where each process ($\mathtt{X}$) is univariate for simplicity, though any of the variables in any of the measures could be trivially made bivariate or more generally multivariate.

In the context of brain dynamics, such a general process might refer abstractly to time-series measurement of the neural activity in a brain region, evolving over time. We can take repeated measurements of, or in other words, sample this process, which corresponds to the 'Applied' angle in Fig. 2.1.1. The sample path, known as a 'realization', of the time series is denoted $\mathtt{x}$ and comprises an ordered collection of individual empirical measurements $\{x_1, x_2, x_3, \ldots, x_T\}$. Within this empirically measured time series (realization), each $x_t \in X_t$ is a realization of the corresponding random variable. In other words, the empirical measurement at x_1 is a specific realization of the random variable X_1.

[1] In the multivariate case, for a process of dimension, m we would use vector notation for the random variable at time t: $\mathbf{X}_t = \{X_{t,1}, X_{t,2}, \ldots, X_{t,m}\}$.

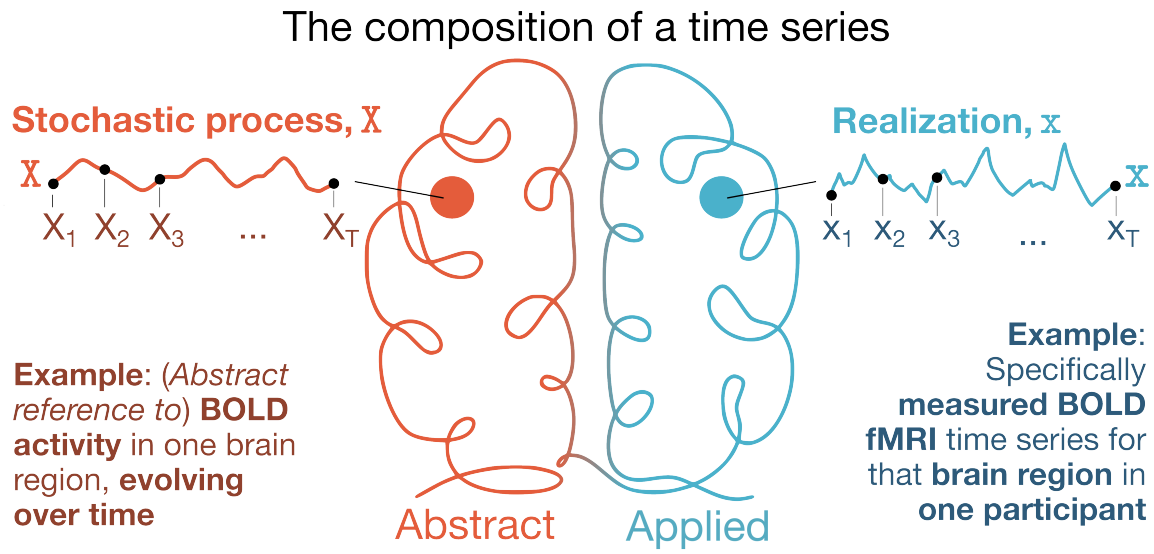


Figure 2.1.1: Decomposing the elements of a time series, in the context of a stochastic process and a specific realization. Here, we start from the more abstract concept of a stochastic process, $\mathtt{X}$, which is an ordered collection of random variables $\{X_1, X_2, X_3, \ldots, X_T\}$ where T is the total length of the process. For example, in the context of brain dynamics, the stochastic process $\mathtt{X}$ might refer abstractly to measurement of neural activity in a brain region, evolving over time. In parallel, on the applied end, a particular observation of our stochastic process is known as a realization, denoted $\mathtt{x}$. This realization is an ordered collection of individual empirical measurements $\{x_1, x_2, x_3, \ldots, x_T\}$. Continuing the above example, the realization $\mathtt{x}$ might reflect the measured blood oxygen level-dependent (BOLD) functional magnetic resonance imaging (fMRI) time series in the given brain region in one participant. *Brain line art image adapted from Adobe Stock Image #614193934 by user* ***Mikita Maryasau*** *(licensed under Adobe Education License).*

This distinction between the underlying stochastic process $\mathtt{X}$ and an individual empirical realization $\mathtt{x}$ is important, particularly when interpreting measured time series data. In most cases—for example, in human functional neuroimaging—we typically only have access to a single realization of a process, so we implicitly assume that this time series is ergodic. Here, 'ergodic' means that the realization ($\mathtt{x}$) will equiprobably traverse all parts of the space in which the process ($\mathtt{X}$) moves. In other words, over sufficient time, the time-averaged behavior of realization $\mathtt{x}$ will converge to the true ensemble-averaged behavior of the underlying process $\mathtt{X}$. Crucially, this allows us to treat the realizations $x_1, x_2, x_3, \ldots, x_T$ (of what would otherwise be distinct random variables) as multiple realizations x of a single, now time-independent random variable X associated with process $\mathtt{X}$. Thus, the ergodic assumption enables application of information-theoretic measures to the empirically observed time series variable X, since we can compute averages over the values of that observed time series (realization) as a computational method for estimating averages over the process distribution.

We define a 'discrete' random variable as one that exhibits a countable number of distinct values, while a 'continuous' random variable exhibits an uncountable number of possible values. Each outcome within the finite set for a discrete variable is associated with a probability mass, and a probability density function (PDF) links probabilities with these finite outcomes. We introduce and interpret the mathematical notation for each measure throughout this piece, which is generally predicated on the probability mass $p(x_t = a_i)$ for a discrete variable. This distribution captures the probability of obtaining the i-th value $x_t = a_i$ at time

t across all time points $\{1 \ldots T\}$. For a continuous variable, the probability mass of obtaining a value $b > x_t > a$ is an integral over the probability density function $f(x)$: $p(b > x_t > a) = \int_a^b f(x)dx$. We assume that the PDF is defined for the time series variable at hand—or variables, where the measure is multivariate. A maximum likelihood estimator (MLE) can be used to estimate the probability mass from observed empirical data for discrete variables.[2] As the information-theoretic measures we consider are generally constructed from sums and differences of entropies, their MLE follows directly from the PDF. We primarily consider discrete random values in conveying the conceptual properties of information-theoretic time-series measures, though we extend to continuous in our illustrative case study with resting-state brain activity in each section of this work.

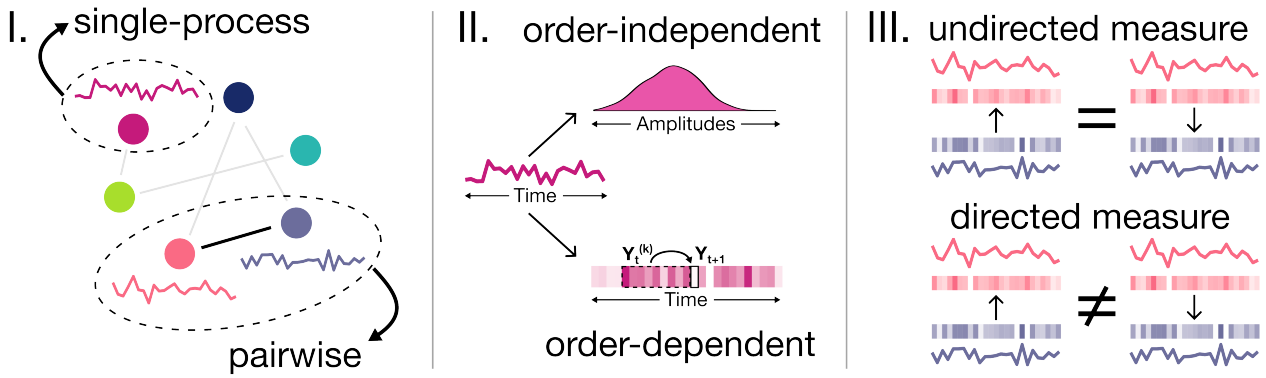


Figure 2.1.2: Three key characteristics distinguish information-theoretic time-series measures covered here. Each panel illustrates a conceptual characteristic that separates classes of information-theoretic time series measures covered in this review. **I.** Single-process versus pairwise: whether the measure describes properties of a single time series (single-process) or relationships between two time series (pairwise). **II.** Order-independent versus order-dependent: whether temporal ordering is taken into account, distinguishing measures based on static distributions (order-independent) from those sensitive to dynamics and/or time-lagged structure (order-dependent). **III.** Undirected versus directed: for the case of pairwise measures specifically, whether the measure captures symmetric relationships (undirected) or directional influence (directed).

With these definitions in hand, we can categorize the existing literature of information-theoretic time-series measures into a sensible and interpretable 'taxonomy'. Our taxonomy is organized based on three key characteristics, as schematically depicted in Fig. 2.1.2 and detailed below:

I. Single-process versus pairwise: A single-process measure operates on a single process (e.g., the activity of a single brain region) while a pairwise measure captures dependencies between two processes (e.g., the activity in two brain regions). Higher-order interactions (computed on three or more processes) measures are out of scope for this paper.

II. Order-independent versus order-dependent: An order-independent measure operates solely on the distribution of time series values without considering the order in which the underlying values occurred. By contrast, an order-dependent measure is sensitive to the temporal order in which the values occurred.

III. Undirected versus directed (pairwise measures only): An undirected measure captures equivalent de-

[2]More sophisticated estimators exist (e.g., based on a Bayesian framework) but remain out of scope for this review.

pendence (i.e., symmetric) between processes X and Y. By contrast, a directed measure is computed specifically from X to Y, such that the value from Y to X is not necessarily symmetric.

2.2 A SCHEMATIC OVERVIEW OF KEY MEASURES

In this work, we present a unified schematic and tabular overview of eleven core information-theoretic measures used in time-series analysis. Collectively, this synthesis provides direct links between conceptual theory and the implementation and interpretation to empirical data in an applied setting. By juxtaposing widely used measures like mutual information or transfer entropy with lesser-known yet powerful tools, like stochastic interaction or directed information, we introduce a schematic that reveals a compact landscape of how information flows can be quantified across time and between signals. We implement categorization according to the three core characteristics summarized in Fig. 2.1.2, reflecting a distilled structure underlying much of the literature—a structure that may otherwise be obscured by disciplinary silos or differences in naming conventions. From these three characteristics, we identified six broad groups of measures:

1. Single-process, order-independent measures;
2. Pairwise order-independent measures, undirected;
3. Pairwise order-independent measures, directed;
4. Single-process, order-dependent measures;
5. Pairwise order-dependent measures, undirected; and
6. Pairwise order-dependent measures, directed.

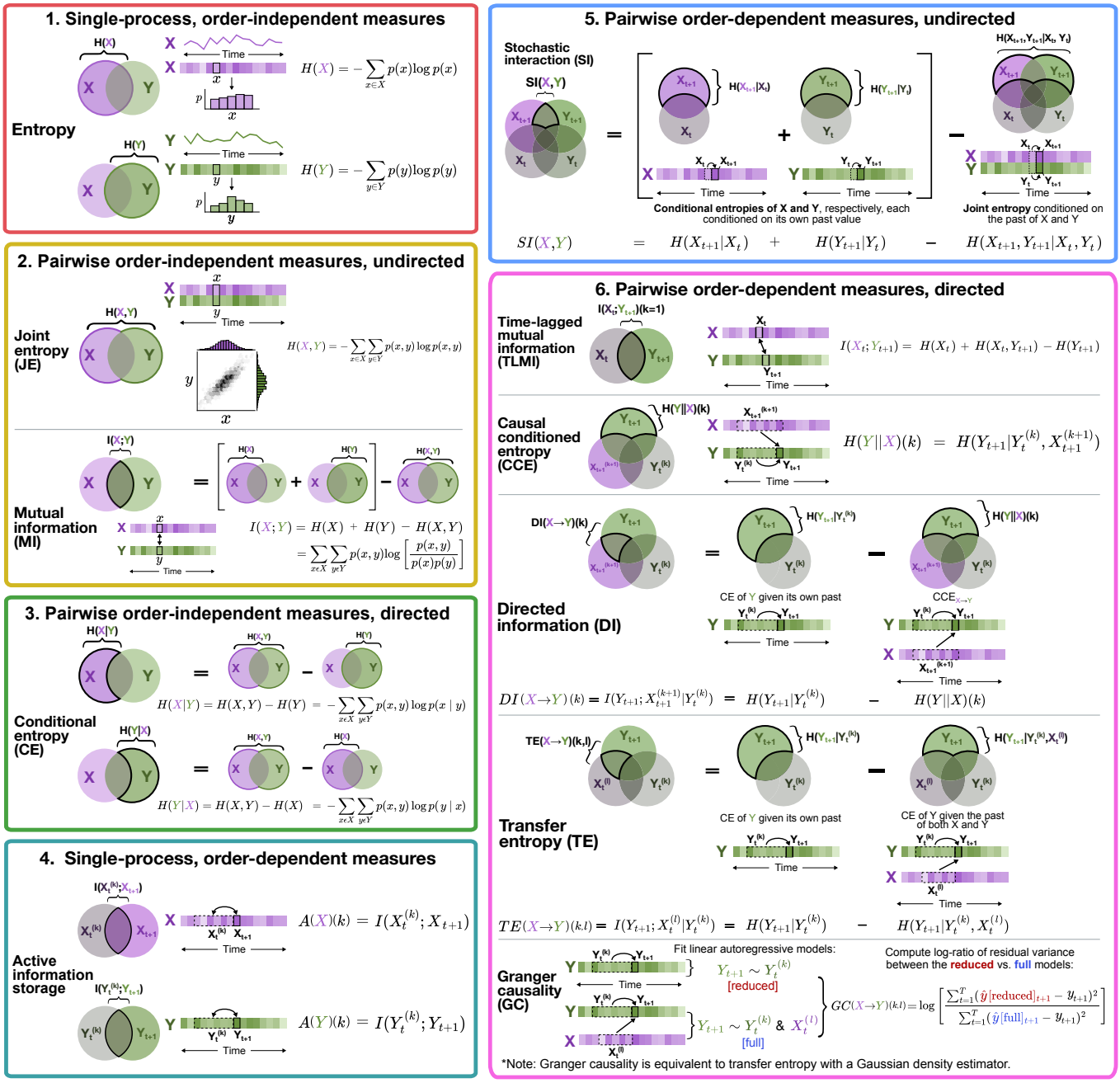


Figure 2.2.1: A unified visual representation of eleven information-theoretic time-series measures, grouped into six categories. Here, we depict the six broad categories of information-theoretic time series measures (single-process and pairwise) in colored sub-panels, labeled from 1 through 6: (1) single-process, order-independent; (2) pairwise order-independent, undirected (noting that joint entropy of two univariate processes is algorithmically equivalent to a bivariate extension of entropy in the case of a bivariate process); (3) pairwise order-independent, directed; (4) single-process, order-dependent; (5) pairwise order-dependent, undirected; and (6) pairwise order-dependent, directed. Within each sub-panel (e.g., '1. Single-process, order-dependent measures'), all of the individual reviewed measure(s) in that category are visually depicted along with representative mathematical formulae. Venn diagrams are used to illustrate entropy-based concepts for the processes X (shaded in purple) and Y (shaded in green). For the order-dependent measures in sub-panels 4 through 6, we use lighter and less-saturated Venn diagram elements to indicate the past history of a time series, while lighter and more saturated elements indicate the present. The eleven reviewed measures are presented in the same order as in Table 2.2.1, which contains a high-level summary for each measure along with the section number in which each measure is described in detail.

We organized these six categories and their constituent measures within color-coded sub-panels in Fig. 2.2.1. For example, '1. Single-process, order-independent measures' only includes one measure (entropy), while '6. Pairwise order-dependent measures, directed' includes five measures. Our aim with this figure is to consolidate a diffuse and disjoint literature into one cohesive visual representation, directly comparing formulations that condition on past and present values of a single process or pair of processes in slightly different ways. We use the popular Venn Diagram schematic (see e.g., Cover and Thomas [146, p. 48]) to illustrate concepts of marginal entropy; joint entropy and mutual information; and conditional entropy, which collectively form the building blocks of more complex pairwise dependency relationships such as stochastic interaction, transfer entropy, and directed information. We note that Venn diagrams provide useful intuition on how sums and differences of these measures are related, though it may be problematic when used for more than three variables since the analogy to set theory is loose, and where the (presumably positive) areas on the diagrams may be associated with negative quantities ([146, p. 75][147, p. 143][158]). With that said, all of the measures considered here (and their areas on the Venn diagrams) are non-negative, and emphasize their purpose to guide intuition on sums and differences of the quantities.

To complement the visual representation in Fig. 2.2.1, we have summarized key aspects of each information-theoretic measure covered in this paper in Table 2.2.1 below. This table consolidates information about the directionality, temporal dependence, and scope (single-process vs. pairwise) for each measure. Importantly, the figure and table invite the reader to view these measures not as isolated techniques, but as variations on a common theme: the reduction of uncertainty about some variable (e.g., Y) by conditioning on others (such as X, or the past of Y, or both). Information theory provides a powerful language for this task through its flexibility in capturing both linear and nonlinear dependencies independently of the generative form of the underlying data. In practice, there are only so many 'sensible' ways to quantify coupling and/or memory in time series data—and Fig. 2.2.1 makes these explicit, highlighting the conceptual logic behind each approach. The goal here is to scaffold a birds-eye view that not only clarifies distinctions, but also helps the reader see the broader unity in this diversity of methods. Of note, all information-theoretic measures covered here can be computed from empirical data using a combination of two open-source software packages: the *Java Information Dynamics Toolkit* (JIDT) [159] and the *python toolkit of statistics for pairwise interactions* (pyspi) [92, 160].

Measure	Notation	Direction	Order sensitivity	Plan-language summary of what measure captures
Entropy	$H(X)$	N/A	Order-independent	How surprising (i.e., uncertain) the samples of process X are on average.
Joint entropy	$H(X, Y)$	Undirected	Order-independent	How surprising (i.e., uncertain) the paired observations of two processes, X and Y, are on average. Note that joint entropy of two univariate processes is algorithmically equivalent to a bivariate extension of entropy in the case of a bivariate process.
Mutual information	$I(X; Y)$	Undirected	Order-independent	The amount of information that samples of process (X) provide about those of another (Y), and vice versa.
Conditional entropy	$H(Y \mid X)$	Directed	Order-independent	The uncertainty remaining for the observations in one process (Y) after simultaneously observing another (X).
Active information storage	$A(X)$	N/A	Order-dependent	The amount of information provided by the past values of process X about its present value.
Stochastic interaction	$\mathrm{SI}(X, Y)$	Undirected	Order-dependent	How much more uncertain the future values of X and Y become when they are treated as independent ('disconnected') sources of dynamics, rather than as a single coupled system ('fully interconnected').
Time-lagged mutual information	$I(X_n; Y_{n+1})$	Directed	Order-dependent	The amount of information shared between the present value of one process (Y) and a past or time-lagged instance of another (X).
Causally conditioned entropy	$H(Y \| X)$	Directed	Order-dependent	The amount of uncertainty remaining in the present value of one process (Y) after observing its own past as well as the past and present of another (X).
Directed information	$\mathrm{DI}(X \to Y)$	Directed	Order-dependent	The amount of information provided by the past and present of a source process (X) about the present of a target process (Y), beyond that which is already explained by the past of Y on its own.
Transfer entropy	$\mathrm{TE}(X \to Y)$	Directed	Order-dependent	The amount of information provided by the past of a source process (X) about the present of a target process (Y), beyond that which is already explained by the past of Y on its own.
Granger causality	$\mathrm{GC}(X \to Y)$	Directed	Order-dependent	The additional predictive power that the past of a source process (X) contributes to the present value of a target process (Y), beyond that of a linear model fit on the past of Y on its own. While GC is not explicitly an information-theoretic measure, it is equivalent to TE computed with a Gaussian density estimator and is therefore included within the scope of this review.

Table 2.2.1: A high-level overview of the eleven information-theoretic time-series measures covered in this review. For each of the eleven information-theoretic time-series measures, this table summarizes their notation, directionality (for pairwise measures), time-ordering sensitivity, what the measure captures, and the section in which the measure is covered in detail. Measure names are colored with the same palette as in Fig. 2.2.1 to denote in which of the six broad categories each measure sits.

2.3 ILLUSTRATIVE FUNCTIONAL NEUROIMAGING CASE STUDY

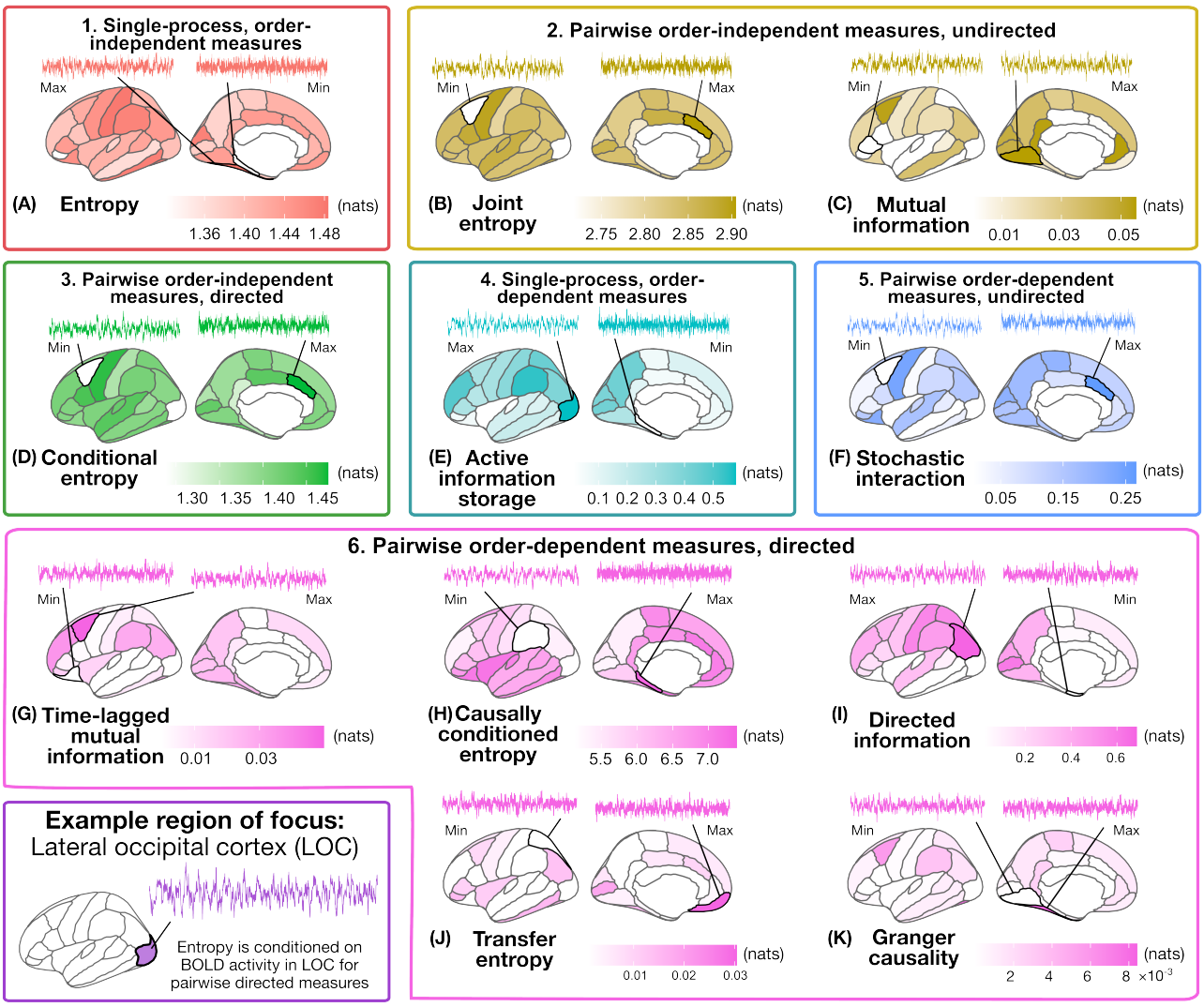


Figure 2.3.1: Visualizing information-theoretic time-series measures on the cortical surface. To provide a concrete demonstration of the application of information-theoretic time-series statistics, we computed the eleven measures from resting-state BOLD fMRI data from one example participant in the HCP S1200 release [161]. For each of the eleven information-theoretic measures, belonging to six categories in total (see Fig. 2.2.1 and Table 2.2.1), the resulting measure values are projected onto the left cortical surface of the brain. In the case of the two single-process measures, entropy and active information storage, the values are computed and shown for each individual region. For the nine pairwise measures, we set the lateral occipital cortex (LOC) as the seed region (shown in purple at the bottom left) and computed the pairwise measures to each of the other 33 regions in the left cortex; the values shown for each brain region reflect the value from the lateral occipital cortex to the corresponding region for the indicated measure. For all measures, we highlight the region with the minimum and maximum value and depict the corresponding fMRI time series.

As an illustrative case study, we focus on functional neuroimaging—comprised of multivariate time series recordings of the activity in various brain areas—to provide context for the implementation and interpretation

of each measure covered in the present work. Indeed, there are myriad applications for information theory in neuroscience across both recording scales and modalities, including neuronal spike trains [162, 163], local field potentials [164], and macroscale functional magnetic resonance imaging (fMRI) dynamics [165, 166]. We include an illustrative example of how and why each measure might be used to map information structure, examining the resting-state blood oxygen level-dependent (BOLD) fMRI time series from one participant in the Human Connectome Project [161] S1200 release (ID '298051', selected at random). For each of the eleven information-theoretic time-series measures depicted in Fig. 2.2.1 and summarized in Table 2.2.1, the computed values are projected onto the left hemisphere cortical surface in Fig. 2.3.1, with one brain map per measure. We maintain the same color-coding and nomenclature as in Fig. 2.2.1 and Fig. 2.3.1 for consistency. The spatial distributions in Fig. 2.3.1 provide intuitive examples of the kinds of signal dynamics each measure captures. This demonstration is purely to showcase the computation and interpretation of the eleven information-theoretic measures covered in this review—rather than drawing any biological conclusions based on a single-participant exemplar. All fMRI data were preprocessed as per earlier work [167] and further described in the following section.

2.4 METHODS FOR EMPIRICAL NEUROIMAGING DATA

For the empirical neuroimaging case study, we focused on resting-state fMRI (rs-fMRI) data from one randomly selected participant (ID '298051') from the S1200 release of the Human Connectome Project [161] (specifically, this participant is included in the 100 unrelated individuals subset). Of note, all imaging data included in the present work were already preprocessed in previous work [167, 168]. All resting-state volumes in this dataset were acquired using a gradient-echo, echo planar image (EPI) sequence with the following parameters: TR/TE = 720/33.1 ms, slice thickness = 2.0 mm, 72 slices, 2.0 mm isotropic voxels, frames per run = 1,200. All data were previously preprocessed in Fallon et al. [167], with a brief overview provided as follows. The rs-fMRI time series were linearly detrended, and the mean signal from white matter, cerebrospinal fluid, and global signal was regressed out. Low-frequency fluctuations ($f < 8 \times 10^{-3}$ Hz) were removed using a high-pass filter, applied to the EPI data using a fast Fourier transform. The average voxelwise time series was summarized per parcel using the Desikan-Killiany 68-region atlas [169], noting that we focused on the left hemisphere in this work (i.e., 34 cortical regions) for simplicity. Information-theoretic measures were computed from the time series using the *JIDT* [159] library, either directly (for the single-process measures) or embedded within the *pyspi* [92] toolbox. We applied continuous estimators to the time series, opting for the Kozachenko–Leonenko (`kozachenko`) for measures that are constructed from the univariate or joint entropy—which uses a nearest-neighbor approach—and usually the Kraskov–Stögbauer–Grassberger (`kraskov`) estimator for the other measures, which combines estimates based on mutual information with 4 nearest neighbors.

2.5 CONCLUSIONS AND FUTURE DIRECTIONS

Collectively, we have presented a unified and accessible guide to these eleven key measures for analyzing time series data, with a particular focus on their applications in computational neuroscience. Through a novel schematic overview and a systematic tabular summary, we have sought to clarify conceptual distinctions between measures—such as directionality, temporal dependence, and dimensional scope—and to

bridge terminology across disparate literatures. Our aim has been not only to consolidate fundamental definitions, but also to offer an integrative perspective that enables researchers with a wide range of background knowledge depths to navigate this methodological landscape with greater fluency. While our examples and emphasis here have focused on neural time-series data, we note that all measures are broadly applicable across diverse scientific domains. Resources like the UEA/UCR Time Series Classification repository [170, 171] provide compelling examples of diverse time-series datasets—from beef spectrograms to yoga pose angles to goose vocalizations—that could benefit from this conceptual framework to yield new insights into how information-theoretic measures relate to each other when computed on diverse empirical data. Each measure in this guide offers a unique lens on dynamics such as uncertainty, dependence, and predictability. Developing an intuitive grasp of how these measures interrelate and differ lays the foundation for richer, more integrative analyses of complex systems like the brain. We hope this guide serves as a starting point, upon which future work may extend this foundation by incorporating additional measures, refining estimation techniques, or exploring emerging applications.

3
Unifying pairwise interactions in complex dynamics

Fun, communication, connection.
Ryan Howard, The Office, Season 7 Episode 9

In Chapter 2, we introduced the mathematical formulation and biological interpretation of diverse information-theoretic time-series measures for functional neuroimaging data. Several of these measures were implemented in the *pyspi* (for '**Py**thon toolkit of **S**tatistics of **P**airwise **I**nteractions') toolkit, which we now introduce in this chapter—comprising the novel implementation of over 200 statistics of pairwise interactions (SPIs) in a unified algorithmic implementation. To focus on my direct contributions to the corresponding published manuscript as second author, this thesis chapter is comprised of excerpts pertaining to the second part of the project, in which I identified and analyzed two distinct functional neuroimaging case studies to systematically evaluate the ability of different pairwise interaction measures to meaningfully distinguish between classes in empirical data. These two neuroimaging case studies correspond to electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), respectively, both of which are usually analyzed with a single measure chosen subjectively and/or heuristically based on literature precedent. However, this chapter demonstrates that novel measures of pairwise interactions can outperform conventional measures for a given dataset, including information-theoretic measures introduced in Chapter 2, such as causally conditioned entropy and directed information. The full manuscript is included as the Appendix 2 of this thesis, which provides a broad review and empirical similarity analysis across over 200 SPIs, introduced across various theoretical traditions, to develop a thorough ontology of pairwise interaction statistics in complex systems.

This chapter comprises excerpts from work published as: Cliff, O. M., Bryant, A. G., Lizier, J. T., Tsuchiya, N.,& Fulcher, B. D. Unifying pairwise interactions in complex dynamics. *Nature Computational Science* (2023), doi: 10.1038/s43588-023-00519-x.

Abstract

Scientists have developed hundreds of techniques to measure the interactions between pairs of processes in complex systems. But these computational methods—from contemporaneous correlation coefficients to causal inference methods—define and formulate interactions differently, using distinct quantitative theories that remain largely disconnected. Here we introduce a large assembled library of 237 statistics of pairwise interactions[1] and assess their behavior on 1053 multivariate time series from a wide range of real-world and model-generated systems. Our analysis highlights commonalities between disparate mathematical formulations of interactions, providing a unified picture of a rich interdisciplinary literature. Using three real-world case studies, we then show that simultaneously leveraging diverse methods can uncover those most suitable for addressing a given problem, facilitating interpretable understanding of the conceptual formulations of pairwise dependence that drive successful performance. Our results and accompanying software enable comprehensive analysis of time-series interactions by drawing on decades of diverse methodological contributions.

3.1 INTRODUCTION

A fundamental question in science is how complex dynamics can be characterized by measuring the interactions within a distributed system. To address this question, many approaches have been developed to measure different types of pairwise interactions from dynamical data. For example, in neuroimaging, functional connections between pairs of brain regions are quantified through statistical correlations, which mark changes in human behaviors [172] and differ in neurological diseases [173]. In Earth systems science, pairwise causal models have been used to infer mechanistic drivers of natural processes, from the influence of sea-surface temperature on sardine and anchovy populations [23] to the atmospheric drivers of air circulation [22]. And economic analysts have studied the cointegration of paired non-stationary time series, such as stock-market indices and their associated future contracts, to infer a significant coupling for building econometric models [21].

3.2 RESULTS

LEVERAGING DIVERSE METHODS TO ADDRESS SCIENTIFIC PROBLEMS

Our results above illustrate the rich diversity of scientific methods for quantifying pairwise interactions. This diversity suggests that, when quantifying pairwise interactions for a given application, there is potential to compare across the scientific literature of statistics of pairwise interactions (SPI)s to: (i) select the best-performing SPI in an unbiased, data-driven way; and (ii) leverage a synergistic combination of multiple complementary SPIs to better capture complex underlying interactions in MTS. Here we provide a simple demonstration of this strategy (referred to as 'highly comparative' due to the broad methodological comparison involved [89, 174]) to three MTS classification problems using three open datasets: (1) *Smartwatch activity* dataset (Fig. 3.2.1A), where the aim is to classify one of four behavioral states (walking, running, resting, or playing badminton) from six-sensor smartwatch MTS (a 3-axis accelerometer and 3-axis gyro-

[1] The *pyspi* library is actively maintained and continues to grow in scope, with 284 measures as of June 2025.

scope) [175]; (2) *EEG state* dataset (Fig. 3.2.1D), where the aim is to distinguish positive versus negative slow cortical potentials from single-subject electroencephalogram (EEG) data (originally used to move a cursor up or down on a computer screen) [171, 176]; and (3) *fMRI film* dataset (Fig. 3.2.1G), where the aim is to classify resting and film-watching conditions from functional magnetic resonance imaging (fMRI) data [177] (see *Methods* for additional details on each dataset). To investigate the performance of different SPIs on these tasks, we represented each MTS as a set of features corresponding to all pairwise interactions between its constituent processes and compared their classification performance using a linear support vector machine (SVM) with cross validation (see *Methods* for details).

We observed a wide range of SPI performance in each case study, ranging from null performance up to high and statistically significant performance: 27–92% accuracy on the smartwatch activity dataset (Fig. 3.2.1B), 48–69% accuracy on the EEG state dataset (Fig. 3.2.1E), and 41–95% accuracy on the fMRI film dataset (Fig. 3.2.1H). Many SPIs displayed significant classification performance on each dataset (permutation test, Bonferroni-corrected $p < 0.05$): 213 SPIs (smartwatch activity), 165 SPIs (EEG state), and 67 SPIs (fMRI film). This wide range of observed SPI performance on all three datasets demonstrates the crucial importance of selecting an SPI that is able to capture the relevant types of interactions underlying a given dataset.

To understand the types of interactions that characterize the labeled classes of MTS, we analyzed and interpreted the highest-performing SPIs on each dataset (full results are in Supplementary Data 1, Sheet 3)[2]. To provide a simple demonstration of this process, here we focus on the top-performing individual SPI in each case study. In the smartwatch activity dataset, the top-performing SPI was causally conditioned entropy (CCE) using a Kozachenko–Leonenko estimator (92% accuracy; SPI label `cce_kozachenko`, see Supplementary Section 1.4.7 for details). Its high performance is driven in part by its ability to capture the coupling from the wrist's z-rotation to its x-translation, which differs strongly between running and walking, as shown in Fig. 3.2.1C. This tells us that elbow movement in the x-direction is more informative of subsequent z-rotation (as the wrist rotates outward in the z-direction) in walking than in running.

[2]Supplementary data sheets are not included as part of this Thesis, but can be accessed online in the electronic version of the manuscript.

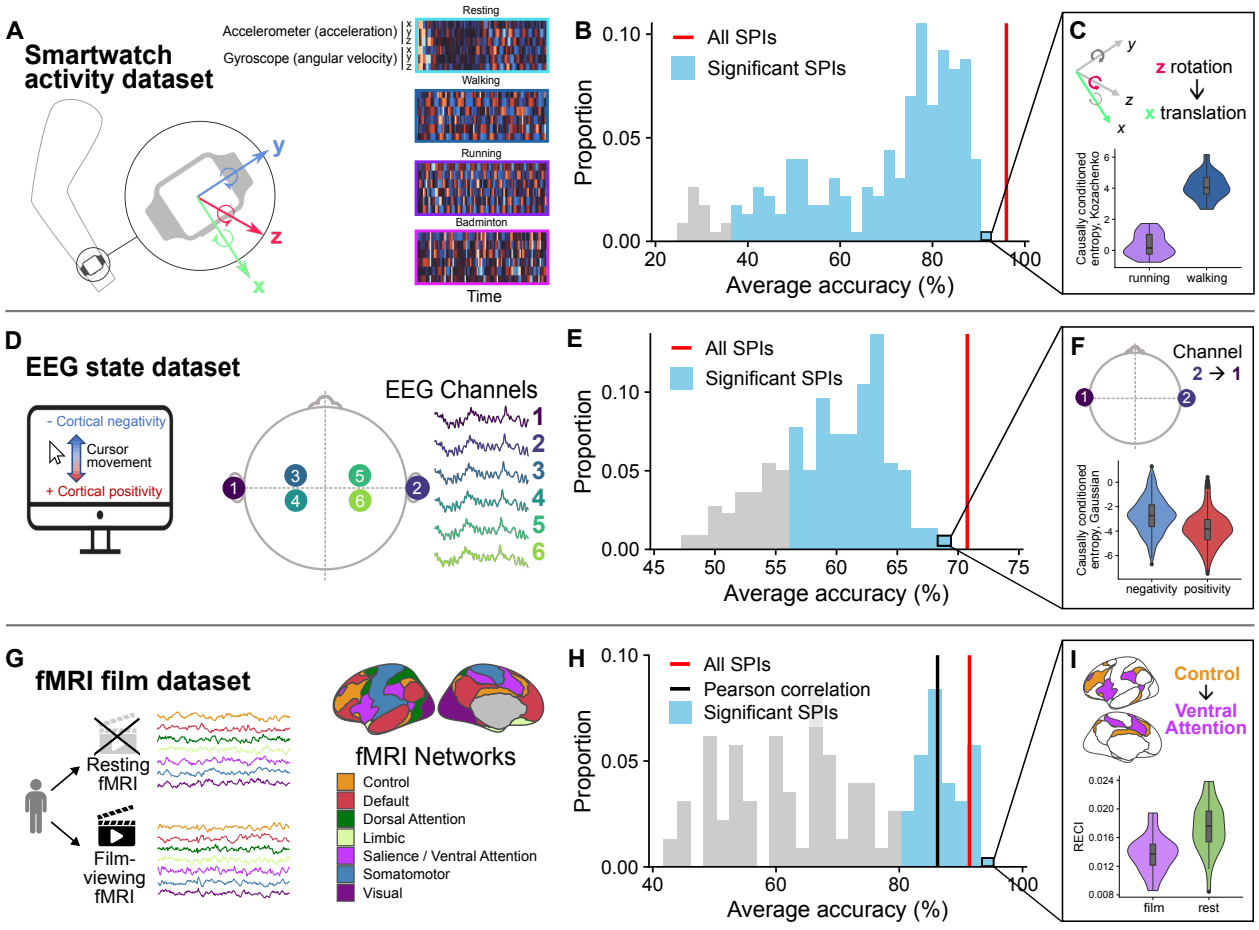


Figure 3.2.1: A comprehensive library of SPIs can be used to accurately classify and understand differences in human movement and neural activity datasets. **A**. In the smartwatch activity dataset, we aimed to determine which of four activities a participant is performing ('resting', 'walking', 'running', or 'badminton') from smartwatch accelerometer recordings. An example MTS from each of the four classes is shown as a process $\times$ time heat map. **B**. Distribution of average classification accuracy (over train–test resamples) across 228 SPIs, where all pairwise interactions are used as the basis for classification for each SPI. The 213 SPIs with significant accuracy are shaded blue (one-sided permutation test, Bonferroni-corrected $p < 0.05$). Concatenating feature vectors from all SPIs into a single classification model yields an average accuracy of 96% (red). **C**. A violin plot showing the distribution of the top-performing SPI—causally conditioned entropy with a Kozachenko–Leonenko density estimator—between the wrist's x-axis translation and z-axis rotation, indicating a stronger interaction during walking (mean $\pm$ SD $= 4.1 \pm 0.8$, $N = 20$) than running (0.4 ± 0.7, $N = 20$). In each violin plot in this figure, the annotated box plot shows the box center and extent as the median, 25th, and 75th percentiles; whiskers extend for 1.5 $\times$ interquartile range and outliers are shown as black dots. **D**. In the EEG state dataset, we aimed to classify positive versus negative cortical activity states using data from six EEG channels, for which the electrode locations and example time series are depicted. **E**. Distribution of average classification accuracy across all 219 SPIs, including the 165 significant SPIs (shaded blue), and combination of all SPIs (71% accuracy, red). **F**. The top-performing SPI, a causally conditioned entropy, is visualized from EEG Channels 2 to 1 in cortical negativity (-2.7 ± 1.4, $N = 282$) versus positivity (-3.8 ± 1.4, $N = 279$). **G**. In the fMRI film dataset, we aimed to classify rest versus film-viewing using fMRI time series from seven brain networks. **H**. Distribution of classification accuracy across all 227 SPIs, including the 67 significant SPIs (shaded blue) and annotating the performance of Pearson correlation coefficient (86%, black) and the combined set of all 227 SPIs (91%, red) for comparison. **I**. The top-performing SPI, regression error-based causal inference, RECI, is shown from the control to ventral attention networks in film-viewing (0.014 ± 0.003, $N = 29$) and rest (0.018 ± 0.004, $N = 29$) conditions.

The top-performing SPI for the EEG state dataset was, as above, a causally conditioned entropy, but using a Gaussian density estimator (69%; SPI label `cce_gaussian`, see Supplementary Section 1.4.7 for details). Its performance was driven in part by its ability to capture the increased coupling from EEG Channel 2 to 1 (from near the right ear to near the left ear) in cortical negativity versus positivity states, as shown in Fig. 3.2.1F. Other statistics designed to capture directed information flow (in a way that includes instantaneous interactions in the presence of feedback) also performed well on this task, including directed information with a Gaussian density estimator (69% accuracy; SPI label `di_gaussian`, see Supplementary Section 1.4.8). While prior applications of SPIs derived from directed information theory are limited [178, 179], our highly comparative analysis suggests them as high-performing methods for measuring EEG coupling alongside other novel candidates for further investigation, including the direct directed transfer function evaluated over high frequencies (67%; SPI label `ddtf_multitaper_mean_fs-1_fmin-0-25_fmax-0-5`, see Supplementary Section 1.5.9) and the Hilbert–Schmidt Independence Criterion (66%; SPI label `hsic`, see Supplementary Section 1.2.5). Of the classical methods for quantifying EEG connectivity, some are recapitulated as high performers by our data-driven analysis, including mean directed coherence across various frequency bands [180] (all 67% accuracy; see Supplementary Section 1.5.10), while others exhibited surprisingly low accuracy, such as algorithmic variants of partial directed coherence [181] (between 55% and 64%).

Finally, for the fMRI film dataset, the top-performing SPI was regression error-based causal inference (95% accuracy; SPI label `reci`, see Supplementary Section 1.3.4 for details). Its high performance is driven in part by its ability to capture the stronger coupling from the control network to the ventral attention network during film-watching compared to rest (Fig. 3.2.1I). The dominant way of measuring coupling in whole-brain fMRI is to use the Pearson correlation coefficient [53, 182] (annotated in Fig. 3.2.1H), which exhibits strong and statistically significant classification accuracy on this problem (86%). However, our data-driven approach highlights 30 alternative SPIs with higher performance (88% to 95%, listed in Supplementary Data 1, Sheet 3). These high-performing methods include alternative types of covariance (e.g., minimum covariance determinant, 93% accuracy; SPI label `cov_MinCovDet`, see Supplementary Section 1.1.1) and precision estimates (e.g., using Ledoit–Wolf shrinkage, 91%; SPI label `prec_LedoitWolf`, see Supplementary Section 1.1.2) that better deal with non-Gaussian bivariate distributions. Other high performers include information-theoretic SPIs (like conditional entropy, joint entropy, and mutual information using Gaussian density estimators, all 91% accuracy, see Supplementary Sections 1.4.1–1.4.3) and directed SPIs which distinguish asymmetric coupling (e.g., the top-performing regression error-based causal inference, `reci`). Compared to the typically subjective process of selecting an appropriate method to analyze a given dataset, the highly comparative approach demonstrated here highlights the most useful scientific methods automatically, facilitating interpretable understanding of the conceptual formulations of pairwise dependence that drive successful performance.

Since different types of systems involve different types of interactions between measured processes, we expected different SPIs to perform well across the three datasets. Indeed, we found that an SPI with high performance on one problem does not imply its high performance on other problems. For example, some SPIs performed well only on a single dataset, like dynamic time warping with an Itakura parallelogram (SPI label `dtw_constraint-itakura`, see Supplementary Section 1.2.7) which was a top performer on the fMRI film dataset (91%), but showed weaker performance on the smartwatch activity dataset (78%)

and null performance on the EEG state dataset (53%). But some SPIs did perform strongly across all three datasets, such as the cross distance correlation (SPI label `dcorrx_maxlag-10`, see Supplementary Section 1.2.2), which ranked among the top 10 SPIs for all three problems (90%, 66%, and 91%, for the smartwatch activity, EEG state, and fMRI film datasets, respectively). Moreover, different algorithmic variants of causally conditioned entropy were top performers for the smartwatch activity (`cce_kozachenko`, 92%), EEG state (`cce_gaussian`, 69%), and fMRI film (`cce_kernel_W-0.5`, 90%) datasets (although we note a strong dependence of the density estimation approach on CCE performance). We also found that grouping SPIs based on the fourteen data-driven modules (identified by their similarity of behavior on real data, Fig. 2) better captured their relative performance on these tasks than the six literature categories (from Fig. 1), as shown in Supplementary Fig.S3, suggesting our modular representation as a useful one for understanding differential SPI performance on a given task[3].

Relative to investigating individual SPIs one at a time, we finally aimed to investigate the value of drawing on multiple SPIs simultaneously. We developed a combined representation of the pairwise dependence structures captured by all SPIs, allowing us to simultaneously represent each MTS using a large and diverse set of pairwise dependency structures (through feature concatenation, as described in *Methods*: 'Classification'). Although this approach represents each MTS in a much higher-dimensional space than the individual SPI representation analyzed above (with associated challenges for robust classifier fitting), we expected it to outperform the individual best SPI on datasets involving multiple types of interactions, such that simultaneously leveraging multiple SPIs provides complementary and useful information about class differences. Relative to the top-performing individual SPI, our combined SPI approach improved classification accuracy on the smartwatch activity dataset (to 96%) and the EEG state dataset (to 71%), shown as vertical red lines in Figs 3.2.1B and E. On the fMRI film dataset, it yielded slightly lower performance (91%) than the top individual SPI, `reci` (95%), suggesting that the associated interactions are well-captured by a single, well-chosen SPI on this dataset (i.e., multiple SPIs do not provide an advantage sufficient to overcome the challenges of fitting a classifier in a higher-dimensional space). By simultaneously drawing on a wide range of SPIs, the simple statistical approach demonstrated here can quantify multiple complementary types of interactions from MTS data (and is likely to yield improved accuracy through optimization, see *Discussion*).

3.3 DISCUSSION

As different time-varying systems contain different statistical relationships between their elements—e.g., that capture instantaneous or time-delayed responses, linear or nonlinear interactions, condition on the past, allow for non-constant delays, or infer directional coupling—our results demonstrate that a highly comparative approach can be used to detect the most informative types of SPIs for capturing the relevant types of interactions underlying a given dataset. Unlike many machine-learning approaches to MTS classification, which are challenging to interpret [175], this approach connects scientists to interpretable theory that shapes understanding of the most important types of pairwise interactions in a dataset, following recent undertakings to unify interdisciplinary literatures on summary statistics ('features') for univariate time series [89, 90, 183] and complex networks [174]. This broad, comprehensive methodological comparison stands in contrast to a more conventional approach in which the data analyst manually selects a method, a practice

[3]The three figures referenced in this paragraph are included in Appendix 2.

that leaves open the possibility that alternative methods may provide clearer interpretation, better performance, or computational efficiencies. The three classification case studies analyzed here provide a simple demonstration of the procedure, automatically highlighting high-performing SPIs from across the literature and providing an interpretable understanding of the relevant interactions in each dataset. We observed a wide range of performance in all cases, highlighting the importance of careful selection of SPIs for a given application. In the EEG state dataset, for example, our analysis flagged high-performing SPIs consistent with common methodological practice in EEG analysis (thus recapitulating existing domain knowledge) and others with surprisingly poor performance, as well as flagging novel high-performing SPIs as promising candidates for methodological innovation in the field. Future work investigating which types of SPIs are best suited to which types of problems will yield new insights into the interactions underlying different complex systems, and is likely to uncover additional novel applications of SPIs to new problems.

3.4 METHODS: CLASSIFICATION CASE STUDIES

3.4.1 DATASETS

Smartwatch activity dataset.

The smartwatch activity dataset is derived from the `BasicMotions` problem in the University of East Anglia (UEA) MTS classification repository [171]. Each MTS includes six sensors—a 3-axis accelerometer and a 3-axis gyroscope—recorded for 10 s at 10 Hz, yielding 1000 time points. There are 20 MTS in each of four classes (resting, walking, running, or playing badminton), for a total of 80 MTS in the dataset. This dataset was recently analyzed in a large MTS classification challenge [175] in which the data was split 50/50 into training and test subsets with 30 stratified repeats. The baseline classifier (based on dynamic time warping) achieved 95.25% accuracy; the best algorithm (HIVE–COTE) achieved 100% accuracy [175].

EEG state dataset.

The EEG state dataset corresponds to the `SelfRegulationSCP1` problem in the UEA MTS classification repository [171] and was originally published in Birbaumer et al. [176]. Electrical activity was measured from six EEG channels in one participant as they were instructed to move a cursor up or down on a computer screen by generating negative or positive slow cortical potentials, respectively. The physical placement of these EEG electrodes is depicted in Fig. 3.2.1D. Cortical activity (measured in μV) was recorded for 3.5 s at 256 Hz, yielding 896 time points in each channel per MTS, with 282 negativity trials and 279 positivity trials. This dataset was analyzed in Ruiz et al. [175], assessing accuracy using 30 stratified train–test splits of 268 training and 293 test samples. In the original comparison, the baseline DTW algorithm achieved an accuracy of 81.81% and the best algorithm (TapNet) achieved 95.68% [175].

fMRI film dataset.

The fMRI film dataset is derived from a functional connectome fingerprinting study examining individual signatures of cortical activity in N=29 individuals at rest or while watching a film [177]. In this dataset, blood-oxygen-level-dependent (BOLD) signals were recorded in 114 parcellated cortical regions with a repetition time (TR) of 813 ms (a sampling rate of 1.23 Hz) for either 1200 frames (resting) or between 952–1000

frames (film-watching). Each trial type (rest versus film-watching) was repeated four times per participant. Here we analyzed pre-processed data obtained from Betzel et al. [184]. For consistency and simplicity, we examined the first rest and first film-watching session per participant; across all participants and trials, only the first 947 frames contain real data, so we restricted our analysis to these time ranges. To retain a comparable number of processes as the first two classification case studies, we averaged BOLD signals from the 114 original brain regions into the seven functional networks from Yeo et al. [185], as depicted in Fig. 3.2.1G. We compare the performance of the Pearson correlation coefficient, used to construct functional connectivity matrices in the original publication, Byrge and Kennedy [177], to our library of SPIs.

3.4.2 CLASSIFICATION

For all three case studies, our simple approach to SPI-based classification involved computing the matrix of pairwise interactions (MPI)—6×6 for the smartwatch activity and EEG state datasets, and 7×7 for the fMRI film dataset—for each z-scored MTS and repeating for each SPI. We then used the elements of these matrices as features for a linear support vector machine (SVM) classifier. Note that we used the most recent version of *pyspi* (v0.4.0) to compute SPIs for the classification case studies, which included some improved implementations of some SPIs. Features were extracted from each MPI differently for directed and undirected SPIs (see Supplementary Data 1, Sheet 1): for undirected SPIs (for which the corresponding MPI is symmetric), we used the upper triangular entries as features, whereas for directed SPIs, we used all non-diagonal elements as features. As a preprocessing step, for each case study, we removed any SPI that had invalid entries (due to numerical issues) in any of the MPIs, or gave constant results across all MTS (see Supplementary Data 1, Sheet 2 for a list of omitted SPIs). This yielded a set of 228 SPIs for the smartwatch activity problem, 219 SPIs for the EEG state problem, and 227 SPIs for the fMRI film problem. For the analysis involving combining all SPIs into a single classifier, this yielded a total of 4755 features for the smartwatch activity dataset, 4659 features for the EEG state dataset, and 6743 features for the fMRI film dataset.

The linear SVM was implemented using default settings from Python's `scikit-learn` [186] as part of a classification pipeline that involved z-score feature normalization (fitted on training data and applied to unseen test data). The very simple methodological choices made here allowed us to focus on demonstrating the key conceptual types of analyses made possible by drawing on a diverse set of SPIs, aiding transparency while acknowledging that more complicated statistical methodologies are likely to improve the classification performance quoted here. We implemented 30 class-stratified train–test splits for cross-validation with the same proportions implemented in [175] using the `StratifiedShuffleSplit` function for the smartwatch activity and EEG state datasets. Since there are N=29 individuals in the fMRI film dataset, we implemented leave-one-individual-out cross-validation, such that each classifier was trained with the rest and film scans of N=28 participants and tested on the rest and film fMRI scans of the left-out participant.

We measured classification performance using total accuracy for all three case studies. Statistical significance was estimated using a permutation testing approach whereby 100 null models were fitted (using randomly shuffled class labels) and evaluated using the same cross-validation classification procedure described above. The observed classification performance for each SPI was then compared with the pooled null distribution of all SPIs, yielding p-values that were then adjusted for multiple comparisons by controlling

the family-wise error at 0.05 using the method of Bonferroni [187]. The performance metric for the union of all SPIs was similarly compared with its corresponding null permutation distribution to yield a single p-value per classification problem.

4

A data-driven approach to identifying and evaluating connectivity-based neural correlates of consciousness

This scented candle . . . represents the eternal burning of competition, or something.
Jim Halpert, The Office, Season 2 Episode 3

In Chapters 2 and 3, we introduced the theoretical motivation and practical implementation for comparative time-series analysis of the brain as a complex system. In this chapter, we expand upon this methodological framework by applying it to one of the greatest open challenges in modern neuroscience [188]: quantitatively identifying and evaluating plausible neural correlates of consciousness that can map between subjective experience and measurable neural activity. This chapter was inspired by a global competition[1] to identify a candidate measure(s) of functional connectivity that could meaningfully distinguish between competing theories of consciousness: Integrated Information Theory (IIT) [189, 190] and Global Neuronal Workspace Theory (GNWT) [191, 192]. The competition was organized by the COGITATE Consortium after mediating an adversarial collaboration between theorists from both camps, using pre-registered hypotheses and a shared experimental paradigm [150]. Here, we review our first prize-winning entry to the COGITATE Consortium's challenge, presented at the BIOMAG 2024 Conference in Sydney, Australia, and subsequent extension to a full-length manuscript. This chapter introduces our systematic analysis of the Consortium's openly-shared magnetoencephalography (MEG) dataset using 246 functional connectivity measures (from the *pyspi* software, introduced in Chapter 3) between regions predicted to underlie conscious vision by IIT and/or GNWT.

This chapter is in submission as: Bryant, A. G., & Whyte, C. J. A data-driven approach to identifying and evaluating connectivity-based neural correlates of consciousness. *bioRxiv* preprint (2025), doi: 10.1101/2025.04.06.646695.

[1] https://www.arc-cogitate.com/biomag-2024

Abstract

Identifying the neural correlates of consciousness remains a major challenge in neuroscience, requiring theories that bridge between subjective experience and measurable neural correlates. However, theoretical interpretation of empirical evidence is often post hoc and susceptible to confirmation bias. Building upon the adversarial collaboration mediated by the COGITATE Consortium, we present a generalizable approach for the data-driven identification, evaluation, and theoretical modeling of connectivity-based neural correlates of consciousness. Using the same magnetoencephalography (MEG) dataset and accompanying pre-registered hypotheses from the COGITATE Consortium, we systematically compared 246 functional connectivity (FC) measures between regions predicted to underlie conscious vision by Integrated Information Theory (IIT) and/or Global Neuronal Workspace Theory (GNWT). We identified a family of FC measures based on the barycenter—tracking the 'center of mass' between two signals—as the top-performing stimulus decoding measures that generalize across regions central to predictions of both IIT and GNWT. To interpret these findings within a theoretical framework, we developed neural mass models that recapitulate the neural dynamics hypothesized to underlie conscious perception by each theory. Comparing simulated barycenter values from these models against empirically measured MEG data revealed that the GNWT-based model, featuring delayed ignition dynamics, better captured observed connectivity patterns than the IIT-based model, which relied on highly synchronized sensory dynamics. Beyond dataset-specific conclusions and limitations, we introduce a framework for systematically identifying and testing candidate neural correlates of consciousness in an unbiased and interpretable manner.

4.1 INTRODUCTION

Characterizing the neural correlates of consciousness remains one of the great open challenges in modern neuroscience. At the heart of this challenge is the search for a theory of consciousness capable of mapping between subjective experiences and their observable neural correlates [188]. Although the field sees an increasing number of publications each year, theoretical interpretation is primarily performed posthoc; moreover, design choices of the methodology itself are highly predictive of which theory the study's findings will support [193]. These biases have cultivated a fragmented literature that lacks a cohesive methodology, in terms of both experimental paradigms and theoretical interpretation.

In a collective endeavor to move beyond current theoretical silos and confirmation bias, researchers from competing theoretical camps have jointly entered into adversarial collaborations that directly test their pre-registered predictions in shared empirical paradigms, conducted by theory-neutral third party experimental groups. This adversarial collaboration, led by the COGITATE Consortium, focused on the predictions of two prominent theories of consciousness: integrated information theory (IIT) and global neuronal workspace theory (GNWT). IIT identifies consciousness with the emergent cause–effect structure of physical systems [190]. According to IIT, a system is conscious to the degree that the system as a whole exhibits

greater informational content than the sum of information contained within all sub-partitions of the system. In contrast, GNWT [191, 192] identifies consciousness with the availability of information in the 'global workspace' [194], a recurrent network with long-range excitatory projections connecting prefrontal cortex (PFC) and parietal cortices.

Within the domain of conscious vision in particular, IIT and GNWT yield contrasting predictions regarding stimulus decodability, the location and temporal dynamics of conscious vision, the role of pre-stimulus activity, and the nature of inter-areal coupling. IIT postulates that consciousness is a persistent structure, with conscious perception requiring phase synchronization between low-level visual areas and category-selective visual areas (lateral occipital cortex [195] and fusiform face area [196]). Together with regions in the parietal cortex, this network is often referred to as the 'posterior hot zone'—and notably does not include the PFC, which IIT predicts is only necessary for cognitive aspects of consciousness, such as introspection and self-report [197]. By contrast, GNWT posits that stimulus information becomes conscious when it enters the global workspace through a late wave of neural signal propagation from sensory cortices to PFC. This wave comprises a nonlinear ignition event, independent of task demands, through which the global workspace broadcasts information between otherwise isolated processes throughout the brain. In the experiments conducted by the COGITATE Consortium [198], the results of which were recently made available [150], these predictions were tested through a shared visual perception paradigm. Specifically, participants were shown a stream of images belonging to one of four categories. Across multiple functional neuroimaging modalities, the Consortium reported varying degrees of evidence supporting these two theories based on region-specific decoding and the timing of conscious percepts. However, neither theory's predictions about the relationship between inter-areal coupling (assessed via functional connectivity) and stimulus decodability were borne out—raising the question of whether there exists a functional connectivity (FC) measure that can discriminate between the neural correlates of visual experience predicted by IIT and/or GNWT.

Of note, the Consortium only examined two measures of inter-areal connectivity: pairwise phase consistency [199] and dynamic FC ([200]; using the Pearson correlation coefficient) between respective pairs of theory-predicted brain regions. Many other measures of inter-areal coupling could, in principle, distinguish between the predictions of each theory in the relevant region–region pairs. For example, in previous work, we systematically compared diverse coupling types—for example, synchronous vs. asynchronous, directed vs. undirected, and time vs. frequency domain-based [79, 92]. The results validated commonly used measures like the Pearson correlation coefficient across neuroimaging modalities, while simultaneously highlighting the potential utility of previously unexplored types of measures of pairwise interaction like directed information. Moreover, the theories put forward by both IIT and GNWT were intrinsically qualitative, making it difficult to precisely quantify the evidence for either theory's predictions.

In this work, we address these limitations in the quantitative testing of competing theories by presenting a

unified and data-driven approach for identifying, evaluating, and theoretically interpreting candidate neural correlates of consciousness. Based on the aims of the COGITATE Consortium [150] pertaining to inter-areal coupling, we present a highly comparative approach to search for inter-areal coupling (i.e., FC) measures that reliably quantify the neural correlates of conscious vision from magnetoencephalography (MEG) imaging data. This comprehensive suite of 246 pairwise coupling measures [92] includes those typically employed in functional neuroimaging, including the Pearson correlation coefficient, pairwise phase consistency, power envelop correlation [201, 202], and ϕ^* [203]—enabling direct comparison to under-explored alternatives. To interpret the FC results within the theoretical frameworks of IIT and GNWT, respectively, we then develop two neural mass models [204]— implementing the stimulus-evoked dynamics hypothesized by each theory to be necessary for conscious perception. Crucially, implementing the hypothesized dynamics for each brain region in computational models allows us to explicitly compute exemplar time series and quantitatively compare the predictions of each theory based on the top-performing FC measures. In sum, we present a data-driven approach for identifying the FC measures that maximally encode stimulus-specific information, and then implement simple and interpretable neural models to derive theoretical insight from the data-driven FC measures.

4.2 MATERIALS AND METHODS

4.2.1 NEUROIMAGING DATA ACQUISITION AND TASK PARADIGM

All data was obtained from the COGITATE Consortium [150] as prepared for a functional connectivity data science challenge (cf. https://www.arc-cogitate.com/biomag-2024) [205]. MEG data were downloaded in BIDS format for 100 participants, N=94 of which had the requisite anatomical scans to perform source localization. This subset of N=94 participants included N=54 females (23.1 $\pm$ 3.5 years) and N=40 males (22.3 $\pm$ 3.6 years) with no known neurological or psychological conditions and no age differences between males and females (Wilcox rank-sum test, P=0.19). All imaging acquisition and behavioral task details for this dataset are provided in Consortium et al. [150]. Briefly, participants viewed supra-threshold visual stimuli comprised of faces, objects, letters, or false fonts, for variable durations (0.5 s, 1.0 s, or 1.5 s) at different rotation angles. Participants were instructed to press a button when either of two specific stimuli were presented, such that stimulus events were classified as 'Relevant target' (i.e., the correct stimulus to respond to), 'Relevant non-target' (i.e., all other stimuli belonging to the same category as one of the target stimuli), or 'Irrelevant' (i.e., a stimulus belonging to a different category to both target stimuli). We included all stimulus types in our analysis, though we excluded the 'Relevant target' setting due to potential motor confounds. Additionally, for simplicity, we focused exclusively on the 1.0 s ('intermediate') stimulus duration to compare neural dynamics during one second of stimulus presentation ('onset') and the subsequent post-stimulus second ('offset'). Participants were conscious of the presented stimuli in both task-relevant and irrelevant trials, and they demonstrated high hit rates and low false alarm rates, eliminating behavioral

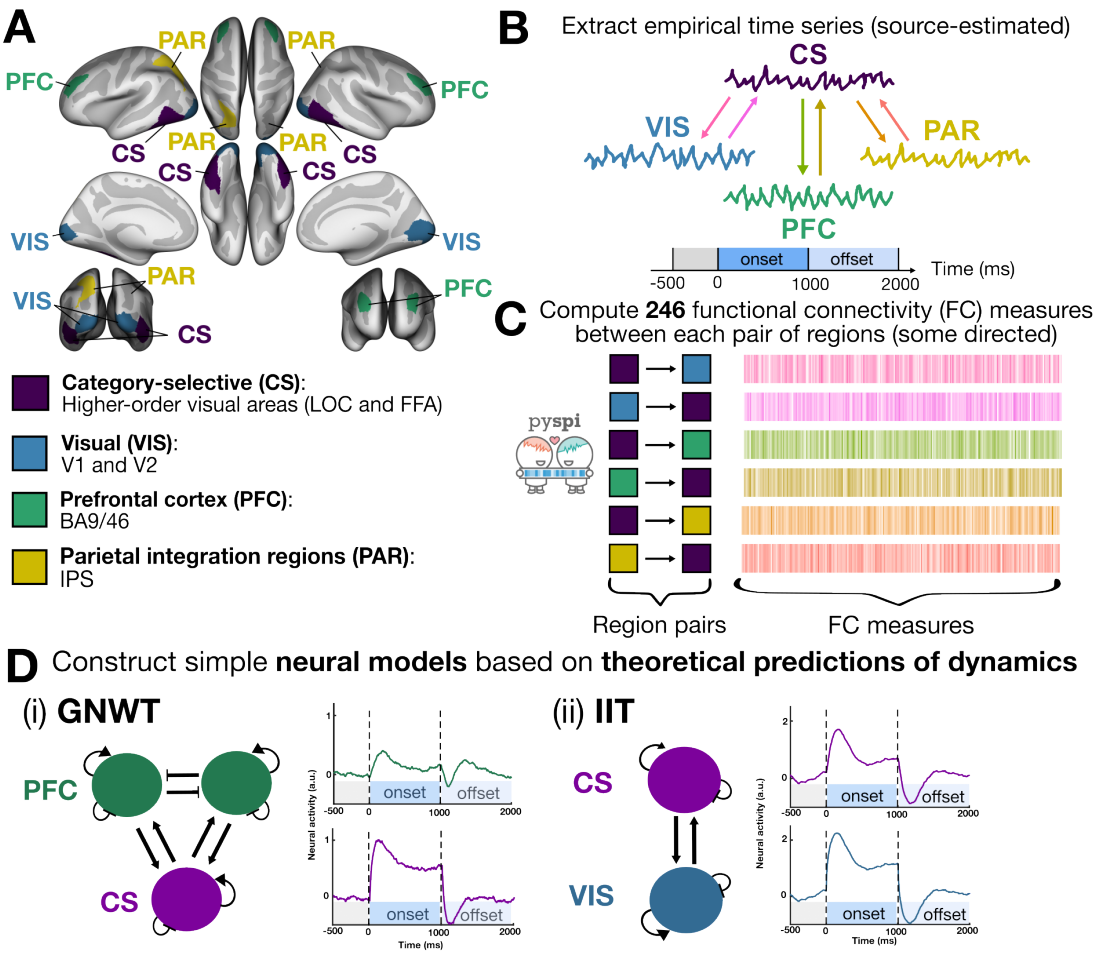


Figure 4.2.1: In this study, we systematically compared different ways of quantifying functional coupling from MEG time series derived from regions hypothesized to support the IIT and/or GNWT theories of visual perception and consciousness. A. We selected four regions of interest that are depicted on the inflated cortical surface in `fsaverage` space: Category-selective (CS), comprised of the higher-order visual lateral occipital cortex (LOC) and fusiform face area (FFA) regions; Visual (VIS), comprised of V1 and V2; Prefrontal cortex (PFC), comprised of Brodmann areas 9 and 46; and Parietal integration regions (PAR), comprised of the intraparietal sulcus (IPS). **B.** We extracted the average source-localized MEG time series across all epochs per event-related field type in each region, considering both stimulus onset (first 1000ms) and the subsequent offset period (second 1000ms). **C.** For each functional pathway (i.e., each directed region–region pair), we computed 246 statistics of pairwise interactions (SPIs) using the *pyspi* library. **D.** Neural mass models designed to recapitulate the dynamics of each theoretically relevant region based upon the preregistered predictions of GNWT and IIT set out in Consortium et al. [150]. The dynamics of the GNWT-based model are governed by two key factors: (1) winner-take-all competition between the two PFC neural masses leading to transient 'ignition' events at stimulus onset and offset; and (2) PFC masses have a slower intrinsic timescale leading to delayed peak in PFC activity. The dynamics of the IIT-based model are likewise governed by two key factors: (1) strong recurrent connectivity between CS and VIS masses; and (2) similar intrinsic timescales, which together yield highly synchronous activity at stimulus onset that is maintained for the duration of the stimulus presentation period which then rapidly and synchronously decays at stimulus offset.

accuracy as a potential confound [150].

4.2.2 NEUROIMAGING PREPROCESSING

Anatomical preprocessing

High-resolution T1-weighted structural volumes were preprocessed with the same methods and accompanying code outlined in Consortium et al. [150]. Briefly, cortical surfaces were reconstructed using FreeSurfer's `recon-all` pipeline (cf. https://surfer.nmr.mgh.harvard.edu/). Scalp surfaces were also reconstructed and combined with an inner skull surface to create a single-shell boundary elements model (using the *MNE* module in Python [206]), from which the solution was used to create a volumetric forward model spanning the full brain volume with a 5 mm grid. We selected four brain regions of interest that are listed in Table 4.2.1 and are depicted schematically in a brain map in Fig. 4.2.1A. To delineate these brain areas in subject space, we resampled the Schaefer et al. [207] 100-parcel atlas (mapping to the 7-Network atlas from Yeo et al. [185]) to each participant's reconstructed cortical surfaces using the `mri_surf2surf` function from FreeSurfer.

MEG preprocessing

MEG data were preprocessed using the same methods previously employed by the COGITATE Consortium for MEG data [150, 208], which included signal-space separation with a Maxwell filter to minimize environment artifacts [209] and FAST-ICA to remove cardiac muscle artifacts [210]. Data were partitioned into 3.5 s epochs, comprised of 1 s pre-stimulus to 2.5 s post-stimulus onset, and epochs were filtered according to gradiometer, magnetometer, and/or artifact properties as described in Consortium et al. [150]. Source modeling was performed using the dynamic statistical parametric mapping method [211]. A common spatial filter was generated from the sum of a noise covariance matrix for the baseline period (-0.5 to 0 s) and a covariance matrix for the active time window period (0 to 2 s). The baseline covariance matrix was used to spatially pre-whiten the data to combine magnetometer and gradiometer data. For each epoch, we fit and applied an inverse solution to extract the time series from the relevant regions (cf. Sec 4.2.2), averaging across all channels within each parcel. To enhance the signal-to-noise ratio [212], we averaged across trials for each event-related field type (i.e., for each stimulus type and task relevance condition) to yield one average time series per brain region for the given trial type in each participant. As we focused on the 1.0 s stimulus duration, we compared the 1 s window of stimulus presentation ('onset') with the following 1 s post-stimulus period ('offset') for each trial type.

4.2.3 DATA-DRIVEN QUANTIFICATION OF INTER-AREAL COUPLING

We sought to comprehensively quantify diverse types of functional coupling between our four regions. We employed the novel framework for highly comparative analysis of pairwise interactions, `pyspi` (v1.1.0) [92], computing a total of 262 FC measures from the MEG time series between each pair of regions (e.g., CS $\rightarrow$ PAR). As part of the standard `pyspi` workflow, all time series were z-scored along the time axis after which

Table 4.2.1: The individual brain regions from the 100-parcellation atlas from Schaefer et al. [207] comprising our regions of interest.

	Left hemisphere	**Right hemisphere**
Category-Selective (CS)	7Networks_LH_DorsAttn_Post_1-lh	7Networks_RH_Vis_3-rh
Primary visual (VIS)	7Networks_LH_Vis_4-lh	7Networks_RH_Vis_4-rh
	7Networks_LH_Vis_5-lh	7Networks_RH_Vis_5-rh
Prefrontal Cortex (PFC)	7Networks_LH_SalVentAttn_PFCl_1-lh	7Networks_RH_Cont_PFCl_3-rh
Parietal integration (PAR)	7Networks_LH_DorsAttn_Post_3-lh	7Networks_RH_DorsAttn_Post_4-rh

linear detrending was applied with the `detrend` function from the `scipy.signals` module [213]. Some of the FC measures are undirected, such that values are symmetric in both directions between region–region pairs, whereas others are directed (and therefore asymmetric). For quality control, we filtered the 262 FC measures down to those that yielded non-zero variance and real, non-NaN values across region–region pairs for at least 90% (i.e., N=85) of participants, yielding a final set of 246 FC measures. As described in Cliff et al. [92], these FC measures comprise diverse types of pairwise interactions, from basic covariance estimates to multi-metric distances to frequency-based coupling.

4.2.4 FITTING CLASSIFIERS

For our primary classification aim ('stimulus type'), we fit a separate model for each region–region pair, task relevance condition, and stimulus period ('on' during the first 1000 ms from stimulus onset and 'off' during the second 1000 ms). For our supplementary aims ('task relevance' and 'cross-task' classification), we still partitioned the data by region–region pair (e.g., VIS $\rightarrow$ CS) and stimulus presentation period ('on' or 'off') into separate classifiers. In all cases, for each FC measure (e.g., power envelope correlation), we fit a binary logistic regression classifier using *scikit-learn* (v1.6.1) [214] to predict either the stimulus type or task relevance from the associated FC measure values (from the given region–region pair). For the stimulus pair classification (primary aim) and task relevance classification (first supplementary aim), we evaluated out-of-sample performance using stratified 10-fold cross-validation with the `StratifiedGroupKFold` function from *scikit-learn*. This allowed us to train each model on a pair of stimuli (or pair of relevance conditions) from one subset of participants and evaluate the model on that pair of stimuli or relevance conditions from another (unseen) subset of participants. For the cross-task classification, each model was trained on the pair of stimuli from all N=94 participants in e.g., 'Relevant non-target' context, and tested on the same pair of stimuli from all N=94 participants in the 'Irrelevant' context (and vice versa). Classification performance was assessed as the mean accuracy across the test subsets per model and is presented throughout this report as the mean cross-validated accuracy $\pm$ 1 SD. As a robustness check, we also fit a linear support vector machine (SVM) classifier for each FC measure and region–pair and evaluated the cross-validated accuracy to confirm that similar performances were achieved with the logistic regression classifiers.

4.2.5 THEORY-DRIVEN NEURAL MODELING

To quantitatively link each theory with the empirically observed neural activity (measured with MEG), we constructed two simple neural mass/rate-based models [204, 215, 216] based on the neural activity predicted for each core region (PFC, CS, and VIS) defined by the COGITATE Consortium [150]. The dynamics of each model were governed by the following system of rate-based equations

$$\begin{aligned} \tau_r \frac{dr}{dt} &= -r + f(Wr + u_{stim} - \gamma a + b) \\ \tau_a \frac{da}{dt} &= -a + r, \end{aligned} \tag{4.1}$$

with state variables $r, a \in \mathbb{R}^{N\times 1}$ representing the aggregate (i.e., coarse-grained) neural activity and (hyperpolarizing) adaptation current (respectively), of each cortical region hypothesized to play a role in conscious perception (for similar models applied to the visual system, see Wilson [217]). For simplicity, the transfer function $f(x)$ is a rectified linear unit. The matrix $W \in \mathbb{R}^{N\times N}$ parameterized the strength of inter-areal and self(within-mass) connectivity; the visual stimulus was encoded in the one-hot vector $u^{stim} \in \mathbb{R}^{N\times 1}$ which, as in the experimental paradigm, lasted 1000 ms of simulation time; τ_r and τ_a are scalar time constants; γ is a scalar constant controlling the strength of the adaptation current; and b is a scalar bias parameter representing the background level of arousal.

Parameter values (supplied in 4.2.2) were manually selected to recapitulate—as precisely as possible—-the time series predictions for each region that were pre-registered by IIT and GNWT theorists [150]. Model simulations and analysis were implemented in custom python (version 3.9.21) code. We solved (4.1) numerically using an Euler–Maruyama integration scheme ($\Delta t = 1$ ms), adding zero mean Gaussian noise ($\sigma = 0.005$) to both the neural activity and adaptation current.

Table 4.2.2: Parameter values for the simple IIT- and GNWT-based neural mass models. Parameter values are in arbitrary units unless stated otherwise. These parameters were used construct models as schematically depicted in Fig. 4.2.1D.

	IIT	**GNWT**
W	$\begin{array}{c} \\ VIS \\ CS \end{array} \begin{array}{c} \begin{array}{cc} VIS & CS \end{array} \\ \begin{pmatrix} 0.1 & 0.75 \\ 0.75 & 0.1 \end{pmatrix} \end{array}$	$\begin{array}{c} \\ CS \\ PFC_1 \\ PFC_2 \end{array} \begin{array}{c} \begin{array}{ccc} CS & PFC_1 & PFC_2 \end{array} \\ \begin{pmatrix} 0.05 & 0.75 & 0.75 \\ 0.25 & 0.25 & -2.75 \\ 0.25 & -2.75 & 0.25 \end{pmatrix} \end{array}$
γ	0.5	1
τ_r	τ_{VIS} = 20, τ_{CS} = 40 (ms)	τ_{CS} = 40, τ_{PFC} = 80 (ms)
τ_a	400 (ms)	400 (ms)
b	0.85	0.85

GNWT predicts that conscious perception should be marked by an 'ignition' event shortly after stimulus onset, as information from high-level visual areas (i.e., CS) updates the contents of the global workspace in PFC through long-range recurrent projections. After this initial ignition event, activity in PFC returns close to baseline. At stimulus offset, when the contents of consciousness change again, GNWT predicts that there should be another 'ignition' event in PFC—updating the contents of the workspace. The GNWT-based model, therefore, consists of three strongly adapting neural masses, one representing CS and two representing the competitive interactions in PFC responsible for the update dynamics at stimulus onset and offset. The stimulus is delivered to CS, which then projects to the two PFC masses—which compete in a winner-take-all mode to inhibit one another. After stimulus onset, strong adaptation reduces the activity of the winning (i.e., dominant) PFC mass. At stimulus offset, the previously suppressed mass (which is not subject to activity-dependent adaptation) then supersedes and inhibits the dominant mass, generating the hypothesized ignition update dynamics in PFC (cf. Fig. 4.2.1D[i]). To generate a single PFC time series, analogous to what would be observed in the MEG data, we averaged the simulated time series of the two PFC masses.

In contrast, according to IIT, conscious contents are maintained in posterior cortical structures throughout the duration of a conscious event via short-range reciprocal projections between low-level visual cortices (i.e., VIS) and higher-level category-selective (CS) cortices generating synchronous activity between sensory regions, maintained for the duration stimulus presentation period (i.e. the duration of the visual experience). The IIT-based model, therefore, consists of two weakly adapting neural masses representing VIS and CS respectively, with strong reciprocal projections. The stimulus is delivered to VIS and is maintained in a high-activity attractor state jointly constituted by VIS and CS throughout stimulus presentation, as depicted schematically in Fig. 4.2.1D(ii).

We note that these models should not be regarded as neuronal implementations of IIT or GNWT. Rather, they are simple phenomenological models (i.e., quantitative summaries) of the neural dynamics hypothesized to underlie conscious perception by each theory. Such models allow us to derive quantitative predictions about inter-areal coupling by analyzing the simulated neural data using the FC measures identified in our data-driven analysis of MEG data. Specifically, we compared time-series features computed from the model-simulated time series with those computed from empirically observed MEG signals using two measures: (1) the Kullback–Leibler (K-L) divergence [218] and (2) the Wasserstein distance [219].

The K-L divergence, denoted $D_{\mathrm{KL}}(P \mid\mid Q)$, captures the amount of additional information (in units of bits) needed to reconstruct probability distribution P when it is approximated by another distribution Q.

$$D_{\mathrm{KL}}(P \mid\mid Q) = \sum_{x \in X} P(x) \log \left(\frac{P(x)}{Q(x)} \right), \tag{4.2}$$

where X is the set of possible outcomes, and $P(x)$ and $Q(x)$ represent the probabilities assigned to each

outcome x by distributions P and Q, respectively. Here, we use the K-L divergence to compute the information 'cost' of approximating the empirical distribution of FC values with simulated distributions derived from each model. Specifically, we compared empirical (P) and model-derived (Q) distributions of time-series features comprised of vectors of discrete observations. These discrete distributions were converted into 100-bin probability distributions with equal bin widths. K-L divergence was computed using custom code in python (version 3.9.21).

To complement the K-L divergence (which is not a formal distance metric and may be biased by outliers), we additionally computed the Wasserstein distance [219]—also known as the 'earth mover's distance'—which, in the discrete case, minimizes both the number of samples which must be moved and the distance those samples are moved for probability distributions P and Q to match (in units of total work). The Wasserstein distance was computed using the `wasserstein_distance` function from the *scipy* package (version 1.13.1) [213].

4.2.6 CODE AND DATA AVAILABILITY

All MEG data analyzed in this study is openly available upon registration at https://doi.org/10.17617/1.wqa3-wk71 [205]. All code needed to reproduce our analyses and visuals is freely available in our GitHub repository (https://github.com/anniegbryant/MEG_functional_connectivity) [220].

4.3 RESULTS

As shown in Fig. 4.2.1A, we focus here on regions predicted by one or both theories to mediate stimulus- and/or task-specific FC: category-selective visual areas (CS; common to both IIT and GNWT), early visual areas (VIS, comprised of V1 and V2; exclusive to IIT), and lateral prefrontal cortex (PFC, comprised of Brodmann areas 9 and 46; exclusive to GNWT). Due to the combinatorial explosion that can arise from highly comparative analysis, we included an additional parietal region (PAR, comprised of the intraparietal sulcus), as the parietal cortex plays an important role in consciousness in both theories (i.e., the parietal cortex is a part of the posterior substrate of consciousness according to IIT [190], and a part of the 'global workspace' according to GNWT [192]). By including PAR as an additional region, we gained a further constraint on high-performing functional connectivity measures without biasing our search toward either theory. The label names for each area in the Schaefer et al. [207] 100-parcellation atlas are listed in Table 4.2.1. We then comprehensively quantify MEG-derived FC between CS and each of the three other regions (taking the average time series across all epochs per event type and participant) using 246 bivariate time-series measures that span an interdisciplinary literature—from spectral properties to information-theoretic measures, using the `pyspi` ('Python toolkit of statistics for pairwise interactions') library [92]. Finally, we used the top-performing FC measures to analyze simulated time series from two theory-derived neural models, allowing us to interpret the empirical results in light of the pre-registered predictions of IIT and GNWT.

Based on the original predictions for inter-areal coupling delineated by the COGITATE Consortium [150], we defined one core classification aim and two supplementary classification aims. In our main analysis targeting the core classification aim, we evaluate whether any FC measure in `pyspi` could distinguish between stimulus types (e.g., 'face' vs. 'object') in either the 'Relevant non-target' and/or 'Irrelevant' settings, in any region–region pair and during either stimulus presentation ('on') or offset ('off'). Our secondary analyses focused on the supplementary classification aims to explore whether FC measures distinguish domain-independent task relevance ('Relevant' vs. 'Irrelevant') and generalize from one relevance type to another in the same stimulus comparison. In all cases, we employed logistic regression classifiers with a grouped cross-validation strategy, reporting the mean $\pm$ standard deviation (SD) out-of-sample accuracy (see Methods, Sec 5.4.3 for more details).

4.3.1 MULTIPLE PROPERTIES OF INTER-AREAL COUPLING CAN DISTINGUISH BETWEEN STIMULUS TYPES ACROSS REGION–REGION PAIRS

For the primary stimulus-decoding analysis, we first considered the overall range of classification accuracy values across all FC measures and experimental paradigm combinations (e.g., stimulus onset vs. offset, relevant non-target vs. irrelevant category). As shown in Fig. 4.3.1, many FC measures classified stimuli with 60-67$\%$ cross-validated accuracy across region–region pairs and category relevance settings. This range sits within the context of prior visual stimulus-decoding analyses based on inter-areal connectivity derived from electroencephalography (EEG) time series [221] or combined EEG and functional magnetic resonance imaging (fMRI) data [222]. The presence of relatively high-performing FC measures in all stimulus pairs suggests there may be conserved signatures of neural activity in the functional pathways of interest while processing different visual stimuli. However, we note the marked variance in classification performance exhibited among FC measures across both region–region pairs and stimulus pairs (as shown in Fig. 4.3.1)—underscoring the heterogeneity in FC patterns across individuals, region–region pairs, and stimulus types.

In the first supplementary classification aim, we probed the neural signatures of task relevance in a domain-general manner, testing whether any FC measure could distinguish 'Relevant non-target' versus 'Irrelevant' trials that combined all stimulus types. Surprisingly, none of the FC measures surpassed 56% accuracy with any region–region pair (cf. Fig. S1)[2], indicating that inter-areal coupling is less distinctive between domain-general task relevance settings. Our second supplementary classification aim was to test 'cross-task' transfer learning (as in Consortium et al. [150]), in which we trained each classifier on FC measure values for a given stimulus pair (e.g., 'face' vs. 'object') during the 'Relevant non-target' setting (i.e., training on all N=94 participants) and tested the classifier in the 'Irrelevant' setting—and vice versa. In line with the predictions of GNWT, cross-task classification yielded similar accuracy distributions across FC measures and region–region pairs (including CS $\leftrightarrow$ PFC) as in our primary stimulus type classification analysis (cf.

[2]All supplementary figures are included in Appendix 3.

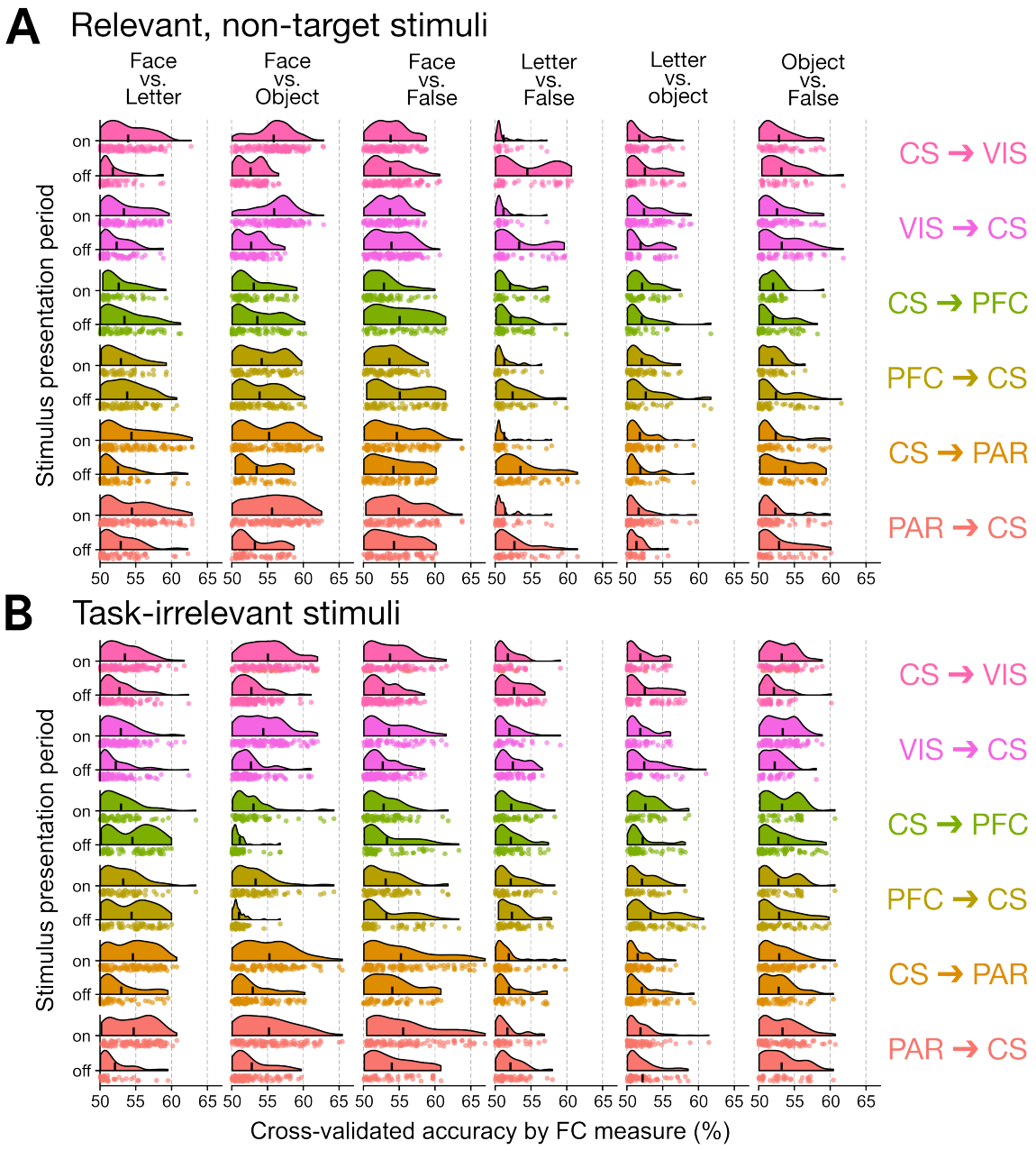


Figure 4.3.1: Many FC measures distinguish stimuli with $\geq 60\%$ accuracy across region–region pairs, particularly for face stimuli. A. Each raincloud plot shows the distribution of classification accuracy values for each SPI in the corresponding stimulus pair (column) and region–region pair (row), partitioned by relevant non-target stimulus presentation period ('on' vs. 'off'). Each dot represents one SPI, and the vertical lines represent the mean cross-validated accuracy across all FC measures in the corresponding distribution. The x-axis was truncated at 50% accuracy, such that FC measures yielding lower than 50% accuracy are not shown here. **B.** Same as in **A**, for task-irrelevant stimuli.

Fig. S2); therefore, we focus on the primary stimulus decoding classification analysis hereafter.

4.3.2 STATISTICS DERIVED FROM THE BARYCENTER BEST DISTINGUISH STIMULI, PARTICULARLY FACES

Having confirmed that multiple FC measures between the examined regions can distinguish between stimulus categories, we now zoom in to focus on the highest-performing measures. This allows us to make general inferences about the types of pairwise dynamics that contain the most information about visual stimulus categories. As shown in Fig. 4.3.2A, we identified the maximum classification accuracy across all FC measures, stimulus presentation periods ('on' or 'off'), and category relevance settings ('Relevant non-target' or 'Irrelevant') for each stimulus pair (e.g., 'Face' vs. 'Object') and region–region pair (e.g., VIS $\rightarrow$ CS). Across region–region pairs, the maximum classification accuracy values ranged between 62-67%, with all of the top values corresponding to face stimuli and another category. However, there is marked variability in classification performance across cross-validation folds (indicated by the shaded ribbons in Fig. 4.3.2A); therefore, while we report mean accuracy $\pm$ 1 SD, we draw inferences mainly from relative performance trends rather than absolute accuracy values.

To better understand the types of dynamics captured by the top-performing FC measure classes, we focused on the top two classification measures highlighted in Fig. 4.3.2A, corresponding to 'Face' vs. 'False font' in the CS $\leftrightarrow$ PAR and 'Face' vs. 'Object' in the CS $\leftrightarrow$ PFC regions, respectively (n.b., the bidirectional arrows indicate symmetric, undirected FC measure connections). In Fig. 4.3.2, we compare the feature distributions among all N=94 participants for the top-performing FC measure per region–region pair. Of note, coupling between the two region–region pairs maximally distinguished stimulus categories based on different geometric formulations of the barycenter—a property originally developed in astrophysics to measure the center of mass between two orbiting celestial bodies [223], which extends here to capture the center of the pair of time series using a given distance minimization metric [224]. For CS $\leftrightarrow$ PAR, the top-performing FC measure is the mean of the squared barycenter, derived from Euclidean geometry during stimulus onset in the task-irrelevant setting ($67.0 \pm 8.3\%$ accuracy). We schematically depict the computation of this FC measure using empirical CS and PAR time series from an example participant in Fig. 4.3.2B(i). The time average of the squared Euclidean barycenter was nearly twice as high during 'False font' stimulus processing (0.59) than during 'Face' stimulus processing (0.30) for this example participant, which reflects the greater MEG fluctuation synchrony between the CS and PAR regions throughout the 1000 ms 'False font' stimulus period than during the 'Face' period.

In a similar vein, the maximum (over time) of the squared barycenter derived using dynamic time warping (DTW) averaging [225] yielded peak 'Face' vs. 'Object' classification performance for the CS $\leftrightarrow$ PFC connection ($64.3 \pm 9.4\%$ accuracy). We depict the computation of this FC measure using the empirical CS and PFC time series from an example participant in Fig. 4.3.2B(ii). Unlike in Euclidean space (as depicted in 4.3.2B[i]), DTW geometry permits dilations and/or shifts across time for distance minimization between the two time series. The maximum of the squared DTW-based barycenter provides insight into the variability in

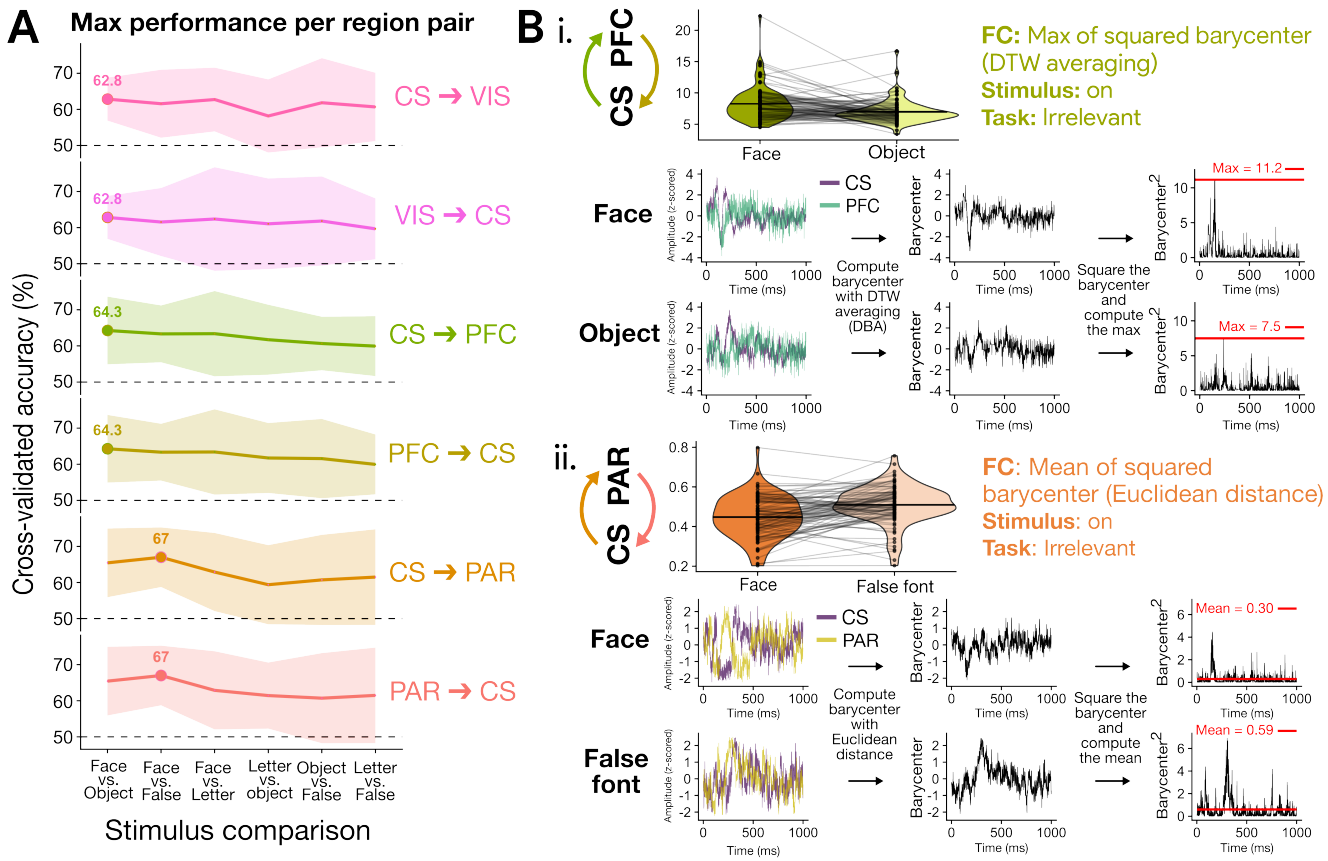


Figure 4.3.2: Examining the top-performing SPI per stimulus comparison highlights strong performance of barycenter statistics. A. For each region–region pair (rows) and stimulus comparison (columns), we plot the mean cross-validated classification accuracy for the top-performing SPI (lines), ± 1 SD (shaded ribbons) across test folds, out of all 246 FC measures, relevance conditions ('Relevant non-target' or 'Irrelevant'), and stimulus presentation periods ('on' or 'off'). The peak classification accuracy is annotated per region–region pair. **B.** (i) The top-performing SPI for CS ↔ PFC—the maximum of the squared barycenter with DTW averaging—is schematically depicted. The violin plots show the SPI value distributions for task-irrelevant 'face' versus 'object' trials during stimulus onset across all N=94 participants. Below, line plots schematically depict the computation of this SPI from 'face' or 'object' trial-averaged MEG time series in an example participant. (ii) As in (i), the top-performing SPI for CS ↔ PAR—the mean of the squared barycenter with Euclidean geometry—is schematically depicted. The violin plots show the SPI value distributions for task-irrelevant 'face' versus 'false font' trials during stimulus onset across all N=94 participants. Below, line plots schematically depict the computation of this SPI from 'face' or 'false font' trial-averaged MEG time series in an example participant (n.b., a different participant than in (i)).

the underlying regions' local activity, as periods of higher MEG amplitude will be more heavily weighted in the temporal alignment and therefore exert greater influence on the barycenter. Indeed, the squared DTW-based barycenter peaked at a magnitude of 11.2 during 'Face' stimuli—attributable to a sharp downward spike in both CS and PFC MEG amplitude at approximately 150 ms that is archetypal for face processing [226, 227]—compared to 7.5 during 'Object' stimuli, in which MEG amplitudes were less variable and fluctuated on a slower timescale between the two regions.

4.3.3 BARYCENTER STATISTICS DOMINATE THE TOP STIMULUS DECODING RESULTS, ALONG WITH BASIC COVARIANCE AND SPECTRAL PROPERTIES

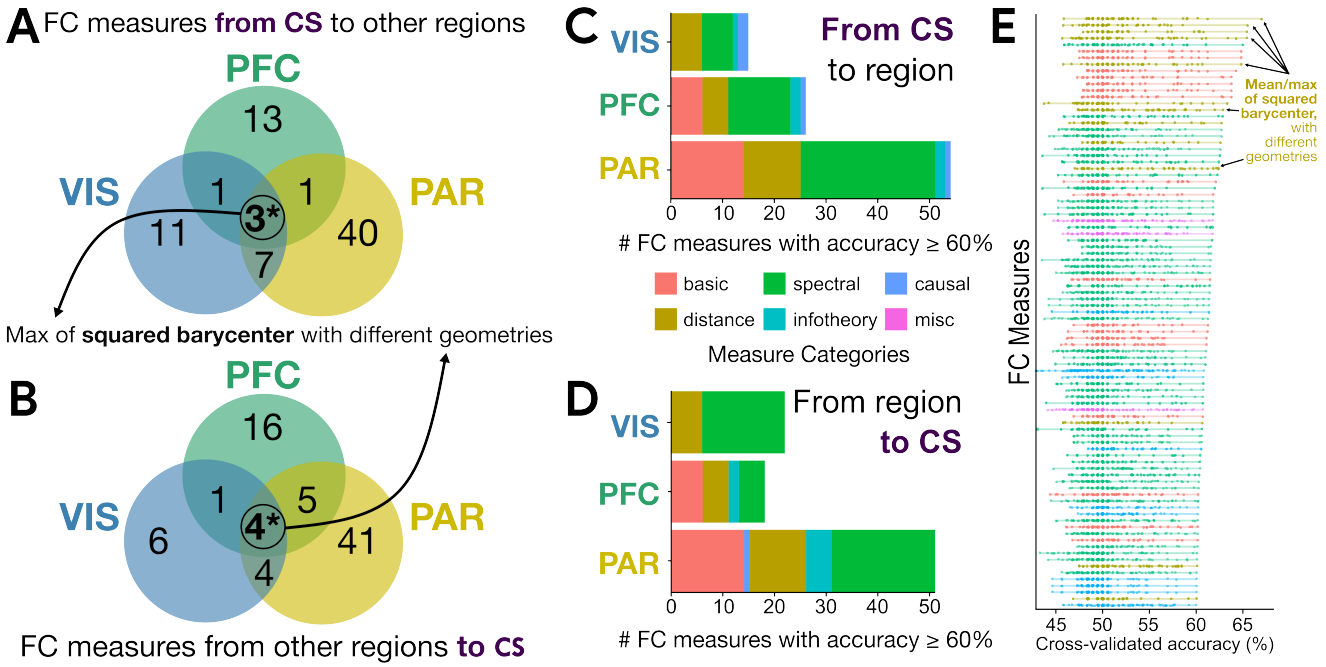


Figure 4.3.3: FC measures that detect distance-based, information-theoretic, and frequency-focused coupling distinguish between stimulus types across all region–region pairs. **A,B**. Here, the number of FC measures yielding at least 62% cross-validated classification accuracy in one or more stimulus pairs are tabulated for each region as either the source (**A**) or target (**B**) region the corresponding interaction. Ellipses within the Venn diagram are color-coded to indicate the corresponding region, following the brain maps depicted in Fig. 4.2.1A. The circled '1' at the center indicates that one SPI distinguished between at least one stimulus pair with $\geq 60\%$ accuracy both from and to all regions: the maximum of the squared barycenter, derived using soft-DTW geometry. **C,D**. For the FC measures tabulated in **A** and **B**, here we show the number of FC measures by literature category per region as the source (**C**) and target (**D**). **E**. For each of the 95 FC measures summarized in **A-B**, the cross-validated classification accuracy values are plotted for all 144 parameter combinations (region–region pairs, stimulus comparisons, onset vs. offset, and relevance settings). FC measures are ranked in descending order of the maximum accuracy and colored by literature category, as in **C** and **D**. The mean or maximum of the squared barycenter with different geometries are annotated for reference.

In order to better understand the regional specificity of the top-performing FC measures, we next investigated the overlap in the number and type of FC measures that distinguished stimulus types well between region–region pairs. Since connectivity with the CS region is shared between IIT and GNWT, we focus on FC measures directed to and from the CS region, respectively—noting that undirected FC measures will yield bidirectionally symmetric values. We filtered the same classification results presented in Fig. 4.3.1A down to those with $\geq 60\%$ mean cross-validated accuracy and tabulated the number of resulting FC measures projecting either from or to CS (e.g., from VIS, PFC, and PAR to CS). As shown in Fig. 4.3.3A-B, some FC measures uniquely distinguished task stimuli in only one functional pathway (e.g., 13 FC measures were unique to PFC $\rightarrow$ CS connectivity). However, we found three FC measures that distinguished at least one stimulus-pair to and from CS to all three other regions, all corresponding to the maximum of the squared

barycenter derived with different geometries (Euclidean, DTW, and a variation on DTW geometry known as soft-DTW [228]). This suggests that the maximum of the squared barycenter captures stimulus-relevant information (for specific stimulus types) in all functional pathways of interest to IIT and/or GNWT, poising it as a useful measure with which the theories may be quantitatively evaluated and contrasted.

All FC measures can be categorized into one of six literature-based domains ('basic', 'distance', 'spectral', 'info-theory', 'causal', or 'miscellaneous'; cf. Cliff et al. [92]), so we next compared the FC measure type composition for the top-performing measures shown in Fig. 4.3.3A-B. As shown in Fig. 4.3.3C-D, undirected FC measures corresponding to distance measures (including the barycenter) and frequency-based coupling (e.g., power envelop correlation) comprised the top-performing statistics for all three functional pathways both to and from the CS region. In addition, statistics capturing basic covariance structure (including the Pearson correlation coefficient) comprised many of the top-performing FC measures between CS $\leftrightarrow$ PFC and CS $\leftrightarrow$ PAR, but not CS $\leftrightarrow$ VIS. Although there is high-level similarity among FC measures that distinguish stimuli across these functional pathways, the relatively low overlap shown in Fig. 4.3.3 indicates that the particular FC measures differ between functional connections. The overlap between CS $\leftrightarrow$ VIS and CS $\leftrightarrow$ PFC corresponds to the mean barycenter computed with soft-DTW geometry, while the FC measure overlap between CS $\leftrightarrow$ PFC and CS $\leftrightarrow$ PAR is driven predominantly by spectral properties like power envelop correlation and imaginary coherence (more detail on these FC measures is shown in Fig. S3).

In addition to characterizing the maximum decoding accuracy for each FC measure across region–region pairs (which represents a 'best-case' stimulus discrimination scenario), we also examined the full performance range for each of the 95 FC measures contained in Figs 4.3.3A-B (i.e., all FC measures with at least $\geq 60\%$ cross-validated accuracy result). Fig. 4.3.3E shows all decoding accuracy values per FC measure, tabulated across 144 combinations of stimulus contrasts, region–region pairs, task relevance settings, and presentation periods (onset vs offset). The highest maximum accuracy values correspond to the mean and maximum of the squared barycenter derived from different geometries, as annotated in Fig. 4.3.3E—though even these FC measures fail to surpass chance classification performance (i.e., 50%) in many settings. This visualization demonstrates the variability in decoding accuracy exhibited by all 95 examined FC measures, none of which outperformed chance in all 144 experimental parameter combinations.

4.3.4 TESTING PREDICTIONS VIA DATA-DRIVEN ANALYSIS OF THEORY-DRIVEN NEURAL MODELING

In the previous sections, we identified a multitude of FC measures that met our core stimulus decoding classification aim across region–region pairs. In particular, FC measures derived from the barycenter generalized across all of the relevant region–region pairs relevant to both IIT and GNWT. Barycenter-based measures are fast to compute, simple to interpret, and novel to functional neuroimaging—prompting further investigation with modeling. To derive theoretical insight into the neural processes predicted to underlie

conscious vision from our novel FC measure, we constructed two neural mass models [204] tailored to recapitulate each theory's pre-registered predictions (see Consortium et al. [150]) for neural activity in each region (for details see Sec. 4.2.5 and schematic depiction in Fig. 4.2.1B).

Importantly, because each theory generates predictions about coupling between different region–region pairs (CS $\leftrightarrow$ VIS for IIT, and CS $\leftrightarrow$ PFC for GNWT), our goal in this section is to evaluate the relative success (rather than the overall variance explained) of each theory-derived model in successfully approximating the empirical data. The models are stimulus-agnostic and are built to reproduce the empirical Euclidean barycenter values averaged over all stimulus types. We focused on the stimulus-averaged dynamics rather than the stimulus-specific dynamics, as neither GNWT nor IIT provided stimulus-specific predictions. Rather, they specified predictions about the qualitative features of inter-areal coupling (e.g. synchronized activity between CS and VIS) hypothesized to underlie conscious perception. In addition, to avoid a mismatch between simulated and empirical barycenter values caused by changes in dipole polarity across regions (cf. Fig. S4A-B), we specifically computed the absolute value (i.e., magnitude) of both the empirical and simulated time series before computing the Euclidean barycenter in this section. As a robustness test, we show in Fig. S5 that the stimulus classification results based upon the barycenter are not systematically biased by this extra preprocessing step.

Our model simulation process is schematically depicted in Fig. 4.3.4A. Using the IIT- and GNWT-based models, we simulated 1000 trials with dynamical noise and then subtracted the average neural activity before stimulus onset (i.e., the 'baseline') from each trial (Fig. 4.3.4A[i-ii]). We approximated noise in the measurement process by adding zero-mean independent Gaussian noise to each trial (Fig. 4.3.4A[iii]). Each trial was divided into 1000 ms segments corresponding to stimulus-on and stimulus-off periods for analysis. We show results obtained using $\sigma = 1.0$ added noise in the main text, with results from the full range of $0.5 \leq \sigma \leq 1.0$ depicted in Fig. S7. Within each stimulus-on and stimulus-off epoch, we computed the maximum of the squared barycenter across Euclidean and non-Euclidean geometries (Fig. 4.3.4A[iv]). For clarity, we show the results for the maximum of the squared barycenter derived from Euclidean geometry. Results are qualitatively comparable for barycenter measures computed using DTW, soft-DTW, and subgradient descent- (SGD-)DTW geometries (cf. Fig. S6).

Briefly, as schematically represented in Fig. 4.3.4B(i), GNWT predicts that visual consciousness is dependent on neural signal propagation from CS to PFC regions, leading to a late 'ignition' of the global workspace after stimulus onset and again at stimulus offset. The GNWT-based model has two key features responsible for the ignition-like updating dynamics in PFC at stimulus onset and offset: (1) the winner-take-all competition between the two PFC masses, and (2) the slow timescale of the PFC neural masses ($\tau_{PFC} = 80$ ms) relative to the comparably fast CS neural mass ($\tau_{CS} = 40$ ms). This leads to a peak in PFC activity ~ 200 ms after stimulus onset—approximately 150 ms after the peak in CS activity, which peaks ~ 50 ms after stimulus onset.

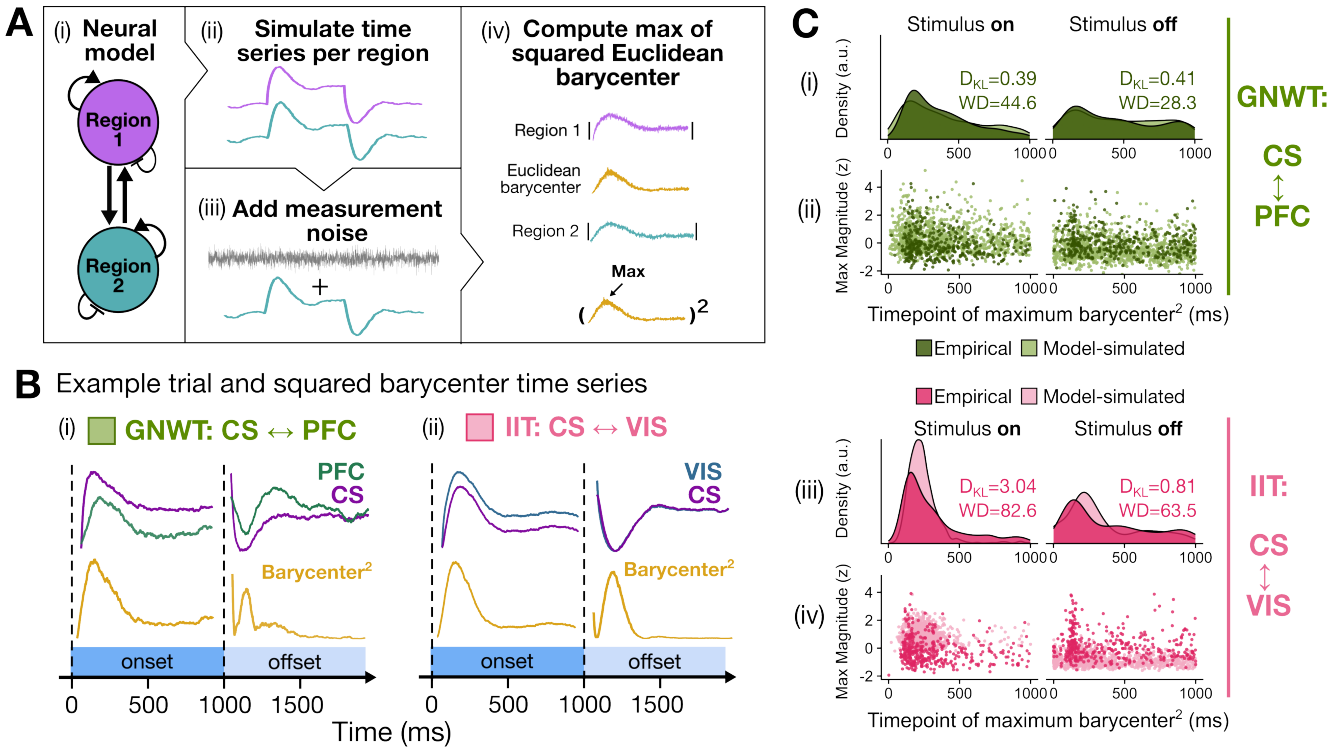


Figure 4.3.4: Comparing model-based predictions for barycenter with results from empirical COGITATE MEG data. **A** Graphical illustration of model analysis pipeline. For both models we simulated 1000-trials with noise, and after baseline subtraction and added zero mean Gaussian noise to each time point of the simulation to mimic the effect of measurement noise before computing the max of the squared barycenter for each trial. For comparability with the empirical data the barycenter was computed after taking the absolute value of the time series avoiding issues of dipole polarity reversal across regions. **B**. Illustration of single trial neural activity and barycenter times series for GNWT- and IIT-based neural mass models. Because the of offset between CS and PFC masses, and the winner-take all dynamics, the timing of the peak barycenter values is more variable across trials in the GNWT-based model than in the IIT-based model. **C.** Here, we examine the timing and magnitude of the squared Euclidean barycenter maximum in both empirical and model-simulated time series (with noise $\sigma = 1$) for both IIT and GNWT during stimulus onset and offset. First, the timing of the barycenter maximum for each (i) CS↔PFC and (iii) CS↔VIS bivariate time series is summarized via kernel density estimation for each stimulus presentation period and theory. The empirical density distribution is dark-shaded and the model-simulated density distribution is light-shaded in each case. The Kullback-Leibler divergence [218], $D_{KL}(\text{Empirical} \mid\mid \text{Model})$, is shown for each theory and stimulus presentation period, in units of bits, along with the Wasserstein distance (WD) in units of total work. (ii, iv) The same maximum squared Euclidean barycenter values comprising the density plots in (i) and (iii) are plotted, with the maximum timing on the x-axis and the magnitude of the corresponding squared Euclidean barycenter on the y-axis. The same color scheme is used as in (i) and (iii).

In terms of the (squared) barycenter, which tracks the center of mass of the two time series, the offset in peak activity between the two masses reduces the amplitude of the (squared) barycenter. The relatively low amplitude of the peak PFC signal also means that in the presence of measurement noise, the maximum of the barycenter is broadly distributed throughout the stimulus-on epoch. At stimulus offset, both the CS and PFC masses show a brief dip in activity due to the accumulated adaptation current. The winner-take-all competition between the two PFC masses then generates a secondary bump of activity approximately

300 ms after stimulus offset, at which point the previously suppressed mass takes over. Similar to the stimulus-on period, the offset in the peak activity between the CS and PFC masses, combined with measurement noise, yields a broad distribution of maximum (squared) barycenter values across the stimulus-off period. In contrast, as schematically represented in Fig. 4.3.4B(ii), IIT predicts that the conscious experience of a visual stimulus involves synchronized and sustained activity between CS and VIS (i.e., 'posterior hot zone') throughout stimulus presentation. The key features driving the dynamics of the IIT-based model are the strong reciprocal connectivity between the VIS and CS masses, and the comparably fast and similar timescale of the two masses ($\tau_{VIS} = 20$ ms, $\tau_{CS} = 40$ ms). Because of the similar timescale and strong reciprocal connectivity, both masses increase (and decrease) synchronously following stimulus onset and offset. During the stimulus presentation period, activity is maintained in a stable high-activity attractor state. The synchronized peaks and troughs in activity at stimulus onset and offset generate large peaks in the (squared) barycenter amplitude, yielding a high-density concentration of maximum timing values at the beginning of the epochs across trials.

Although qualitative comparison of the empirical and simulated distributions shows that both the GNWT- and IIT-based roughly approximate the empirical data, quantitative comparison of the simulated and empirical distribution of maximum squared Euclidean barycenter timings shows that the GNWT-based model provides a better approximation to the empirical data than the IIT-based model with lower K-L divergence values in both stimulus-onset ($D_{\mathrm{KL}}[\mathrm{GNWT}](P \mid\mid Q)$ = 0.39 bits vs. $D_{\mathrm{KL}}[\mathrm{IIT}](P \mid\mid Q)$ = 3.04 bits) and offset ($D_{\mathrm{KL}}[\mathrm{GNWT}](P \mid\mid Q)$ = 0.41 bits vs. $D_{\mathrm{KL}}[\mathrm{IIT}](P \mid\mid Q)$ = 0.81 bits) periods. This result is also reflected in the Wasserstein distance at stimulus-onset ($W_{\mathrm{GNWT}} = 44.6$ vs. $W_{\mathrm{IIT}} = 82.6$) and stimulus-offset ($W_{\mathrm{GNWT}} = 28.3$ vs. $W_{\mathrm{IIT}} = 63.5$). In other words, simulated maximum squared barycenter values from the GNWT-based model provided a better approximation to the empirical data than those derived from the IIT-based model in terms of both information cost (lower K-L divergence values) and Wasserstein distance. Importantly, this pattern is robust across various noise levels (cf. Fig. S7) and task relevance settings (cf. Fig. S8). We show in Fig. S4F(ii) that the PFC does indeed show a secondary bump in activity after stimulus onset, as GNWT predicts. Based on these results, we argue that the delayed ignition dynamics implemented in the winner-take-all connectivity of the GNWT-based model provide a better approximation to the empirical data than the tightly synchronized sensory dynamics implemented in the IIT-based model. This conclusion is, of course, dependent on the model-based implementation of the theoretical predictions and may be subject to revision.

4.4 DISCUSSION

Here, we presented an approach for data-driven identification, evaluation, and theoretical interpretation of connectivity-based neural correlates of consciousness. Our work is built upon the datasets and pre-registered hypotheses generated by the COGITATE Consortium adversarial collaboration that tested the predictions of IIT and GNWT [150]. We focused here on functional connectivity with the aim of finding

a measure(s) that provided reliable stimulus-type decoding—reasoning that if a neural process underlies the content of consciousness, variation in that process should encode information relevant to the contents of conscious perception. Through the most comprehensive comparison performed to date, we systematically evaluated 246 different measures of pairwise coupling, derived from an interdisciplinary time series literature [92], computed between regions hypothesized to drive conscious vision in IIT and/or GNWT.

For each region–region pair, we identified multiple FC measures that decoded stimulus type with >60% cross-validated (out-of-sample) accuracy. Importantly, in line with the predictions of GNWT, task relevance did not generally modulate stimulus-type decoding accuracy across FC measures and region–region pairs (including CS $\leftrightarrow$ PFC). Classifiers trained on a pair of stimuli in the task-relevant setting generalized to the same stimuli in the task-irrelevant setting, and vice versa, through a cross-validation paradigm. Top-performing measures included those commonly used to quantify activity from MEG data, such as the power envelop correlation [201, 202], as well as those novel to functional neuroimaging applications, such as the barycenter derived from Euclidean and non-Euclidean geometry. Indeed, the top-performing time series features that provided >60% stimulus decoding accuracy across region–region pairs common to both GNWT (CS $\leftrightarrow$ PFC) and IIT (CS $\leftrightarrow$ VIS), were based upon the maximum of the squared barycenter. The barycenter tracks the center of mass of two time series, and is sensitive to both the amplitude and relative phase of each time series. The maximum of the squared barycenter, therefore, identifies the time point of maximum covariance of both amplitude and phase. Since this class of features maximally decoded stimulus types, we reasoned that statistics derived from the barycenter form strong candidates for encoding information relevant to conscious visual perception.

We next sought to derive theoretical insight into the neural processes hypothesized to underlie conscious vision by comparing empirical barycenter values to simulated barycenter values derived from two simple neural mass models [204] tailored to recapitulate the pre-registered predictions (see Consortium et al. [150]) for what neural activity should be observed in each region by IIT and GNWT. The winner-take-all competition between the two PFC masses, coupled with their slow timescale, yielded a broad distribution of maximum (squared) barycenter timings across the stimulus onset period and stimulus offset period. In contrast, in the IIT-based model, the CS and VIS masses exhibit a similar timescale and strong reciprocal connectivity, such that both masses increased (and decreased) synchronously at stimulus onset and offset. This generated tightly peaked distributions of maximum squared barycenter values at the beginning of each onset and offset period. Quantitatively, simulated barycenter values from the GNWT-based model provided a better approximation for the empirical data than the IIT-based model, in terms of both information cost (lower K-L divergence values [218]) and the Wasserstein distance (also known as 'earth mover's distance' [219]). It can therefore be argued that the delayed ignition dynamics implemented in the GNWT-based model provide a better approximation to the neural processes generating the observed data than the highly synchronized sensory dynamics implemented in the IIT-based model.

When combined with the result that stimulus decoding from CS $\leftrightarrow$ PFC FC measures generalized between task-relevant and task-irrelevant conditions, our findings lend more credence to the predictions of GNWT than the predictions of IIT. Whereas GNWT is agnostic about the role of synchrony in sensory cortex, IIT predicts that ignition-like dynamics in PFC reflect task demands rather than conscious perception. The success of the GNWT model in quantitatively accounting for the data across task-relevance conditions is therefore contrary to the predictions of IIT. Again, however, we note that this result is conditional upon the model-based implementation of hypotheses from GNWT and IIT, and may be subject to revision should another model—still consistent with the hypotheses—be put forward. In practice, theory testing (and refinement) is an iterative, rather than absolute, process [229, 230], so we regard this as a strength, rather than weakness, of our approach. In addition, the predictions set out by Cogitate Consortium [150] are not derived from the core of IIT and GNWT, and therefore may be revised whilst leaving the core tenets of each theory intact [229]. This is not to say that the accuracy of the predictions should not increase or decrease our credence in each theory; rather, this is just to point out that we should not expect to falsify [231] IIT or GNWT outright on the basis of the results.

Before turning to a more general discussion of the approach taken in this paper, we highlight technical limitations that should be considered when interpreting this work and our findings. Even among the top-performing FC measures (including statistics derived from the barycenter), there is variability in stimulus decoding performance across different experimental parameters. Much like the 'no free lunch' theorem in machine learning [93, 232], our findings do not suggest a singular FC measure that will universally perform well across all stimulus decoding contexts. It remains to be seen, therefore, whether statistics based upon the barycenter generalize as top performers beyond the current context. Additionally, we note that our models are phenomenological and should be thought of as quantitative summaries of the neural dynamics hypothesized to underlie conscious perception by IIT and GNWT—rather than full implementations of the theories themselves.

There are also potential limitations specific to our MEG-based connectivity analyses. For example, we averaged across all epochs within a given event-related field type for each participant, such that each of the N=94 participants was represented by a single data point per FC measure in a given classification analysis. While the temporal averaging was performed to enhance the signal-to-noise ratio and reduce computational load [212], this design might also obscure relevant inter-trial differences and/or inter-individual heterogeneity in neural responses to the same stimulus type and experimental setting. Additionally, we employed the same preprocessing pipeline as in Consortium et al. [150], which employed the widely-used MNE-Python workflow [233]. However, issues remain regarding proper source localization, environmental/biological artifact removal, and latency variability, the latter of which poses issues for interpreting asynchronous (e.g., lagged) FC measures [234]. Field spread is also a concern due to the difference in spatial proximity between regions—for example, VIS and CS regions are much closer together than VIS and PFC—which also means that VIS and CS regions are more likely to exhibit similar MEG time series due to spatial autocor-

relation that decays with distance [235]. We do note, however, that this would bias VIS and CS towards greater synchrony (rather than less) such that this limitation does not affect the theoretical conclusions we draw above.

Beyond the conclusions and limitations of the current study, we present a roadmap for identifying and evaluating candidate neural correlates of consciousness in a data-driven manner that integrates with theory-driven modeling to derive quantitative and interpretable insights. Importantly, the data-driven nature of this approach means that one need not make *a priori* assumptions about which time series feature(s) will be best suited for the process of interest. Once informative time-series features are identified, modeling plays an indispensable role in mapping between the neural processes hypothesized to underlie the observed dynamics and the empirically measured time series features themselves. The models we specified here are deliberately as simple as possible, ensuring interpretability, and are also specifically constructed to simulate barycenter values, which are sensitive to the relative amplitude and phase of the two time series. Of note, given the rich literature in quantitative neural theory [215, 216], it will almost always be possible to build a computational model of the neural process of interest that has a sufficient level of complexity to capture the features picked out by a highly comparative data-driven approach. For example, had our data-driven evaluation of FC measures identified a frequency-based measure as the top performing feature (such as the power envelop correlation), a more complex modeling approach that explicitly incorporates frequency-based interactions between neural masses (e.g., Mejias et al. [236]) would have been more appropriate.

One key advantage of formal modeling is that even if the predicted time-series features are not identified as informative, if the model includes the full suite of neural processes hypothesized by a theory to underlie a cognitive/perceptual process of interest, the model can still be used to inform the interpretation of the time series features that are identified as informative. Although it is not always possible, models should ideally be built before data is collected. Indeed, a key limitation of the current study is that we built two models based upon preregistered *qualitative* hypotheses. Because these hypotheses are specified in words rather than equations, there will always be an imprecise mapping between formal models and the hypotheses themselves. It is therefore critical for the progression of the field that we move towards formulating hypotheses in quantitative terms that can directly link to formal models—which provide a concrete starting point for revisions that may better explain the observed empirical data. Here we fit the parameters of each model by hand, an important next step in the analysis pipeline will therefore involve leveraging previous work on model fitting and model comparison to fit and evaluate families of theory-derived models [237, 238].

To our knowledge, this is the first study to systematically compare hundreds of FC measures for decoding visual stimuli across task relevance settings. We focused here on bivariate coupling measures using the `pyspi` toolbox [92], though this approach is readily extendable to highly comparative analysis of both univariate time series [89, 90, 239] (to focus on local dynamics of a single region) or higher-order interactions

comprising three or more regions, a rapidly growing area of research [25, 26]. Combining across different scales of dynamics, from local to higher-order, could paint a fuller picture of the rich neural complexity driving a biological phenomenon like conscious visual perception [79]. In addition, an important avenue for future work will be to quantitatively compare measures that distinguish between conscious and unconscious stimuli, allowing us to isolate the processes that make a stimulus conscious—rather than focusing on the neural process underlying consciously perceived stimuli, as we did here. Importantly, our data-driven approach also generalizes across neuroimaging modalities and, when combined with neural modeling, offers clear insights into the processes hypothesized to generate empirically observed brain dynamics. Large openly available datasets are becoming more and more common in consciousness science, while the field is concurrently moving towards more quantitative neural theories [240–245]. There is, therefore, a unique opportunity for collaborations between experimentalists, data analysts, and neural modelers to systematically build, and quantitatively compare, measures and theories of consciousness.

ACKNOWLEDGMENTS

We are grateful to the COGITATE Consortium for their Herculean effort in setting up the adversarial collaboration, and for making the data openly available. In addition, we would like to thank our supervisors Ben Fulcher and Mac Shine for their guidance and encouragement throughout this project.

5

Extracting interpretable signatures of whole-brain dynamics through systematic comparison

It's never the [measure] you most suspect. It's also never the [measure] you least suspect, since anyone with half a brain would suspect them the most.
Dwight K. Schrute, The Office, Season 6 Episode 10

The empirical analyses presented thus far have focused on a single level of component granularity in isolation, particularly that of pairwise coupling between regions. Without the principled comparison of candidate features within, much less across, a representation type, brain-imaging studies across diverse and important applications—from functional fingerprinting [246] to targeted stimulation of a neural circuit [42, 43]—may rely on sub-optimal, subjective choices that could obscure the most relevant (and potentially most parsimonious) dynamical structures. To address this gap, in this chapter, we introduce a systematic approach to capture informative dynamical structure from neural time-series data that compares across a broad range of interpretable analysis methods, bridging across individual components of a system (e.g., activity within individual regions) *and* interactions between components (e.g., the functional connectivity between region–region pairs). This comprehensive framework encompasses five different representations with increasing complexity, from the localized activity of a single brain region up to the distributed activity of all brain regions and their pairwise interactions. We apply this method to quantitatively compare brain dynamics in each of four distinct neuropsychiatric disorders relative to neurotypical controls. This chapter embodies the central thesis themes of parsimony and biological interpretability by prioritizing simple linear classifiers and examining performance gains relative to dimensionality and complexity with each representation.

This chapter is published as: Bryant, A. G., Aquino, K., Parkes, L., Fornito, A., & Fulcher, B. D. Extracting interpretable signatures of whole-brain dynamics through systematic comparison. *PLOS Computational Biology* (2024), doi: 10.1371/journal.pcbi.1012692.

ABSTRACT

The brain's complex distributed dynamics are typically quantified using a limited set of manually selected statistical properties, leaving the possibility that alternative dynamical properties may outperform those reported for a given application. Here, we address this limitation by systematically comparing diverse, interpretable features of both intra-regional activity and inter-regional functional coupling from resting-state functional magnetic resonance imaging (rs-fMRI) data, demonstrating our method using case–control comparisons of four neuropsychiatric disorders. Our findings generally support the use of linear time-series analysis techniques for rs-fMRI case–control analyses, while also identifying new ways to quantify informative dynamical fMRI structures. While simple statistical representations of fMRI dynamics performed surprisingly well (e.g., properties within a single brain region), combining intra-regional properties with inter-regional coupling generally improved performance, underscoring the distributed, multifaceted changes to fMRI dynamics in neuropsychiatric disorders. The comprehensive, data-driven method introduced here enables systematic identification and interpretation of quantitative dynamical signatures of multivariate time-series data, with applicability beyond neuroimaging to diverse scientific problems involving complex time-varying systems.

5.1 INTRODUCTION

Quantifying the dynamical structures in a complex system like the brain can clarify relationships across scales—for example, from individual neuronal populations to macroscopic brain states. In the brain, as with other complex systems, there are myriad ways to represent different types of dynamical patterns, many of which involve discretizing the brain into 'regions' [1, 2]. Neuroimaging modalities like functional magnetic resonance imaging (fMRI) measure brain activity within voxels that are typically aggregated into such brain regions, yielding a region-by-time multivariate time series (MTS). While an MTS can be analyzed using statistics derived from a vast theoretical literature on time-series analysis [183], only a limited set of methods are typically used to summarize fMRI data. For example, most resting state-fMRI (rs-fMRI) studies have examined functional connectivity (FC) between pairs of brain regions by computing the Pearson correlation coefficient [10, 11], a zero-lag cross-correlation that assumes a joint bivariate Gaussian distribution, which may not always hold with rs-fMRI data [11, 82].

An MTS comprised of rs-fMRI data can be summarized at multiple levels, including: (i) individual regional dynamics; (ii) coupling or communication between pairs of regions across distributed networks; and (iii) higher-order interactions amongst multiple regions [24]. The choice of how to represent the rich dynamical structure in a brain-imaging dataset—e.g., intra-regional activity (properties of the fMRI signal time series for a given region) or inter-regional coupling (statistical dependence of the fMRI signal time series for two regions)—and which statistical properties, or 'features', to measure is typically made manually by a given researcher. Of the few studies to directly compare alternative time-series features for functional connectivity, or (less commonly) intra-regional activity, the examined set of statistics is usually limited in scope and size [13–15]. Intra-regional activity is often overlooked in favor of inter-regional functional connectivity, in part because of the hypothesis that neuropsychiatric disorders arise from disruptions to inter-regional coupling and integration, with local dynamics alone insufficient to explain or predict diagnosis [35, 247–

251]. However, this hypothesis has yet to be evaluated systematically, and conclusions about the lack of localized disruptions are largely based on previous work that has examined regional properties based on graph theory (e.g., degree centrality or node strength for a seed region in an FC matrix) [252, 253] or a handful of time-series features like the fractional amplitude of low-frequency fluctuations (fALFF) [4, 5] or regional homogeneity (ReHo) [6]—which could miss more nuanced changes to the blood oxygen-level dependent (BOLD) dynamics of a given brain region. Indeed, quantifying intra-regional activity can yield whole-brain maps of localized disruption, aiding interpretability and opening the door for new types of questions that are inaccessible using pairwise FC, such as how a given region responds to the targeted stimulation of another region [42].

Addressing the challenge of tailoring an appropriate methodology to a given data-driven problem, 'highly comparative' feature sets have recently been developed to unify a comprehensive range of algorithms from across the time-series literature, enabling broad and systematic comparison of time-series features (that would previously have been impractical). For example, the *hctsa* library includes implementations of thousands of interdisciplinary univariate time-series features [89, 90], and the *pyspi* library includes hundreds of statistics of pairwise interactions [92]. These algorithmic libraries have since been applied to systematically compare feature-based representations of diverse complex systems, from arm movement during various types of exercise [92] to light curves of different types of stars [254]. While it has been posited that rs-fMRI is too noisy and poorly sampled in time for more sophisticated time-series analysis methods (e.g., that have been applied to EEG data) [76–78], emerging applications of both *hctsa* and *pyspi* to rs-fMRI data have highlighted the utility of venturing beyond standard linear time-series analysis for both intra-regional [83] and inter-regional dynamics [84]. These highly comparative time-series feature approaches can distill the most informative biomarkers for a given application, such as distinguishing fMRI scans obtained at rest vs while watching a film [92] or comparing neurophysiological dynamics with cortical microarchitecture [239] or structural connectivity [255]. Importantly, time-series analysis methods developed outside of the neuroimaging space (such as dynamic time warping, originally developed for spoken word recognition [88]) have been shown to usefully quantify patterns in fMRI dynamics [85–87], underscoring the benefit of examining a comprehensive and interdisciplinary range of analysis methods.

To date, prior highly comparative analyses have focused on either intra-regional activity or inter-regional coupling separately, which fails to address which of the two—or indeed their combination—is optimal for a given problem. This question of which dynamical properties to capture from a complex dynamical system, and which algorithms might best capture them, is common to many data-driven studies, leaving open the possibility that alternative ways of quantifying dynamical structures may be more appropriate (e.g., yield better performance, clearer interpretation, or computational efficiency) for a given problem at hand. To address this gap, here we introduce a systematic method to evaluate multiple data-driven univariate time-series features, pairwise interaction statistics, and their combination, which is generally applicable to data-driven problems involving distributed time-varying systems. In the context of fMRI MTS, combining properties of intra-regional activity and inter-regional coupling is motivated by findings that the two representations synergistically improve classification performance across a variety of clinical settings including schizophrenia [35], Alzheimer's disease [36], nicotine addiction [37], and attention-deficit hyperactivity disorder [256]. Our method is applied to rs-fMRI for case–control classification of four neuropsychiatric disorders as an exemplary case study: schizophrenia (SCZ), bipolar I disorder (BP), attention-deficit hyperactivity disorder

(ADHD), and autism spectrum disorder (ASD).

Neuroimaging-based diagnosis of psychiatric disorders is an area of great interest, as diagnosis is currently based on behavioral criteria [119] and is hindered by patient heterogeneity [257, 258] as well as inter-rater reliability issues [259, 260]. Increasingly large open-source rs-fMRI datasets spanning neuropsychiatric disorders like schizophrenia [261] and autism spectrum disorder [262, 263] have co-evolved with the publication of increasingly complex classifiers like deep neural networks, which employ intricate algorithmic architectures that can be challenging to interpret [124, 264, 265]. Despite the improvements to in-sample classification performance often afforded by this increased complexity, core goals of neuroimaging-based classification studies include: (1) unpacking the neuroanatomical and functional underpinnings of a given disorder; and (2) developing a generalizable classification framework that aids the diagnosis of future (unseen) patients [119–123]. Here, we implement linear support vector machine (SVM) classifiers to combine features based on clear algorithms derived from interdisciplinary literature, aiming to compare and contrast the types of underlying dynamical processes that distinguish brain activity across disorders—rather than absolute classification performance [42, 266]. Broadly, insights into the dynamical processes that are altered can deepen understanding of mechanisms underlying disease pathogenesis and progression, potentially yielding clinically translatable biomarkers [267, 268]. This study is accompanied by a repository that includes code to reproduce all presented analyses and visualizations [269].

5.2 RESULTS

Starting with an rs-fMRI blood oxygen level-dependent (BOLD) signal MTS, in the form of brain regions sampled over time as shown in Fig. 5.2.1A(i-ii), we quantified both intra-regional dynamics (Fig. 5.2.1A[iii]) and pairwise coupling (Fig. 5.2.1A[iv]) using interpretable time-series features derived from a rich interdisciplinary scientific literature. We employed a case study comprising four neuropsychiatric disorders across two cohorts (as schematically depicted in Fig. 5.2.1B and summarized in Table 5.4.1). To represent intra-regional BOLD activity fluctuations, we computed 25 univariate time-series features that included the *catch22* feature set [91], as schematically depicted in Fig. 5.2.1C. The *catch22* features were distilled from an initial library of over 7000 candidate features [90] to concisely capture diverse properties of local dynamics, such as distributional shape, linear and nonlinear autocorrelation, and fluctuation analysis [91] (as described further in Methods Sec. 5.4.2). We also included three additional features: the mean and standard deviation (SD)—based on our previous findings that these first two moments of a time series distribution can be highly informative in a given application [271]—and fALFF as a benchmark statistic for localized rs-fMRI dynamics [4, 272]. Inter-regional FC was summarized using 14 statistics for pairwise interactions (SPIs) derived as a representative subset from an initial library of over 200 candidate SPIs in the *pyspi* library [92], including the Pearson correlation coefficient. This set of SPIs includes statistics derived from causal inference, information theory, and spectral methods, which collectively measure a variety of coupling patterns (directed vs. undirected, linear vs nonlinear, synchronous vs lagged) that paint a fuller picture of inter-regional communication patterns [13, 85, 92]. The full list of univariate and pairwise time-series features and brief descriptions are provided in Tables 5.4.2 and 5.4.3.

In order to understand how well localized intra-regional dynamical properties and/or inter-regional FC could

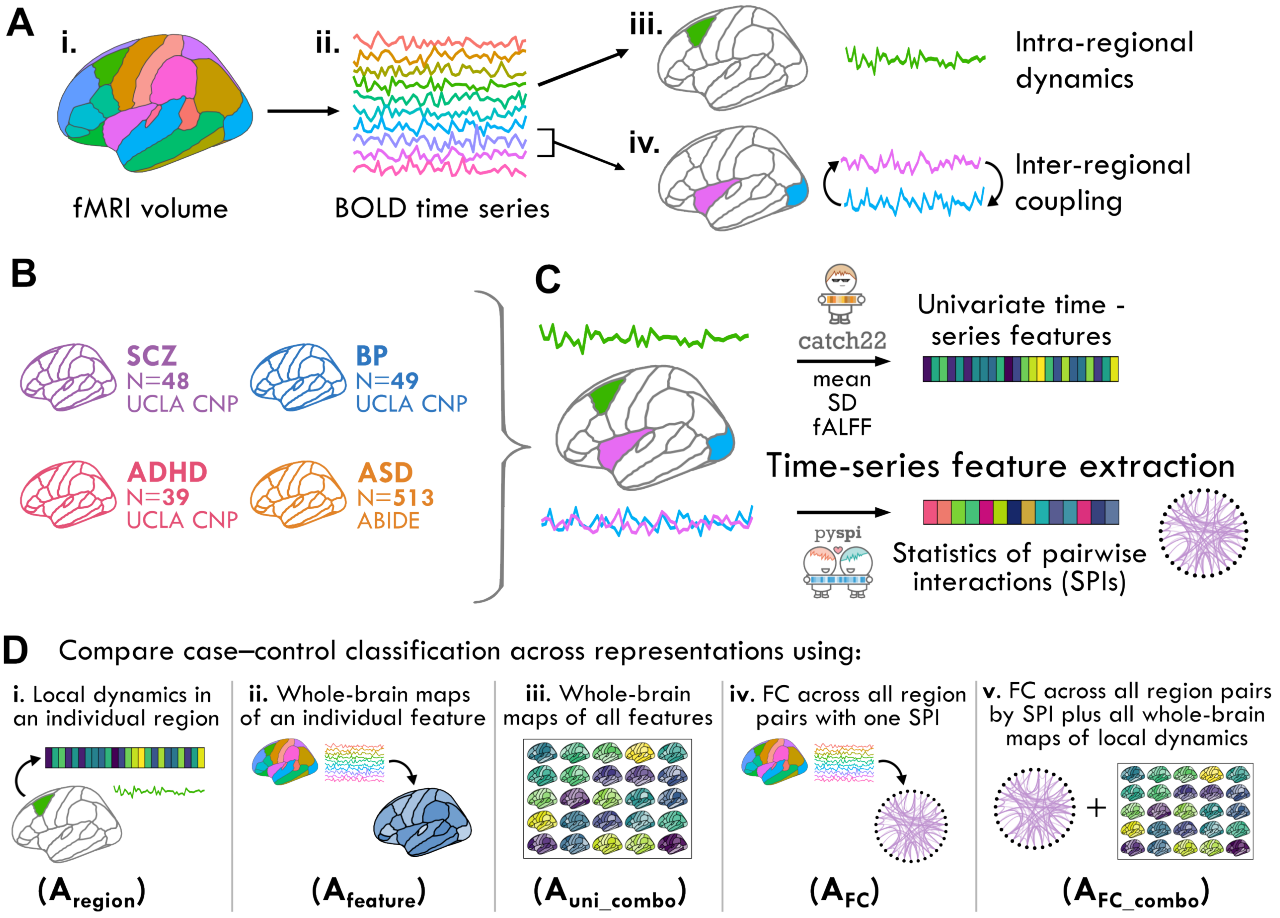


Figure 5.2.1: In this study, we systematically compared different ways of quantifying dynamical patterns from resting-state fMRI time series, focusing on statistics of local regional dynamics and pairwise coupling across four neuropsychiatric disorders. **A.** For a given resting-state fMRI volume (**i**), the cortex and subcortex are divided into individual regions from which the voxel-averaged BOLD signal time series is extracted (**ii**). Two key ways of quantifying dynamical patterns from these data are to: (**iii**) measure properties of the dynamics of individual brain regions (green); or (**iv**) to compute statistical dependencies between pairs of regions (pink and blue). **B.** To evaluate the performance of different types of dynamical representations of an fMRI time-series dataset (for identifying disease-relevant changes in neural activity), we included four neuropsychiatric exemplars originating from two open-access datasets: the UCLA CNP LA5c study [261] and the ABIDE I/II studies [262, 263, 270]. Each of the two cohorts also included cognitively healthy controls for comparison (N = 116 for UCLA CNP, N = 578 for ABIDE). **C.** For each type of dynamical structure that we extracted from an fMRI dataset (i.e., for each feature-based 'representation' of the data), we computed interpretable time-series features that encapsulate a diverse range of activity properties. Local dynamical properties were quantified from each brain region using a set of 25 time-series features (the *catch22* feature set [91] along with the mean, SD, and fALFF). Interactions between all pairs of regions were quantified using a set of 14 statistics of pairwise interactions (SPIs), comprising a representative subset from the *pyspi* package [92]. Values for both univariate time-series features and sets of pairwise coupling strengths from SPIs were captured in feature vectors that encapsulate a given type of interpretable dynamical structure(s). **D.** We evaluated case–control classification performance for each neuropsychiatric disorder using linear SVM classifiers fit to each of five distinct ways of representing resting-state fMRI properties: (i) all 25 univariate time-series features in an individual region, A_{region}; (ii) whole-brain maps of an individual time-series feature, $A_{feature}$; (iii) whole-brain maps of all 25 univariate time-series features, A_{uni_combo}; (iv) functional connectivity (FC) networks across all region pairs using one SPI, A_{FC}; and (v) A_{FC} along with all 25 univariate time-series features computed from all brain regions (A_{uni_combo}), termed A_{FC_combo}.

distinguish cases from controls, we systematically compared five distinct representations of BOLD activity dynamics, labeled A_{region}, $A_{feature}$, A_{uni_combo}, A_{FC}, and A_{FC_combo}, as illustrated in Figs 5.2.1D(i)–(v):

1. A_{region} (Fig. 5.2.1D[i]) represents each rs-fMRI time series using a set of features capturing 25 dynamical properties of an individual brain region. Strong classification performance in this representation suggests localized, region-specific changes in neural activity in the corresponding disorder.

2. $A_{feature}$ (Fig. 5.2.1D[ii]) represents each rs-fMRI time series using a given time-series feature that captures a single dynamical property from all brain regions (82 for SCZ, BP, and ADHD, and 48 for ASD). Strong classification performance in this representation suggests that a specific property of intra-regional resting activity is altered in one or more part(s) of the brain in the corresponding disorder.

3. A_{uni_combo} (Fig. 5.2.1D[iii]) represents each rs-fMRI time series using a set of features combining the 25 time-series properties evaluated across all brain regions. For this representation to out-perform A_{region} and/or $A_{feature}$ individually, it suggests changes to different time-series properties across different regions that could not be fully captured by either reduced representation.

4. A_{FC} (Fig. 5.2.1D[iv]) represents each rs-fMRI time series using a set of features that capture all pairs of inter-regional coupling strengths computed using a single SPI. Strong classification performance in this representation suggests that a specific type of inter-regional coupling is altered amongst one or more brain region pairs in the corresponding disorder.

5. A_{FC_combo} (Fig. 5.2.1D[v]) represents each rs-fMRI dataset as a set of features that capture all pairs of inter-regional coupling from a given SPI (as A_{FC}) as well as 25 time-series properties of all individual brain regions (as A_{uni_combo}). For this representation to outperform A_{FC}, it suggests the presence of complementary and disease-relevant information in both localized intra-regional activity and pairwise coupling that is not captured with FC alone.

We first discuss each analytical representation individually to examine the types of biological and methodological insights each representation affords in *Results* sections 5.2.1-5.2.5, then we systematically compare the performance of cross-validated classifiers based on each of the five representations in *Results* section 5.2.6. In all analyses, classification performance was evaluated using cross-validated balanced accuracy (i.e., the arithmetic mean of sensitivity and specificity, cf. [273]) measured using linear SVM classifiers; see Methods Sec. 5.4.3 for details on classifier selection.

5.2.1 ANALYSIS A_{REGION}: CHARACTERIZING LOCALIZED CHANGES TO INTRA-REGIONAL DYNAMICS ACROSS DISORDERS

With Analysis A_{region}, we propose that if the BOLD activity in a single brain region is disrupted across patients relative to controls, the intra-regional dynamical properties of that region can meaningfully distinguish cases from controls in the corresponding disorder (as depicted in Fig. 5.2.1D[i]). Such a result would contrast with previous studies positing that neuropsychiatric disorders like schizophrenia and ADHD are characterized by inter-regional dysfunction across distributed networks, rather than spatially localized disruptions [249, 272, 274, 275]. By reducing the complexity of an individual's rs-fMRI data down to 25

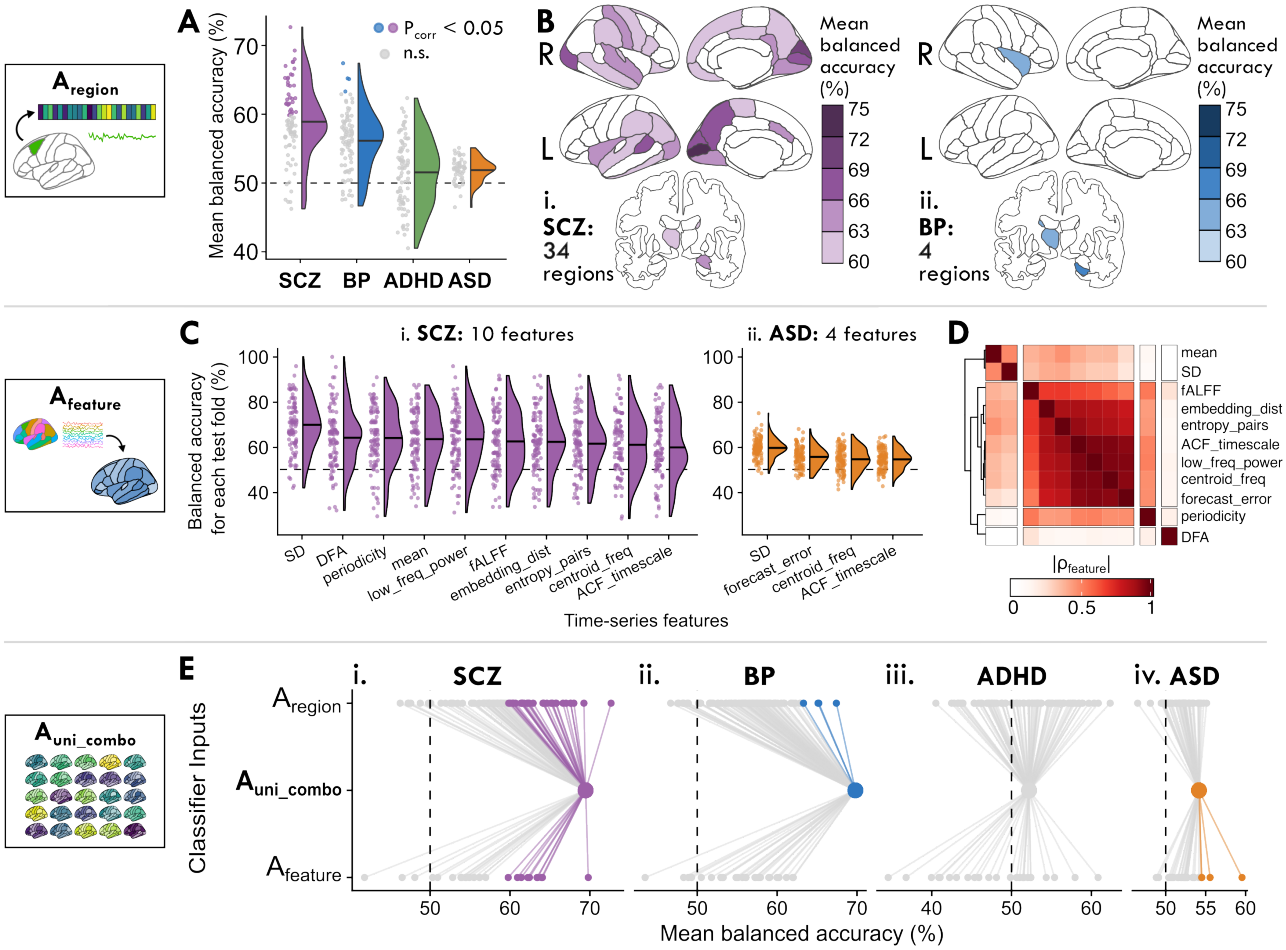


Figure 5.2.2: Local properties of BOLD dynamics within individual regions distinguish SCZ and BP cases from controls, and classification performance improves by combining all regions and time-series features. A. We investigate how the dynamics of a given individual brain region can distinguish cases from controls per disorder in A_{region} (cf. Fig. 5.2.1D[i]). The balanced accuracy is shown for all brain regions as a raincloud distribution, where each point corresponds to the mean classification performance (over 100 cross-validated test folds) of a single region. Points are colored to indicate a statistically significant balanced accuracy metric (Benjamini–Hochberg $P_{\text{corr}} < 0.05$, corrected across 82 brain regions for SCZ, BP, and ADHD and 48 brain regions for ASD) or grey for $P_{\text{corr}} > 0.05$. The horizontal line within each half-violin indicates the mean balanced accuracy across all brain regions for the corresponding disorder. **B.** The mean case–control classification balanced accuracy is shown for each brain region in Desikan–Killiany brain maps for SCZ (**i**) and BP (**ii**). Only statistically significant brain regions ($P_{\text{corr}} < 0.05$, corrected across 82 regions) are shaded, and the total number of significant brain regions are indicated for SCZ (34) and BP (4). **C.** We evaluate how a given dynamical property measured across the entire brain can distinguish cases from controls per disorder $A_{feature}$ (cf. Fig. 5.2.1D[ii]). For the 11 time-series features which yielded brain maps that significantly distinguished cases from controls in SCZ and/or ASD ($P_{\text{corr}} < 0.05$, corrected across 25 time-series features), the balanced accuracy is shown across all 100 test folds per feature (such that each point indicates one fold). **D.** The feature similarity score, $|\rho_{\text{feature}}|$, between each pair of the 11 features from **C** is visualized as a heatmap, revealing four distinct clusters of features based on their values measured across all brain regions from all control participants, along with the UCLA CNP SCZ group and ABIDE ASD group. **E.** We examine how the combination of all 25 time-series features measured from all brain regions can distinguish cases from controls across disorders in A_{uni_combo} (cf. Fig. 5.2.1D[iii]). To compare the performance of this combination classifier with that of individual regions and tir es features, the mean balanced accuracy is shown for all brain regions (upper row, labeled A_{region}), the c ation classifier (middle row, labeled A_{uni_combo}), and time-series features (bottom row, labeled $A_{feature}$). Lines are included as a visual guide to aid comparison. Colored dots indicate a classification model with significant balanced accuracy ($P_{\text{corr}} < 0.05$); gray dots indicate models that are not significant ($P_{\text{corr}} > 0.05$).

time-series features that encapsulate the dynamical properties of an individual region, we can interpret the results through a spatial map of disrupted fMRI dynamics in each disorder. As shown in Fig. 5.2.2A, multiple brain regions exhibited dynamics that significantly distinguished cases from controls (defined as Benjamini–Hochberg corrected $P_{\text{corr}} < 0.05$, one-tailed permutation test) in SCZ (34/82 individual regions) and in BP (4/82 regions), but not in ADHD or ASD; see Table S1 for full brain region balanced accuracy results[1]. The spatial distribution of intra-regional differences in underlying dynamics (assessed via cross-validated balanced accuracy) is shown for each brain region in Fig. 5.2.2B(i) and (ii) for SCZ and BP, respectively; results for ADHD and ASD are shown in Fig. S1C, though we focus on SCZ and BP here[2].

In SCZ, BOLD dynamics of medial occipital brain regions were the most informative of diagnosis, as shown in Fig. 5.2.2B(i). Specifically, against a background of half of all brain regions yielding significant classification performance, dynamical signatures of activity in the left pericalcarine cortex best classified SCZ cases from controls, with $72.7 \pm 11.2\%$ (mean $\pm$ standard deviation) balanced accuracy (one-tailed permutation test, $P_{\text{corr}} = 2 \times 10^{-4}$, corrected across 82 brain regions), followed closely by the right cuneus ($69.3 \pm 12.6\%$, $P_{\text{corr}} = 2 \times 10^{-3}$) and left cuneus ($68.0 \pm 11.0\%$, $P_{\text{corr}} = 3 \times 10^{-3}$; see Fig. S1A for individual region test-fold distributions). The relatively high performance of these medial occipital structures was unique to SCZ (compared to the other three disorders), which may reflect an SCZ-specific signature of localized alterations to visual areas. By contrast, subcortical structures like the right hippocampus ($67.4 \pm 14.4\%$, $P_{\text{corr}} = 0.02$) and left thalamus ($65.1 \pm 11.9\%$, $P_{\text{corr}} = 0.03$) distinguished BP cases from controls with the highest balanced accuracy. Among the significant brain regions in SCZ and BP, only one brain region (the left thalamus) was shared between the two disorders (cf. Fig. S1B), suggesting different spatial distributions of localized alterations in each disorder. We confirmed that the case–control classification performance across brain regions in SCZ and BP was not biased by overall region volume or by gray-matter volume differences between cases versus controls, as had been previously posited [35] (cf. Fig. S2 and Table S2). We also confirmed that the intra-regional dynamics of all significant brain regions distinguished SCZ and BP cases from controls better than a classifier trained solely on age and sex for each disorder (results shown in Table S3), indicating the intra-region dynamical features are capturing disorder-relevant differences beyond these basic demographic features (cf. Fig. S3E; Table S3). However, we found that including age, sex, and head motion (quantified as mean framewise displacement, or FD) improved classification performance for some brain regions in SCZ and BP—particularly for SCZ, for which mean FD was significantly higher than controls (cf. Fig. S3C,D). While this does not necessarily indicate that intra-regional dynamics are biased by these confounding variables, one should interpret the biological ramifications of region-specific findings with caution given the ability of confound variables to distinguish diagnostic groups.

5.2.2 ANALYSIS A_{FEATURE}: ASSESSING CHANGES IN BRAIN-WIDE MAPS OF INDIVIDUAL DYNAMICAL PROPERTIES ACROSS DISORDERS

In this section, we asked whether an individual time-series feature of intra-regional dynamics measured across all brain regions could statistically distinguish cases from controls. Strong classification performance in this representation would suggest that the given property is altered in one or more brain regions across cases of the corresponding disorder. To address this, we fit a separate classification model for each time-

[1] Supplementary tables are not reproduced in this thesis, but can be accessed in the electronic manuscript version.
[2] Supplementary figures are included in Appendix 4.

series feature, computed across all brain regions, as schematically depicted in Fig. 5.2.1D(ii). As shown in Fig. 5.2.2C, between SCZ and ASD, brain-wide maps of 11 out of the 25 intra-regional time-series features yielded significant balanced accuracies in at least one of the four disorders (one-tailed permutation test $P_{\text{corr}} < 0.05$, corrected across 25 features; see Table S4 for all balanced accuracy results). Notably, the average feature classification performance was generally lower in ASD than in SCZ, and after adjusting for multiple comparisons, no time-series features significantly distinguished BP or ADHD cases from controls. Three features distinguished both SCZ and ASD cases from controls with significant balanced accuracies: the BOLD SD, the frequency corresponding to the centroid of the power spectral density ('centroid_freq'), and the first $1/e$ crossing of the linear autocorrelation function ('ACF_timescale'). By contrast, seven additional time-series features individually distinguished SCZ cases, and one additional feature distinguished ASD cases from controls; the extent of feature overlap is visually summarized in Fig. S4A.

In order to summarize the types of time-series properties that were informative of SCZ and/or ASD diagnosis, we computed a similarity index $|\rho_{\text{feature}}|$ (defined as the absolute Spearman rank correlation between feature values measured from all brain regions in all control participants together with the UCLA CNP SCZ and ABIDE ASD groups) between all pairs of the 11 total significant features in Fig. 5.2.2C (see Table S5 for a list of $|\rho_{\text{feature}}|$ values for all pairs of features). As shown in Fig. 5.2.2D, we found one cluster of features that are generally sensitive to the linear self-correlation structure of a time series. This feature cluster notably included fALFF, which exhibited similar behavior across the dataset as the first $1/e$ crossing of the linear autocorrelation function ('ACF_timescale', $|\rho_{\text{feature}}| = 0.65$), the power in the lowest 20% of frequencies ('low_freq_power', $|\rho_{\text{feature}}| = 0.63$), and the centroid of the power spectral density ('centroid_freq', $|\rho_{\text{feature}}| = 0.58$). Some features in this cluster are sensitive to nonlinear structure in the time series, such as the entropy computed over probabilities of patterns over two time-steps ('entropy_pairs'), which significantly distinguished SCZ cases from controls. The 'periodicity' feature is also derived from the linear autocorrelation function, capturing the first local maximum of the ACF that meets criteria described in Wang et al. [276], displaying similar behavior to the features in the linear autocorrelation cluster (all $0.4 \leq |\rho_{\text{feature}}| \leq 0.55$). By contrast, the 'DFA' feature (which implements detrended fluctuation analysis to estimate the timescale at which the sharpest change in the scaling regime occurs [277]) was largely unrelated to all ten other features (all $|\rho_{\text{feature}}| < 0.2$), and was the second highest performing feature in SCZ alone ($64.1 \pm 13.3\%$, $P_{\text{corr}} = 8 \times 10^{-3}$). In general, the linear features shown in Fig. 5.2.2D collectively exhibited better classification performance in SCZ and ASD than most nonlinear properties we examined, such as the time reversibility statistic 'trev' that measures nonlinear autocorrelation structure (all mean balanced accuracy values <57% across disorders).

The SD and mean stood out as other high-performing features, which is notable given their sensitivity to the raw rs-fMRI BOLD time series distributions—in contrast to the other 23 features we evaluated, which were all computed from z-score normalized time series to capture properties of the underlying dynamics. Consistent with prior work [271], the BOLD SD was the top-performing feature for both SCZ ($69.8 \pm 11.7\%$, $P_{\text{corr}} = 1 \times 10^{-4}$) and ASD ($59.5 \pm 4.4\%$, $P_{\text{corr}} = 1 \times 10^{-6}$); by contrast, the mean rs-fMRI BOLD signal only distinguished SCZ cases from controls ($63.5 \pm 12.7\%$, $P_{\text{corr}} = 0.01$). BOLD SD has previously been reported as a powerful biomarker across diverse applications [278], including in case–control classification for SCZ [279, 280] and ASD [281]. However, brain-wide BOLD SD has also been quantitatively linked to head motion in the scanner [282, 283], and both SCZ and ASD groups showed greater head movement than

controls (cf. Fig. S5A and Table S6). We therefore computed the Pearson correlation coefficient between mean framewise displacement (FD) and BOLD SD to assess their linear relationship, finding significant associations at the whole-brain level for ASD but not for SCZ (cf. Fig. S5B and Table S7)—suggesting different underlying biological mechanisms driving neural signal variability in the two disorders.

In summary, the $A_{feature}$ analysis revealed that brain-wide maps of the BOLD SD and properties sensitive to linear autocorrelation structure could distinguish cases from controls in SCZ and ASD, performing well compared to more complex and/or nonlinear statistics. This supports the general use of linear time-series features—like the centroid frequency of the power spectral density ('centroid_freq') or timescale of the autocorrelation function ('ACF_timescale')—to parsimoniously summarize relevant properties of rs-fMRI dynamics, while also highlighting the potential utility of the SD to clarify disorder-relevant differences in BOLD signal variance measured across brain regions.

5.2.3 ANALYSIS A_{UNI_COMBO}: COMBINING MULTIPLE BRAIN REGIONS AND TEMPORAL PROPERTIES TO REPRESENT WHOLE-BRAIN FMRI DYNAMICS ACROSS DISORDERS

In contrast to the previous two reduced representations of intra-regional dynamics across the brain—in either a single region (A_{region}) or of a single time-series property ($A_{feature}$)—we next asked if a more comprehensive representation that integrates multiple time-series properties of resting-state activity measured from all brain regions could improve case–control classification performance. This entails a much higher-dimensional representation of brain-wide dynamics (2050 dimensions for the UCLA CNP cohort: SCZ, BP, and ADHD, and 1200 dimensions for the ABIDE ASD cohort). Higher-dimensional input spaces are more prone to overfitting due to greater complexity from more free parameters (described as the 'curse of dimensionality' [125]), and potentially from noise accumulation in the setting of noisy features [284]. To account for potential overfitting, we applied 10-repeat 10-fold cross-validation and we report out-of-sample performance to evaluate the generalizability of a model's predictive performance [285]. An improvement in performance (assessed via balanced accuracy) in this representation would therefore suggest that multiple properties of intra-regional activity measured across multiple spatial locations are important for characterizing a given disorder, in a way that cannot be reduced to a single region or single time-series feature. To investigate this question, we concatenated all intra-regional features computed from the whole brain into a single 'combination' classifier model, A_{uni_combo}, as depicted schematically in Fig. 5.2.1D(iii).

Figure 5.2.2E shows the mean balanced accuracy using this comprehensive representation of local dynamical properties across all brain regions as singular points in the middle row (A_{uni_combo}) for each disorder (see Table S8 for balanced accuracy results). We found that the A_{uni_combo} representation distinguished cases from controls particularly well for SCZ ($69.5 \pm 12.0\%$, one-tailed permutation test $P = 1 \times 10^{-7}$, Fig. 5.2.2E[i]) and BP ($69.8 \pm 11.6\%$, $P = 2 \times 10^{-8}$, Fig. 5.2.2E[ii]), as indicated by the points located far from chance performance (shown as dashed lines at 50%). The strong performance in SCZ is consistent with the large number of individual brain regions (34) and intra-regional time-series features (10) that yielded significant classification balanced accuracies on their own (cf. Fig. 5.2.2B,C). By contrast, only four individual brain regions (and zero individual time-series features) significantly distinguished BP cases from controls, suggesting that BP is better characterized by the combination of multiple types of alterations to BOLD dynamics in multiple brain areas. This combination classifier did not statistically differentiate ADHD cases from con-

trols ($49.8 \pm 8.7\%$, $P = 5 \times 10^{-1}$, Fig. 5.2.2E[iii]), which is consistent with the lack of significant individual brain regions in A_{region} (Sec 5.2.1) or time-series features in $A_{feature}$ (Sec 5.2.2). We tested the possibility that the null performance of the A_{uni_combo} classifier in ADHD may be at least partially attributable to noise accumulation [284], though we found no improvement in classification performance with a PCA-reduced feature space (cf. Fig. S6B) nor with different classifier types (cf. Fig. S7 and Table S9).

The combination classifier did significantly distinguish ASD cases from controls, albeit with lower balanced accuracy than with SCZ or BP ($53.6 \pm 4.7\%$, $P = 1 \times 10^{-10}$, Fig. 5.2.2E[iv]). We found similar results when we restricted our analysis to either of the two largest individual ABIDE sites (cf. Fig. S8B and Table S10), indicating that the lower performance cannot be entirely ascribed to the heterogeneity of the large multi-site ABIDE sample. We did find modest improvements in ASD classification performance using a nonlinear radial basis function (RBF) kernel as part of our robustness analysis (cf. Fig. S7A, Table S9), indicating that nonlinear feature-space boundaries better distinguished ASD cases from controls. In this high-dimensional feature space, this result is consistent with a larger dataset enabling more complex classification approaches to better capture signal, in a way that would be at higher risk of overfitting in a smaller dataset.

In summary, the A_{uni_combo} approach to quantifying fMRI dynamics allowed us to combine multiple properties of intra-regional dynamics from multiple brain regions in a way that can distinguish cases from controls, particularly for SCZ and BP, demonstrating the utility of combining whole-brain maps of multiple time-series features for a given classification problem.

5.2.4 ANALYSIS A_{FC}: COMPARING FEATURES OF INTER-REGIONAL FUNCTIONAL CONNECTIVITY ACROSS DISORDERS

While the previous three analyses focused on the BOLD dynamics of individual regions, in this section, we asked how patterns of distributed communication *between* pairs of brain regions assessed via functional connectivity (FC) capture case–control differences in each of the four disorders. We quantified pairwise communication as time-series dependencies, venturing beyond the traditional linear, contemporaneous Pearson correlation-based functional connectivity to compare its ability to distinguish cases from controls relative to thirteen other statistics of pairwise interactions (SPIs) from the *pyspi* library [92]. In total, these fourteen SPIs collectively measure different types of inter-regional coupling, including nonlinear, time-lagged, and frequency-based interactions, derived from literature including causal inference, information theory, and spectral methods (see Methods, Sec. 5.4.2 and Table 5.4.3 for more details). By comparing multiple different types of functional coupling, we aimed to investigate whether the Pearson correlation coefficient is comparatively a top performer, or if alternative metrics would be more sensitive to case–control differences in a given disorder. As depicted schematically in Fig. 5.2.1D(iv), each SPI was computed between the rs-fMRI time series of all region–region pairs per participant, and we concatenated values across all region–region pairs as classifier inputs for each SPI (see Methods, Sec. 5.4.3). Supplying the full region–region pair space to the linear SVM allows the classifier to learn global properties of brain connectivity through linear combinations of region pairs, like the brain-wide mean correlation [286].

We found that for SCZ and ASD, representing rs-fMRI dynamics through pairwise coupling significantly distinguished cases from controls—for five SPIs in SCZ ($P_{\mathrm{corr}} < 0.05$, one-tailed permutation test, corrected

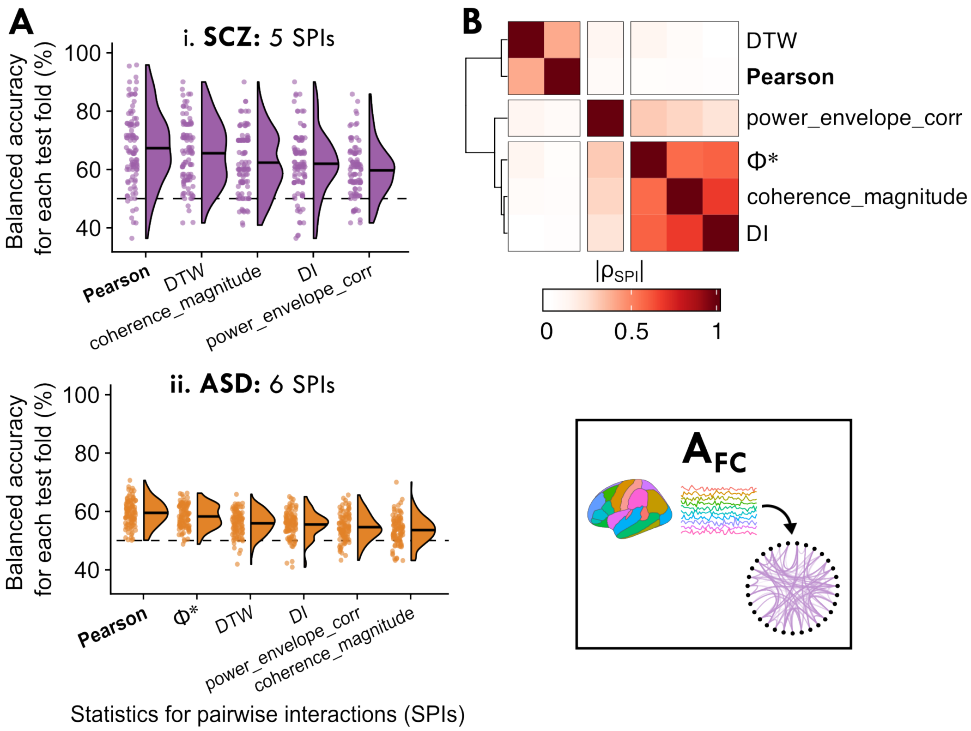


Figure 5.2.3: Representing brain activity as the set of all pairwise functional connectivity strengths, A_{FC}, can significantly distinguish cases from controls, with the classical Pearson correlation coefficient (capturing linear contemporaneous coupling) a top performing metric. A. For the 6 SPIs features which yielded brain maps that significantly distinguished cases from controls in SCZ and/or ASD ($P_{\text{corr}} < 0.05$, corrected across 14 SPIs), the balanced accuracy is shown across all 100 test folds per SPI (such that each point indicates one fold). The Pearson correlation coefficient is annotated in boldface for easier reference. **B.** The SPI similarity score, $|\rho_{\text{SPI}}|$, is visualized between each pair of the 6 SPIs from **A** as a heatmap, revealing three clusters of SPIs with similar behavior on the dataset (based on their outputs across all region–pairs and in all control participants, as well as the UCLA CNP SCZ group and ABIDE ASD group). As in **A**, the Pearson correlation coefficient annotation is shown boldface.

for 14 comparisons using the method of Benjamini–Hochberg) and six SPIs in ASD. The test fold–wise balanced accuracy distribution is shown for each SPI in SCZ and ASD in Fig. 5.2.3A (see Table S11 for all balanced accuracy results). By contrast, none of the fourteen SPIs significantly classified BP or ADHD cases from controls; the distribution for the top five performing SPIs in BP is visualized in Fig. S9B, although we focus on SCZ and ASD in this section. Note that while some SPIs showed case–control differences in brain-wide average FC values (cf. Fig. S10), the global mean FC values performed similar or worse to classifiers supplied with the full region–region pair matrix (cf. Fig. S11).

The significant performance of multiple individual SPIs suggests that inter-regional communication patterns were disrupted in both SCZ and ASD in a way that pairwise FC metrics are generally well suited to capture, consistent with prior work [287, 288]. Notably, the Pearson correlation coefficient performed well (similarly to the raw covariance matrix, cf. Fig. S12), yielding the highest classification balanced accuracy for SCZ ($67.3 \pm 12.6\%$, $P_{\text{corr}} = 1 \times 10^{-4}$, corrected across 14 SPIs) and ASD ($59.5 \pm 4.4\%$, $P_{\text{corr}} = 1 \times 10^{-9}$) (it was also among the best SPIs for BP, although this was not statistically significant; cf. Fig. S9B[ii]). In order to better understand the types of SPIs that performed well for SCZ and ASD, we computed an empirical SPI similarity index ($|\rho_{SPI}|$) using the same method as implemented in A_{feature} above (i.e., computed across all controls as well as SCZ and ASD participants; cf. Sec. 5.2.2). As shown in Fig. 5.2.3B (and listed in Table S12), the Pearson correlation coefficient behaved most similarly to dynamic time warping (DTW, similarity index $|\rho_{SPI}| = 0.40$), which incorporates shifts and dilations between a pair of time series [88, 289] and has previously been suggested as a promising alternative to standard correlation for rs-fMRI analysis [85]. Other top-performing SPIs included directed information ('DI', [290]) and Φ^* [203], which captured different types of pairwise coupling compared to the Pearson correlation coefficient and DTW, as shown in Fig. 5.2.3B, and warrant further investigation in future work. We found similar results when assessing multi-diagnosis classification performance in the UCLA CNP cohort (cf. Fig. S13). Ultimately, the Pearson correlation coefficient was both the top-performing SPI based on case–control classification and the most parsimonious, underscoring the utility of measuring linear, synchronous inter-regional coupling under a Gaussian assumption to capture disorder-relevant pairwise dependence structures from rs-fMRI.

5.2.5 ANALYSIS $A_{\text{FC\_COMBO}}$: INTEGRATING LOCAL REGIONAL DYNAMICS WITH PAIRWISE COUPLING INTO ONE CLASSIFICATION MODEL ACROSS DISORDERS

Having demonstrated that multiple statistical representations of fMRI time series—based on intra-regional dynamics and inter-regional coupling—can distinguish cases from controls in each disorder, we next asked whether a unified representation combining intra-regional and inter-regional properties would offer complementary information about rs-fMRI activity that could better distinguish cases from controls. To test this, for each SPI, we concatenated the FC matrix (A_{FC}) with the full region $\times$ univariate feature matrix ($A_{\text{uni\_combo}}$), forming the basis for analysis $A_{\text{FC\_combo}}$ as depicted schematically in Fig. 5.2.1D(v). We then evaluate how well each SPI *plus* the full $A_{\text{uni\_combo}}$ matrix of local dynamics could distinguish cases from controls in each disorder.

As shown in Fig. 5.2.4A, combining local dynamics with pairwise coupling allowed us to classify cases from controls with significant balanced accuracy for all 14 SPIs in SCZ, BP, and ASD ($P_{\text{corr}} < 0.05$, correcting for 14 SPIs; with the exception of phase lag index, 'PLI', for BP; see Table S13 for full results). None of the

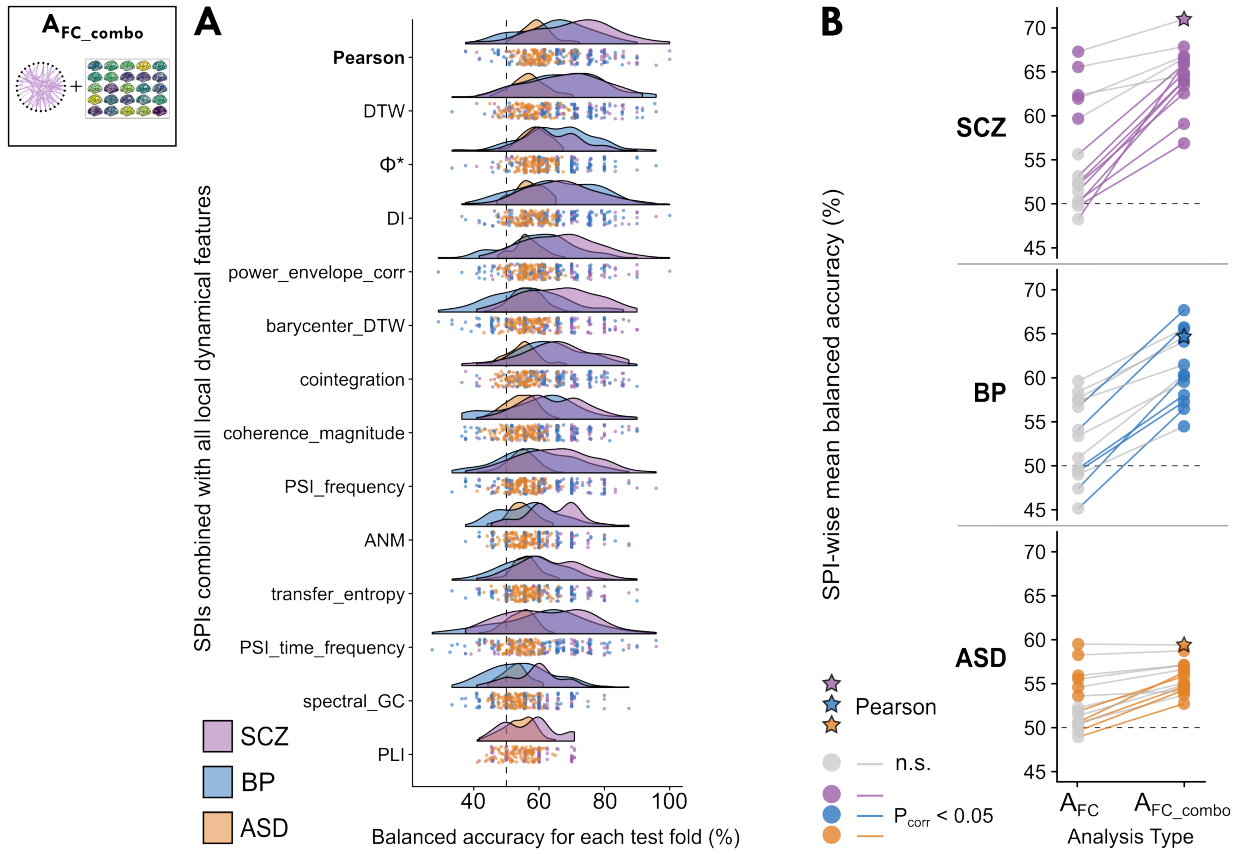


Figure 5.2.4: Adding statistics of local regional dynamics enhances case–control classification performance for pairwise FC metrics. Here we evaluate the effect of including brain-wide maps of local regional dynamics with pairwise coupling data for each SPI on case–control classification performance. By concatenating the full univariate region $\times$ feature matrix from A_{uni_combo} with the full set of brain region pairs per SPI from A_{FC}, this yielded a set of 14 linear SVM classifiers termed A_{FC_combo}. **A.** The balanced accuracy is shown in raincloud plots for all 100 test folds per SPI (such that each point indicates one test fold). The Pearson correlation coefficient is annotated in boldface for easier reference. **B.** The mean balanced accuracy is shown for each SPI on its own (left, A_{FC}) and with the inclusion of univariate region $\times$ feature data (right, A_{FC_combo}). Points are shaded in with color to indicate whether the corresponding balanced accuracy was significant ($P_{\text{corr}} < 0.05$, corrected across 14 SPIs) or not significant (gray) relative to the corresponding null distribution (cf. Sec. 5.4.3). Each line corresponds to one SPI to visually guide comparison across representation types. Lines are shaded darker to indicate a significant difference in SPI performance with the inclusion of localized dynamics (corrected two-sided t-test $P_{\text{corr}} < 0.05$, corrected across 14 SPIs; cf. Methods, Sec. 5.4.3). Stars indicate the balanced accuracy for the Pearson correlation coefficient for easier reference.

A_{FC_combo} models yielded significant balanced accuracies for ADHD after correcting for multiple SPIs (with $P_{\text{corr}} < 0.05$). The Pearson correlation coefficient remained a top-performing SPI (as in A_{FC}, cf. Fig. 5.2.3B), yielding classification performance of $71.0 \pm 12.4\%$ in SCZ ($P_{\text{corr}} = 2 \times 10^{-8}$) and $59.4 \pm 4.5\%$ in ASD ($P_{\text{corr}} = 2 \times 10^{-8}$), as shown in Fig. 5.2.4A. The A_{FC_combo} classifier using the Pearson correlation coefficient was also a top performer in BP ($64.7 \pm 10.9\%$, $P_{\text{corr}} = 8 \times 10^{-5}$), although DTW, Φ^*, and DI exhibited marginal improvements over the Pearson correlation (cf. Fig. 5.2.4A). The relatively tight distributions over test folds (shown as violin plots in Fig. 5.2.4A) highlight the consistency of the relatively high performance of these top SPIs across out-of-sample test cases in the three disorders, particularly in SCZ and BP.

Despite involving a considerable increase in feature-space dimensionality, adding intra-regional univariate dynamics from A_{uni_combo} substantially increased the number of SPI-based models (from A_{FC}) that could distinguish cases from controls in SCZ, BP, and ASD. This is particularly evident for BP, for which no SPIs meaningfully distinguished cases from controls on their own (cf. Figs 5.2.3A, S9B)—and yet, when we also included brain-wide maps of localized activity features in A_{FC_combo}, all SPIs but PLI allowed us to classify BP cases from controls with significant balanced accuracy. In order to summarize the extent to which combining intra-regional dynamics with inter-regional coupling improved classification performance, we directly compared the mean classification performance with each SPI in A_{FC} versus A_{FC_combo} in Fig. 5.2.4B. This visualization revealed that all SPIs better distinguished cases from controls with the inclusion of localized dynamics, including those that yielded statistically null performance on their own in A_{FC}. We quantified these improvements in case–control classification performance in A_{FC_combo} using two-tailed T-tests, corrected for resampled cross-validation (see Methods Sec. 5.4.3, [291]; all T-test results are presented in Table S14). As shown by the colored lines in Fig. 5.2.4B, the improvement with A_{FC_combo} was significant (after correcting for 14 SPI comparisons per disorder) for 9 SPIs in SCZ, 6 SPIs in BP, and 5 SPIs in ASD.

The addition of intra-regional time-series features did not just improve the performance of significant A_{FC} models, but also elevated the performance of many non-significant SPIs to classify cases from controls well across disorders. For example, additive noise modeling ('anm')—a measure of nonlinear dependence that tests causality from one brain region to another with a Gaussian process [292]—did not distinguish cases from controls well on its own (all balanced accuracies <52% averaged across test folds), but the inclusion of local univariate dynamics boosted its performance to $64.0 \pm 8.6\%$ in SCZ, $57.3 \pm 9.4\%$ in BP, and $54.7 \pm 4.3\%$ in ASD (all balanced accuracy $P_{\text{corr}} < 3.0 \times 10^{-3}$ and $T_{\text{corr}} > 2.4$). Similar improvements were observed across disorders for the barycenter ('barycenter'), a measurement of the center of mass of the rs-fMRI BOLD time series between the pair of brain regions, with the barycenter magnitude reflecting the extent of dynamic coupling [224, 225]. While most of these significantly improved SPIs yielded mid-ranking balanced accuracies in A_{FC_combo}, in BP, the two top-performing SPIs—DTW ($67.7 \pm 12.0\%$, $P_{\text{corr}} = 2 \times 10^{-6}$) and Φ^* ($65.7 \pm 11.0\%$, $P_{\text{corr}} = 2 \times 10^{-8}$)—exhibited among the largest improvements with the inclusion of intra-regional univariate dynamics (both $T_{\text{corr}} = 2.7$, $P_{\text{corr}} = 0.03$). However, SPIs that performed well for SCZ and ASD on their own (A_{FC}), such as the Pearson correlation coefficient, generally did not show large margins of improvements with the inclusion of local univariate dynamics (all $T_{\text{corr}} < 1.7$ and $P_{\text{corr}} > 0.09$)—suggesting that such high-performing SPIs capture relevant case–control information that overlaps more with that of local univariate dynamics than low-performing SPIs.

In summary, we found that including intra-regional time-series features generally improved the case–control

classification performance of SPIs, particularly for those which did not separate cases from controls well on their own in A_{FC}. This underscores the benefits of a unified representation that combines local intra-regional dynamics and pairwise inter-regional coupling to simultaneously capture complementary aspects of brain activity changes across clinical settings.

5.2.6 IDENTIFYING WHICH REPRESENTATION TYPE(S) ARE OPTIMALLY SUITED TO CAPTURE CASE–CONTROL DIFFERENCES IN EACH DISORDER

In previous sections, we evaluated a different number of models within each representation type—e.g., 25 intra-regional features in $A_{feature}$ versus 1 model for A_{uni_combo} versus 14 SPIs for A_{FC}. These differences in model numbers (and model complexity) from A_{region} through to A_{FC_combo} make it challenging to fairly compare the relative strengths of each representation in summarizing case–control differences in rs-fMRI dynamics. To address this, we implemented a strategy to build a single model incorporating the statistical information of a given representation type, corresponding to the five ways of quantifying dynamical properties from an rs-fMRI dataset evaluated above. For each cross-validation fold, we identified the top-performing model based on in-sample training performance (A_{region} or $A_{feature}$) or performance in a reduced feature space of the first 10 PCs (see Methods, Sec 5.4.3). This strategy enabled us to directly compare classification performance on unseen data from just one model per representation across test folds.

In Fig. 5.2.5, we show the distribution of classification balanced accuracy values across the 100 test folds per disorder as raincloud plots. As shown in Fig. 5.2.5(iii), A_{uni_combo} exhibited the best overall classification performance for SCZ (Fig. 5.2.5A) and BP (Fig. 5.2.5B); all results for this analysis are in Table S15. This is particularly evident for BP, in which selecting one brain region or one intra-regional time-series feature yielded an average classification balanced accuracy of 54.4% or 52.2%, respectively—yet the combination of local dynamics across brain regions in A_{uni_combo} yielded an average 69.8% balanced accuracy. Selecting one brain region per test fold yielded 65.8% balanced accuracy in SCZ, which is surprisingly high for a disorder generally characterized by distributed dysfunction [35, 247, 248], and this was surpassed by the A_{uni_combo} average of 69.5% balanced accuracy. By contrast, in ASD, selecting one representative intra-regional time-series feature per test fold outperformed all other representations, yielding 59.5% balanced accuracy on average—an effect driven primarily by the BOLD SD, which was selected in 99 out of 100 cross-validation folds, consistent with results shown for ASD in A_{region} (Fig. 5.2.2C,D). Since A_{uni_combo} included all of the BOLD SD data from $A_{feature}$, the drop in classification performance is surprising, underscoring how uniquely informative the SD was for ASD—with the caveat that this effect may be driven in part by head motion (cf. Fig. S5).

Comparing these univariate models with A_{FC} (Fig. 5.2.5iv) and A_{FC_combo} (Fig. 5.2.5v), we found that picking the best representative SPI in A_{FC_combo} (69.4% average balanced accuracy) yielded comparable balanced accuracy to that of A_{uni_combo} (69.5%) in SCZ. This suggests that for any given SPI in A_{FC_combo}, the local univariate dynamics are a key driver in the improved classification performance for SCZ. A similar relationship was observed in BP (Fig. 5.2.5B), in which A_{uni_combo} (69.8% average balanced accuracy) actually surpassed the classification performance of the representative subset of SPIs in A_{FC_combo} (65.4% average). Moreover, for ASD (Fig. 5.2.5D), the representative intra-regional feature subset—predominantly comprised of the BOLD SD—performed similarly to that of A_{FC} (59.4% average balanced accuracy) and

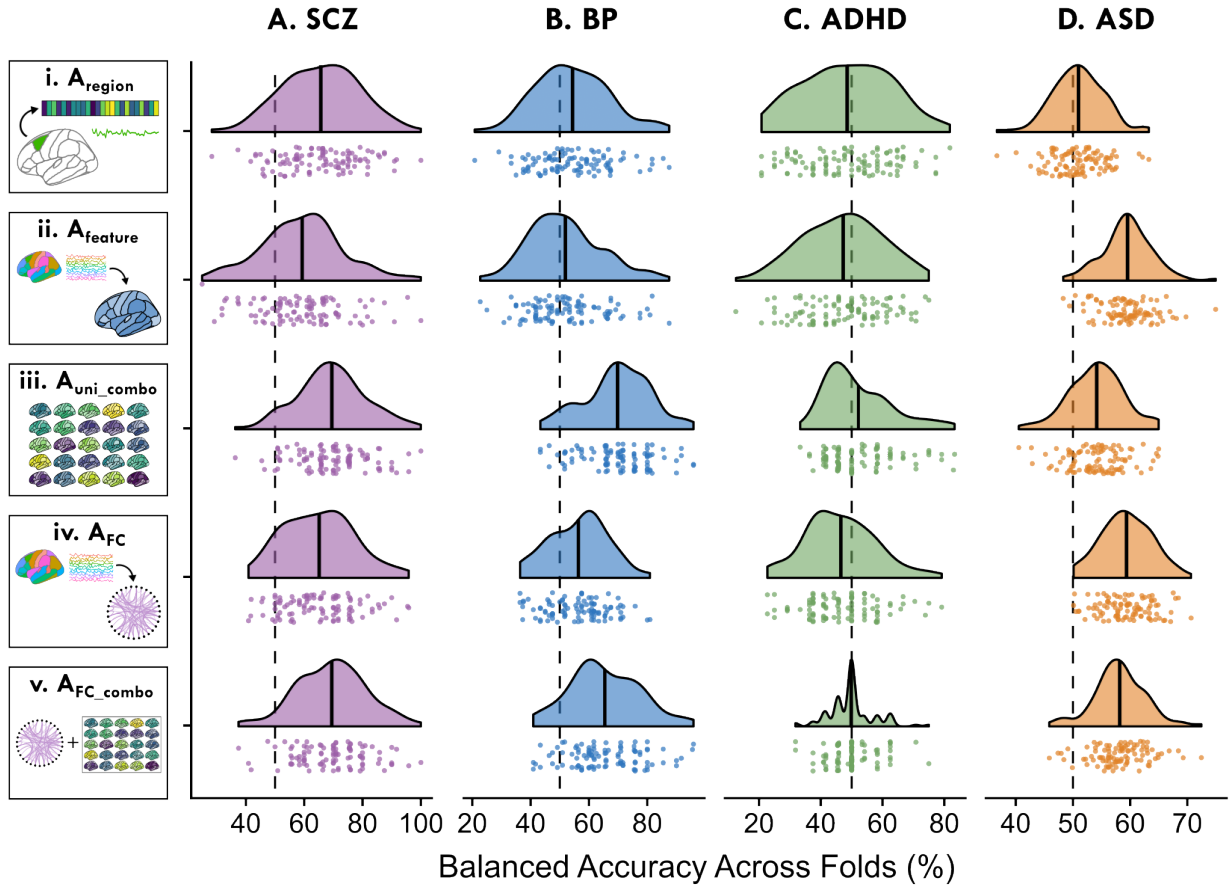


Figure 5.2.5: Direct comparison of model performance across representations and neuropsychiatric disorders. For SCZ (**A**), BP (**B**), ADHD (**C**), and ASD (**D**), we show the distribution of the best-performing model for each of 100 test folds across the five representations. In A_{region} (**i**) and $A_{feature}$ (**ii**), we identified the model that yielded the top in-sample training balanced accuracy per fold and retained its out-of-sample test balanced accuracy. In A_{uni_combo} (**iii**), as there was only one model for comparison, the distribution shows the 100 test folds for this model. For A_{FC} (**iv**) and A_{FC_combo} (**v**), we first performed a PCA and retained the first 10 PCs per SPI, then identified which PC-space SPI yielded the top in-sample training balanced accuracy per fold. Upon identifying that SPI, the full feature-space SPI was evaluated with the linear SVM classifier to measure the out-of-sample balanced accuracy for the corresponding test fold. For each half-violin plot, the horizontal line indicates the mean balanced accuracy across all 100 test folds for the given representation type (row) and disorder (column).

A_{FC_combo} (58.2%). This comparison highlights the utility of localized, intra-regional dynamics in distinguishing cases from controls across disorders, with room for future work optimizing feature selection and classifier parameters in the case of models with larger feature spaces (e.g., A_{FC} and A_{FC_combo}).

5.3 DISCUSSION

There are myriad ways in which a data analyst can extract interpretable feature-based representations of dynamical structures contained in a multivariate time series, like those measured with rs-fMRI. From this statistical smorgasbord, a given researcher typically chooses a set of dynamical properties to study for a given problem subjectively, such as a set of pairwise Pearson correlation coupling strengths or fALFF in a given set of brain regions. The lack of systematic methodological comparison thus leaves it unclear from any given study whether these chosen dynamical properties are optimal, or whether alternative—and potentially simpler and more interpretable—statistics may outperform those reported. To address these concerns, here we introduced a systematic comparison of feature-based representations of rs-fMRI time series, based on localized intra-regional dynamics, pairwise inter-regional coupling, and their combination, allowing us to systematically capture and compare different facets of rs-fMRI dynamics. Our results demonstrate the benefits of such comparison by identifying the most parsimonious and informative types of structures that are relevant for a given application, revealing disorder-specific signatures across neuropsychiatric groups. Our work provides a methodological foundation for systematically invoking representative features from a rich interdisciplinary literature on time-series analysis [89, 92] to determine the most appropriate way(s) of summarizing interpretable dynamical structures in MTS datasets. This approach is generalizable to a wide range of problems across neuroimaging modalities and applications, as well as to a vast array of science and industry problems in which complex time-varying systems are measured and analyzed.

The five statistical representations of MTS evaluated here can be investigated individually to gain insights into the types of local dynamics and pairwise dependencies that are optimally suited for a given application. For example, by comparing model performance within A_{region} (Results Sec. 5.2.1), we found that the dynamics of many individual brain regions could distinguish cases from controls in SCZ, and to a lesser extent in BP. This is a striking finding, given that such complex neuropsychiatric disorders are believed to arise from distributed dysfunction across brain networks rather than localized changes to individual regions [247, 249, 275, 293]—which suggests that our expanded breadth of rs-fMRI time-series features combined into a unified statistical representation for each brain region is informative of disorder-specific alterations. Examining dynamics within individual brain regions enables spatial interpretability through visualizing brain-wide maps of classification performance, providing a clear region-by-region picture of activity disruptions. The ability to characterize changes in BOLD dynamics at the level of individual regions has been key to addressing questions about regional differences in the response to spatially targeted brain stimulation [42, 43]. It also provides a clearer way to test molecular hypotheses about regional disruption in disorders, which can be compared with rich multimodal region-level atlases spanning morphometry, cortical hierarchy, and multiomics [44, 45, 83, 294] to more deeply characterize the physiological underpinnings of disease-relevant changes in future work. This systematic approach recapitulated prior results of individually informative brain regions—like medial occipital regions in SCZ [295–297] and subcortical regions in BP [298, 299]—while also identifying novel region-specific changes in local dynamics that generate new hypotheses for

future work. While we did not find significant region-specific patterns of BOLD alterations in ADHD or ASD, this does not preclude the possibility that different parcellation schemes and/or different intra-regional features (beyond the 25 examined here) could better elucidate disorder-specific alterations [300].

Looking across the features we compared in this study, our findings generally support the use of linear time-series analysis techniques that are commonly used for rs-fMRI data analysis, both at the level of individual brain regions and at the level of pairwise coupling—while simultaneously identifying novel high-performing metrics that warrant future investigation, including directed information to capture asymmetric information flows. In $A_{feature}$ (Sec. 5.2.2), we found that whole-brain maps of individual intra-regional dynamical properties related to linear autocorrelation (including Fourier power spectrum structure, like the centroid of the power spectrum) were top performers in distinguishing SCZ and ASD cases from controls (see Fig. 5.2.2C). For example, the 'ACF_timescale' statistic distinguished ASD cases from controls while the fALFF did not (despite their high empirical correlation, as has been noted in previous work [255]), demonstrating that algorithms derived from theoretically similar foundations can perform differently in a given application. Interestingly, the strong performance of this timescale feature might be related to previous work showing changes to intrinsic neural timescales in SCZ [301, 302] and ASD [302, 303]. The nonlinear intra-regional features we examined (such as the time reversibility statistic 'trev') did not enable us to distinguish cases from controls in any disorder, consistent with the view that rs-fMRI BOLD dynamics (which are noisy and sparsely sampled in time) are well approximated by a linear stochastic process such that methods aiming to capture more complex (e.g., nonlinear) dynamical structures may not be beneficial at this timescale, as has recently been proposed [74, 304].

The BOLD SD stood out relative to the other top-performing intra-regional features, as it is sensitive to the raw time-series values—while all other features (besides the mean) were computed after z-score normalization, in order to focus on underlying dynamics of the time series in a way that does not depend on the measurement scale. We found that whole-brain maps of the BOLD SD were informative of SCZ and ASD diagnosis relative to controls (see Fig. 5.2.2C), expanding upon our previous work showing that the BOLD SD outperformed other univariate time-series features for SCZ [271]. Changes to BOLD signal variability have previously been described in both SCZ [279, 280] and ASD [281], with prior work underscoring its utility and reliability as an rs-fMRI statistic [305]. More broadly, regional BOLD SD alterations have been reported in settings ranging from healthy aging [306, 307] to Alzheimer's disease [308], and it is hypothesized that BOLD signal variance is linked to functional integration [309] as well as numerous molecular and cytoarchitectural properties [278]. While we demonstrated that BOLD SD was generally unrelated to head movement in SCZ in Results Sec 5.2.2, we are cautious with our interpretations for ASD given the strong positive associations between SD and head movement (a non-neural confound) in the ABIDE cohort; this warrants further investigation and clarification in a future study.

When we combined multiple properties of intra-regional dynamics across the whole brain in A_{uni_combo}, we found improved case–control classification performance relative to either representation on its own (A_{region} or $A_{feature}$) in SCZ and BP—consistent with disruptions to brain dynamics that are both spatially distributed and multifaceted [250, 268, 310]. By contrast to the improved performance with the expanded A_{uni_combo} representation for SCZ and BP, reduced representations (i.e., individual brain regions or individual time-series features) better distinguished ASD cases from controls (see Figs 5.2.2E, 5.2.5D). In ASD, the BOLD

SD was the dominant high-performing feature among the 25 we compared, emerging as the top-performing metric in 99 out of 100 training validations (see Fig. 5.2.5D). Although the BOLD SD was contained within the combined region $\times$ feature matrix evaluated in the combination classifier A_{uni_combo}, increasing the number of input attributes to a classifier does not necessarily improve its performance due to overfitting beyond the 'latent dimensionality' of the dataset, defined as the number of meaningful variables inferred from the data that capture underlying essential patterns [311, 312]. This suggests that aside from the BOLD SD, the other local univariate time-series features contributed less disease-relevant information to the linear classifier, thereby demonstrating that systematic comparison can uncover potential model simplifications to select a more parsimonious feature-based representation.

Through our comparison of SPIs in A_{FC} (see Results Sec. 5.2.4), we found that pairwise coupling strengths also served as an informative way to represent dynamical structures related to case–control differences. Out of the 14 evaluated SPIs, we found that the Pearson correlation coefficient was a top-performing statistic in both SCZ and ASD, suggesting that linear time-series analysis methods are overall well-suited for capturing the salient dynamical properties of rs-fMRI MTS [83, 255, 313]. Our comparisons also highlighted interesting alternative statistics with different behavior but similarly high performance to the Pearson correlation coefficient, such as DI [290, 314] and Φ^* [203, 315]. These SPIs involve more conceptually and computationally complex formulations of pairwise interactions that are seldom applied to fMRI datasets, although we have previously found that alternative metrics like DI out-performed the Pearson correlation coefficient in an fMRI-based problem [92]. While there was little evidence that increasing the complexity in an FC measurement beyond the Pearson correlation coefficient substantially improved SCZ or ASD classification here, we did observe modest improvements in BP classification using each of DTW, Φ^*, and DI (with the inclusion of local univariate dynamics) instead. The alternative FC metrics examined here could be explored in future work using data with higher temporal resolution and signal-to-noise ratios (e.g., from MEG [239, 316])—a setting in which more complex types of interactions may be measurable (e.g., that are nonlinear and/or time-lagged).

Consistent with complex and spatially distributed disruptions to rs-fMRI activity, we found that all 14 SPIs better distinguished cases from controls for SCZ, BP, and ASD when we also included brain-wide maps of all 25 intra-regional time-series features in A_{FC_combo} (see Sec. 5.2.5). This finding is in line with previous work demonstrating that combining local and pairwise properties of rs-fMRI data can synergistically improve classification performance [35–38], supporting the notion that intra-regional activity and inter-regional coupling provide complementary information about disorder-relevant brain dynamics. While we cover an extensive space of dynamical statistics here, future applications might combine local dynamics with inter-regional coupling using network properties, local–global coupling, geometric embedding techniques like regional homogeneity (ReHo), or higher-order interactions. For example, Ponce-Alvarez et al. [39] and Tanabe et al. [317] demonstrated the benefits of quantifying 'glocal' measures [318] of synchrony between individual regions and whole-brain networks for distinguishing wakefulness from anesthetic states.

Future applications of this systematic framework might also consider more nuanced data fusion techniques beyond simple matrix concatenation to properly combine heterogeneous input data types (e.g., multiple brain regions; local dynamics versus pairwise coupling) into one classifier [40, 41]. Our modular approach incorporates incremental complexity from individual brain regions up to integrated whole-brain maps of local

dynamics and pairwise coupling, such that results are interpreted based on the overall performance of a given model. While we did not explore feature importance scores due to intrinsic issues with multicollinearity among input features, future work could incorporate dimensionality reduction techniques to mitigate this collinearity and specifically interpret individual feature weights (e.g., linear SVM coefficients) to compare relative contributions of different brain regions and/or temporal signatures to a given decision boundary. More broadly, techniques to distill down the feature space will be of great utility to analyzing complex system MTS in general, both in terms of computational demand and mitigating noise accumulation.

While it is not straightforward to compare classification performance across disorders given differences in sample size and acquisition sites [284], we observed that SCZ cases were distinguished from controls with higher balanced accuracy across representations than the other three disorders. By contrast, for ADHD, none of the five representations we evaluated yielded significant case–control balanced accuracies. Inverse probability weighting boosted classifier performance beyond predicting the majority class in many models in ADHD (cf. Fig. S14A), although our robustness analysis across classifier types and hyperparameter tuning suggests that poor classification performance is attributable to high heterogeneity in a small sample size. Indeed, the ADHD sample (N=39) was smaller than that of SCZ (N=48), BP (N=49), or ASD (N=513), though previous studies have reported conflicting findings about the effect of sample size on classification performance [252, 319, 320]. It is possible that the ADHD group is comprised of individuals with less severe symptoms than that of SCZ or BP—which could decrease classification performance with rs-fMRI features [321]—although we did not explicitly incorporate ADHD assessment or medication data in the scope of this analysis. In support of this possibility, there was no difference in head motion between the ADHD cases and controls, which is surprising for a disorder generally characterized by hyperkinesis [322] and reports of greater head motion in the scanner [323, 324]. Multiple publications have reported that head motion in the scanner is linked to symptom severity in ADHD [324–326], and Aoki et al. [327] found in a meta-analysis that few neuroimaging studies detected significant ADHD–control alterations in the absence of group head motion differences.

We found several intra- and inter-regional properties of rs-fMRI dynamics that significantly distinguished ASD cases from controls, although the corresponding balanced accuracy values were generally lower than that of SCZ and BP. There are a few potential explanations for why ASD cases were less distinguishable from healthy controls than SCZ or BP in general. The ABIDE dataset is approximately ten times larger than any of the disorder–control combined groups in the UCLA CNP cohort [262, 263], and prior studies have demonstrated an inverse relationship between sample size and classification performance using ABIDE data [328–332]. Notably, we included ASD and control participants from all ABIDE sites and did not explicitly account for imaging site in our classification analyses. While site-specific effects appear to be minimal in this dataset [333] and non-disorder factors like imaging protocol or scanner seem to contribute less to variance in resting-state activity than key brain regions [334], we did compare classifiers restricted to participants from each of the two largest ABIDE sites, finding minimal differences to those trained on the full ABIDE dataset. However, future work could evaluate the effect of multi-site integration techniques such as ComBat harmonization [335, 336] on intra- and inter-regional feature performance. Another difference with the ASD group relative to the other disorders (from the UCLA CNP cohort) is that homotopic region pairs were consolidated into one bilateral region as part of the competition data provided by Traut et al. [270], which could obscure relevant hemisphere-specific changes to localized dynamics in ASD [337, 338]. It is possible

that these populations exhibited changes to neural functional architecture that were more heterogeneous and individual-specific [319, 339, 340], making it difficult to identify an effective SVM decision boundary across representations—underscoring the value of ongoing work with normative modeling [341].

The goal of this study was to compare within and across interpretable representations of rs-fMRI dynamics, with a focus on results obtained with the linear SVM classifier for simplicity. We explored alternative classifiers that implement SVM with an RBF kernel or L1 ('LASSO') regularization, random forest, or gradient boosting (cf. Fig. S7) and did not observe general performance improvements, although it is possible that combining other nonlinear classifiers with hyperparameter optimization could more sensitively discriminate among features in a given disorder if nonlinear boundaries are present. Maximizing case–control classification performance was not an explicit aim of this study, although we note that our findings sit within the range of recent rigorous large-scale classification studies [132, 133, 342]. Future work could expand upon this generalizable framework to optimize performance in a given application setting.

In the '$p \gg n$' [343] large feature space setting with A_{uni_combo}, A_{FC}, and A_{FC_combo}, alternative regularization may be beneficial—which could be accomplished through more exhaustive hyperparameter tuning and/or different regularization approaches, such as graphical lasso [344]. Future work could also systematically evaluate ways to reduce (spatial) dimensionality in a biologically informative manner (e.g., condensing region pairs into canonical functional networks [185] or applying similarity network fusion [41, 345]). While this work presents a systematic methodological framework that is flexible and generalizable across domains, future studies aiming to identify disorder-specific biomarkers will need to evaluate their findings in external validation datasets to thoroughly assess the validity of a given brain region or time-series feature. Moreover, future work could extend beyond the binary case–control classification paradigm to examine transdiagnostic versus disorder-specific properties of resting-state dynamical structure [346, 347].

Here we focused on five key ways of systematically comparing intra-regional and inter-regional features from an fMRI dataset, but many other representations of spatiotemporal data could also be investigated in future work. This includes quantifying properties of the networks defined by pairwise FC matrices (cf. highly comparative graph analysis [174]), statistics of spatiotemporal patterns (like spirals and traveling waves [348, 349]), and higher-order (beyond pairwise) dependence structures [25]. For example, statistics such as metastability (based on the standard deviation of the Kuramoto order parameter) could quantify properties of integration and segregation across brain states [350]. More broadly, future work could relax the picture of interacting spatially localized and contiguous brain regions, towards spatially distributed modes, extracted as components through dimensionality reduction [30, 31] or geometrically [27, 28]. We also focused here on representative subsets of the univariate time-series analysis literature (the *catch22* subset of over 7000 features in the *hctsa* feature library [90]) and pairwise dependence literature (14 representative SPIs from the full *pyspi* library of over 200 SPIs [92]). Note that neither the univariate features nor the SPIs measured here were specifically tailored to neuroimaging applications; for example, the *catch22* set was derived based on classification performance across 128 diverse univariate time series datasets (including beef spectrograms and yoga poses) [351]. Indeed, given the demonstrated biological heterogeneity in neuropsychiatric disorders [352], effective exploration of larger time-series feature spaces could identify subtypes within a given disorder with distinctive dynamical profiles [353]—establishing the foundation for data-driven nosology [354]. Subsequent work could consider ways to incorporate the full sets of local

and pairwise time-series properties or to derive reduced subsets with algorithmic approaches tailored to a particular dataset. Broadening the scope of comparison in these ways—of both types of dynamical structures and algorithms for quantifying them—comes with associated issues of statistical power required to reliably pin down specific effects that future work will need to carefully consider.

5.4 METHODS

An overview of our methodology is illustrated in Fig. 5.2.1. We extracted the BOLD time series from each rs-fMRI volume, which were analyzed at the level of individual brain regions and pairs of brain regions (Fig. 5.2.1A). We included participants from four neuropsychiatric disorder groups derived from two main cohort studies, as depicted in Fig. 5.2.1B and summarized in Table 5.4.1. For each participant, we computed a set of time-series features reflecting the local dynamics of a single region's BOLD time series and coupling between a pair of brain regions (Fig. 5.2.1C). We compared the performance of these univariate and pairwise features separately, and in combination, using linear classifiers for each disorder—specifically testing five different ways of capturing dynamical structure, as shown in Fig. 5.2.1D and outlined in the following.

5.4.1 NEUROIMAGING DATASETS AND QUALITY CONTROL

UCLA CNP

Raw BOLD rs-fMRI volumes consisting of 152 frames acquired over 304 seconds were obtained from the University of California at Los Angeles (UCLA) Consortium for Neuropsychiatric Phenomics (CNP) LA5c Study [261]. This included cognitively healthy control participants as well as participants diagnosed with schizophrenia (SCZ), attention-deficit hyperactivity disorder (ADHD), and bipolar I disorder (BP). Details of diagnostic criteria and behavioral symptoms have been described previously, along with imaging acquisition details [261]. Imaging data were preprocessed using the *fmriprep* v1.1.1 software [355] and the independent component analysis-based automatic removal of motion artifacts (ICA–AROMA pipeline [356], as described previously [357]). Additional noise correction was performed to regress out the white matter, cerebrospinal fluid, and global gray-matter signals using the ICA–AROMA + 2P + GMR method [357–359]. Time series of 152 samples in length were extracted from the noise-corrected BOLD rs-fMRI volumes for 68 cortical regions [169] and 14 subcortical regions [360] spanning the left and right hemispheres.

ABIDE

Preprocessed BOLD rs-fMRI time-series data consisting of between 49 and 433 time points (mean: 184 $\pm$ 63 time points) were obtained from the Autism Brain Imaging Data Exchange (ABIDE) I & II consortium [262, 263], working with a dataset specifically aggregated for an international classification challenge by [270] that included healthy controls and participants with autism spectrum disorder (ASD). Imaging data were preprocessed by the classification challenge authors [270, 333] using the standard pipeline from the '1000 Functional Connectomes Project', which includes motion correction, skull stripping, spatial smoothing, segmentation, and noise signal regression [333]. We opted to use the preprocessed time series corresponding to the Harvard–Oxford cortical atlas, which is published with the FSL software library [361] and includes 48

regions, noting that homotopic region pairs were pre-consolidated into singular bilateral regions.

Quality control

After preprocessing, we identified six participants from the UCLA CNP dataset (IDs: sub-10227, sub-10235, sub-10460, sub-50004, sub-50077, and sub-70020) for whom the ICA–AROMA + 2P + GMR pipeline yielded constant zero values across all timepoints and brain regions. These six participants (N=3 Control, N=2 SCZ, N=1 ADHD) were excluded from all further analyses. Head movement in the scanner was evaluated using framewise displacement (FD) calculated from the six framewise head-motion parameters (corresponding to rotation and translation in the x, y, and z planes) using the method of Power et al. [362]. The mean FD was compared between each disorder and the corresponding control group using a Wilcoxon rank-sum test, implemented in R with the `wilcox_test` function from *rstatix* (version 0.7.2) We found that SCZ, BP, and ASD—but not ADHD—cases exhibited significantly higher mean FD relative to control participants (all $P < 0.05$), consistent with prior studies [363–365]. In order to mitigate the potential confounding effect of head motion on the dynamical BOLD properties with putative neural causes, we applied the movement-based exclusion criteria described in [358] (labeled as 'lenient' criterion in that study), such that any participant with mean FD > 0.55 mm was excluded from further analysis. This step excluded two participants from the UCLA CNP study (N=1 SCZ, N=1 Control) and 59 participants from the ABIDE study (N=36 ASD, N=23 Control).

Participant summary statistics

After performing quality control, we retained between N = 39 and N = 578 participants per clinical group across the two studies, with summary statistics provided in Table 5.4.1 and demographic distributions visualized in Fig. S3. Participants in the UCLA CNP dataset were scanned at two different sites [261] and participants in the ABIDE dataset were scanned across seventeen sites [333]. We did not include site information as a covariate in our classifiers, nor did we perform cross-site alignment, to simply focus on how our method performs in a larger heterogeneous clinical sample. As a robustness analysis, we did examine the performance of A_{region}, A_{feaure}, and A_{uni_combo} within the two largest ABIDE sites individually (Sites #5 and #20) to compare classification performance within versus across sites.

5.4.2 TIME-SERIES FEATURE EXTRACTION

Univariate features

For a given brain region, the BOLD time series can be summarized using a set of univariate time-series features. We opted to use the *catch22* library of univariate properties [91], which are listed in Table 5.4.2. The *catch22* features were drawn from the broader *hctsa* [89, 90] library, designed to be a highly explanatory subset of time-series features for generic time-series data mining problems [351] (though not defined specifically for neuroimaging datasets). To compare the performance of these features with a standard biomarker for quantifying localized resting-state BOLD activity, we also computed the fractional amplitude of low-frequency fluctuations (fALFF), which measures the ratio of spectral power in the 0.01–0.08 Hz range to that of the full frequency range [4]. fALFF is considered to be an index of spontaneous activity in individual brain regions and is fairly robust to the noise, scanner differences, and low sampling rates typical of

rs-fMRI data [5, 366, 367]. The *catch22* features were computed in R using the 'theft' package (version 0.4.2; [368]) using the `calculate_features` function with the arguments `feature_set = "catch22"`, which z-scores each time series prior to feature calculation. We also calculated the mean and standard deviation (SD) from the raw BOLD time series by including the `catch24 = TRUE` argument, which yielded a total of 24 univariate time-series features. We computed fALFF in Matlab as per the code implementation included in [255]. Collectively—the *catch22* features, mean, SD, and fALFF—comprise 25 univariate features that encapsulate a diverse set of statistical time-series properties with which to compare across different clinical exemplars. These 25 time-series features were computed across all brain regions and participants, yielding a three-dimensional array in the form of $N \times R \times F$—comprised of N participants, R brain regions, and F time-series features.

Pairwise features

To summarize different types of pairwise coupling, or functional connectivity (FC), across pairs of brain regions, we used a subset of 14 SPIs from the *pyspi* library of pairwise interaction statistics [92]. Each of these SPIs is listed along with its code name in 'pyspi' in Table 5.4.3, along with information about the literature category and directionality per SPI. These 14 SPIs contain a single representative from each of the 14 data-driven modules of similarly performing SPIs described in [92]. It includes the most commonly used FC metric for fMRI time series, Pearson correlation coefficient (denoted as 'cov_EmpiricalCovariance' in *pyspi*), enabling direct comparison of its performance to the other thirteen evaluated SPIs, which span nonlinear coupling (e.g., additive noise modeling), frequency-based coupling (e.g., power envelope correlation), asynchronous coupling (e.g., coherence magnitude), time-lagged coupling (e.g., dynamic time warping), and directed coupling (e.g., directed information). The 14 SPIs were computed from the BOLD time series for each pair of brain regions using the *pyspi* package (version 0.4.0) in Python [160]. Given the large number of MTS matrices to process (across all participants), we distributed *pyspi* computations on a high-performance computing cluster. Each SPI yielded a matrix of pairwise interactions between all brain regions, with a total of 6642 possible pairs for UCLA CNP (from R=82 brain regions) and 2256 pairs for ABIDE (from R=48 brain regions), after omitting self-connections. As some SPIs (like the Pearson correlation) are undirected, the resulting FC matrix is symmetric, and only the set of coupling values for all unique brain-region pairs (corresponding to the upper half of the FC matrix) is retained for these SPIs. Across all SPIs, *pyspi* computations yielded a three-dimensional array in the form of $N \times P \times S$, comprised of N participants, P region pairs (noting that this value will differ for directed versus undirected SPIs), and S SPIs.

5.4.3 CASE–CONTROL CLASSIFICATION

We sought to comprehensively compare the ability of diverse intra-regional time-series features and inter-regional coupling strengths to separate healthy controls from individuals with a given clinical diagnosis. Univariate, intra-regional time-series features are sensitive to local, region-specific disruptions in neural activity, whereas inter-regional SPIs are sensitive to disrupted coupling between pairs of brain regions. We first asked whether case versus control separation would be better with: (i) multiple dynamical properties measured from a single brain region; (ii) a single time-series property measured across brain regions; or (iii) the combination (the full set of time-series properties across all brain regions). We then asked

how pairwise FC measurements could capture case–control differences, and whether including our set of diverse univariate features would enhance SPI-wise classification performance. To compare the ability of univariate and/or pairwise features to separate case versus control individuals, we evaluated five different classification pipelines as enumerated in the following subsections, labeling each analysis as 'A_X' where X captures a short description of the analysis. These are listed as follows:

1. A_{region}: The performance of individual brain regions given all F univariate time-series features, assessed via classifiers fit to $N \times F$ (participant $\times$ feature) matrices as depicted in Fig. 5.2.1D(i). $R = 82$ models are fit with this representation for SCZ, BP, and ADHD, and 48 models for ASD.

2. $A_{feature}$: The performance of individual univariate time-series features measured from all brain regions, assessed via classifiers fit to $N \times R$ (participant $\times$ region) matrices as depicted in Fig. 5.2.1D(ii). $F = 25$ such models are fit for all four disorders.

3. A_{uni_combo}: The performance of all F univariate time-series features computed across all R brain regions, assessed via classifiers fit to $N \times RF$ (participant $\times$ brain-region—feature) matrices as depicted in Fig. 5.2.1D(iii). Only one such model is fit per disorder.

4. A_{FC}: The performance of the set of coupling strengths measured from all brain region pairs using an individual SPI, assessed via classifiers fit to $N \times P$ (participant $\times$ region–pair) matrices, as depicted in Fig. 5.2.1D(iv). $S = 14$ such models are fit for each disorder.

5. A_{FC_combo}: The performance of an individual SPI measured from all brain region pairs with the addition of all F univariate time-series features, assessed via classifiers fit to $N \times PRF$ (participant $\times$ region-pair–region–univariate-feature) matrices as depicted in Fig. 5.2.1D(v). $S = 14$ such models are fit for each disorder.

Classifier fitting and evaluation

Each of the analyses enumerated in the previous sections (e.g., A_{region}) refers to a way in which we extracted a set of interpretable time-series features from an fMRI dataset to capture different types of dynamical statistics. We sought to compare how these feature-based representations could distinguish cases from controls in each neuropsychiatric disorder using a simple measure of classification performance. Classification was evaluated using linear SVMs [369], which comprise a simple classical machine learning model that can handle large input feature spaces using regularization [370].

The UCLA CNP cohort is characterized by class imbalances (such that there are more control participants than those with a given neuropsychiatric disorder), so we applied inverse probability weighting, which weights each sample according to the inverse frequency of its corresponding class to increase the impact of the minority class on the decision boundary [371, 372]. Additionally, we evaluated classifier performance using balanced accuracy [273], which measures the average between classifier sensitivity and specificity to reflect accuracy in both the majority and minority classes, as recommended for neuroimaging data with class imbalances [372].

Classifiers were evaluated using stratified 10-fold cross-validation (CV) such that the hyperplane was defined based on the training data and used to predict the diagnostic labels of the unseen test data for each

fold. In order to reduce outlier-driven skews in time-series feature distributions, each training fold was normalized using the scaled outlier-robust sigmoidal transformation described in [90] and the same normalization was applied to each test fold to prevent data leakage [373]. We confirmed that this reduced more distributional skew of many univariate and pairwise time-series features compared to z-score normalization, as shown in Fig. S15. To ensure that results were not driven by any one specific CV split, each 10-fold CV procedure was repeated 10 times [374–376], yielding 100 out-of-sample balanced accuracy metrics per model. We report the mean $\pm$ SD tabulated across the 100 out-of-sample balanced accuracy metrics per model. As a robustness analysis, we subsequently compared the performance of each model (e.g., the left pericalcarine cortex in A_{region}) with the same 10-repeat 10-fold cross-validation as in the main analysis using each of: a linear SVM with L1 ('LASSO') coefficient regularization [377] and squared-hinge loss, an SVM with a radial basis kernel, a random forest ensemble, or a gradient boosted decision tree ensemble. The standard linear SVM with hinge loss and without L1 coefficient regularization yielded the highest classification performance overall (cf. Fig. S7A), and as this linear SVM represented the most parsimonious and interpretable of the evaluated models, we focus on classification metrics from this model throughout the paper. We also evaluated the effect of nested cross-validation to tune the regularization parameter, C, across values ranging between [0.0001, 100]—as well as the inclusion (or exclusion) of inverse probability weighting—finding that this also did not generally improve performance beyond our *a priori* parameter settings of $C = 1$ and applying inverse probability weighting (cf. Fig. S7B and Table S16).

The main linear SVM classifiers were implemented in Python with the `SVC` function from *scikit-learn* [378]) with a default value of 1 for the regularization parameter, C, and with the `class_weight = 'balanced'` argument for inverse probability weighting. The scaled outlier-robust sigmoidal transformation was applied using a custom Python script adapted from the Matlab-based `mixedSigmoid` normalization function from *hctsa* [89, 90]. We implemented 10 repeats of 10-fold CV using the `RepeatedStratifiedKFold` function from *scikit-learn* as part of a 'pipeline', which also automatically applied normalization based on training fold parameters and evaluated balanced accuracy for each test fold. We set the random number generator state for the `RepeatedStratifiedKFold` function such that all compared models were evaluated on the same resampled sets of test folds (for a total of 100 test folds). While the balanced accuracy is only computed for the default decision threshold for each classifier, we also computed the area under the receiving operator characteristic curve (AUC) for each test fold to evaluate model performance across thresholds. Models achieving greater than 50% balanced accuracy (i.e., better than chance) showed similar performance between balanced accuracy and AUC (Pearson's R between 0.77–0.97), so we focus on balanced accuracy in this work to align with the binary predictions of diagnostic group [372]. The other classifiers in our robustness comparison were also implemented with *scikit-learn*, using the `RandomForest`, `SVC`, `RandomForest`, or `GradientBoostingClassifier` functions, and nested cross-validation was performed with the `RandomizedSearchCV` function.

Dimensionality reduction and feature selection

While A_{region} and $A_{feature}$ both entail models with more samples than features, the other three representations—A_{uni_combo}, A_{FC}, and A_{FC_combo}—involve models with more features than samples, which risks overfitting to accumulated noise [125]. To test for such overfitting, we evaluated the change in performance after reducing the SVM feature space with principal components analysis (PCA), focusing on analysis A_{uni_combo} as a

starting point. PCA was implemented using the `PCA` function from the *FactoMineR* package in R (version 2.9) [379]. Since classification models based on individual brain regions yielded the lowest-dimensional classifiers (A_{region}: 25 time-series features per model), we selected the first 25 principal components (PCs) from the region $\times$ feature PCA for each case–control comparison group for consistency. Scores for these 25 PCs are provided in TableS17, and the first two PCs are plotted for each disorder in Fig. S6A. Each case–control comparison was evaluated using these 25 PCs with the same repeated 10-fold classification procedure as described above.

To fairly compare model performance across representations, we also performed data-driven feature selection for all representations except for A_{uni_combo}. For A_{region} and $A_{feature}$—the two representations where $p < n$—we implemented a strategy in which for each of the 100 folds (10 folds resampled 10 times), we identified the brain region (for A_{region}) or time-series feature (for $A_{feature}$) that yielded the best in-sample training fold balanced accuracy, and retained its corresponding out-of-sample test balanced accuracy. For example, for A_{region} in SCZ, in fold $\#1$ out of 100, the left pericalcarine might yield the highest in-sample balanced accuracy, so we would retain the out-of-sample balanced accuracy for the left pericalcarine for this fold. For fold $\#2$, the right cuneus might yield the highest in-sample balanced accuracy, so we would retain the right cuneus out-of-sample balanced accuracy, and so on for the rest of the folds. In the event of a tie for a given fold, we computed the mean out-of-sample balanced accuracy between the tied brain regions. As only one model was evaluated for A_{uni_combo}, we simply examined the distribution of 100 out-of-sample balanced accuracy values.

By contrast, A_{FC} and A_{FC_combo} both involve $p \gg n$ feature matrices that increase the potential for overfitting, so multiple SPIs yielded 100% in-sample balanced accuracy for each training fold— precluding the utility of this method for selecting one representative property of inter-regional coupling per fold. As an alternative selection criterion, we fit a PCA for each SPI (or SPI with local univariate properties for A_{FC_combo}) and retained the first 10 PCs. Within each training fold, we supplied these 10 PCs as input feature matrices to the same linear SVM classifier as above, in order to identify the SPI which (in PC-space) yielded the highest in-sample training balanced accuracy. Upon identifying that top-performing SPI, we used its full feature-space matrix as the input to a linear SVM classifier to measure its out-of-sample balanced accuracy for the corresponding fold. In the event of a tie for a given fold, we computed the mean out-of-sample balanced accuracy between the tied SPIs (assessed in the full feature space). This selection method provided a heuristic for selecting the top SPI in a lower-dimensional space while preserving input information for the out-of-sample classification metrics we evaluated across representations.

Statistical significance

To make statistical inferences about how a given classifier performed relative to chance level, we calculated the probability of observing a given mean balanced accuracy relative to a null distribution. For each model, we fit 1000 SVM classifiers with randomly shuffled diagnostic class labels using the same 10-repeat 10-fold CV as described above—yielding 1000 null balanced accuracy estimates per model. After confirming that all evaluated models yielded approximately normal null distributions centered at 50% (i.e., chance or predicting all the same class) balanced accuracy (Fig. S16), we derived the mean and standard deviation from each null distribution. Instead of fitting a full non-parametric permutation test for each specific model (with e.g.

10,000 or more permutations), we instead imposed a Gaussian distribution with two moments (i.e., mean and SD) to approximate the cumulative distribution function per model. Each observed balanced accuracy metric was compared with its corresponding cumulative distribution function using the `pnorm` function in R to obtain a one-tailed P-value, capturing the probability of obtaining a null balanced accuracy greater than or equal to the observed balanced accuracy. The cumulative distribution function step considerably reduced computational expenses, enabling the measurement of very small P-values that would be computationally prohibitive to detect in a full nonparametric treatment considering the number of models tested throughout this study. In order to correct for multiple comparisons within each disorder, we applied Benjamini–Hochberg correction [380] to control the false discovery rate at the $\alpha = 0.05$ level. All Benjamini–Hochberg corrected P-values are denoted as P_{corr} (unless otherwise specified) and the number of comparisons are indicated as appropriate.

Corrected T-statistics

To quantify the change in classifier performance for pairwise FC matrices with versus without the inclusion of univariate region $\times$ feature data (from $\mathsf{A}_{\mathsf{uni\_combo}}$), we compared the balanced accuracy distributions across all 100 test folds for the corresponding two classifier inputs per SPI. The standard T-test for group means is suboptimal in this case, as the participant overlap across test folds in the two compared models violates the assumption of independent samples [381]. We instead implemented a corrected two-tailed test statistic T_{corr} designed specifically for repeated k-fold cross-validation, as defined in [382] based on the original corrected T-statistic described in [383]. This correction was applied using the `repkfold_ttest` function from *correctR* package in R (version 0.2.1) [291].

Confound analysis based on age, sex, and head motion

The proportion of males versus females was compared between each disorder and the corresponding control group using a χ^2 test, implemented in R with the `chisq_test` function from *rstatix* (version 0.7.2). We compared age and mean framewise displacement (FD) distributions for each disorder and the corresponding control groups using Wilcoxon rank-sum tests, implemented in R with the `wilcox_test` function from *rstatix* (version 0.7.2). To evaluate how well these confound variables could predict clinical diagnosis on their own, we fit the same 10-repeat 10-fold linear SVM classification pipeline to sex data (with male encoded as 0 and female encoded as 1), age data, and mean FD data. Results from this robustness analysis are depicted in Fig. S3D. We used these classification metrics as a baseline from which we could assess the improvement in case–control distinction using BOLD-derived time-series features. In addition to the performance of these confound variables on their own and that of all the models we evaluated through our main analysis, we additionally evaluated the classification performance of models trained on BOLD time-series features plus confound variables as a robustness analysis. These classification metrics are shown side-by-side in Fig. S3E.

Volumetric analysis of group differences

In order to test whether the balanced accuracy of a given brain region showed a linear relationship with gray matter volume differences in case–control comparisons, we compared the number of voxels in each

brain region per participant in the UCLA CNP cohort (as volumetric data from the Harvard–Oxford cortical atlas is not available for the ABIDE ASD cohort). We used the *fmriprep* anatomical derivative 'aparc+aseg' parcellation and segmentation volumes that were mapped to MNI152 space using nonlinear alignment and tabulated the number of voxels in each region. Of note, these segmentation volumes correspond to the Desikan–Killiany–Tourville atlas [384] in which voxels for three brain regions (frontal pole, temporal pole, and banks of the superior temporal sulcus) were redistributed into other regions—thus precluding volumetric analyses for those three regions. For each brain region, we regressed mean volume (measured in the number of voxels) on diagnosis for each case–control comparison using an ordinary least squares model, with the resulting β coefficients listed in Supplementary Table S2. We examined the magnitude of the β coefficient estimated for each brain region to evaluate whether volumetric differences relative to controls were related to that region's balanced accuracy. The Pearson correlation coefficient was computed to measure the linear association between β coefficient magnitude and classification performance for all 79 brain regions per disorder. Additionally, to evaluate whether the overall volume of a region was related to how well its resting-state activity distinguished cases from controls, we compared the mean number of voxels per region across participants per disorder with the disorder-wise mean balanced accuracy per region using the Pearson correlation coefficient.

5.4.4 DATA VISUALIZATION

All figures in this study were compiled using R (v. `4.3.2`) using the *ggplot2* package (v. `3.5.1`; [385]) unless otherwise specified. Brain maps were visualized using the *ggseg* package (v. `1.6.5`; [386, 387]) along with the *ggsegHO* package (v. `1.0.2.001`) for the Harvard-Oxford cortical atlas. Raincloud plots were generated with the `geom_violinhalf` function from the *see* package (v. `0.8.3`; [388]). Heatmaps with dendrograms were created using the `ComplexHeatmap` function from the *ComplexHeatmap* package (v. `3.18`, [389, 390]). Venn diagrams were created using the `venn.diagram` from the *VennDiagram* package (v. `1.7.3`; [391]).

5.4.5 CODE AND DATA AVAILABILITY

All data used in this study is openly accessible, with preprocessed data provided in Zenodo [392]. UCLA CNP rs-fMRI imaging data can be downloaded from OpenNeuro with accession number ds000030 (https://openfmri.org/dataset/ds000030/). ABIDE rs-fMRI regional time series can be downloaded from Zenodo [270]. All code used to compute univariate and pairwise time-series features, perform classification tasks, and visualize results is provided on GitHub and Zenodo [269].

AUTHOR CONTRIBUTIONS

B.D.F. jointly conceived the study with A.F. B.D.F. supervised the project. K.A. and L.P. preprocessed data and provided advice for data analysis. A.G.B. analyzed data and wrote the original manuscript draft with B.D.F. All authors discussed results and interpretations and provided commentary throughout manuscript revision.

ACKNOWLEDGMENTS

The authors thank Kieran Owens and Mac Shine for their feedback and suggestions for the manuscript and all participants and researchers involved in creating and openly sharing the fMRI datasets analyzed here. The authors also thank Patricia Tran and Preethom Pal for their contributions to related prior research projects.

FUNDING AND COMPETING INTERESTS STATEMENT

L.P. was supported by the National Institute Of Mental Health of the National Institutes of Health under Award Number R00MH127296. K.A. is an employee and shareholder of BrainKey Inc., a medical image analysis software company, with all contributions to the present work made prior to commencing this position.

TABLES

Table 5.4.1: Key demographics for participants included in this study after quality control exclusion processes.

Study	Diagnosis	*N*	% Female (*N*)	Age; Mean (SD)
ABIDE	ASD	513	14.6 (75)	17.5 (10.5)
	Control	578	26.0 (152)	17.1 (9.5)
UCLA CNP	ADHD	39	48.7 (19)	31.6 (10.1)
	BP	49	42.9 (21)	35.3 (9.0)
	SCZ	48	25.0 (12)	36.6 (9.0)
	Control	116	46.6 (54)	31.2 (8.7)

Table 5.4.2: The univariate time-series features computed for each brain region to capture intra-regional dynamics. For each univariate time-series feature (shown in the 'Feature name' column), a brief description is given for the property the corresponding feature captures.

Feature name	Feature description
ACF_first_min	First minimum of the autocorrelation function (ACF)
ACF_timescale	First 1/e crossing of the ACF
AMI_timescale	First minimum of the automutual information function (linear version)
AMI2	Histogram-based automutual information (lag 2; 5 bins)
centroid_freq	Frequency corresponding to the centroid of the power spectral density
DFA	Detrended fluctuation analysis (low-scale scaling)
embedding_dist	Goodness of exponential fit to embedding distance distribution
entropy_pairs	Entropy of successive pairs of values in symbol-discretized time series
fALFF	Ratio of spectral power in the 0.01-0.08Hz range to the full frequency range
forecast_error	Error of 3-point rolling mean forecast
high_fluctuation	Proportion of high incremental changes in the series
low_freq_power	Power in the lowest 20% of frequencies
mean	Mean of the time series values
mode_10	Mode of a 10-bin histogram of time-series values
mode_5	Mode of a 5-bin histogram of time-series values
outlier_timing_neg	Negative outlier timing (higher = later in time series)
outlier_timing_pos	Positive outlier timing (higher = later in time series)
periodicity	First local maximum of the ACF following Wang et al. [276] criteria
rs_range	Rescaled range fluctuation analysis (low-scale scaling)
SD	Standard deviation of time series
stretch_decreasing	Longest stretch of decreasing values
stretch_high	Longest stretch of above-mean values
transition_variance	Transition matrix column variance
trev	Time reversibility (nonlinear autocorrelation measure)
whiten_timescale	Change in autocorrelation timescale after incremental differencing

Table 5.4.3: The 14 statistics for pairwise interactions (SPIs) computed across brain region pairs. For each statistic of pairwise interactions (SPI; shown in the 'SPI' column), a brief description is given for the type of interaction that the corresponding SPI captures.

SPI	**Description**
PSI_frequency	Phase slope index (PSI) of the frequency domain, measures information flow based on complex-valued coherency
ANM	Additive noise model, tests for directed nonlinear dependence of x –>y
DI	Directed information, measures information flow from one time series to another without any time-lags
transfer_entropy	Transfer entropy, measures information flow from one time series to another with Takens time-delay embedding
phi_star	Phi-star, integrated information proxy that measures information lost when two time series are disconnected
spectral_GC	Spectral Granger causality, frequency domain equivalent to Granger causality (nonparametric)
PLI	Phase lag index, measures phase synchronization across the full frequency range
PSI_time_frequency	PSI of the time-frequency domain, measures information flow based on complex-valued coherency
barycenter_DTW	Barycenter, the univariate time series that captures the center of mass between two time series
DTW	Dynamic time warping, measures the minimum distance between two (potentially dilated) time series
power_envelope_corr	Power envelope correlation, measures correlation between power envelopes of two time series
coherence_magnitude	Coherence magnitude, measures mean coherence between two time series for full frequency range
cointegration	Cointegration, tests if the linear combination of two time-series has a lower integration order
Pearson	Pearson correlation coefficient – computed as empirical covariance, which is equivalent as we z-score data

6

Mapping functional, structural, and molecular correlates of homotopic connectivity

Left brain, right brain, or duality of man...
Michael Scott, The Office, Season 2 Episode 6

Amidst the complex interactions within and between the two hemispheres of the cerebral cortex, one particularly intriguing phenomenon is the presence of homotopic functional connectivity (HoFC). HoFC refers to the synchronized activity between homologous regions in the left and right hemispheres, and forms a critical component of the brain's structural and functional architecture. While several studies have highlighted HoFC as one of the strongest features of the functional topology of the brain, even defying typical constraints based on spatial distance and structural wiring between regions, there is still relatively little known about the precise mechanisms that orchestrate HoFC. To date, most research has quantified HoFC using the Pearson correlation coefficient, which (as discussed in Chapters 1 and 5) is a linear measure of contemporaneous coupling between a given region–region pair. In this chapter, we sought to comparatively quantify HoFC through a multitude of functional connectivity measures from the *pyspi* library (cf. Chapter 3, [92]), including linear, nonlinear, and time-lagged measures from multiple disciplines.

However, in quantifying over 200 FC measures in all homotopic region–region pairs in the atlas from Desikan et al. [169] in $N = 100$ Human Connectome Project (HCP) participants [161], we found that a large subset of measures exhibited very similar patterns of HoFC to that of the Pearson correlation coefficient. As shown in Figure 6.0.1A, one subset of measures spanning time- and frequency-based coupling (both directed and undirected), mutual information, time-series alignment, regression-based fits, and distance measures all exhibited Spearman rank correlation values of $|\rho| > 0.8$ with the Pearson correlation coefficient (many in excess of $|\rho| > 0.9$). For example, one of the information-theoretic measures that emerged as particularly informative across applications in Chapters 3 and 5 is directed information with a Gaussian density estimator (first introduced in Chapter 2). While directed information is conceptually and algorithmically distinct from the Pearson correlation coefficient, across all homotopic region–region pairs and individuals, they were very highly correlated (Spearman ρ=0.96), as shown in Fig. 6.0.1B. Consistent with our emphasis

on interpretability through methodological parsimony, we therefore progressed with the Pearson correlation coefficient as the representative HoFC measure in this chapter. This choice was made not only for its simplicity and interpretability, but also because it is a well-established measure in the literature, allowing for direct comparisons with previous findings. In this chapter, we explore Pearson correlation-based HoFC to investigate potential structural, functional, and molecular correlates that may underlie this phenomenon.

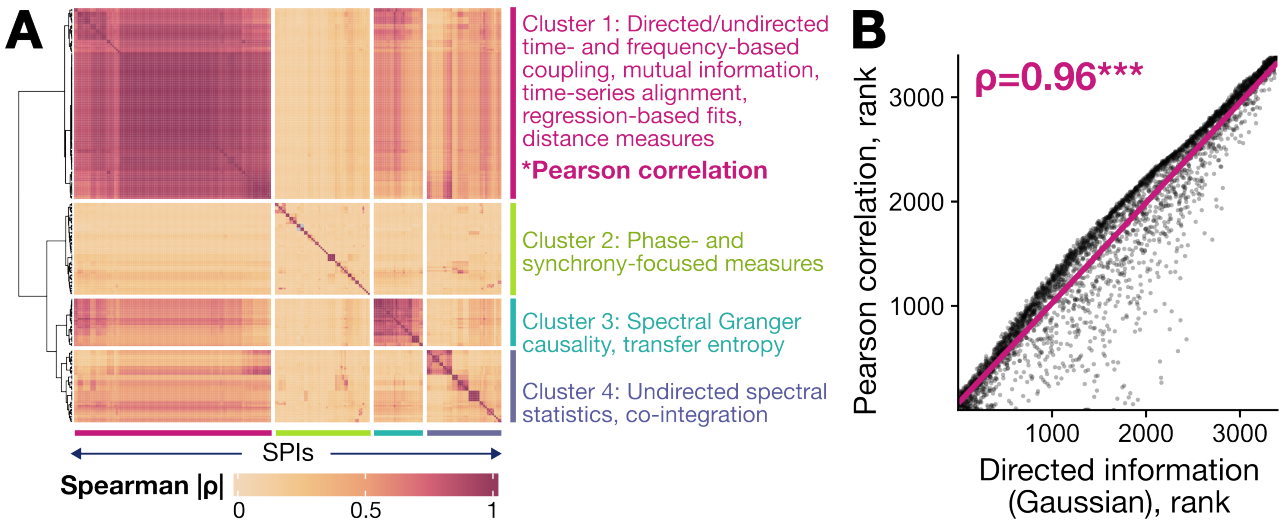


Figure 6.0.1: Many measures yield highly correlated homotopic functional connectivity (HoFC) values, including directed information. A. This heatmap depicts the absolute Spearman rank correlation (ρ) between each pair of functional connectivity measures (SPIs; with 228 in total, from the *pyspi* library [92]) computed across all homotopic region–region pairs in the atlas from Desikan et al. [169] (as in Chapters 2 and 5). The dataset comprises resting-state fMRI data from 100 unrelated participants in the S1200 release of the Human Connectome Project [161]. The measures are sorted along rows and columns based on hierarchical clustering of the correlation matrix using the algorithm from Ward Jr [393], with the dendrogram on the left side indicating the clustering structure (with a manually defined threshold of four clusters). **B.** The resulting values (converted to ranks per measure) are compared between the Pearson correlation coefficient (y-axis) and directed information with a Gaussian density estimator (x-axis) for each homotopic region–region pair per individual, for a total of 34$\times$100 = 3400 pairs. The Spearman rank correlation coefficient (ρ) is annotated.

This chapter is closely related to a manuscript in submission as: Bryant, A. G., Shine, J. M., & Fulcher, B. D. Mapping functional, structural, and transcriptomic correlates of homotopic connectivity. *bioRxiv* preprint (2025), doi: 10.1101/2025.06.10.657943.

Abstract

Homotopic functional connectivity (HoFC)—the synchronous activity between homologous regions of the left and right cortical hemispheres—is a hallmark of inter-hemispheric brain architecture, yet its biological underpinnings remain incompletely understood. Here, we characterize spatial variation in resting-state HoFC and its relation to dominant natural axes of anatomical, functional, and transcriptomic organization across the human cerebral cortex. We show that the regional distribution of HoFC, from the lowest in association areas (especially limbic) to the highest in primary sensorimotor regions, is preserved in neuropsychiatric disorders in a separate neuroimaging cohort (including schizophrenia, bipolar I disorder, and attention-deficit hyperactivity disorder) despite disease-associated perturbations in magnitude. Regional variation in HoFC is not fully explained by properties of physical embedding or vascular factors alone; rather, it is associated with cortex-wide resting-state synchrony and thalamic connectivity, as well as expression of layer 5 intra-telencephalic pyramidal neuron markers. These findings suggest that homotopic coupling is related to both functional network dynamics and specific neuroanatomical substrates, consistent with shared subcortical drive, highlighting candidate mechanisms that may support resting-state interhemispheric coupling. Collectively, this work advances our understanding of the biological principles and mechanisms underlying cortical homotopy.

6.1 INTRODUCTION

The left and right hemispheres of the cerebral cortex exhibit a spectrum of lateralized functions and integrated processes that collectively underpin behaviors ranging from sensory perception to higher-order cognition. Coordination between the hemispheres is characterized in part by functional connectivity between 'homotopic' regions of the brain—homologous areas in the left and right hemispheres—which comprises a core feature of macroscopic brain organization. Indeed, converging evidence indicates that in the resting-state brain (i.e., in the absence of a cognitive task), homotopic regions show greater synchrony than intra-hemispheric or heterotopic region–region pairs [41, 394–398]. Such inter-hemispheric coupling emerges in embryonic development [399] and is maintained throughout the lifespan, with healthy aging-related increases in sensorimotor areas and decreases in higher-order associative regions [7]. Homotopic functional connectivity ('HoFC') has been documented across species—including rodent [400–402], zebra finch [397], and macaque [403, 404]—and is among the most reproducible and temporally robust features observed in resting-state functional magnetic resonance imaging (fMRI) [404, 405], as well as in rodent wide-field calcium imaging [52]. While the pronounced functional coupling between homologous regions is conserved across development and evolution, its relationship with macroscale properties of brain structural and functional organization remains incompletely understood.

Computational modeling analyses have suggested that properties of both geometric embedding (e.g., physical distance or number of streamlines between a pair of regions) and homophilic attraction (e.g., similarity in gene expression or cytoarchitecture) jointly shape the cortical connectome [406]. Within this network-based framework, homotopic connections coincide with greater similarity in gene-expression patterns, receptor density profiles, cortical laminarity, and hemodynamic connectivity than heterotopic connections [41]. HoFC appears to be physically supported by dense structural interconnections via callosal fibers [407], which are

considerably more numerous between homotopic than non-homotopic regions in the human brain [408]. Homotopic connections appear somewhat anomalous in terms of cortical geometry, however. For example, Salvador et al. [409] demonstrated that HoFC does not decrease as a function of distance between the two homotopic regions, which has been independently validated in subsequent studies [404, 410]—suggesting that HoFC does not adhere to the general inverse association between resting-state functional connectivity (FC) magnitude and inter-regional distance [49, 409]. Moreover, while the majority of callosal fibers innervate homotopic regions (as opposed to intra-hemispheric or heterotopic region–region pairs) [408], Uddin et al. [411] reported preserved HoFC following complete commissurotomy, in which the corpus callosum is surgically severed in the case of treatment-resistant epilepsy. Since a commissurotomy disrupts monosynaptic callosal projections, the extant HoFC is hypothesized to be supported (in part) through alternate polysynaptic routes—potentially via the anterior commissure, which is known to connect homotopic regions in the temporal cortex [412]. Indeed, O'reilly et al. [413] reported that inter-hemispheric coupling in the rhesus monkey was largely preserved after severing callosal projections, as long as the anterior commissure was left intact. Resting HoFC has been reported in mammals without a corpus callosum by ontogeny [397], further pointing to alternative (or potentially supplementary) physical mechanisms underpinning homotopic connectivity. Collectively, this raises important questions about the neurophysiological mechanisms that underpin inter-hemispheric connectivity throughout the brain.

One of the dominant axes of macroscopic organization in the cortex follows an organized and gradual spatial transition of functional connectivity properties [71, 414]. The first principal gradient of functional connectivity, derived in Margulies et al. [414], places primary sensorimotor regions at one end and higher-order transmodal and association regions at the other [415]—aligning with a canonical 'functional hierarchy' in the cerebral cortex [416]. This spatially ordered functional gradient naturally maps onto the topological hierarchy of the cortex [71], originally described to encapsulate synaptic architecture in the primate visual cortex [417, 418]—where 'topological' refers to the number of synaptic steps away from the retina. The brain's topological organization is hypothesized to mechanistically underpin hierarchical functional specialization [419, 420]. Emerging evidence suggests that HoFC also systematically varies along this principal functional gradient; for example, Stark et al. [394] characterized the regional heterogeneity in HoFC, identifying a spectrum ranging from the lowest HoFC (transmodal association regions) up to the highest HoFC (primary sensory-motor regions). The authors posit that the higher linear correlation between BOLD activity in primary sensory regions reflects a 'default state' of inter-hemispheric synchrony, which decreases in favor of lateralized hemisphere-specific processing as information is passed up the putative functional hierarchy [71, 414]. Other groups have reported similar patterns from primary unimodal to heteromodal to association cortices [421, 422]—though inter-hemispheric synchrony may ramp up among higher-order regions in the context of a complex task [423, 424].

Amid this regional heterogeneity, we currently lack a comprehensive account of how region-specific variation in HoFC interfaces with broader macroscopic cortical organization—a gap that must be bridged to better understand inter-hemispheric communication in the resting brain, in both health and disease. Here, we focus on regional HoFC variation across the cortex in the resting state, aiming to investigate how the spatial distribution of HoFC relates to macroscale gradients of functional, anatomical, and transcriptomic variation across the cortex. Hypothesizing that resting HoFC is linked to a region's position along the dominant axes of functional and anatomical organization in the cortex, respectively, we specifically compared

regional HoFC with brain-wide maps capturing anatomical hierarchy, resting-state functional connectivity, and transcriptomic variation. We also characterize the robustness of regional variation in HoFC magnitudes in the context of distinct neuropsychiatric disorders (in an independent neuroimaging cohort), where disruptions to inter-hemispheric coupling are frequently reported [425–434], and examine how homotopic coupling relates to cortex-wide resting-state functional connectivity within the ipsilateral and contralateral cortices. Additionally, we explore several potential physiological bases for cortical HoFC derived from physical embedding, structural connections, and vascular innervation, along with molecular underpinnings and thalamocortical connectivity. Collectively, our analyses contribute to our understanding of the structural, functional, and molecular correlates of homotopic connectivity in the human cerebral cortex, suggesting promising avenues for future hypothesis-driven investigation.

6.2 RESULTS

In order to quantitatively investigate macroscale patterns of inter-hemispheric coupling, we characterize the spatial distribution of homotopic functional connectivity (HoFC) across the human cerebral cortex. We analyze openly available data from the Human Connectome Project (HCP) [161], with preprocessed group-consensus structural and functional connectomes provided by the ENIGMA Consortium [435] and described in depth previously [436]. Briefly, Larivière et al. [435] first computed individual structural and functional connectomes for each of $N = 207$ participants; for the functional connectome, negative connections were set to zero and Fisher's r-to-z transformation [437] was applied (as in previous graph-theoretical analyses [438, 439]), before averaging across participants to yield a consensus connectome. More details about this dataset are included in the Methods (Sec. 6.5.1). We first assess how the spatial variation in HoFC magnitude relates to maps of different properties of macroscopic cortical organization along anatomical, functional, and transcriptomic gradients included in the *neuromaps* library [440] (in Sec. 6.2.1). After characterizing these relationships, we then evaluate the robustness of the spatial variation in HoFC magnitudes to three neuropsychiatric disorders in which inter-hemispheric synchrony is reportedly disrupted, querying the UCLA Consortium for Neuropsychiatric Phenomics dataset [261] (in Sec. 6.2.2). To interpret HoFC in the context of the broader topological landscape of the cerebral cortex, we additionally compare the average functional network strength with HoFC magnitude (Results, Sec. 6.2.3). Finally, we investigate potential molecular and connectomic underpinnings of the spatial variation in HoFC by comparing the map to canonical transcriptomic markers of layer 5 pyramidal neurons (using the Allen Human Brain Atlas [441]) and to thalamic connectivity (Results, Sec. 6.2.5).

6.2.1 HOMOTOPIC FUNCTIONAL CONNECTIVITY VARIES ALONG A FUNDAMENTAL AXIS OF CORTICAL ORGANIZATION

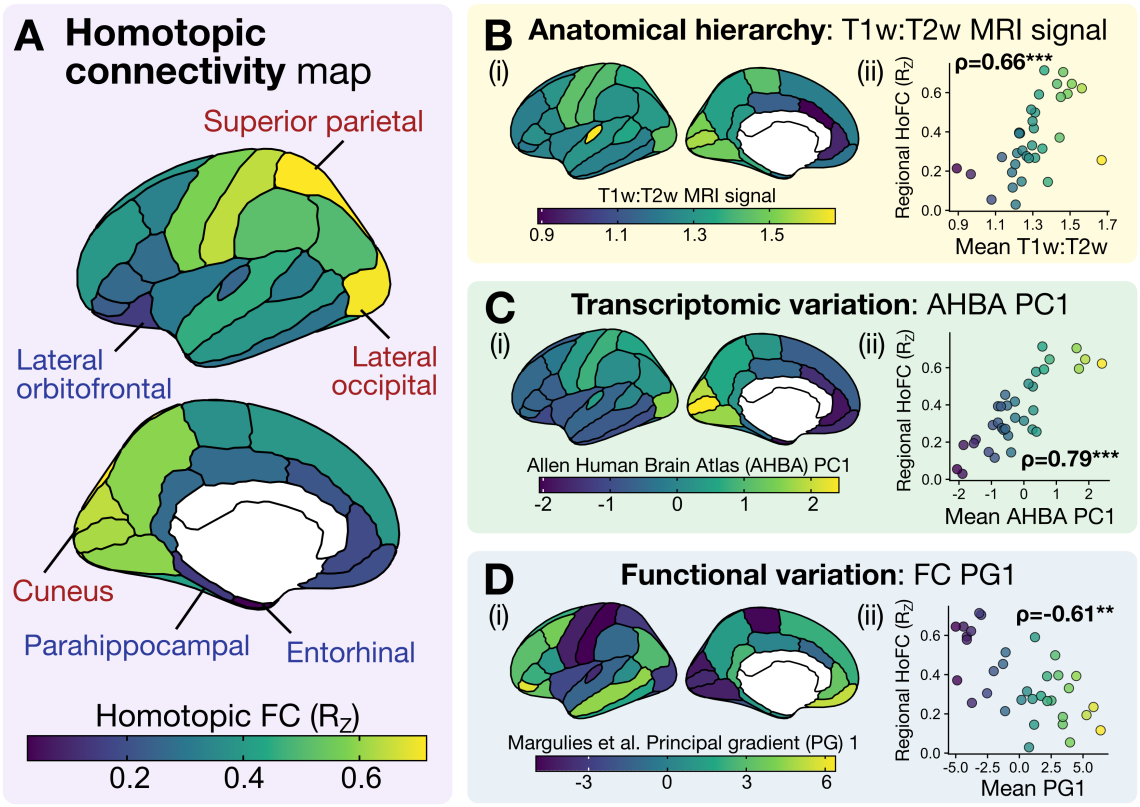


Figure 6.2.1: Regional homotopic functional connectivity (HoFC) variation aligns with a fundamental axis of macroscale cortical organization. A. The group consensus z-transformed Pearson correlation coefficient (R_Z) between homotopic region pairs is depicted on the cortical surface. Regions with the highest HoFC are annotated with red text (superior parietal, lateral occipital, and cuneus) and regions with the lowest HoFC visible from medial/lateral views are annotated with blue text (lateral orbitofrontal, parahippocampal, entorhinal). **B.** (i) The mean T1w:T2w MRI signal is plotted for each region (averaged between the two hemispheres) in the same parcellation atlas (from Desikan et al. [169]) as the HoFC maps in **A**. (ii) The mean T1w:T2w MRI signal (x-axis) is compared with the HoFC value (y-axis) per region, with the Spearman rank correlation (ρ) annotated. **C.** (i) The mean first principal component of gene expression from the Allen Human Brain Atlas (AHBA PC1) is plotted for each region (averaged between the two hemispheres). (ii) As in **B**(ii), the AHBA PC1 value (x-axis) is compared with the HoFC value (y-axis) per region. **D.** (i) The mean first principal gradient (PG1) of functional connectivity (FC) from Margulies et al. [414] is plotted for each region (averaged between the two hemispheres). (ii) As in **B**(ii), the PG1 value (x-axis) is compared with the HoFC value (y-axis) per region. $^{***}, P < 0.001$; $^{**}, P < 0.01$; single-hemisphere spin test with 10,000 permutations [442].

We first characterized the spatial distribution of HoFC magnitudes across the cerebral cortex from a consensus connectome derived from $N = 207$ individuals from the HCP S1200 release (see Methods, Sec. 6.5.1). For each cortical region in the parcellation atlas from Desikan et al. [169] (34 regions per hemisphere), we examined the z-scored Pearson correlation coefficient (R_Z) between the left and right hemispheres from the group consensus connectome. This approach yielded a cortex-wide map of HoFC magnitudes, allowing us to quantitatively compare regional HoFC gradients with other properties of macroscale cortical organization. As shown in brain maps in Fig. 6.2.1A, we observed a clear gradient of HoFC magnitudes across

the cortex, from the lowest values in the entorhinal cortex ($R_{\mathrm{Z}} = 0.05$) and temporal pole ($R_{\mathrm{Z}} = 0.03$; not visible in Fig. 6.2.1A) to the highest in the superior parietal cortex ($R_{\mathrm{Z}} = 0.71$) and lateral occipital cortex ($R_{\mathrm{Z}} = 0.70$). This pattern is consistent with previous analyses using the HCP dataset [421, 422] as well as different imaging cohorts [394, 410, 421], supporting prior findings that resting-state HoFC is highest in primary motor, visual, and somatosensory regions and lowest in association cortices.

Margulies et al. [414] demonstrated that the first principal gradient (PG1) of resting-state functional connectivity (FC) systematically varies across the cortical sheet in a way that tracks the topological hierarchy of laminar projections, ranging from primary sensory regions to transmodal to association cortices [71]. Here, we focus on properties of concerted variation in macroscale cortical organization and their relation to a 'fundamental axis' of the cerebral cortex [416], with 'hierarchy' referring specifically to the anatomical organization of laminar feed-forward and feed-back patterns of inter-areal anatomical connections (originally identified in non-human primate cortex with histological tract-tracing) analysis [418, 443, 444]. Specifically, we examine three distinct types of graded spatial variation across the cortical sheet using curated reference maps from the `neuromaps` toolbox [440]. Given the proposed links between HoFC and white matter microstructural properties [395], we first queried the ratio of T1-weighted to T2-weighted (T1w:T2w) signal from structural MRI, which is sensitive to intracortical myelin content (as well as several other microstructural properties of neurons) [445]. Burt et al. [446] subsequently demonstrated that T1w:T2w is a reliable non-invasive proxy for the anatomical hierarchy of the cortex. We additionally examined transcriptional organization across the cortex using the first principal component (PC1) of brain-wide gene expression from the Allen Human Brain Atlas, (AHBA) [441], which captures maximal variance in gene expression and provides a proxy for cell-type distributions across the cortex [446–448]. Lastly, we queried the FC PG1 to capture maximal variance in resting-state FC the cortical sheet, positioning primary sensorimotor regions at one end and higher-order association regions at the other [414].

We compared the regional HoFC variation (as shown in Fig. 6.2.1A) with the mean value in the corresponding area of each map of macroscale cortical organization, averaging between the left and right hemispheres for each in the latter case. Statistical significance was evaluated using a non-parametric single-hemisphere spin test [442] with 10,000 permutations (see Methods for details, Sec. 6.5.3). Of note, the T1w:T2w and AHBA gradients exhibit a positive correlation in their respective spatial variations [446] and a negative correlation with that of PG1 magnitudes; in other words, T1w:T2w and AHBA values are generally negatively correlated with the anatomical hierarchy of the cortex, while PG1 values exhibit a positive association. As depicted in Figs 6.2.1B-D, we found that the spatial variation in HoFC magnitude is significantly correlated with all three maps of macroscale cortical organization—in a manner that suggests HoFC is inversely associated with the anatomical hierarchy in the cortex. Specifically, we observed a positive correlation to the T1w:T2w map ($\rho = 0.66$; $P_{\mathrm{spin}} = 5 \times 10^{-4}$, single-hemisphere spin test with 10,000 permutations) and transcriptomic AHBA PC1 map ($\rho = 0.79$; $P_{\mathrm{spin}} = 2 \times 10^{-4}$) and a negative correlation to the PG1 map ($\rho = -0.61$; $P_{\mathrm{spin}} = 1 \times 10^{-3}$). Regions with the highest-magnitude resting HoFC ($R_{\mathrm{Z}} \geq 0.6$) exhibited among the highest values of T1w:T2w MRI signal (Fig. 6.2.1B), corresponding to lower positions along the anatomical hierarchy. These same high-magnitude HoFC regions also sit at the apex of AHBA PC1 gene expression values (Fig. 6.2.1D), which tracks a molecular gradient in which higher values correspond to greater microglial and CA1 pyramidal neuron density and downregulation of cell signaling and modification pathways [449]. Similar alignment between T1w:T2w and AHBA PC1 gradients was reported in Burt et al.

[446], suggesting these two properties converge toward a common axis of anatomical and transcriptomic regional variation. Lastly, the high-magnitude HoFC regions exhibited among the lowest functional PG1 values (Fig. 6.2.1C), which maps to the primary unimodal end (rather than association) of the dominant axis of functional connectivity described in Margulies et al. [414]. Collectively, our results suggest that HoFC spatial variation tracks primary axes of anatomical hierarchy and macroscale organization across the cortical sheet.

6.2.2 SPATIAL VARIATION IN HOMOTOPIC FUNCTIONAL CONNECTIVITY IS PRESERVED ACROSS NEUROPSYCHIATRIC DISORDERS

We next posed two follow-up questions: (1) How reproducible is the spatial variation in resting-state HoFC in another young and clinically normative population?; and (2) Is this macroscale HoFC gradient altered in the context of disease? For the latter, we focused on neuropsychiatric disorders, which are generally attributed to dysfunction across distributed brain networks [247, 249, 275, 293], with widespread alterations to functional coupling throughout the cerebral cortex. Mounting evidence indicates that HoFC is reduced across the brain in a multitude of neuropsychiatric disorders, including anxiety and depression [425–427], autism spectrum disorder [428–430], bipolar disorder [431], and schizophrenia [432–434], suggesting that reduced HoFC may serve as a transdiagnostic biomarker [450]. However, it remains to be clarified whether this general reduction in HoFC exhibits characteristic spatial changes in a given disorder or manifests diffusely throughout the cortex. To address both of these questions, we investigated the spatial distribution of resting-state HoFC magnitudes in participants in a neurotypical control group ($N = 116$) or those diagnosed with one of three neuropsychiatric disorders included in the UCLA Consortium for Neuropsychiatric Phenomics (CNP) study [261]: schizophrenia (SCZ; $N = 48$), bipolar I disorder (BP; $N = 49$), and attention-deficit hyperactivity disorder (ADHD; $N = 39$). More information about data acquisition and preprocessing with this cohort is provided in the Methods (Sec. 6.5.1).

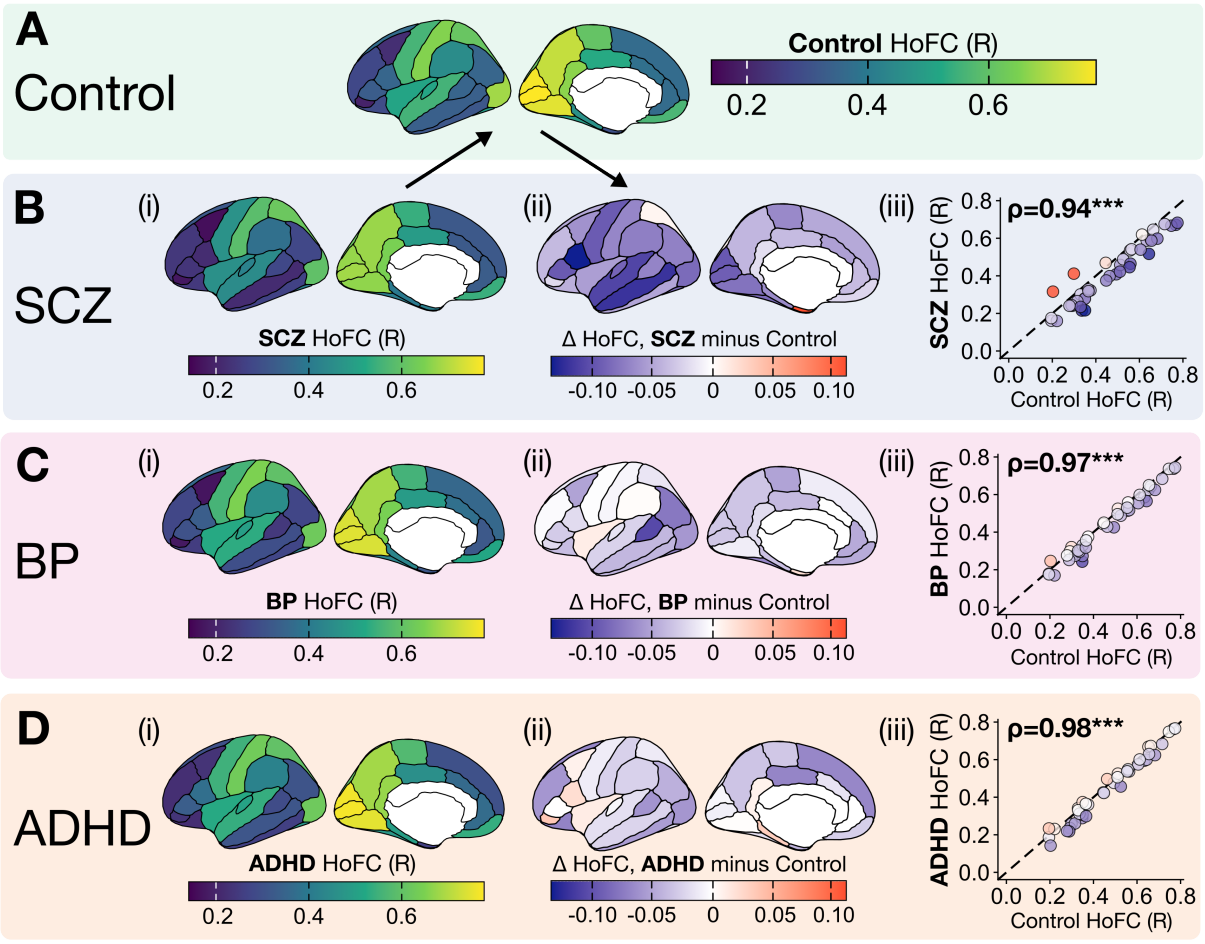


Figure 6.2.2: The macroscale HoFC regional gradient is preserved across neuropsychiatric disorders, despite heterogeneous disease-associated perturbations in inter-hemispheric coupling. A. The regional HoFC gradient is depicted on the cortical surface for control participants (N=116) in the UCLA CNP cohort. HoFC values reflect group-averaged Pearson correlation coefficient values (R); note that correlation values were not z-scored in the preprocessing pipeline for this dataset. **B.** (i) The regional HoFC gradient is depicted on the cortical surface for schizophrenia (SCZ) participants (N=48) in the UCLA CNP cohort. (ii) For each brain region, the difference in HoFC (Δ HoFC, in units of R) between SCZ versus control participants is plotted on the cortical surface. (iii) The regional HoFC values are plotted for control (x-axis) versus SCZ (y-axis) participants, with the Spearman rank correlation (ρ) indicated in the top left corner. Points reflect brain regions (total of 34) and are colored as in (ii), and the identity line with slope=1 is shown as a black dashed line. **C.** For bipolar I disorder (BP, N=49), the regional HoFC gradient (i), difference relative to controls (ii), and scatter plot compared with control values (iii) are plotted as with SCZ in **B**. **D.** For attention-deficit hyperactivity disorder (ADHD, N=38), the regional HoFC gradient (i), difference relative to controls (ii), and scatter plot compared with control values (iii) are plotted as with SCZ in **B**.

First, comparing the HoFC map from neurotypical controls in the UCLA CNP cohort (cf. Fig. 6.2.2A) to that of the HCP cohort (cf. Fig. 6.2.1A) demonstrated general concordance between the two maps, with a Spearman correlation of $\rho = 0.62$ ($P = 5 \times 10^{-4}$, single-hemisphere spin test with 10,000 permutations [442]). The direct spatial comparison is shown in more detail in Fig. S1, with overall visual agreement in the pattern of higher HoFC in primary sensorimotor regions and lower HoFC in higher-order transmodal

and association areas[1]. Next, having demonstrated consistent spatial variation in HoFC magnitudes in this independent control cohort, we investigated the spatial variation in HoFC in each of the three disorders. As shown in Figs 6.2.2B(i)-D(i), the distribution of HoFC magnitudes across the cortex was visually similar between each neuropsychiatric disorder cohort (averaged across participants) and that of the control group—although magnitudes were visually lower in SCZ in particular. To directly examine the nature of HoFC changes across the cortical sheet in these three disorders, we computed the mean difference in HoFC R values between each disorder and the control cohort per region, yielding a $\Delta R = R_{\text{Disorder}} - R_{\text{Control}}$ value for each region. These ΔR values are plotted on the cortical surface in Figs 6.2.2B(ii)-D(ii), revealing that most regions exhibited reduced HoFC to some degree relative to controls, consistent with previous reports [431–434]. The most pronounced HoFC differences appeared in SCZ, in which three cortical areas showed a change of $\Delta R < -0.1$: pars opercularis ($\Delta R = -0.13$), middle temporal gyrus ($\Delta R = -0.12$), and superior temporal gyrus ($\Delta R = -0.11$). By contrast, most regions in BP and ADHD showed smaller differences in HoFC ($-0.05 \leq \Delta R \leq 0.05$), though the banks of the superior temporal sulcus ('bankssts') exhibited more pronounced HoFC reduction in BP with $\Delta R = -0.11$. Overall, the ΔR distributions reflect heterogeneous and relatively low-magnitude HoFC perturbations in each disorder.

Against the backdrop of generally reduced HoFC (to varying degrees) in SCZ, BP, and ADHD, we next quantified the extent to which the HoFC spatial gradient observed in the neurotypical control cortex is altered in each of these neuropsychiatric disorders. In Figs 6.2.2B(iii)-D(iii), we compared the HoFC (R) per cortical region in each diagnostic group (x-axis) versus that of the UCLA CNP control cohort (y-axis, from Fig. 6.2.2A). The spatial gradient of HoFC was largely preserved in all three disorders, with Spearman's $\rho = 0.93$ in SCZ, $\rho = 0.97$ in BP, and $\rho = 0.98$ in ADHD. While most regions fell below the identity line (dashed black in Fig. 6.2.2D), meaning $\Delta R < 0$, computing the line of best fit per disorder (not shown) yielded slopes between $m = 0.91$ (SCZ) and $m = 1.02$ (ADHD)—suggesting that despite heterogeneous disease-associated alterations across the cortical sheet, the overall regional variation in HoFC appears to be robust and reproducible across cohorts.

[1] Supplementary figures are included in Appendix 5.

6.2.3 HOMOTOPIC CONNECTIVITY IS LINKED TO BROADER RESTING-STATE FUNCTIONAL NETWORK TOPOLOGY

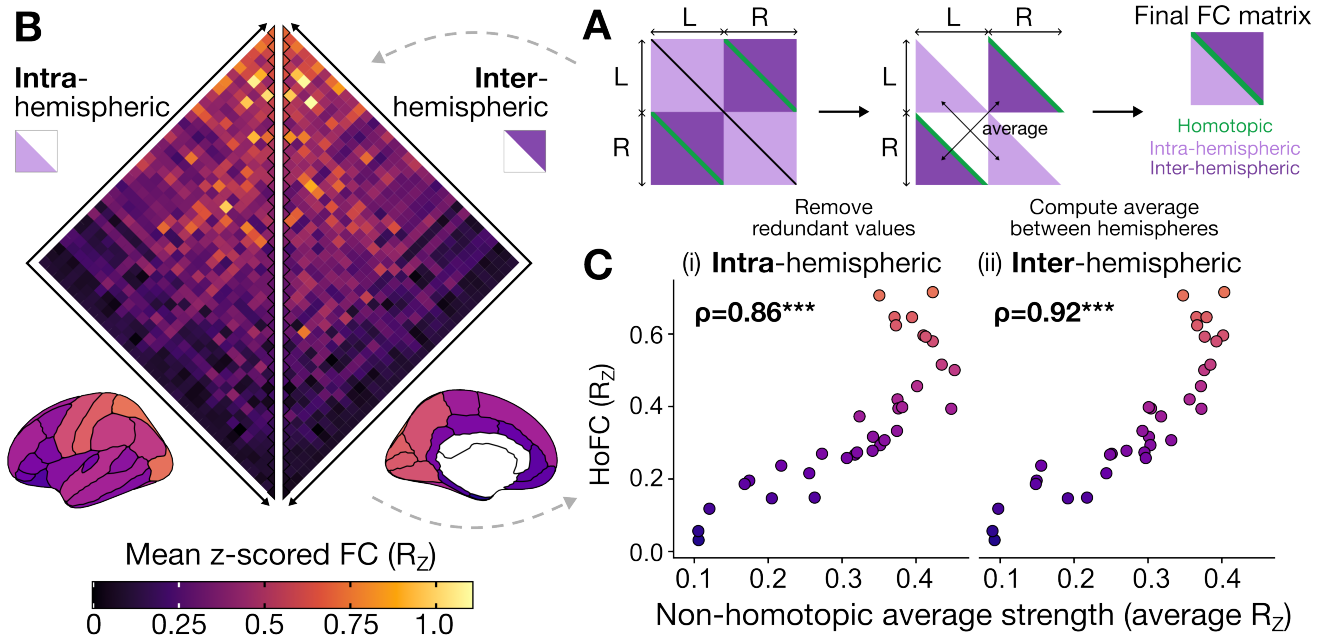


Figure 6.2.3: Homotopic functional connectivity (HoFC) is linked to regional degree within and between hemispheres in the functional connectome. **A.** Our approach to compare hemisphere-averaged connections within versus across hemispheres is depicted schematically. Starting with the full 68×68 functional connectome (with the atlas from Desikan et al. [169]), we first divided the connectome into four quadrants (L ↔ L, L ↔ R, R ↔ L, and R ↔ R). Within each quadrant, correlation values above and below the diagonal line are symmetric (and therefore, redundant), so only the lower four triangles were retained. To generalize across left and right hemispheres, we then computed the average intra-hemispheric connections (by averaging over L ↔ L and R ↔ R per region–region pair) and average inter-hemispheric connections (by averaging over R ↔ L and L ↔ R per region–region pair). This yielded one final matrix, with homotopic connections along the diagonal, and one triangle for intra- and inter-hemispheric functional connections, respectively. **B.** The resulting group-averaged functional connectome (hemisphere-averaged from the process depicted in **A**) is shown, with values representing the mean z-scored Pearson correlation across all individuals in the HCP cohort from ENIGMA (see Methods, Sec. 6.5.1). The left triangle depicts intra-hemispheric connectivity, while the right triangle depicts inter-hemispheric connectivity. The nodes along the center line indicate homotopic connections, and the same HoFC values shown in the heatmap are also plotted on the cortical surface below to guide visual interpretation with the same color scale. **C.** For each cortical region, the average FC value (R_Z) to all other regions is computed within the (i) ipsilateral hemisphere or (ii) contralateral hemisphere and plotted on the x-axis versus the HoFC value for the corresponding region on the y-axis. In each case (intra- or inter-hemispheric), the Spearman rank correlation (ρ) is indicated at the top left corner. $^{***}, P < 0.001$; single-hemisphere spin test with 10,000 permutations [442].

While prior work has demonstrated that homotopic region pairs exhibit overall greater linear correlation in BOLD activity than intra-hemispheric or heterotopic connections [41, 395–398, 404], to our knowledge, the relationship between HoFC and general properties of functional connectivity in the resting brain has not been investigated previously. To address this gap, we sought to clarify how regional HoFC magnitude relates to average node strength in the functional connectome, both within and across hemispheres. We aggregated the connections per region–region pair into a single group FC matrix, with intra-hemispheric

connections in one matrix triangle and inter-hemispheric connections in the other, as depicted schematically in Fig. 6.2.3A. The resulting group-averaged functional connectome is plotted in Fig. 6.2.3B, which visually demonstrates a striking level of symmetry between intra-hemispheric functional connections (left triangle) and inter-hemispheric connections (right triangle) across all region–region pairs. Regions are ordered according to their mean HoFC magnitude, revealing a spectrum ranging from overall higher FC values (both within and across hemispheres) in regions with higher HoFC to lower cortex-wide FC values in regions with lower HoFC. For each region, we computed its overall average (non-homotopic) nodal strength within the intra- and inter-hemispheric functional networks, defined as the average functional connectivity (in R_{Z}) between each region and each of the other 33 regions in the atlas from Desikan et al. [169]. As shown in Fig. 6.2.3C, the average strength per region was strongly correlated with HoFC magnitudes, both within ($\rho = 0.86$; $P_{\mathrm{spin}} = 1 \times 10^{-4}$, single-hemisphere spin test with 10,000 permutations) and across hemispheres ($\rho = 0.92$, $P_{\mathrm{spin}} = 1 \times 10^{-4}$). This indicates that the macroscale gradient in resting-state HoFC also mirrors that of average FC strength generally, both within and across hemispheres.

6.2.4 HOMOTOPIC FUNCTIONAL CONNECTIVITY IS SPATIALLY RELATED TO VARIATION IN VASCULAR INNERVATION, BUT NOT GEOMETRIC EMBEDDING

In order to better understand plausible candidates for the physiological underpinnings of HoFC in health and disease, we next investigated properties related to structural connectivity, physical embedding, and vascular pathology. Structural connectivity within the mammalian cortex follows an exponential distance rule, such that the probability of a physical connection between two areas and the strength of the connection both decrease exponentially with increasing physical distance [451, 452]; this effect has been replicated in human structural connectivity analysis [453, 454]. Resting-state FC generally exhibits a similar spatial constraint [409, 455, 456], and the structure–function coupling magnitude varies along the anatomical hierarchy of the cortex [60, 457, 458]. Cortico-cortical HoFC strength is also reportedly linked to structural connectivity integrity between region–region pairs [395, 459]—as is the case in resting-state FC in general [49]. However, prior work also indicates that HoFC does not decrease with inter-region distance, with greater connectivity between homotopic regions than would be expected based on their physical distance [404, 409, 410]—perhaps due to the specificity of cross-hemisphere connectivity. In light of these findings, it has yet to be clarified whether HoFC is associated with cortical variation in structural architecture. To test this, we compared the resting HoFC gradient in Fig. 6.2.1A with: (1) anatomical connectivity, quantified as the log-transformed number of streamlines estimated using diffusion tensor imaging; and (2) spatial proximity, quantified as the Euclidean distance between the centroid vertex of each region on the cortical surface. The group-consensus structural connectome analyzed here is from the same participants as with the functional connectome, all preprocessed in Larivière et al. [435]; more details are provided in the Methods (Sec. 6.5.2).

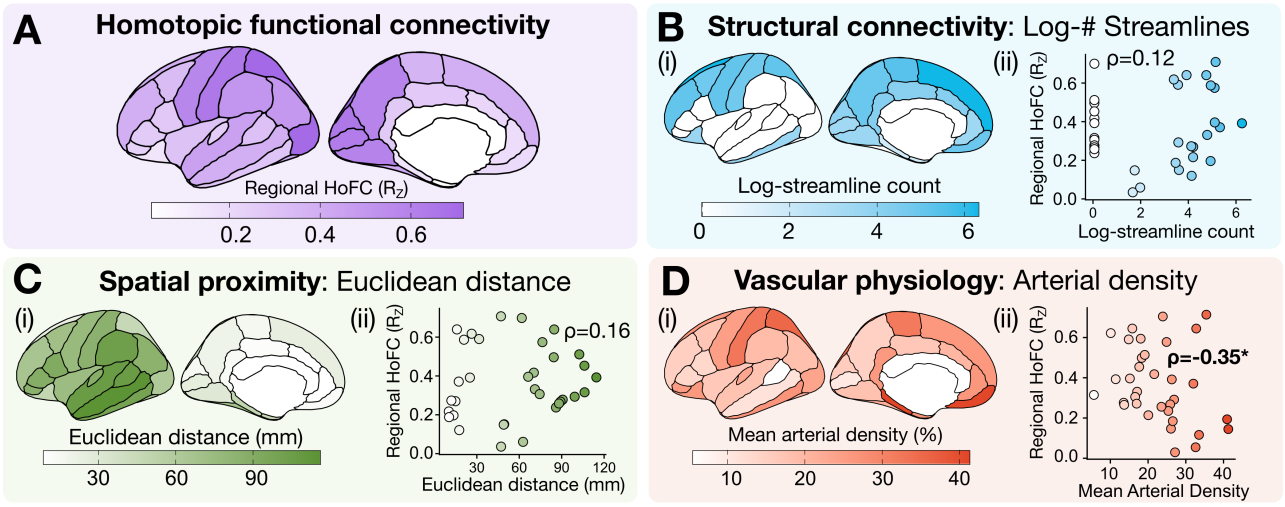


Figure 6.2.4: Regional HoFC is not statistically associated with structural connectivity or Euclidean distance, but it is inversely associated with arterial density. A. The same HoFC gradient from Fig. 6.2.1A is replicated here to guide visual interpretation, with values projected onto the cortical surface. **B.** (i) The log-transformed streamline count between homotopic region pairs is depicted on the cortical surface. As part of the consensus connectome procedure in Larivière et al. [435], region–region pairs with low and/or highly variable streamline counts were thresholded out and the log-count set to zero. (ii) The mean log-streamline count connecting each homotopic region pair (x-axis) is plotted against the HoFC for each corresponding region (y-axis), with the Spearman rank correlation (ρ) annotated. **C.** (i) The Euclidean distance between the centroid vertex of each homotopic region pair is depicted on the cortical surface. (ii) As in **B**(ii), the mean Euclidean distance between each homotopic region pair (x-axis) is plotted against the HoFC for each corresponding region (y-axis). **D.** (i) The mean arterial density in each region [460] is plotted on the cortical surface, averaged between the two hemispheres per region. (ii) As in **B**(ii), the mean hemisphere-averaged arterial density (x-axis) is plotted against the HoFC for each corresponding region (y-axis). $^*, P_{\text{spin}} < 0.05$; single-hemisphere spin test with 10,000 permutations [442].

As shown in Fig. 6.2.4B, the cortical gradient of resting HoFC was not significantly correlated with structural connectivity, estimated as the $\log_{10}$-transformed streamline count ($\rho = 0.12$, $P_{\text{spin}} = 0.5$, single-hemisphere spin test with 10,000 permutations [442]). Of note, the preprocessed structural connectome was generated with a distance-dependent consensus threshold approach [461] to retain edges reliably detected across the $N = 207$ participants while preserving individual-level connection length distributions. In the resulting structural connectome provided in Larivière et al. [435], for 12 of the 34 regions in the atlas from Desikan et al. [169], homotopic anatomical connections were not reliably detected across participants after adjusting for streamline length, such that the estimated streamline count was approximated as zero after log-transforming all other streamline counts (corresponding to the 12 regions plotted along the zero line on the x-axis in Fig. 6.2.4B). We confirmed in a robustness analysis that in the absence of thresholding, with a subset of 100 structural connectomes (out of the $N = 207$ in the ENIGMA-provided dataset [435]) preprocessed with a similar pipeline [167], there was no significant association between anatomical connectivity and HoFC across the cortex ($\rho = 0.18$, $P_{\text{spin}} = 0.31$; Fig. S2). Similarly, there was no significant association between spatial proximity and HoFC magnitude ($\rho = 0.16$, $P_{\text{spin}} = 0.3$), as shown in Fig. 6.2.4C. Some of the closest regions along the cortical midline (separated by $\leq$20 mm) showed relatively low HoFC ($R_{\text{Z}} \leq 0.4$); in contrast to regions like the postcentral gyrus, which exhibited among the highest HoFC ($R_{\text{Z}} = 0.64$), despite the left and right regions sitting 85.3 mm apart on average (in the `fsaverage` template surface). These findings

indicate that properties of anatomical wiring and physical embedding do not account for regional variation in HoFC strength—suggesting that the spatial HoFC gradient may be related to additional factors that are (at least partly) dissociable from direct structural architecture.

As another possible physiological mechanism that might contribute to homotopic BOLD connectivity, we also examined properties of the cerebral vasculature. BOLD fMRI intrinsically captures neurovascular coupling by measuring the ratio of deoxygenated to oxygenated hemoglobin over time in each voxel [462]. Prior work indicates that homotopic regions exhibit highly synchronous oscillations in vessel diameter [463], and Drew et al. [464] posit that vascular geometry serves as a physiological substrate for HoFC in addition to inter-hemispheric anatomical projections. Moreover, emerging work suggests that asymmetries in capillary transit time heterogeneity [465] and/or venous draining [466] are associated with attenuated BOLD fMRI connectivity between homotopic regions. Given these findings linking vascular innervation properties with HoFC, we reasoned that spatial HoFC variation may relate to differences in the type and density of vascular innervation across the cortical surface. To investigate this, we queried the probabilistic atlases for arterial and venous density in Bernier et al. [460] in the same parcellation space. Arteries carry oxygenated hemoglobin from the heart into the brain; they are generally 1-5 mm in diameter and dilate within a few hundred milliseconds of neural activity onset [467]. By contrast, veins carry de-oxygenated hemoglobin from the brain back into the peripheral vasculature; veins are generally similar in diameter to arteries, though they exhibit dilations much smaller in volume and on a slower timescale than arteries (i.e., tens of seconds after activity onset) [462, 468].

As shown in Figs 6.2.4D(i,ii), arterial density was mostly concentrated around medial temporal and frontal regions as well as the lateral parietal cortex, and exhibited a significant negative association with the regional variation in HoFC magnitude ($\rho = -0.35$, $P_{\text{spin}} = 7 \times 10^{-3}$, single-hemisphere spin test with 10,000 permutations). This negative association suggests that regions with greater arterial supply tend to exhibit lower HoFC magnitude, which is consistent with emerging work demonstrating that proximity to the nearest artery is inversely associated with a region's weighted degree centrality (i.e., strength) within the functional connectome [469]. By contrast, venous density was concentrated around visual and inferior frontal/temporal cortices and was not associated with HoFC ($\rho = -0.08$, $P_{\text{spin}} = 0.8$; Fig. S3). Collectively, the significant (albeit relatively low-magnitude) negative correlation between arterial density and HoFC magnitude suggests that cerebral vascularization may be related to inter-hemispheric functional connectivity.

6.2.5 REGIONAL HOMOTOPIC CONNECTIVITY VARIATION MIRRORS PATTERNS OF INTRATELENCEPHALIC NEURON DENSITY AND THALAMIC COUPLING

While the demonstrated spatial variation in HoFC magnitude is robust across cohorts and neuropsychiatric disorders, the lack of association with structural connectivity and spatial proximity invites the question: in addition to vascular innervation, what other physiological mechanisms may support resting-state HoFC? We return to seminal work in the rodent visual cortex, in which Miller and Vogt [470] traced callosal fibers to demonstrate that homotopic connections project specifically to neurons in layer 5 of the cortex—compared with heterotopic fibers, which project throughout layers 2 to 5. Although callosal fibers seem to mediate HoFC, homotopic regions still exhibit coupling in cases of corpus callosum agenesis [471] and surgical severance [411], suggesting alternate (and potentially compensatory) mechanisms—for example, polysy-

naptic connections or common external drive from subcortical structures, such as the thalamus or basal forebrain [404, 472].

Emerging evidence suggests that excitatory glutamatergic layer 5 neuron density closely tracks the first principal gradient of functional connectivity [473] and that neurons in layer 5 project to the contralateral cortex more so than other layers [52]. We cannot directly trace layer-specific fibers in any of the participant cohorts examined here; we can, however, examine transcriptomic markers of layer- and connectivity-specific layer 5 neuron subtypes in the brain. Baker et al. [474] thoroughly characterized subtypes of neocortical pyramidal neurons with somata in layers 5 and 6, broadly distinguishing between intra-telencephalic neurons ('IT', which exclusively project axons to other cortical regions) versus extra-telencephalic neurons ('ET', which also project to subcortical and brainstem structures). Within IT and ET classes, neuronal subtypes exhibit diverse sublaminar localization, dendritic arborization, axonal projection targets, and transcriptomic signatures [475, 476]. Here, we focused on IT and ET neurons with somata in layer 5 specifically, selecting one representative marker gene per subtype from Baker et al. [474]: *SATB2* for IT neurons and *CTIP2* for ET neurons. Previous work has identified *SATB2* as a key regulator for axonogenesis, synapse formation, and the differentiation of neurons with both callosal and subcortical projections [477–479]. The transcription factor *CTIP2* is expressed in subcerebral projection neurons and is required for axonal fasciculation, outgrowth, and pathfinding [480, 481]. Co-expression of these two genes is critical for proper axonal specialization in layer 5 neurons [482, 483]—with *SATB2* and *CTIP2* driving differentiation of callosal and corticospinal projection neurons, respectively.

In order to quantitatively compare HoFC magnitudes with expression of these key transcriptomic mediators of inter-hemispheric connectivity, we queried the AHBA specifically for *SATB2* and *BCL11B* (the alternative gene name for *CTIP2*), interpolating the cortical microarray expression values in the same parcellation atlas as the functional connectome [169] (see Methods, Sec. 6.5.2 for more details). This analysis is motivated by recent work investigating the genetic correlates of resting-state HoFC in the healthy brain [484] and in the context of schizophrenia [434], both of which identified hundreds of genes associated with HoFC that collectively contribute to functions including cell–cell communication and synaptic transmission. We depict the brain maps for *SATB2* and *CTIP2* expression (taking the average between left and right hemispheres per region; normalized to arbitrary units) for IT and ET layer 5 neurons, respectively, in Fig. 6.2.5A-B(i). Comparing regional HoFC values with the mean normalized expression of these two genes across all 34 regions in Fig. 6.2.5B revealed that regional *SATB2* expression is positively correlated with R_{Z} ($\rho = 0.49$; $P_{\mathrm{spin}} = 1 \times 10^{-3}$, single-hemisphere spin test with 10,000 permutations [442]). By contrast, *CTIP2* expression was not significantly correlated with HoFC magnitudes across the cortex (Spearman $\rho = -0.28$; $P_{\mathrm{spin}} = 0.1$).

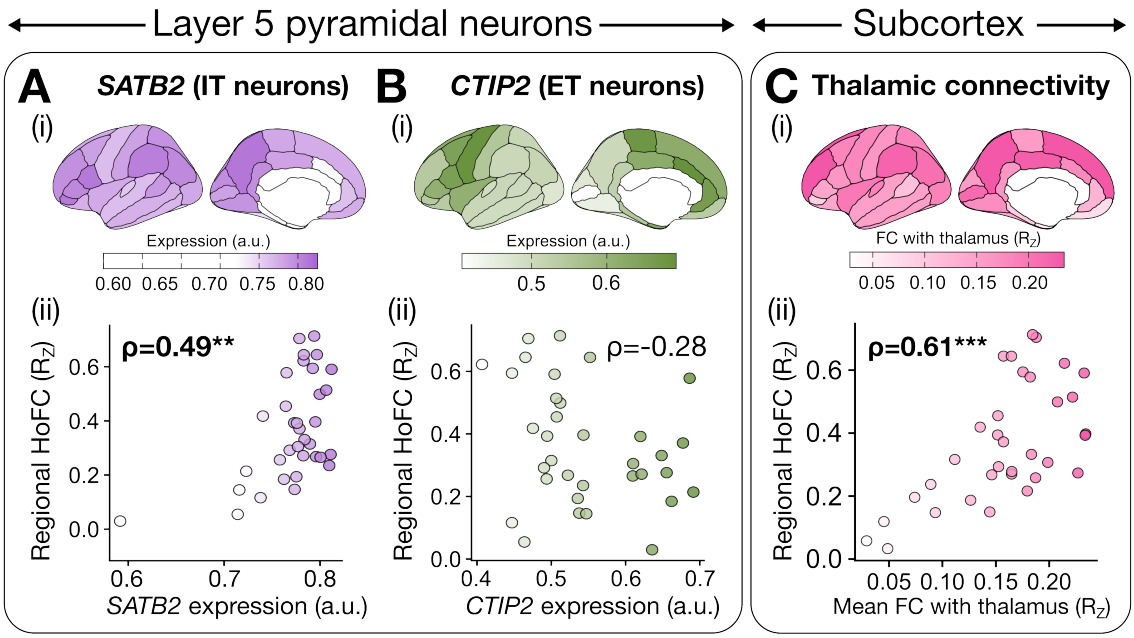


Figure 6.2.5: Homotopic connectivity magnitude is positively associated with expression of *SATB2*, a marker of intra-telencephalic layer 5 pyramidal neurons, and with resting-state thalamic connectivity. Two genes were investigated as potential molecular underpinnings of HoFC: (**A**) *SATB2* and (**B**) *CTIP2*. For both genes, the mean normalized expression value between the left and right hemisphere (queried from the AHBA [441]) is projected onto the cortical surface in (i). These mean values per region are then compared on the x-axis in **A**(ii) and **B**(ii) with the HoFC value per corresponding region on the y-axis, with the black lines indicating the line of best fit. In both cases, the Spearman rank correlation (ρ) is indicated. **C.** (i) For each brain region, the mean FC with the bilateral thalamus (i.e., averaged for both intra- and inter-hemispheric connections in both hemispheres) is depicted on the cortical surface. (ii) The same values from (i) are compared on the x-axis with the HoFC value per region on the y-axis, with the black line indicating the line of best fit and ρ indicating the Spearman rank correlation value. $^{***}, P_{\text{spin}} < 0.001$; $^{**}, P_{\text{spin}} < 0.01$; single-hemisphere spin test with 10,000 permutations [442].

Higher *SATB2* expression in a given region in the adult brain may reflect greater molecular support for interhemispheric callosal projections through more *SATB2*+ (IT) neurons—which generally exhibit somata in layers 5 and 6 and project within the cortex, striatum, and claustrum [474]. Layer 5 pyramidal neurons are key components of cortico-cortical circuits and are known to project to subcortical structures, including the thalamus, as well as contralateral cortical targets via callosal fibers [485]. Ruffini et al. [486] propose that shared inputs from the thalamus may directly contribute to HoFC, in which case the functional synchrony would arise from the same input from common upstream influences rather than direct interhemispheric communication. Despite such potential theoretical links, the extent to which thalamocortical connectivity supports or constrains HoFC during rest remains poorly understood. As a first step to explore this potential relationship, we quantified the association between regional HoFC and thalamocortical functional connectivity by computing the average FC between each pair of homotopic regions and the thalamus across both hemispheres. As shown in Fig. 6.2.5C, we observed a positive correlation between HoFC magnitude and bilateral thalamic connectivity ($\rho = 0.61$; $P_{\text{spin}} = 7 \times 10^{-4}$, spin test with 10,000 permutations [442]), suggesting that regions more tightly coupled with the thalamus also exhibit stronger interhemispheric correlation. These findings collectively suggest that HoFC in the cerebral cortex is spatially related to both layer 5

IT pyramidal neuron gene expression and to thalamo-cortical functional connectivity in the resting state, suggesting these are promising candidate biological mechanisms for further investigation in future HoFC research.

6.3 DISCUSSION

In this study, we provided a comprehensive account of how resting-state homotopic functional connectivity (HoFC)—a key feature of inter-hemispheric organization—varies across the human cortex and relates to structural, functional, and transcriptomic properties. We have shown that regional resting-state HoFC is not spatially homogeneous, but rather exhibits a characteristic regional variation that, in turn, aligns with the dominant axis of anatomical hierarchy and macroscale functional and transcriptomic organization across the cortical sheet (Fig. 6.2.1). This spatial gradient was robust to heterogeneous disease-associated perturbations in region-level HoFC (in a separate neuroimaging cohort [261]), most of which corresponded to relatively low-magnitude decreases across the cortex (Fig. 6.2.2). Regions that were more functionally connected within and across hemispheres in the resting state also exhibited greater homotopic coupling, suggesting that HoFC and overall functional connectivity strength co-vary along a similar gradient. Furthermore, we demonstrated that regional HoFC variation in the cortex cannot be fully explained by structural architecture or vascular factors (Fig. 6.2.4), but instead appears to track neuroanatomical features such as layer 5 intra-telencephalic (IT) neuron gene expression and thalamic connectivity (Fig. 6.2.5). Together, these findings suggest that homotopic connectivity relates to a confluence of functional network dynamics and specific neuroanatomical substrates, providing new insights into multimodal correlates of resting-state inter-hemispheric integration.

Our findings are consistent with previous reports of inverse associations between HoFC and position along a fundamental axis [416] of macroscale cortical organization, from primary sensorimotor to association regions [394, 421, 422], though the specific mechanisms mediating this relationship will need to be directly tested in the future. For example, prior work has established that primary sensory/motor cortices are more heavily myelinated [487] and exhibit higher T1w:T2w signal ratios [488]. The T1w:T2w ratio is sensitive to multiple microstructural properties of the gray matter [445, 489] and is a hypothesized (inverse) correlate of the anatomical hierarchy of laminar differentiation across the cortex [446]. We observed a significant positive association between HoFC and T1w:T2w gradients across the cerebral cortex, suggesting that HoFC decreases with ascending position along the putative anatomical hierarchy. Of note, Shafiei et al. [239] reported that the gradient of T1w:T2w values—along with that of AHBA PC1 values—is strongly associated with the dominant axis of resting-state neurophysiological fMRI dynamics across the cerebral cortex. Collectively, our findings are congruent with the 'default state' of inter-hemispheric synchrony proposed by Stark et al. [394], with the highest HoFC observed in primary sensorimotor regions. The tighter bilateral coupling between these primary sensorimotor regions may suggest greater resting-state symmetry in support of sensory perception and motor coordination, potentially via shared inputs from thalamic drive [486].

The preserved resting-state HoFC gradient in a separate neurotypical cohort and in three neuropsychiatric disorder groups (SCZ, BP, and ADHD) is surprising, given the inter-individual variability in resting-state functional connectivity [490], which is particularly pronounced in neuropsychiatric disorders [491]. This sug-

gests that while most regions exhibited decreased HoFC with heterogeneous magnitudes across the three disorders relative to neurotypical controls, such disease-related perturbations were outweighed by the dominance of the overall HoFC spatial gradient. In a recent meta-analysis, Yao and Kendrick [450] reported that HoFC changes in ADHD were too small and/or variable to be reliably included, instead identifying common HoFC reductions in visual and motor regions in SCZ along with autism spectrum disorder, anxiety, depression, and addiction (n.b., BP was not included in that specific cross-disorder analysis). Prior work in SCZ populations has demonstrated that HoFC reductions are associated with auditory verbal hallucinations [492], social withdrawal [434], and cognitive deficits [493]. To our knowledge, this is the first analysis to demonstrate that the macroscale HoFC gradient is robust to subtle disease-associated alterations at the regional level. Of note, the magnitude of HoFC reductions was most pronounced in SCZ, followed by BP, and least pronounced in ADHD—which is consistent with previous findings that resting-state fMRI dynamics were more distinctively altered in SCZ and BP than in ADHD in this cohort [79]. However, future investigation in an external dataset(s) will be needed to more fully explore this preliminary finding in the context of key covariates, including symptom severity and treatment usage.

Having identified a robust gradient of HoFC across the cortex across multiple participant groups and imaging cohorts, we further established that HoFC magnitude is strongly related to the overall regional FC strength, averaged across all of each region's functional connections. While the general distribution of brain-wide homotopic connectivity strengths has been compared to that of heterotopic connections [41, 404], the relationship between HoFC in a given region and its other intra- and inter-hemispheric functional connections has not previously been characterized, to our knowledge. Collectively, our functional network analysis suggests that HoFC is higher in regions that are more functionally connected to the rest of the cortex (i.e., higher node average strength; cf. Fig 6.2.3). In other words, regions that show greater overall functional connectivity within the ipsilateral hemisphere, and are therefore more functionally central, also exhibit functional synchrony with the contralateral hemisphere, indicative of a more globally shared signal. Our findings suggest that both HoFC and shared global signal are closely related, with decreasing dominance of both measures in the resting-state brain with ascending position along the anatomical hierarchy and the first principal gradient of functional connectivity.

We investigated potential contributions of vascular innervation and/or structural architecture to the resting-state HoFC gradient, finding that only arterial density was (negatively) associated with HoFC magnitude. It is difficult to disentangle biological versus methodological implications in the context of BOLD fMRI, which is itself sensitive to neurovascular coupling [462] and may be less suited to detect arterial rather than venous dynamics [494]. One potential biological interpretation could be that higher arterial density coincides with greater metabolic lateralization, which would in turn reduce the need for synchronous bilateral activity [51, 495, 496]—though further hypothesis-driven investigation in an external dataset is warranted to clarify the meaning of the inverse spatial association between HoFC magnitude and regional arterial density. The finding that resting-state HoFC magnitude was not statistically associated with structural connectivity (quantified as log-transformed number of streamlines, estimated from diffusion-weighted imaging) is surprising, given the proportion of callosal fibers innervating homotopic region pairs [408, 497] and previous findings that cortical HoFC is affected by structural connectivity integrity [395, 459]. As with functional connectivity in general, the relationship between structure and function is nuanced and highly variable across both spatial and temporal scales; for example, Shen et al. [404] reported that the number of tract-traced fibers gener-

ally correlated with temporal stability of a given homotopic connection, but this pattern was subverted by primary sensory regions (including V1) that do not exhibit direct structural connections. While the quantity of innervating white-matter tracts did not spatially correlate with HoFC magnitude in this analysis, it is worth noting that primary sensorimotor cortices are interconnected via fibers with thicker myelin (i.e., faster conductance) while heteromodal and association areas are interconnected via fibers with thinner myelin (i.e., slower conductance) [497]. It may be that resting-state HoFC is more related to qualities of the structural connections, such as fiber diameter and conductance speed, rather than the overall quantity of fibers [394, 404]; indeed, Mollink et al. [395] reported that HoFC is intrinsically associated with properties of white-matter myeloarchitecture [395].

We also found that *SATB2* expression is positively correlated with regional variation in HoFC magnitude, which is a marker for layer 5 IT pyramidal neurons [474]. This is a departure from previous findings that layer 5 IT neuron density is positively associated with a region's position along the topographical hierarchy of the cortex, as indexed by the first principal gradient of functional connectivity [473]. The difference could arise from sensitivity to different subtypes in layer 5 IT neurons (e.g., cortico-cortical versus cortico-striatal projecting) and laminar-specific asymmetries [498]. Of note, functional connectivity between two brain regions can arise through multiple underlying mechanisms, including the influence of common afferent inputs, shared efferent targets, or reciprocal anatomical connections [50, 52]. In the context of HoFC, Mancuso et al. [51] proposed that the particularly strong BOLD activity correlation observed between homotopic regions in the sensory and motor cortices may be driven by convergent thalamic inputs and/or coordinated corticospinal outputs—a hypothesis that has since been echoed by recent work applying the Ising model to characterize structure–function coupling [486]. Indeed, projections from parvalbumin-staining thalamic neurons are most densely concentrated in granular layer IV of primary sensory cortex (and motor cortex, to a lesser extent [499]), such that variations in thalamic input across the cortex could underpin resting-state homotopic coupling [500]. We found that primary sensory and motor cortices exhibited the strongest HoFC, and that regional HoFC magnitude is positively associated with bilateral thalamic FC, potentially suggesting that common subcortical drive, particularly from the thalamus, may contribute meaningfully to homotopic coordination observed in resting-state networks [404]. In support of this interpretation, Markicevic et al. [43] found that unimodal regions are more sensitive to chemogenetic perturbation of the murine basal ganglia than transmodal regions, which the authors attribute to fewer overall inputs (which are more dominated by feedforward subcortical projections) into unimodal regions compared to transmodal areas.

When interpreting the findings in this paper, one should consider potential limitations that collectively suggest avenues for future investigation. First, while we analyzed two open-access resting-state fMRI datasets— which have been investigated extensively in previous work (HCP and UCLA Consortium for Neuropsychiatric Phenomics)—future replication with an external dataset(s) is needed to validate the biological findings we presented here. Moreover, these datasets were collected with different acquisition parameters and preprocessed using different pipelines, and while the agreement between the two was significant ($\rho = 0.62$), the dependence of HoFC spatial variation on preprocessing methods should be examined. For both BOLD fMRI datasets, we used the atlas from Desikan et al. [169], which is defined based on anatomical boundaries and may not be optimally suited for representing aggregate functional activity [3]. While this atlas was chosen given consistency in nomenclature and topographical position of regions between the left and right hemispheres relative to the other parcellations included as part of the ENIGMA preprocessed

dataset [435] (all based on atlases from Schaefer et al. [207]), future investigation with parcellation atlases tailored specifically for homotopic analysis is warranted [501, 502]. As with any associative brain-map analysis, we also point out that the resting-state fMRI data were obtained from different individuals compared to the T1w:T2w, FC PG1, AHBA PC1, and vascular density maps examined herein. Moreover, we exclusively focused on the Pearson correlation coefficient as an index of HoFC, which is specifically sensitive to linear, synchronous coupling between regions. While the Pearson correlation coefficient is empirically well-suited to the spatial and temporal scales of resting-state BOLD fMRI [79], future work should examine the robustness of the cortical HoFC gradient in different imaging modalities (e.g., magnetoencephalography [503]) and with other measures that can capture different types of coupling, such as time-lagged and/or frequency-based connectivity [79, 92]. The present analysis also focused exclusively on resting-state connectivity, though future work is needed to clarify how cortical HoFC evolves, and potentially reorganizes, with increasing task complexity in healthy participants as well as in the context of neuropsychiatric disorders [51, 404, 407, 504].

6.4 CONCLUSIONS

In this work, we focus on four aspects of homotopic functional architecture that, to the best of our knowledge, have not been addressed previously: (1) How does regional variation in homotopic connectivity interface with macroscale gradients of cortical organization that include anatomical hierarchy, functional connectivity, and vascular innervation? (2) Is the characteristic spatial variation in homotopic connectivity robust between cohorts and clinical groups, particularly neuropsychiatric disorders? (3) Does homotopic connectivity relate to a region's overall strength in the brain-wide functional connectome? and (4) What candidate neuroanatomical substrates (e.g., transcriptomic and/or subcortical factors) might support regional variation in homotopic connectivity in the healthy brain? Our analyses collectively indicate that: (1) the regional gradient of homotopic connectivity is significantly associated with a fundamental natural axis of macroscale cortical organization, along with arterial density, though not with properties of physical embedding or structural connectivity; and that (2) this spatial variation is remarkably similar in an independent neuroimaging cohort that includes SCZ, BP, and ADHD cases, despite heterogeneous disease-associated perturbations. (3) Homotopic connectivity magnitude is strongly associated with a region's overall intra- and inter-hemispheric average strength in the functional connectome at rest. Finally, (4) we identified potential molecular mechanisms of homotopic connectivity related to layer 5 intra-telencephalic pyramidal neuron density and thalamic connectivity. If confirmed in future studies that incorporate independent validation datasets, these findings provide important insights into the functional architecture of inter-hemispheric connectivity in health and disease.

6.5 METHODS

6.5.1 NEUROIMAGING DATASETS

Resting-state fMRI dataset 1: Human Connectome Project 3T fMRI

The first dataset fMRI dataset included in this study is comprised of $N = 207$ unrelated participants ($N = 83$ males, 28.7 ± 3.7 years of age) from the S1200 release of the Human Connectome Project [161], as curated and preprocessed by the ENIGMA Consortium [435]. We used preprocessed group-averaged functional connectomes included in the ENIGMA Toolbox [435], as in previous work [436, 505]. Details on both imaging acquisition and preprocessing are described in Larivière et al. [435] and are briefly summarized as follows. Resting-state volumes in this dataset were acquired using a gradient-echo, echo planar image (EPI) sequence with the following parameters: TR/TE = 720/33.1 ms, slice thickness = 2.0 mm, 72 slices, 2.0 mm isotropic voxels, frames per run = 1,200. All resting-state fMRI data were preprocessed using the 'minimal preprocessing' HCP pipeline [506], which includes corrections for distortion, head motion, and magnetic field bias; skull removal; and intensity normalization. Noise removal was performed using the FIX software to regress out head motion, white matter signal, cardiac pulsation, and signal from arterial and large veins. Corrected fMRI volumes were mapped to MNI152 space and subsequently projected to the cortical surface using a cortical ribbon-constrained mapping algorithm. Normative FC matrices were generated for each participant by computing the pairwise Pearson product-moment correlations (R) between resting-state time series of all cortical regions, setting negative correlation values to zero. To generate a group consensus FC matrix, Fisher's R-to-Z transformation was applied as in previous work [410, 421] and then averaged across all N=207 participants, such that all FC values from this dataset are presented as R_{Z}. We queried the 'aparc' parcellation (corresponding to the 68-region parcellation using the atlas from Desikan et al. [169]) along with 14 subcortical structures using the `load_fc_as_one` function to obtain thalamocortical connectivity R_{Z} values.

Resting-state fMRI dataset 2: UCLA Consortium for Neuropsychiatric Phenomics

In order to characterize regional variation in HoFC in the context of neuropsychiatric disorders, we additionally included resting-state fMRI data from the open-access University of California at Los Angeles (UCLA) Consortium for Neuropsychiatric Phenomics (CNP) LA5c Study [261]. This included cognitively healthy control participants as well as participants diagnosed with schizophrenia (SCZ), attention-deficit hyperactivity disorder (ADHD), and bipolar I disorder (BP). Details of diagnostic criteria and behavioral symptoms have been described previously, along with imaging acquisition details [261]. Resting-state volumes in this dataset were acquired using a T2*-weighted EPI sequence with the following parameters: TR/TE = 2s/30 ms, slice thickness = 4.0 mm, 34 slices, 4.0 mm isotropic voxels, frames per run = 152. This resting-state fMRI dataset was preprocessed in previous work [79, 507] as briefly described in the following. Imaging data were preprocessed using the *fmriprep* v1.1.1 software [355] and the independent component analysis-based automatic removal of motion artifacts (ICA–AROMA) pipeline [356]. Additional noise correction was performed to regress out signal from white matter, cerebrospinal fluid, and global gray matter using the ICA–AROMA + 2P + GMR method [359, 507, 508]. Quality control analysis based on noise regression and head motion excluded eight participants from this resting-state fMRI dataset (N=4 control, N=3 SCZ,

N=1 ADHD). After preprocessing and quality control, resting-state fMRI data were retained for downstream analysis from a total of N=252 participants: N=116 controls, N=48 SCZ, N=49 BP, and N=39 ADHD. More information about these quality control measures and the demographic composition of the final participant cohort is described in Bryant et al. [79]. The 68-region parcellation from Desikan et al. [169] was applied to extract the noise-corrected 152-length time series per region (corresponding to 304 s), averaging across all voxels per region. FC matrices were generated for each participant by computing the pairwise Pearson product-moment correlations (R) between resting-state time series of all cortical region–region pairs. We note that raw Pearson correlation coefficient values (R) were retained, without Fisher's transformation or thresholding, for this dataset. Within each diagnostic group, FC matrices were averaged across participants to yield one group-average FC matrix per group (i.e., one FC matrix each for control, SCZ, BP, and ADHD groups).

6.5.2 EXAMINING BRAIN MAPS OF MACROSCALE FUNCTIONAL AND MOLECULAR ORGANIZATION

Brain maps of anatomical hierarchy, transcriptomic variation, and functional organization

To quantify multiple facets of macroscale organization across the cortical sheet, we examined three brain maps from the neuromaps [440] library: T1w:T2w MRI signal (as a proxy for anatomical hierarchy), AHBA PC1 (as a proxy for cell type density and key transcriptomic variation), and FC PG1 (which orients regions along a principal axis of functional connectivity, from unimodal sensorimotor to higher-order association areas). All map information is provided in the `neuromaps` documentation (https://netneurolab.github.io/neuromaps/listofmaps.html) and briefly summarized in the following. As a proxy of anatomical hierarchy [446], we examined the *hcps1200-myelinmap-fsLR-32k* map of T1-weighted to T2-weighted (T1w:T2w) signal, derived from structural MRI obtained from 417 unrelated participants from the HCP S1200 release [161, 445]. We additionally queried the *abagen-genepc1-fsaverage-10k* map capturing the first principal component (PC) of gene expression in the Allen Human Brain Atlas [441], as computed in Markello et al. [509] to capture the axis of maximal spatial variation in the cortical transcriptome. Lastly, we queried the *margulies2016-fcgradient01-fsLR-32k* map, which quantifies the first principal gradient (PG1) of functional connectivity (FC) from Margulies et al. [414]. Each map was parcellated into the 68-region atlas from Desikan et al. [169] using the `Parcellator` class method, and the average value was computed between hemispheres for each homotopic region pair (e.g., for the cuneus T1w:T2w ratio, the average was computed between the left and right cuneus).

Structural architecture brain maps

We examined two aspects of structural architecture in the cortex: anatomical connectivity and Euclidean distance between homotopic region–region pairs. For anatomical connectivity, we used the group-averaged structural connectome from the same N=207 HCP participants as the fMRI dataset, preprocessed and included in the ENIGMA Toolbox [435]. Details on both diffusion-weighted imaging (DWI) acquisition and preprocessing are described in Larivière et al. [435], and are briefly summarized as follows. DWI volumes were acquired using a spin-echo EPI sequence with the following parameters: TR/TE = 5.52s s/89.5 ms, 1.25 mm isotropic voxels, b-values=1,000/2,000/3,000 s/mm^2, 270 diffusion directions, 18 b0 images. As

part of the HCP 'minimal preprocessing' pipeline [506], b0 intensity normalization was applied along with correction for susceptibility distortion, eddy currents, and head motion. Structural connectivity matrices were generated per participant using *MRtrix* [510] and whole-brain streamlines were reconstructed using the SIFT2 (spherical-deconvolution informed filtering of tractograms) pipeline [511]. A group consensus structural connectome was derived using a distance-dependent thresholding procedure [461] with $\log_{10}$ transformation to reduce variance, yielding a structural connectivity matrix with log-transformed number of streamlines (i.e., fiber density). Streamline counts for region–region pairs in which streamlines were not reliably detected across individuals (after adjusting for length) were set to zero after log-transforming the other streamline counts. We queried the 'aparc' parcellation (corresponding to the 68-region parcellation using the atlas from Desikan et al. [169]) using the `load_sc` function. For our robustness analysis presented in Fig. S2, we analyzed diffusion-weighted MRI data from a subset (N=100) of the individuals in the main analysis, as preprocessed separately (though with a very similar pipeline) and described in Fallon et al. [167]. After reconstructing individual-level structural connectomes, the mean streamline count per edge (i.e., region–region pair) was computed, without any thresholding. As with the main analysis, we applied a $\log_{10}$ transformation to reduce variance, yielding an unthresholded group structural connectivity matrix with log-transformed number of streamlines. Streamline counts of zero, or less than 1 (i.e., negative after $\log_{10}$ transformation), were set to zero after log-transforming the other streamline counts.

To compute the physical distance between each homotopic region–region pair, we identified the centroid vertex (from the pial surface) of each cortical region in the *fsaverage* template space from FreeSurfer. We then calculated the Euclidean distance between the centroid vertices in the left and right hemisphere per region, respectively, using the `cdist` function from the *scipy* package (version 1.13.1) [512] in Python. Since coordinates in the pial surface files per hemisphere (`lh.pial` and `rh.pial`, respectively) are represented in millimeters (mm), the resulting Euclidean distances reflect the mm between each homotopic region–region pair.

Vascular innervation brain maps

We examined two aspects of vascular innervation in the cortex: arterial density and venous density. For both properties, we queried brain maps published in Bernier et al. [460], where the details of imaging acquisition and preprocessing can be found. Briefly, arteries were imaged using time-of-flight angiography (ToF) and veins with susceptibility weighted imaging (SWI), respectively, in a cohort of N=42 healthy individuals (aged 20-31). Arterial and venous density maps were derived from the ToF and SWI maps using a multi-scale Frangi diffusive filter (MSFDF) pipeline [513] which includes denoising, vessel voxel classification, and multiscale vessel enhancement. Vessel centerlines and diameters were extracted and nonlinearly aligned to MNI T1 (0.5mm) space. In order to minimize contamination of the venous tree by arterial signal in SWI, any voxels overlapping with those identified in ToF volumes were excluded. Vascular density maps were computed for gray matter, white matter, and cerebrospinal fluid (CSF) based on voxel-wise vessel segmentation. This yielded voxelwise maps of arterial and venous density, with values ranging from 1-100 that indicate the vessel probability in the corresponding voxel. The 68-region volumetric atlas from Desikan et al. [169] was applied to these voxelwise maps to extract the average arterial and venous density per region, respectively. The average value was computed between hemispheres for each homotopic region pair (e.g., for the cuneus arterial density, the average density was computed between the left and right cuneus).

Transcriptomic brain maps

In order to characterize key molecular correlates of regional HoFC variation, we examined the expression of two layer 5 pyramidal neuron markers: *CTIP2* (*BCL11B*) and *SATB2*. We queried these two genes from postmortem microarray data (averaged across all donors) provided by the Allen Human Brain Atlas (AHBA) [441] using the *abagen* python package (version 0.1.3) [509, 514]. Expression values were summarized per cortical region using the 68-region parcellation from Desikan et al. [169] and then averaged between hemispheres for each homotopic region pair (e.g., for the cuneus *CTIP2* expression, the average expression was computed between the left and right cuneus).

Thalamic connectivity brain map

Lastly, we characterized the spatial variation in functional coupling between each cortical region and the thalamus. We additionally queried the full cortical and subcortical group consensus FC matrix provided by the ENIGMA consortium, which was preprocessed with the same steps as described above in Sec. 6.5.1. For each cortical region, we computed the average z-scored Pearson correlation (R_Z) across four connections—two intra-hemispheric and two inter-hemispheric—using the cuneus as an example: (1) left cuneus and left thalamus; (2) right cuneus and right thalamus; (3) left cuneus and right thalamus; and (4) right cuneus and left thalamus. This enabled comparison of the extent to which each cortical region (out of 34) is functionally coupled with the thalamus on average at the group-consensus level.

6.5.3 STATISTICAL ANALYSES

Brain map comparison with homotopic connectivity gradient

We obtained regional homotopic functional connectivity (HoFC) estimates from the group-averaged functional connectomes from the HCP and UCLA CNP datasets as the correlation between the left and right homologs of each region. To compare the spatial variation in HoFC with different brain maps of macroscopic cortical organization, we computed the Spearman rank correlation (ρ) with the `spearmanr` function from the *scipy* package [512], comparing across a total of 34 regions. In order to evaluate the statistical significance of correlation estimates while accounting for spatial autocorrelation along the cortical sheet, we performed a single-hemisphere spin test with 10,000 permutations using the method of Alexander-Bloch et al. [442]. This method, implemented with the `nulls.alexander_bloch` function in *neuromaps*, rotates one map (e.g., the T1w:T2w map) along the spherical surface of a single hemisphere, which preserves spatial contiguity while disrupting the spatial alignment between the two compared maps. For each permutation, the Spearman correlation is computed between the permuted and original map (here, the HoFC gradient is always the original map), which collectively yields a null distribution. The resulting p-value reflects the proportion of permutations (out of 10,000) with null ρ values equal to or greater in magnitude than the observed empirical correlation, ρ.

Functional network topology quantification

We examined regional HoFC variation within large-scale functional network topology based on the symmetric 68-region cortical FC matrix provided by the ENIGMA consortium. Left and right hemisphere region

labels were collapsed to form a common 'base region' name (e.g., consolidating 'lh-entorhinal' and 'rh-entorhinal' down to 'entorhinal'). Each connection was labeled as 'homotopic' (same region across hemispheres), 'intra-hemispheric' (different regions within the same hemisphere), or 'inter-hemispheric' (different regions in different hemispheres; sometimes referred to as 'heterotopic') based on anatomical pairing. As depicted schematically in Fig. 6.2.3A, we distilled the original group FC matrix down to a 34×34 matrix in which homotopic connections occupy the diagonal, intra-hemispheric connections populate the lower triangle, and inter-hemispheric connections comprise the upper triangle. Regions were ordered in both rows and columns according to their HoFC magnitude, ranked from highest to lowest. In order to summarize the average node strength, we computed the mean of the z-scored FC values (R_Z) between a given region and all 33 other regions in the same hemisphere (for intra-hemispheric connectivity) or the opposite hemisphere (for inter-hemispheric connectivity).

6.5.4 DATA VISUALIZATION

All cortical brain maps were created using the *ggseg* package (version 1.6.6) in R [386]. Scatterplots and heatmaps were generated using the *ggplot2* package [515] (version 3.5.1) and violin plots with the *see* package [516] (version 0.11.0), both in R.

6.5.5 CODE AND DATA AVAILABILITY

All code needed to replicate the analyses and visualizations presented in this paper is openly available on GitHub at https://github.com/DynamicsAndNeuralSystems/Homotopic_FC_HCP/. All data included in this paper is openly available for researchers to access. Preprocessed functional and structural connectivity matrices (derived from HCP data, which is openly available [161]) are included as part of the ENIGMA toolbox at https://enigma-toolbox.readthedocs.io/en/latest/pages/05.HCP/ [435]. Resting-state fMRI data from the UCLA CNP study are available for download at OpenNeuro with accession number ds000030 (https://openfmri.org/dataset/ds000030/). All brain maps are included as part of the *neuromaps* Python package [440] unless otherwise stated; for example, the vascular innervation maps [460] are available for download from the Braincharter GitHub repository at https://github.com/braincharter/vasculature/releases/tag/Atlas_v1.0. AHBA microarray transcriptomic data can be downloaded directly from the Allen Institute (https://human.brain-map.org/) or via the *abagen* Python library [509, 514].

ACKNOWLEDGMENTS

We thank A/Prof Joseph T. Lizier, Joshua M. Tan, Dr Giulia Baracchini, Dr Jayson Jenganathan, and Dr Isabella Orlando for insightful discussions that were very helpful in shaping this project. We also thank the Human Connectome Project and UCLA CNP Project, and their respective participants, for openly sharing this excellent data resource. We also thank the ENIGMA Consortium for openly sharing the preprocessed structural and functional HCP connectomes as part of their toolbox, as well as the Network Neuroscience Lab at the Montreal Neurological Institute for sharing the *neuromaps* toolbox and the many curated brain-maps contained therein. High-performance computing facilities provided by the School of Physics at The University of Sydney supported analyses and results presented here.

7

Benchmarking overlapping community detection methods for applications in human connectomics

I don't need a judge to tell me to keep my community clean.
Ryan Howard, The Office, Season 5 Episode 1

The brain exhibits a spatially distributed and fluid network of complex interactions, yet analysis of this landscape is typically constrained to methods that provide a static, non-overlapping modular decomposition. Specifically, we focus here on 'communities' in the structural connectome, comprised of subnetworks that are more densely connected to each other than to the rest of the brain. Amidst a growing repertoire of overlapping community detection algorithms (OCDAs) that allow regions to flexibly participate in more than one community, only a handful of studies have explored overlapping brain network decompositions, generally using a manually selected algorithm. This leaves open the possibility that the chosen OCDA may not be the most appropriate for the empirical network at hand, since each OCDA makes different assumptions about how communities and overlapping nodes are defined. Bridging this gap to efficiently capture and interpret this overlapping community structure in the brain is vital to advance our knowledge of how the brain integrates information across distributed networks.

In this chapter, we extend the distributed network analysis introduced in Chapter 6 to develop the first data-driven comparative method for selecting an OCDA, tailored to a given empirical network. Applying this pipeline to the right-hemisphere structural connectome of the human cerebral cortex, we demonstrate how identifying an optimally suited OCDA uncovers unique biological insights into the brain's complex community architecture, positioning overlapping nodes at the apex of the cortical hierarchy. Our findings highlight that the top-performing OCDA uncovers rich overlapping patterns of connectivity, with derived network properties that correlate with the principal gradient of functional activity (as analyzed in Chapter 6)—patterns that are obfuscated by a comparable traditional hard partition algorithm. Findings from this work underscore the importance of uncovering meaningful overlapping network partitions that better capture how the brain balances functional integration and segregation on a static backbone of structural connectivity. More broadly, the comprehensive methodology to select from a range of OCDAs, provided in open-source code for wide

accessibility, and the biological insights themselves imbue a new perspective on network structure that is highly relevant to researchers across the intersection of brain structure, function, and network topology.

This chapter is published as: Bryant, A. G.*, Jha, A.*, Agarwal, A., Cahill, P., Lam, B., Oldham, S., Arnatkevičiūtė, A., Fornito, A., & Fulcher, B. D. Benchmarking overlapping community detection methods for applications in human connectomics. *Network Neuroscience*, 1-58, doi: https://doi.org/10.1162/netn.a.39.
* indicates co-first authorship.

Abstract
Brain networks exhibit non-trivial modular organization, with groups of densely connected areas participating in specialized functions. Traditional community detection algorithms assign each node to one module, but this representation cannot capture integrative, multi-functional nodes that span multiple communities. Despite the increasing availability of overlapping community detection algorithms (OCDAs) to capture such integrative nodes, there is no objective procedure for selecting the most appropriate method and its parameters for a given problem. Here we overcome this limitation by introducing a data-driven method for selecting an OCDA and its parameters from performance on a tailored ensemble of generated benchmark networks, assessing 22 unique algorithms and parameter settings. Applied to the human right-hemisphere structural connectome, we find that the 'Order Statistics Local Optimization Method' (OSLOM) best identifies ground-truth overlapping structure in the benchmark ensemble and yields a seven-network decomposition of the right-hemisphere cortex. These modules are bridged by fifteen overlapping regions that generally sit at the apex of the putative cortical hierarchy—suggesting integrative, higher-order function—with network participation increasing along the cortical hierarchy, a finding not supported using a non-overlapping modular decomposition. This data-driven approach to selecting OCDAs is applicable across domains, opening new avenues to detecting and quantifying informative structures in complex real-world networks.

7.1 INTRODUCTION

The complex web of structural connections between brain areas can be represented as a network of nodes (brain regions) and edges (axonal connections) [55, 56]. This representation allows us to quantify and understand properties of the brain's network structure, like its organization into densely interconnected modules. Connectome modules often correspond to segregated groups of brain regions with distinct, functionally specialized roles [57, 62], wherein highly connected hub nodes are thought to play an integrative role in communicating diverse information between them [517]. Accurately delineating network modules is an essential step in understanding the brain's modular organization. The community detection algorithms that have most commonly been used to study brain network modules have assumed a hard partition of nodes into mutually exclusive communities; i.e., each node is assigned to a single community. This non-overlapping modular decomposition can be represented as a discrete categorical label that captures each node's modular affiliation [518].

Applications of community detection to brain networks have ranged from the structural network of individual neurons in *C. elegans* [519] and *Drosophila* [520], to the mesoscopic axonal connections in the mouse [521–523], non-human primate [524], and human [525–527], where inferred network modules align with broad specializations of cognitive function [528]. While these studies support a conceptual model of the brain in which segregated groups of nodes process specialized types of information, non-overlapping partitions present limitations that may yield unintended consequences for inferring network-wide modular organization—for example, collapsing two modules bridged by one or more overlapping nodes into a single module. Indeed, transmodal cortical association areas are defined by their flexible multifaceted functions, and might therefore be more accurately affiliated with multiple modules. A representation that allows nodes to belong to multiple communities may thus provide a more complete picture of the brain's structural and functional organization [69, 70].

Overlapping community detection algorithms (OCDAs) decompose the network into a fixed number of modules, as non-overlapping methods do, but with the additional freedom of allowing each node to belong to multiple communities. A wide range of such methods have been developed [529–534], which may offer a more organic method to characterize the brain's spatially distributed and fluid network of complex interactions with an overlapping modular architecture. For example, Wu et al. [535] applied the clique percolation method [536] to identify five communities based on gray-matter volume correlations, revealing overlap among well-known functional systems, such as decision-making and emotional processing. Other studies have characterized the hierarchical network architecture of the mouse [537] and macaque [538] structural connectome with algorithms that allow overlapping community assignment, though the overlap itself was an auxiliary point in both studies. By contrast, considerably more work has focused on overlap among functional networks, evaluating functional magnetic resonance imaging (fMRI) [69, 539–547] and/or calcium imaging [70] with a given OCDA to characterize overlapping community structure in functional networks. Collectively, findings from these studies converge on the view that brain areas spanning multiple structural and/or functional modules enable inter-network communication ('integration') that still preserves intra-module specialization ('segregation')—highlighting how OCDAs offer nuanced information about the rich overlapping patterns in brain community structure that is obfuscated by hard partitions. Moreover, as explored in Bijsterbosch et al. [546], an overlapping node may play a non-stationary role, communicating exclusively with one region at a time or consistently serving as a fully integrative bridging node that facilitates information transfer across functionally specialized systems.

Each of these applications of overlapping community detection to neuroimaging data began with a subjective selection of the OCDA from a range of possible alternatives—without clear evidence of the superiority of one approach over the other for the problem at hand. This speaks to a general challenge in applying community-detection algorithms: each algorithm makes different assumptions about how modules and overlapping nodes (which are members of multiple communities) are defined. We know that applying an OCDA to a network will always produce *a result*, and that differences in the assumptions between OCDAs yield different assignments of nodes to communities. This multitude of choices for an OCDA algorithm and its parameters thus raises the question: How can we determine which OCDAs may exhibit the most accurate and informative decomposition of a given real-world network?

No single algorithm can perform optimally on all possible data (cf. the 'no-free-lunch' theorem [93]); stated in the present context, no community detection method can work well on all networks [94, 95]. Smith et al. [548] developed a 'Question-Alignment' approach that tailors the optimal community-detection method based on hypothesized community properties, but this requires *a priori* predictions about community structure without a quantitative benchmark for comparison. Other previous work has benchmarked OCDA performance on typically very large synthetic networks (> 1000 nodes) and/or relatively small empirical networks separately [534, 549–554], though algorithmic performance is largely dependent on network topology. As a result, it is difficult to reliably extrapolate previous benchmark comparisons to a network with different characteristics—size, sparsity, topology—such as the brain's structural connectome. Rather than selecting an OCDA subjectively or from a benchmark study that may not generalize to a new complex network type, we need a way to objectively tailor our choice of algorithm to the structure of a given network at hand.

We address this challenge by developing a flexible method for the data-driven selection of an OCDA for

We address this challenge by developing a flexible method for the data-driven selection of an OCDA for a given empirical network. Here, we focus specifically on the right-hemisphere structural connectome as our target empirical network, with the goal of narrowing down the optimal OCDA and its parameters with which to identify overlapping community structures and interpret their biological significance. Our approach first generates an ensemble of synthetic benchmark networks with properties derived directly from the target empirical network—the diffusion MRI-derived structural connectome of the right hemisphere from the Human Connectome Project (HCP) [555]—with known 'ground-truth' overlapping community assignments, and then compares the performance of a range of OCDAs on these networks. After identifying the top-performing algorithm across the benchmark network ensemble, we then analyze and interpret the resulting overlapping modular decomposition in the context of brain network structure and function. To our knowledge, this is the first work to comprehensively examine overlapping communities in the human brain's structural connectome using diffusion MRI data. In total, we introduce a novel approach to tailoring OCDA methods to a given real-world network and systematically quantify the overlapping community structure of diverse complex systems.

7.2 METHODS

Given an empirical network, our goal is to systematically evaluate candidate OCDAs to identify a top-performing candidate for inferring that network's overlapping community structure. We propose a data-driven approach with two main steps, as depicted schematically in Fig. 7.2.1: (i) *benchmark calibration*; and (ii) *performance evaluation*. First, we generate an ensemble of benchmark networks that mimic community structure properties of the empirical network and contain ground-truth module assignments of nodes. Second, for each generated benchmark network in the ensemble, we quantify how well each OCDA identifies the ground-truth overlapping community structure. This highlights algorithms with strong overall performance on the benchmarks, which are then judged to most accurately uncover overlapping community structure—yielding OCDA(s) tailored to the target network.

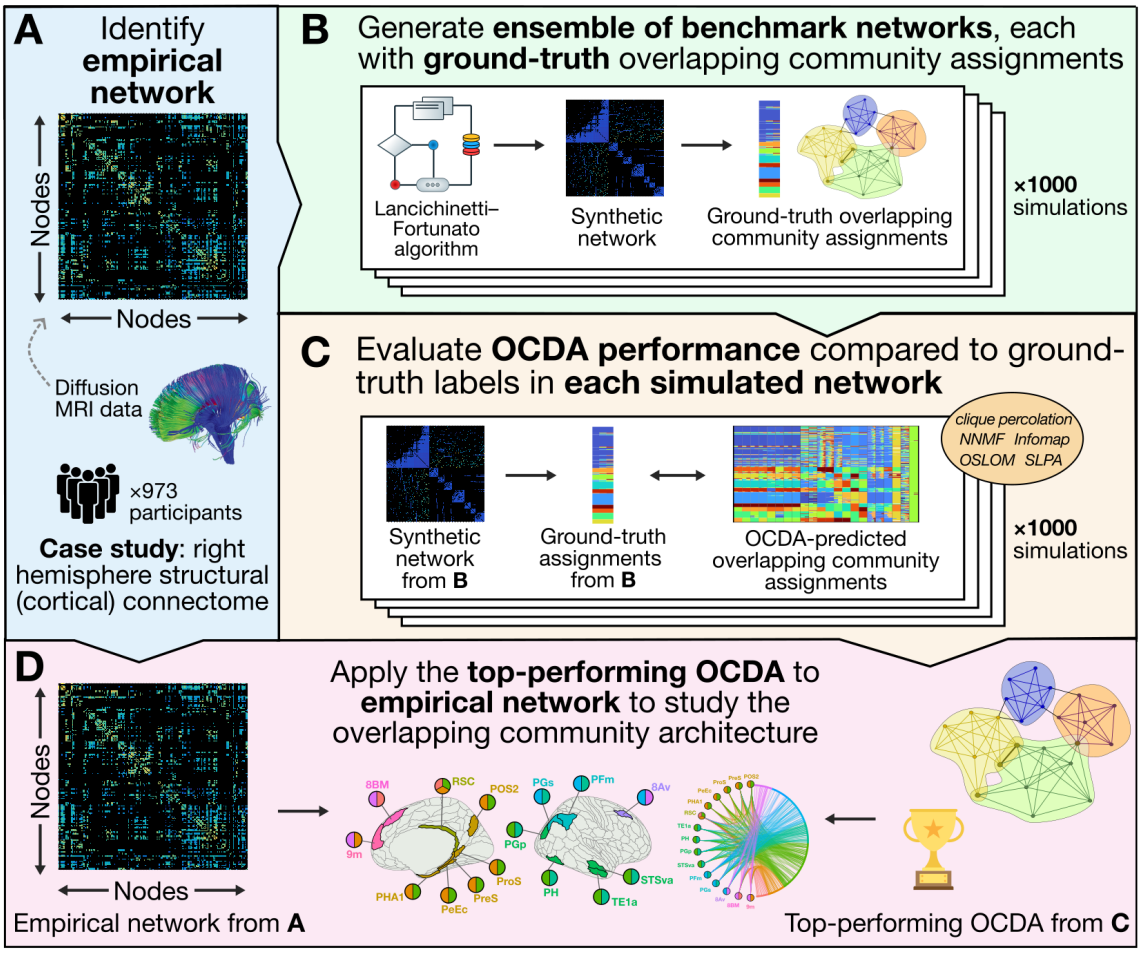


Figure 7.2.1: A data-driven method for tailoring overlapping community-detection algorithms (OCDAs) to empirical networks. **A.** An empirical network is shown as an adjacency matrix. **B.** We use a generative network model [556] to generate an ensemble of networks with known overlapping community assignments, constrained by the properties of the empirical network. **C.** We assess the ability of each OCDA (and parameters) to reproduce the ground-truth overlapping community structure of each synthetic network. Here, we compare five OCDAs: clique percolation [531, 536, 557], non-negative matrix factorization (NNMF), [529], speaker-listener propagation algorithm (SLPA) [532], order statistics local optimization method (OSLOM) [530], and Infomap [558, 559], across 22 parameter variations, quantitatively evaluating their ability to reproduce the known community assignments using Extended Normalized Mutual Information (ENMI). Algorithms with high ENMI across the ensemble are judged to be most appropriate to infer overlapping community structure in the empirical network. **D.** The top-performing OCDA identified in **C** is applied to infer the overlapping community architecture in the empirical network introduced in **A**. *This figure includes images provided under CC BY 2.0 by NICHD NIH (flickr.com/photos/nichd/16672073333) and CC0 by Wikimedia Commons (https://commons.wikimedia.org/wiki/File:Illustration_of_overlapping_communities.svg, https://upload.wikimedia.org/wikipedia/commons/a/a6/Trophy_Flat_Icon.svg).*

7.2.1 PREPROCESSING STRUCTURAL CONNECTOME DATA

For our target empirical network, we apply the systematic OCDA selection method to structural connectome data collected from 973 participants from the Human Connectome Project (HCP) [555]—comprising a subset with complete diffusion MRI data from the broader S1200 release [560]. All participants were healthy and aged between 22 and 35 years and provided written informed consent; ethics was approved by the

Institutional Review Board of Washington University in St Louis [555]. Structural magnetic resonance imaging (MRI) data was acquired using a 3T Siemens Skyra scanner with a customized head coil (100 mT/m maximum gradient strength and a 32-channel head coil) located at Washington University in St. Louis, Missouri. Diffusion-weighted images were acquired using a spin-echo echo planar imaging sequence and processed according to the HCP Diffusion pipeline [506], with additional preprocessing using `MRtrix3` [510, 561] and the `FMRIB` Software Library [361]. Tractography was conducted in each participant's structural MRI (T1) space using second-order integration over fibre orientation distributions (iFOD2), a probabilistic algorithm that improves the quality of tract reconstruction in both highly curved and crossing fibre regions [562]. Anatomically Constrained Tractography (ACT) and Spherically Informed Filtering of Tractograms (SIFT-2) were applied to the tractography data [563] to further improve the biological accuracy of these structural networks. The cortex was parcellated using the 360-region HCP-MMP1 atlas from Glasser et al. [564], comprising 180 cortical regions per hemisphere.

While there is empirical evidence that the probability for a connection between two regions exponentially decreases with the inter-region distance [451], the 'distance bias' in diffusion preprocessing can further decrease this probability—arising in part from the fact that reconstructing longer tracts introduces more steps for algorithmic errors [56, 565, 566]. Given these methodological limitations, we focused on intra-hemispheric connections within a single hemisphere. This allowed us to investigate connectivity properties within a single contiguous and relatively well-reconstructed cortical sheet, while avoiding analyzing inter-hemispheric connections with poor reliability. This choice is a common one among many studies in structural connectomics [28, 567–570]. After confirming that the left and right hemispheres exhibit highly similar edge degree, strength, and variance distributions (cf. Fig. S1)[1], we selected the right hemisphere for our analysis, focusing on intra-hemispheric connectivity between the 180 right-hemisphere regions in the HCP-MMP1 atlas from Glasser et al. [564]. Moreover, as many studies that have performed (non-overlapping) community detection on whole-brain structural networks report bilaterally symmetric solutions [571–574], we believe the inclusion of the left hemisphere would be unlikely to alter the relative performance trends we observe across algorithms.

We constructed a representative group-level connectome from the individual-level data using a thresholding approach, which is common practice to minimize contributions of spurious connections [56, 575] that are typical in diffusion MRI [565, 576]. While sparsification of the starting empirical network is not explicitly required for the benchmark network simulation algorithm of Lancichinetti and Fortunato [556], filtering out noisier edges can improve estimation of parameters used in the algorithm (for example, community size distributions and the mixing parameter, μ). We implemented the consensus-based thresholding approach from [577], retaining only the edges that ranked in the lowest 15% for coefficient of variation (CV) in strength across the entire cohort (N=973). This method retains the most stable edge weights across the entire sample, with edges failing to meet this cutoff threshold set to zero. Of note, one node—*13l*, in the orbitofrontal cortex—exhibited highly variable edge weights (all with CV of 0.3 or higher), and had a total degree of zero (after applying the 0.15 threshold-based consensus). To prioritize consistency in edge weights for community detection, we thus excluded region *13l* from our analysis. In this group-level connectome, edge weights correspond to the thresholded average number of streamlines between the two regions terminating

[1] Supplementary figures are included in in Appendix 6.

within a 5 mm radius of each other. Due to the range in streamline count magnitudes across edges, which spanned multiple orders of magnitude, we log-transformed the right-hemisphere edge weights between region–region pairs before further analysis.

7.2.2 GENERATING BENCHMARK NETWORKS

Once we have an empirical network as a starting point (i.e., the preprocessed consensus right-hemisphere structural connectome in this case), the first step in our comparative method involves generating an ensemble of benchmark networks with ground-truth overlapping community assignments, as depicted in Fig. 7.2.1B. For the subsequent inference to be informative, these generated benchmark networks should have properties as similar as possible to the empirical network. Here, we focus on weighted, undirected networks (corresponding to our intended application of structural brain networks, such as those estimated using diffusion-weighted imaging [578]). We use the stochastic network generation algorithm of Lancichinetti and Fortunato [556] which—for a given set of parameters that constrain the structural properties of the generated benchmarks—produces weighted networks with overlapping community assignments. Our ability to tune the benchmark networks to match the properties of the target empirical network is controlled by the algorithm parameters, which we describe briefly in following.

Across a given number of nodes, N, Lancichinetti and Fortunato [556] benchmark networks are assumed to have a power-law degree distribution, $p_k \sim k^{-\tau_1}$, with average degree, $\langle k \rangle$, and maximum degree, $k_{\max}$. Community sizes are sampled from a power-law distribution with exponent τ_2, with a specific minimum and maximum community count along with a set number of overlapping nodes, O_n. Edge weights are assigned under the constraint of a power-law relationship between degree, k, and strength, s, of network nodes, $s = k^{\beta}$, determined by the parameter β. The extent to which a node's connections (and their weights) are contained within its own communities (or spread across other communities) is controlled through two mixing parameters: (i) μ_t, the average fraction of a node's edges that are external to the communities it belongs to; and (ii) μ_w, the average fraction of a node's aggregate edge weights external to the communities to which it belongs. Given these parameters, benchmark networks are generated by forming links between nodes within and outside their assigned communities across iterative network rewirings.

To generate benchmark networks with properties that match the empirical network as closely as possible, we require a systematic method to fit the generative parameters of the Lancichinetti and Fortunato [556] algorithm. Some parameters that control the basic connectivity properties of the network (N, $\langle k \rangle$ and $k_{\max}$) were computed directly from the empirical network or—in the case of τ_1 and β—were estimated from power-law fitting algorithms [579]. Other parameters governing the community structure (e.g., number and size of communities and overlapping nodes) cannot be inferred from the network but instead require subjective judgment about the type and resolution of community structure that is most appropriate for the application of interest. In the absence of ground truth parameter values grounded in clear domain knowledge (or other priors on the community structure), we sampled these parameters from uniform distributions across reasonable ranges. Specifically, we allowed communities to take any size (from 1–N nodes), and sampled the power-law distribution exponent, $\tau_2 \sim U(2, 4)$, and the number of overlapping nodes, $O_n \sim U(0.1N, 0.2N)$, from uniform distributions. To ensure that connectivity was predominantly within communities, mixing parameters were sampled in ranges: $\mu_t \sim U(0.2, 0.4)$, and $\mu_w \sim U(0.2, 0.4)$. Using this combination of fixed

parameters estimated from the network, $(N, \langle k \rangle, k_{\max}, \tau_1, \beta)$, and randomly sampled parameters representing reasonable ranges of the overlapping modular structure of interest, $(\tau_2, O_n, \mu_t, \mu_w)$, we generated an ensemble of 1,000 benchmark networks, each resulting from a sampled parameter vector.

The Lancichinetti and Fortunato [556] generative network model makes assumptions about the data that may not hold in all cases; for example, the power-law form of degree distribution that this model assumes will not always be a good representation of real network data. As alternative models are developed in the future that can better fit the key structural properties of a real network, these could be substituted into this benchmark generation step to improve the representativeness of the simulated benchmark networks to the real network data.

7.2.3 OVERLAPPING COMMUNITY DETECTION ALGORITHMS

Once we have generated an ensemble of benchmarks that resemble the empirical network (Fig. 7.2.1B), we next compare the performance of OCDAs in reproducing the known community structure (Fig. 7.2.1C). Here we investigated five different OCDAs, each with various parameter options and delineated briefly below.

Clique Percolation [531, 557] generates overlapping communities by clustering a maximal clique network, and then projecting this hard partition of maximal cliques back to an overlapping partition of nodes. We used the algorithmic implementation of clique percolation developed in Shen et al. [536], which introduces a 'quality of network coverage' measure (Q_c) that is optimized based on the maximal clique view. Specifically, a clique is a complete subgraph, and maximal cliques are not subgraphs of any other cliques in a graph. The method first generates a new network, G', in which each node is the maximal clique of the original network, G, and then computes a hard partition of this new network. As some nodes may be members of multiple cliques, this yields overlapping community assignments in the original network; importantly, the implementation in Shen et al. [536] ensures that all nodes are assigned to at least one clique such that there are no isolated nodes, which is a departure from the original method in Derényi et al. [557] and the subsequent extension to incorporate weighted edges in Farkas et al. [580]. The algorithm includes a parameter, k, that sets the minimum size of maximal cliques to be mapped in the new network, G'. Here we compare seven possible values of this clique size parameter, $k = 3, 4, ..., 9$, labeling the resulting algorithms, correspondingly, as: `Clique_3`, ..., `Clique_9`.

Non-Negative Matrix Factorization (NNMF) [529] uses a Bayesian non-negative matrix factorization model to assign a participation score to each node for each community, which is based on the interaction between that node with other members of the community. NNMF assigns every node a soft membership distribution across communities that can be interpreted as a membership probability of each node to each community. Specifically, we row-normalized the non-negative matrix W from the NNMF decomposition $X \approx WH$, yielding the matrix P, where the probability of node i participating in community k is given as

$$P_{ik} = \frac{W_{ik}}{\sum_{k'=1}^{i} W_{ik'}} \,. \tag{7.1}$$

To convert this set of continuous community membership probabilities into a set of discrete labels defining the set of community memberships for a given node, we applied a threshold, p_{th}. That is, the node i is

considered a member of community k if its community membership probability $P_{ik} \geq p_{\text{th}}$. We compared four thresholds, $p = 0.1, 0.2, 0.3, 0.4$, and labeled the resulting algorithms as: `NNMF_10`, `NNMF_20`, `NNMF_30`, and `NNMF_40`, respectively. While the theoretical space of solutions in NNMF is non-unique, the combination of row-normalization and Bayesian regularization in the method from Psorakis et al. [529] promotes consistent and interpretable membership distributions across networks.

Order Statistics Local Optimization Method (OSLOM) [530] is a local method that attempts to find and modify the statistical significance of clusters relative to a global null model in which there is no community structure (the configuration model [581]) by adding/removing neighboring nodes. Thus, OSLOM yields clusters that are statistically unlikely to be found in an equivalent random graph with the same degree sequence; the set of individual clusters found in a network may overlap. In adjusting the significance level, P, from which to assess whether to make local changes to a cluster, the resulting size of clusters is changed (low values of P yield fewer, larger clusters). We compared a range of ten values $P = 0.1, 0.2, ..., 1.0$, denoted as `OSLOM_10`, `OSLOM_20`, ..., `OSLOM_100`, respectively.

Speaker-Listener Label Propagation Algorithm (SLPA) [532] is an extension of the Label Propagation Algorithm (LPA) to the overlapping setting. In LPA, the community assignments of individual nodes are updated iteratively based on the majority label possessed by its immediate neighbors [582]. In the overlapping generalization, SLPA, nodes are not forced to take a single label, but can hold the memory of previously observed labels; the more frequently a node has observed a node label, the more likely it is to then propagate it to its neighbors. After some maximum number of iterations of label propagation, nodes are assigned to communities based on the most frequently observed labels in their memories. Once communities are identified, SLPA removes labels observed less than the fraction r; while r values in the range 0.01 to 0.45 have been used previously [583–586], here we set $r = 0.09$ as it sat within the range of thresholds yielding convergent results in the original SLPA publication [532]. Xie et al. [532] note that the set threshold is exclusively for post-processing (i.e., refining communities after they have been defined), such that the SLPA dynamics are determined solely by network topology and employed interaction rules. We denote this method as `SLPA`.

Infomap [558, 559, 587] is an information-theoretic method operating under the principle that network communities can be identified by compressing the description of the probability flow of random walks. It solves the map equation, which quantifies the description length of a random walker's trajectory through the network, by balancing the information required to describe movements between communities and that needed to move within communities. Infomap involves construction of a 'codebook', meaning a set of labels (or 'codes') assigned to different parts of the network that efficiently capture the trajectories of the random walker within the network. This method was extended to capture overlapping nodes using a generalized map equation that allows these codebooks to allow multiple modules assigned toa single node [587]. We denote this algorithm as `Infomap`.

In our benchmark comparison procedure, outlined in the following sections, the five OCDA methods and corresponding parameter combinations yielded a total of 22 specific OCDA algorithms to evaluate. One important distinction between OCDA types rests on their treatment of edge weights. Some algorithms, such as Clique Percolation and SLPA, require the input network to be binarized—effectively representing the network as a topology of connections, without consideration of connection strengths. This binarization

can result in a loss of information for a network where edge weights carry meaningful structural and/or functional significance [588], as in brain connectivity data. On the other hand, algorithms like OSLOM, NNMF, and Infomap are inherently designed to incorporate the graded structure of edge weights directly into the process of community detection. This methodological divergence has important implications for benchmarking, such that comparing outputs from binarized versus weighting-sensitive algorithms requires careful consideration of how (and if so, the extent to which) the usage of edge weights might affect algorithm performance. We chose to include both types of algorithms in our analysis for a representative cross-section of common OCDAs used across disciplines. To bolster comparability, we opted for implementations with only one (if any) user-defined parameter to systematically vary such that all other conditions are fixed per algorithm. While edge-weight sensitivity may confer a potential advantage in networks with informative edge weight distributions, our goal is to compare all methods under conditions matched as closely as possible (manipulating only the key parameter per algorithm; e.g., k in Clique Percolation) to isolate algorithm-specific differences in overlapping community detection performance.

7.2.4 PERFORMANCE EVALUATION

After applying each OCDA to infer the overlapping community structure in each benchmark network, we next assessed how well each algorithm reproduces the ground-truth overlapping community assignment, as depicted in Fig. 7.2.1C. Quantifying the relative performance of OCDAs is challenging, though several measures have been developed to capture different aspects of the goodness-of-fit relative to a ground-truth community decomposition for a synthetic benchmark network (or ensemble of networks, in our case). Here we used extended normalized mutual information (ENMI) to quantify the overall similarity of two overlapping community assignments. ENMI is an extension of normalized mutual information (NMI), which was introduced in Danon et al. [589] (and further discussed in Newman et al. [590]). Normalized mutual information is defined as

$$I_{\mathrm{norm}}(X;Y) = \frac{I(X;Y)}{[H(X)+H(Y)]/2}\,, \tag{7.2}$$

where $I(X;Y)$ is the mutual information [147] between the random variables X and Y. NMI captures the amount of information shared between X and Y, quantified as the reduction in uncertainty in X while observing Y (and vice-versa; this is a symmetric measure). For a non-overlapping community assignment, X and Y for a given node i are coded as integers, i.e., $x_i = 1, 2, \ldots, |C_X|$, up to the total number of communities in the first partition $|C_X|$, and similarly for y_i.

In the case of an overlapping decomposition, multiple community assignments per node can be captured using a binary encoding, with $(\mathbf{x}_i)_k = 1$ if node i is a member of community k, and $(\mathbf{x}_i)_k = 0$ otherwise. In this case, for each node i, each community k can then be considered as a random variable, e.g., as $X_k = (\mathbf{X})_k$, with random variables for all k communities being captured as $\mathbf{X}$. Probabilities $P(X_k = 1)$, $P(X_k = 0)$, and similarly for Y_l, can be estimated from a given overlapping modular decomposition, allowing information-theoretic computations to be performed on the resulting collections of binary random variables $\mathbf{X}$ and $\mathbf{Y}$. Lancichinetti et al. [591] adapted NMI to apply to this overlapping setting, yielding extended NMI (ENMI), as outlined in Appendix B of Lancichinetti et al. [591]. ENMI ranges from 0 to 1, and given our setting of comparing a ground-truth overlapping community assignment **X** with an algorithm-predicted assignment **Y**, a larger ENMI value indicates superior recovery of the underlying overlapping community

structure in the benchmark network.

ENMI is frequently implemented as a proxy for the goodness of fit with which an algorithm infers the ground-truth overlapping community structure in a given network [592–596], including the study by Xie et al. [532] which introduced the SLPA method included in our comparisons. Lancichinetti et al. [591] demonstrated that ENMI increases monotonically as the community structure becomes more pronounced (based on varying the mixing parameter in the same generative benchmark network algorithm employed in this study). Moreover, while it is not straightforward to evaluate the 'accuracy' of specific node-to-module assignments in a label-matched manner across algorithms, ENMI captures the overall match to ground-truth community structure in a label-invariant manner. Altogether, ENMI is theoretically suitable for the multi-membership nature of overlapping communities, and its implementation makes it flexible for cases where the algorithm introduces extra communities beyond the number of ground-truth communities.

OCDAs introduce additional algorithmic complexity and computational burdens beyond those of a non-overlapping option, like the Louvain method [63], which is the most widely used community-detection method in network neuroscience [64]. In order to evaluate whether an OCDA better recovers community structure in a network with known ground-truth overlaps than a simpler non-overlapping method, we additionally applied the Louvain (non-overlapping) community detection algorithm to each network in the benchmark ensemble. This allowed us to compute the ENMI between the resulting Louvain decomposition and the ground-truth community structure in order to compare if, and by how much, an OCDA improves community detection in simulated networks resembling the brain's right-hemisphere cortical connectome. We then compare the computational runtime for each method across the 1,000 simulated networks in the benchmark ensemble to evaluate performance–computational burden tradeoffs across the methods (including Louvain).

Finally, in order to evaluate how well each OCDA correctly identifies specific overlapping nodes in the benchmark ensemble, we compared the sensitivity and specificity for the ground-truth nodes in each network. In both cases, we binarized all 180 nodes (i.e., regions in the right cortex) according to both the ground-truth assignments (X), and the algorithm-predicted assignments (Y); assigning a value of 1 only if that node belongs to two or more communities. This allowed us to construct a confusion matrix, where 'positive' corresponds to overlapping assignment and 'negative' corresponds to single-community assignment. We then computed the sensitivity and specificity for each OCDA and each simulated network in the benchmark ensemble. These metrics complement the ENMI, which summarizes across all 180 nodes regardless of overlapping status, to focus specifically on the recovery of ground-truth overlapping node identity. A high sensitivity means that the algorithm correctly identified many of the overlapping nodes in the network, while a high specificity means the algorithm did not assign many non-overlapping nodes to multiple communities.

7.2.5 APPLYING OSLOM TO THE EMPIRICAL STRUCTURAL HUMAN CONNECTOME

Since OSLOM was identified as the top-performing OCDA in the benchmark ensemble analysis, we further refined its threshold parameter (P) for the applied case study analysis with the empirical right-hemisphere cortical connectome. We compared modular decomposition at each $P = 0.01, 0.02, \ldots 1$ to identify a suitable range of P values, across which OSLOM identified a consistent number of modules and overlapping nodes. Our empirical selection of a representative P-value threshold for application to the observed right-

hemisphere structural connectome is guided here by decomposition stability, rather than in-sample performance, precisely to minimize the risk of overfitting. Importantly, OSLOM is a stochastic algorithm, such that the resulting decomposition can vary depending on the user-supplied seed for the pseudo-random number generator (which determines which nodes are initially selected as the basis for community evaluation) [530]. Most prior applications of OSLOM to interdisciplinary real-world problems that specify P and algorithm hyperparameters do not specify the random seed used to generate their results (in which case the seed is automatically generated based on the time at analysis, down to the microsecond) [597, 598]. Bruner et al. [599] used the singular pre-specified seed of 73, and Acman et al. [600] examined a handful of random seeds (1, 5, 42, 93, 212) before manually selecting results obtained with an individual seed (42). Gates et al. [543] did not specify whether seeds were set, but they applied OSLOM to the functional brain connectome 10 times per participant, deriving an 'element-centric similarity matrix' based on the similarity of individual OSLOM decompositions between participants.

Recognizing that the robustness of a particular decomposition is essential for inferring generalizable biological significance in the human connectome, we systematically evaluated the variability of OSLOM outputs in the fixed-parameter setting across different random initialization seeds. We verified that OSLOM produced deterministic outputs for a given initialization seed and P-value threshold combination, mitigating concerns about internal stochasticity. For each threshold between 0.01 to 1, we ran OSLOM using 100 different initialization seeds (setting the seed to each of the integers from 1 to 100) and examined the average number of modules and number of overlapping nodes across thresholds. After narrowing the P-value range down to $P = 0.15, 0.16, \ldots 0.35$ based on the stability in the number of modules and overlapping nodes (averaged across seeds; cf. Figs. S2A-B), we then selected a final value of P=0.3 as an intermediate value that yielded top performance in the simulated benchmark analysis.

Having selected $P = 0.3$, we next sought to identify a particular decomposition from the 100 realizations (corresponding to the 100 initialization seeds) that is most similar to (i.e., most 'central') all other decompositions—analogous to selecting a Fréchet in a distance-based framework. We quantified similarity in modular decomposition between each pair of seeds using the ENMI (Fig. S2C, cf. Sec. 7.2.4) for consistency in the comparison measure used to evaluate each OCDA against each simulated network in the benchmark ensemble. We additionally constrained our search to seeds that yielded module sizes between 10 and 60 nodes (inclusively) in order to focus on biologically interpretable region groups. This approach favors the selection of a solution that is most stable (i.e., consistent) across initialization, providing a single representative decomposition that best captures the consensus structure detected by the algorithm (in this case, OSLOM). This similarity analysis identified seed 61 as the most representative, exhibiting a mean ENMI of 0.64 ± 0.11 with all other seeds; for comparison, the mean ENMI across all other seeds was 0.59 ± 0.10. Of note, seeds with higher average ENMI values (including seed 61) also exhibited greater variance in ENMI values; this heteroskedastic relationship is depicted as a scatter plot in Fig. S3A. Although we did not explicitly penalize for variance in our seed selection, the tight coupling between mean and median ENMI values shown in Fig. S3B supports its use as a proxy for 'representativeness'. All subsequent analyses in the empirical right-hemisphere cortical connectome case study were conducted using the modular decomposition derived from this seed–parameter combination. Taken together, while stochasticity in community detection algorithms is challenging, we developed and implemented a systematic approach from which we selected a representative result from OSLOM to subsequently analyze.

7.2.6 COMPARING OVERLAPPING NODES WITH BIOLOGICALLY RELEVANT PROPERTIES

In order to investigate various aspects of how the overlapping community structure relates to the putative functional hierarchy in the cortex, we used principal gradients (PGs) of functional connectivity [414]. We obtained pre-processed maps of the first and second PGs (PG1 and PG2, respectively) from the *neuromaps* python package [440] (version 0.0.5). After transforming the reference maps into `fsaverage` surface space, we analyzed the vertex-wise values for PG1 and PG2 for each of the 180 regions in the HCP–MMP1 parcellation [564]. We summarized the mean values across vertices to obtain region-averaged values for PG1 and PG2, respectively.

7.2.7 COMPARING OVERLAPPING AND NON-OVERLAPPING MODULAR PARTITIONS

To clarify whether the top-performing OCDA can provide additional and unique information beyond that of a non-overlapping partition, we compared network properties from the two modular decomposition types. For a representative non-overlapping partition, we applied Louvain community detection using the method from Rubinov and Sporns [601], first tuning the resolution parameter γ to yield relatively consistent decompositions across 100 iterations, each with the seed set to an integer from $1, 2, \ldots 100$ (as with OSLOM). After sweeping across γ values ranging from $0.5 \leq \gamma \leq 1.5$, we selected $\gamma = 1$ (which happens to be the default in the Brain Connectivity Toolbox [601]) as this forms an elbow in the number of communities with lower variance across seeds than higher γ values (as plotted in Fig. S4A). We selected a representative seed (98) based on mean normalized mutual information (NMI) with all other seeds (total of 100, as shown in a heatmap in Fig. S4B), which yielded a final six-community decomposition—projected onto the right hemisphere cortical surface in Fig. S4C. As a robustness test, we also compared results using $\gamma = 1.5$ to enforce a seven-module (non-overlapping) decomposition for the direct comparison of seven communities between OSLOM and Louvain, with the results shown in Fig. S4D-G.

We computed two frequently-used nodal measures from the Louvain decomposition using the python implementation of the Brain Connectivity Toolbox [601] (`bctpy`, https://github.com/aestrivex/bctpy): (i) within-module strength (z-scored), z, which measures how strongly a node is connected to others within its own community; and (ii) participation coefficient, P, which measures the diversity of a node's connections across network modules [56, 61, 602]. To understand the relative position of each OSLOM-defined overlapping node within the non-overlapping Louvain partition, we plotted all nodes in $z - P$ space, following the characterization in Guimerà and Nunes Amaral [61]. For direct comparison with the top-performing OCDA, we adapted the algorithm for P to allow each overlapping node to be assigned to multiple modules.

7.2.8 DATA VISUALIZATION AND CODE AVAILABILITY

The overlapping community decomposition was visualized within the parcellated cortical surface using the *ggseg* package in R [386]. We used the *NetworkX* package [603] in Python to visualize community structure with a 'circos' layout and an undirected graph network layout in Fig. 7.3.3B. Raincloud plots are visualized using the 'see' [516] and 'ggplot2' [604] packages in R. An open and extendable Matlab implementation of our methodology is available on GitHub[2], which also includes all code to reproduce our main results in a combination of Matlab (and R, for visualization purposes only). The methods

[2]https://github.com/NeuralSystemsAndSignals/OverlappingCommunityDetection_HCP

schematic in Fig. 7.2.1 includes images provided under CC BY 2.0 by NICHD NIH (flickr.com/photos/nichd/16672073333) and CC0 by Wikimedia Commons (https://commons.wikimedia.org/wiki/File:Illustration_of_overlapping_communities.svg, https://upload.wikimedia.org/wikipedia/commons/a/a6/Trophy_Flat_Icon.svg).

7.3 RESULTS

Here, we apply our data-driven algorithm selection method to characterize the overlapping modular structure of the right cortical connectome, derived from 973 individuals in the HCP dataset [555]. As depicted schematically in Fig. 7.2.1, our systematic approach generates an ensemble of benchmark networks with properties that mirror the target empirical connectome, with a ground-truth overlapping structure that enables selection of the top-performing OCDA. After identifying OSLOM as the top-performing algorithm across the benchmark ensemble, we then analyze the overlapping community structure inferred by this method, comparing the resulting decomposition with functional properties to demonstrate the unique insights beyond those that can be inferred from a traditional non-overlapping partition.

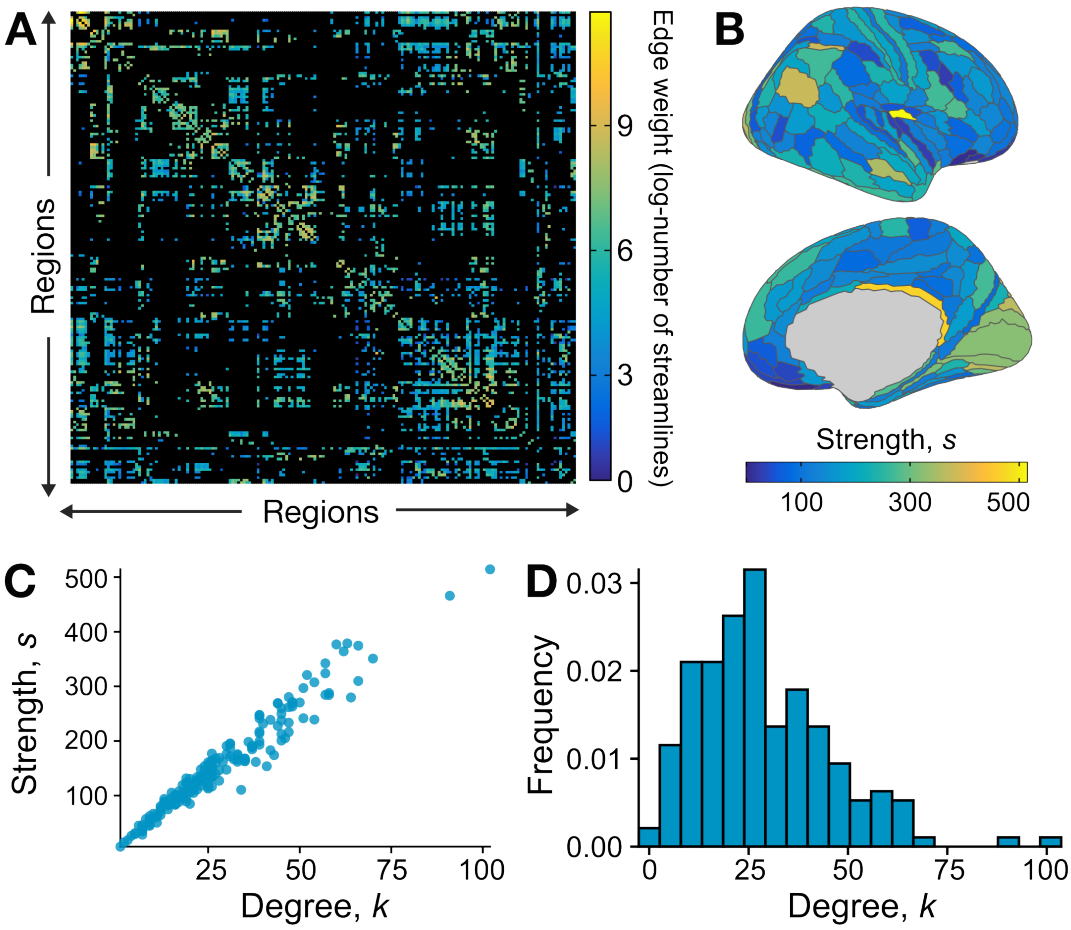


Figure 7.3.1: Network properties of the group-level structural connectome of the right hemisphere of the human cortex. **A.** The empirical structural connectome of the right cortical hemisphere is shown as a weighted 180×180 adjacency matrix, with color representing the log-transformed number of streamlines (connections) between nodes. **B.** The strength, *s*, is computed for each region (as the sum of all log-transformed edge weights) and projected on the cortical surface for the right hemisphere. **C.** Plotting node strength (s) vs. node degree (k, the sum of binarized connections) demonstrates that $s \propto k$—enabling the inference that $s \propto k^{\beta}$, with $\beta = 1$, for Lancichinetti and Fortunato [556] benchmark network generation. **D.** The frequency distribution for degree (k) values across nodes, which does *not* exhibit power-law scaling.

7.3.1 THE CONNECTOME-TAILORED BENCHMARK NETWORK ENSEMBLE HIGHLIGHTS OSLOM AS THE TOP-PERFORMING ALGORITHM

In order to tune parameter values for the generative network algorithm of Lancichinetti and Fortunato [556], we first examined key network properties of the empirical human right-hemisphere structural connectome. We constrained our analysis to intra-hemispheric connections in the right hemisphere, confirming that the left and right hemispheres exhibited highly similar edge degree, strength, and variance distributions (cf. Fig. S1). For the 180-node right-hemisphere human structural connectome, shown in Figs. 7.3.1A-B, we followed the general procedure depicted in Fig. 7.2.1. We first needed to set the benchmark-generation parameters, $(N, \langle k \rangle, k_{\max}, \tau_1, \tau_2, \beta, O_n, \mu_t, \mu_w)$, to match the data, as described in Methods (Sec. 7.2.2). Basic network parameters were computed directly from the network: $N = 180$, $\langle k \rangle = 29$, and $k_{\max} = 102$.

Plotting node degree, k, against node strength, s, shown in Fig. 7.3.1C, revealed $s \propto k$, allowing us to identify the exponent in $s \propto k^{\beta}$, as $\beta = 1$. A limitation of the Lancichinetti and Fortunato [556] algorithm is its assumption of a power-law degree distribution, $p_k \sim k^{-\tau_1}$ which, as shown in Fig. 7.3.1D, does not hold for this structural connectome. Of note, deviations from power-law distributions have been reported in diverse connectomic datasets and preprocessing paradigms [605–607], suggesting that this is not unique to the dataset and/or analytical choices implemented here. We nevertheless set τ as the best power-law fit to this distribution, as $\tau_1 = 2$. Sampling from all other parameters across appropriate ranges (using uniform distributions, as described in Sec. 7.2.2), we generated an ensemble of 1,000 right-hemisphere structural connectome benchmark networks, each with a ground-truth overlapping community assignment. The benchmark ensemble was generated on a MacBook Air with an Apple M2 chip and 16GB RAM, and with the above-described parameter configurations, the 1,000 networks were generated and saved in 224 seconds (3.7 minutes). We evaluated the performance of 22 OCDAs (based on five representative algorithms) on each generated benchmark network using the ENMI metric (cf. Sec. 7.2.4). For each simulated network, we fit each of the 22 OCDAs comprising five core methods (Infomap, SLPA, OSLOM, Clique Percolation, and NNMF), along with the Louvain method [63] as a non-overlapping baseline comparison method. We compare the computational runtime required for each algorithm group in Table 7.3.1, finding that among the OCDAs—all of which took at least one order of magnitude longer than Louvain (0.002 $\pm$0.001 seconds per 180$\times$180 network; median $\pm$ inter-quartile range, IQR)—SLPA was the fastest (0.04 $\pm$0.01 seconds) while OSLOM was the slowest (12$\pm$6 seconds).

Table 7.3.1: Computational runtime for the OCDAs (and Louvain as a baseline) across the 1,000 networks in the benchmark ensemble. All analyses were computed sequentially on a MacBook Air with an M2 silicon chip and 16GB of RAM. Methods are sorted in order of median runtime from smallest to largest. IQR: interquartile range

Method	**Median runtime (s)**	**IQR runtime (s)**
Louvain	0.002	0.001
SLPA	0.04	0.01
Infomap	0.4	0.1
Clique Percolation	0.7	0.4
NNMF	1.1	0.1
OSLOM	12	6

We schematically depict our process for evaluating the decompositions generated by each OCDA across the 1,000 networks simulated with the Lancichinetti and Fortunato [556] method in Fig. 7.3.2A, highlighting one representative benchmark network for illustrative purposes. The 'ground-truth' community structure and overlapping community assignments are shown in Fig. 7.3.2B, along with those predicted by each of the 22 OCDAs. Visually, we see that variations of the OSLOM method best recapitulate the known overlapping community structure of this representative benchmark, though Clique Percolation results (with larger k-clique sizes) also exhibit similar structure to the ground-truth decomposition. To summarize across all 1,000 generated benchmark networks, in Fig. 7.3.2C, we show the ENMI distributions as raincloud plots for each OCDA configuration (excluding OSLOM with parameter settings that yielded redundant decompositions). We also included the ENMI distribution for the Louvain method for comparative purposes, with results also summarized in Table 7.3.2. The three OSLOM algorithms yielded the highest median ENMI values (0.83–0.84, IQR 0.15–0.16), indicating that OSLOM best uncovered the community structure across

human right-hemisphere structural connectome-like benchmark networks out of the methods compared here (cf. Fig. S5). Of note, all parameter configurations of OSLOM yielded higher median ENMI values than the Louvain partition (0.65, IQR 0.31), collectively supporting its utility despite the higher computational burden than other methods. Clique Percolation algorithms exhibited the second-highest median ENMI values (0.55–0.69, IQR 0.19–0.37), though only the larger cliques ($k \geq 6$) yielded higher median ENMI values than the Louvain decomposition. As noted in the Methods (cf. Sec. 7.2.4), Clique Percolation binarizes edge weights while OSLOM retains edge connection strengths, which may be a contributing factor to the performance differences between the two algorithms. Within NNMF, maximal ENMI values were achieved with probability thresholds (p) > 0.1, while Infomap and SLPA exhibited generally weaker performance; all three general OCDA classes yielded lower median ENMI values than the Louvain method.

In terms of recovering overlapping nodes specifically, we also evaluated the sensitivity and specificity of algorithm-predicted overlapping nodes relative to the ground truth in each simulated network. As shown in Fig. S6, NNMF with different thresholds yielded the highest overall sensitivity (meaning it recovered the most ground-truth overlapping nodes out of all evaluated algorithms), but specificity values were considerably lower than for other algorithms—indicating that NNMF generally predicts more nodes to be overlapping than are actually overlapping. Moreover, NNMF also yielded overall considerably lower ENMI values compared to OSLOM, Clique Percolation, and even the non-overlapping Louvain decomposition, suggesting that its strong performance in identifying overlapping nodes does not directly generalize to its overall network community structure inference. OSLOM consistently yielded the second-highest range of sensitivity values (31.6–60.9%) across P-thresholds, while maintaining 100% specificity (with the exception of OSLOM-100, which exhibited a specificity of 96.8$\pm$5.2%). The other three OCDA classes—Clique Percolation, Infomap, and SLPA—all exhibited near-perfect specificity (all $>$97%) yet very low sensitivity to ground-truth overlapping nodes (all $<$17%). Altogether, the benchmark ensemble analysis suggests that OSLOM best recovers the overlapping community structure of the human (right-hemisphere) structural connectome out of the evaluated OCDAs, in terms of overall community identification (ENMI analysis) as well as identification of specific overlapping nodes (sensitivity/specificity analysis). Given these results, we progressed with OSLOM as the preferred overlapping community-detection algorithm to study the structural architecture of this intra-hemispheric cortical connectome dataset.

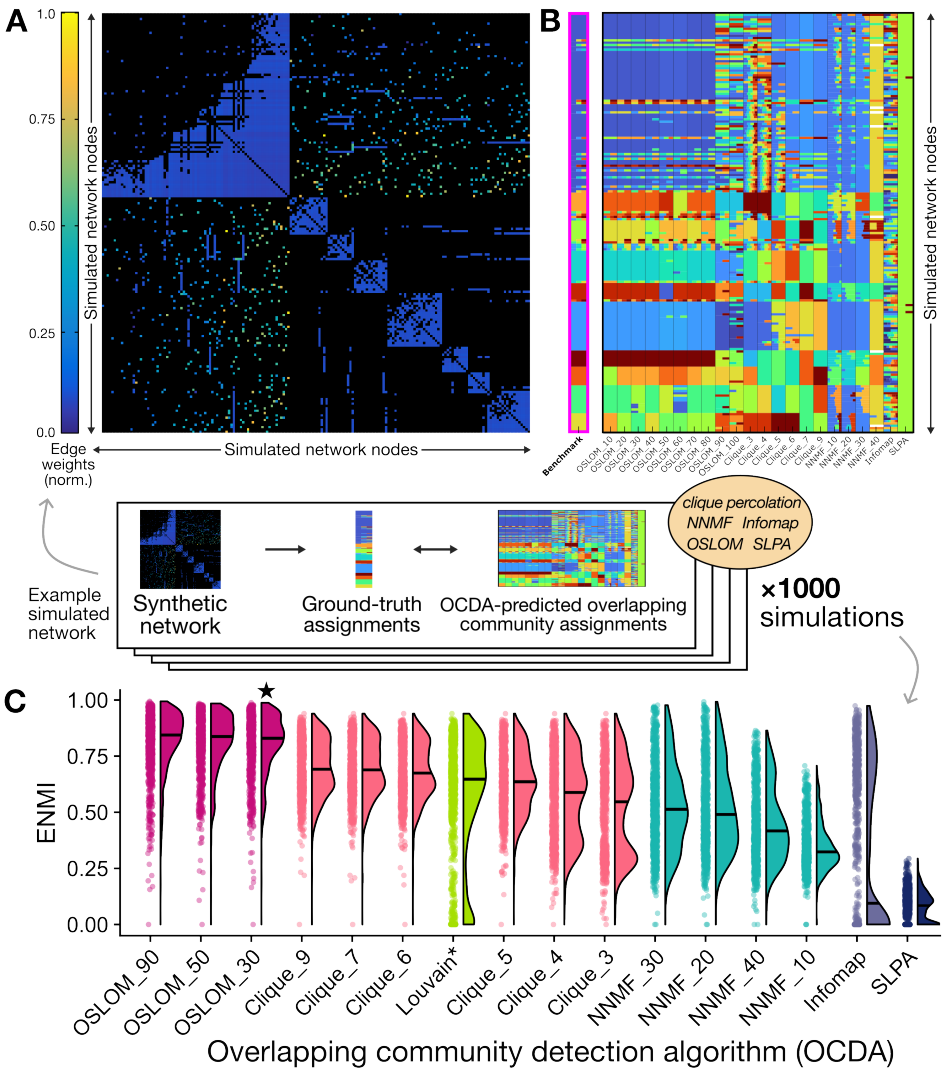


Figure 7.3.2: Overlapping community structure predicted by OCDAs (across 22 different combinations of parameters for five types of OCDA methods) on a representative connectome-like benchmark network. A. A representative simulated network from the benchmark ensemble of 1,000 networks, comprised of 180×180 nodes representing the right-hemisphere HCP-MMP1 regions [564] as with the empirical structural connectome in Fig. 7.3.1A. The same color map is used as with the empirical connectome matrix for visual consistency. Regions are sorted according to the 'Benchmark' (i.e., ground-truth) community assignment, and cells are colored according to (simulated) connection strength, normalized to the unit interval $[0, 1]$ for visualization. **B.** The 'Benchmark' column (highlighted in pink with bold text) represents the ground-truth community structure for the network depicted in **A**, in which each of the 180 nodes is assigned to one or more communities (represented by colors). Nodes belonging to the same community are assigned the same color, and overlapping nodes are assigned multiple colors. The other 22 columns on the right depict predictions of the overlapping community structure from each OCDA. Note that colors designate different communities within each OCDA column, but individual colors are not directly comparable between OCDA results (i.e., between columns). **C.** For each algorithm, ENMI performance distributions are shown as raincloud plots across the generated ensemble of 1,000 benchmark networks. Algorithms are sorted by mean ENMI across the ensemble (shown as black horizontal lines), and algorithms are color-coded as OSLOM (magenta), Clique Percolation (light pink), NNMF (teal), Infomap (light blue), and SLPA (navy blue). ENMI values are also shown for the Louvain method (green) as a baseline non-overlapping method for comparison. Note that OSLOM results are shown here for three selected resolutions ($P = 0.3, 0.5, 0.9$); ENMI distributions are similar for other values of P.

Table 7.3.2: Extended normalized mutual information (ENMI) for the OCDAs (and Louvain as a baseline) across the 1,000 networks in the benchmark ensemble.

Method	Median ENMI	IQR for ENMI
OSLOM_90	0.84	0.15
OSLOM_80	0.84	0.15
OSLOM_60	0.84	0.16
OSLOM_70	0.84	0.15
OSLOM_50	0.84	0.16
OSLOM_40	0.83	0.15
OSLOM_30	0.83	0.15
OSLOM_20	0.82	0.15
OSLOM_10	0.82	0.17
OSLOM_100	0.79	0.17
Clique_9	0.69	0.19
Clique_7	0.69	0.19
Clique_6	0.68	0.19
Louvain*	0.65	0.31
Clique_5	0.64	0.21
Clique_4	0.59	0.32
Clique_3	0.55	0.37
NNMF_30	0.51	0.27
NNMF_20	0.49	0.29
NNMF_40	0.42	0.23
NNMF_10	0.32	0.12
Infomap	0.10	0.70
SLPA	0.08	0.13

7.3.2 OVERLAPPING COMMUNITY DETECTION REVEALS REGIONS THAT BRIDGE MULTIPLE NETWORKS AND SIT HIGH IN THE CORTICAL HIERARCHY

Having identified OSLOM as the top-performing algorithm for uncovering the overlapping community structure in our right-hemisphere structural connectome dataset, we progressed with this algorithm to characterize its overlapping community structure. As described in Sec. 7.2.5, we performed a more comprehensive parameter sweep (across P-value thresholds in finer increments of 0.01 across 100 initialization sweeps), and identified a plateau in the number of modules detected (6 to 7) and the number of overlapping nodes detected (14 to 15) between $0.15 \leq P \leq 0.35$ (cf. Figs. S2A,B). We selected an intermediate value within the stable plateau of $P = 0.3$, which also yielded among the highest ENMI with the ground-truth overlapping community structure in the benchmark analysis (cf. Fig. 7.3.2C). After examining pairwise ENMI value across initialization seeds to identify an optimally representative decomposition, we selected seed 61 (highlighted in the ENMI heatmap in Fig. S2), which exhibited a mean ENMI of 0.64 ± 0.11 with all other examined seeds at $P = 0.3$. This selection of parameters is referred to hereafter as `OSLOM_30`.

As shown on the brain's surface in Fig. 7.3.3A, `OSLOM_30` yielded an overlapping modular decomposition with seven modules, which we labeled as follows: (1) dorsomedial cortex; (2) limbic; (3) visual; (4) ventral attention and visual stream; (5) somatomotor; (6) insula; and (7) frontal pole. Of note, most individual re-

gions within each right-hemisphere structural community are spatial neighbors, with modules thus forming spatially contiguous clusters. This is consistent with prior findings that spatially proximal regions are more likely to exhibit axonal connectivity between each other [451] and share similar connectivity patterns with the rest of the brain [608, 609]. The seven `OSLOM_30`-defined communities show substantial overlap with the seven canonical resting-state functional networks described in [185] (see Fig. S7 for overlap proportions shown as a heatmap), which is consistent with work from Chen et al. [610] showing that anatomical community structure, based on cortical thickness (not structural connectivity), is closely related to functional systems [610].

Relative to non-overlapping methods, a key benefit of OCDAs is their ability to identify overlapping nodes—i.e., brain regions that bridge two or more different communities. Our `OSLOM_30` decomposition identified fifteen overlapping brain regions: *PreS*, *ProS*, *PeEc*, *PHA1*, *POS2*, *RSC*, *TE1a*, *PH*, *STSva*, *PFm*, *PGs*, *8Av*, *8BM*, and *9m*. These overlapping regions are highlighted in bold on the right-hemisphere cortical surface along network edges in Fig. 7.3.3A and the communities to which they were assigned are summarized in Table 7.3.3. Specifically, these overlapping regions were collectively assigned to six pairs of communities (e.g., 'Frontal pole + limbic', 'Somatomotor + ventral attention and visual stream'), and one region was assigned to three communities ('Limbic + visual + dorsomedial'). The intra-hemispheric connectogram (with log-transformed streamline counts, cf. Sec. 7.2.1) in Fig. 7.3.3B verifies that these overlapping regions exhibit structural connections across multiple modules. While the benchmarking analysis presented in Fig. 7.3.2 (based on ENMI and sensitivity and specificity of assignment of overlapping status to nodes) does not directly support the accuracy of specific regions assigned to specific overlapping communities with 'OSLOM', the connectivity patterns in Fig. 7.3.3B indicate that these regions are indeed structurally affiliated with regions that 'OSLOM' assigned to the corresponding two (or three) modules. Examining intra-hemispheric structural connectivity specifically between these fifteen overlapping nodes in Fig. 7.3.3C reveals strong interconnectivity, consistent with the notion that overlapping regions integrate information across segregated brain networks—noting that these regions exhibit higher overall degree on average (cf. Fig. S8A).

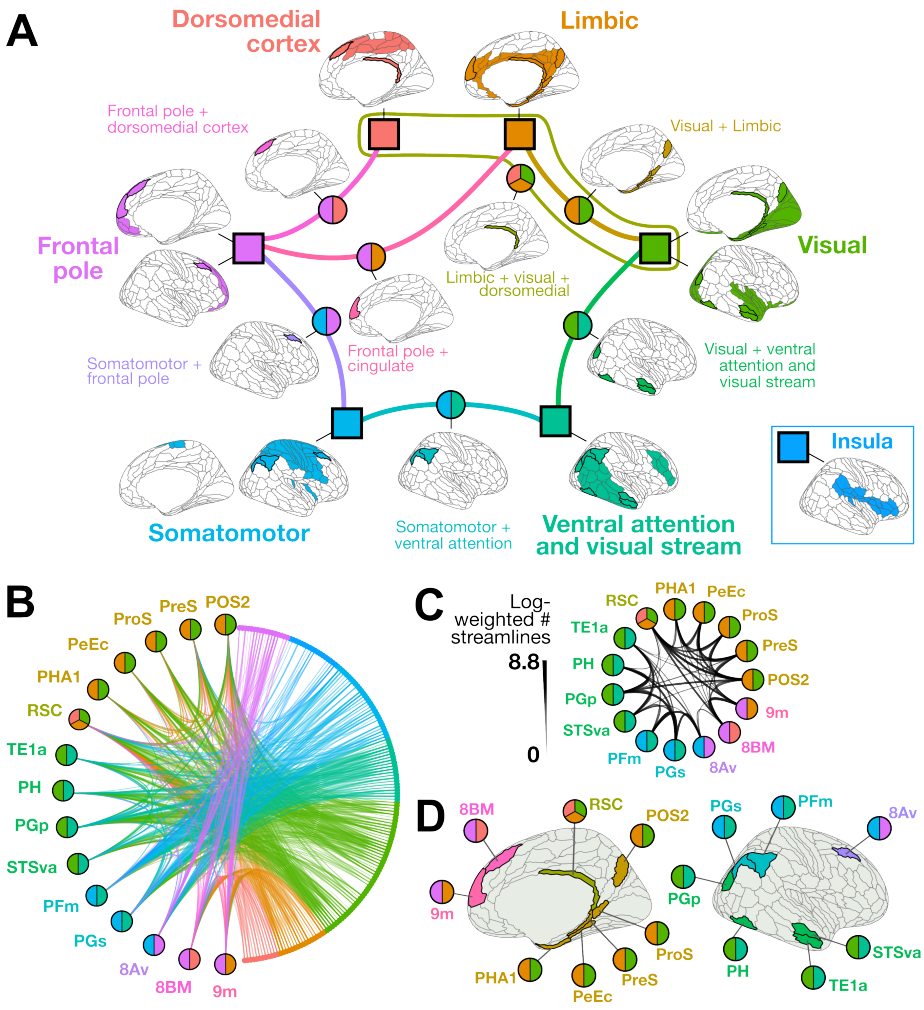


Figure 7.3.3: The overlapping community structure of the human structural connectome (right hemisphere) shows anatomical specificity and cross-network structural connectivity. A. The `OSLOM_3D` algorithm identified seven overlapping communities of the human structural connectome, which are represented with square nodes and corresponding brain maps with the 180-region parcellation from Glasser et al. [564]. Regions colored in a given map at each node were assigned to the corresponding module (and potentially to another module as well). The edges and their corresponding brain maps indicate regions with overlapping community assignments to two or more communities. Overall, the resulting communities are anatomically localized to the following regions: I. Dorsomedial cortex, II. Limbic, III. Visual, IV. Ventral attention + visual stream, V. Somatomotor, and VI. Insula, and VII. Frontal pole. Note that the 'Insula' module is positioned to the side as it does not share any overlapping connections to any of the other communities. **B.** In this connectogram, every node is plotted on the edge of the circle, ordered by assigned structural module. Lines connecting nodes indicate connectivity between the pair of regions, with width and transparency showing the log-transformed streamline count. Fifteen brain regions were assigned to two different communities ('overlapping regions'), distinguished on the left of the plot with two- or three-tone circles used to indicate the networks bridged by the corresponding region. The full structural connectome is summarized between these eleven overlapping regions (left) and the rest of the brain (right), with the latter colored by `OSLOM_3D`-defined community as in **A**. Edges are colored according to the OSLOM community to which the connecting region on the right belongs, with edge weight and transparency mapping to the average number of tracts between the two regions. **C.** The structural connectome between the fifteen overlapping regions is depicted, with line width and transparency indicating the log-transformed streamline count. **D.** Each of the fifteen overlapping nodes is highlighted in color on the brain's surface, annotated with the same multi-tone circles as in **A** to indicate shared community membership.

Figure 7.3.3D plots the overlapping regions on the right-hemisphere cortical surface, revealing that the fifteen nodes are evenly distributed along both anterior–posterior and medial–lateral axes. On average, each overlapping node comprised a larger surface area than a non-overlapping node (unpaired Wilcox rank-sum test, $P = 0.03$; cf. Fig. S8B). Comparing the locations of these overlapping regions with those of the seven structural communities, we find that they sit at the spatial interface between adjacent communities. For example, the 'Somatomotor + ventral attention' overlapping regions, *PFm* and *PGs*, sit at the ventral end of the 'Somatomotor' module and the dorsal end of the 'Ventral attention and visual stream' module. Taken together with the inter-module connectivity depicted in Figs. 7.3.3A,B, this suggests that overlapping nodes can form a structural bridge for rapid integrative processing between the constituent brain networks—in which case these regions could serve as cross-network 'integrators' or 'hubs' [611–613].

Table 7.3.3: The fifteen nodes that `OSLOM_30` identified as overlapping across two or more structural communities in the right-hemisphere cortical connectome. The '+' sign denotes an overlap between the previous and subsequent listed communities.

Brain Region	Description	OSLOM communities bridged
PreS	PreSubiculum	Visual + limbic
ProS	ProStriate Area	Visual + limbic
PeEc	Perirhinal Ectorhinal Cortex	Visual + limbic
PHA1	ParaHippocampal Area 1	Visual + limbic
POS2	Parieto-Occipital Sulcus Area 2	Visual + limbic
RSC	RetroSplenial Complex	Dorsomedial + visual + limbic
TE1a	Area TE1 anterior	Visual + ventral attention and visual stream
PH	Area PH	Visual + ventral attention and visual stream
PGp	Area PGp	Visual + ventral attention and visual stream
STSva	Area STSv anterior	Visual + ventral attention and visual stream
PFm	Area PFm Complex	Somatomotor + ventral attention and visual stream
PGs	Area PGs	Somatomotor + ventral attention and visual stream
8Av	Area 8Av	Somatomotor + frontal pole
8BM	Area 8BM	Frontal pole + dorsomedial
9m	Area 9 Middle	Frontal pole + limbic

Based on the seminal work in Mesulam [71]—and mounting evidence that 'integrative hubs' contribute to multiple functional domains [69, 538, 614, 615]—we hypothesized that overlapping regions that OSLOM assigned to two or more structural communities would sit higher in the putative functional hierarchy of the cortex, such that their cross-community connectivity enables the exchange of information from distributed modules. To test this hypothesis, we first examined the extent to which each overlapping node mapped to each of the seven canonical resting-state functional networks from Yeo et al. [185]. As shown in Fig. 7.3.4A, we found that roughly half of the overlapping nodes map primarily to the higher-order default mode and frontoparietal (cognitive control) systems, which comprise association and paralimbic cortices. By contrast, the remaining overlapping regions correspond primarily to the visual network (a unimodal system) or the ventral attention and limbic networks (heteromodal systems). The differential network composition across the fifteen overlapping nodes suggests a natural bipartition, with each half sitting towards either the lower or the higher end of the cortical functional hierarchy, respectively.

We next quantitatively compared the location of each overlapping node along the cortical hierarchy by examining the first and second principal gradients (PGs) of functional connectivity [414] provided in the *neuromaps* atlas [440] (cf. Sec. 7.2.6). As described in Margulies et al. [414], the first PG tracks a topographical hierarchy in the cortex, anchored by unimodal sensorimotor regions at the base and transmodal association regions at the top, with heteromodal regions supporting domain-general processing along the middle; whereas the second PG distinguishes between primary modalities by separating auditory and sensorimotor from visual regions. As shown in Fig. 7.3.4B, overlapping brain regions identified by `OSLOM_30` exhibit significantly higher first PG (PG1) values—averaged across all vertices per region—than non-overlapping counterparts ($P = 0.004$, Wilcox rank-sum test). This supports the notion that regions assigned to two or more intra-hemispheric structural communities sit higher in the topographical hierarchy of the cortex on average—with the caveat that external validation will be needed to confirm this, given the inherent limitations in the accuracy of specific community label assignments. Although there was no statistically significant difference in PG2 values between overlapping versus non-overlapping nodes ($P = 0.14$, Wilcox rank-sum test), plotting nodes in PG1–PG2 space in Fig. 7.3.4C revealed that overlapping regions sit within the visual–transmodal axis, with a notable lack of regions in the auditory region (characterized by low PG1 and high PG2 values). This supports the notion that regions bridging two or more structural communities generally sit higher in the topographical hierarchy of the cortex.

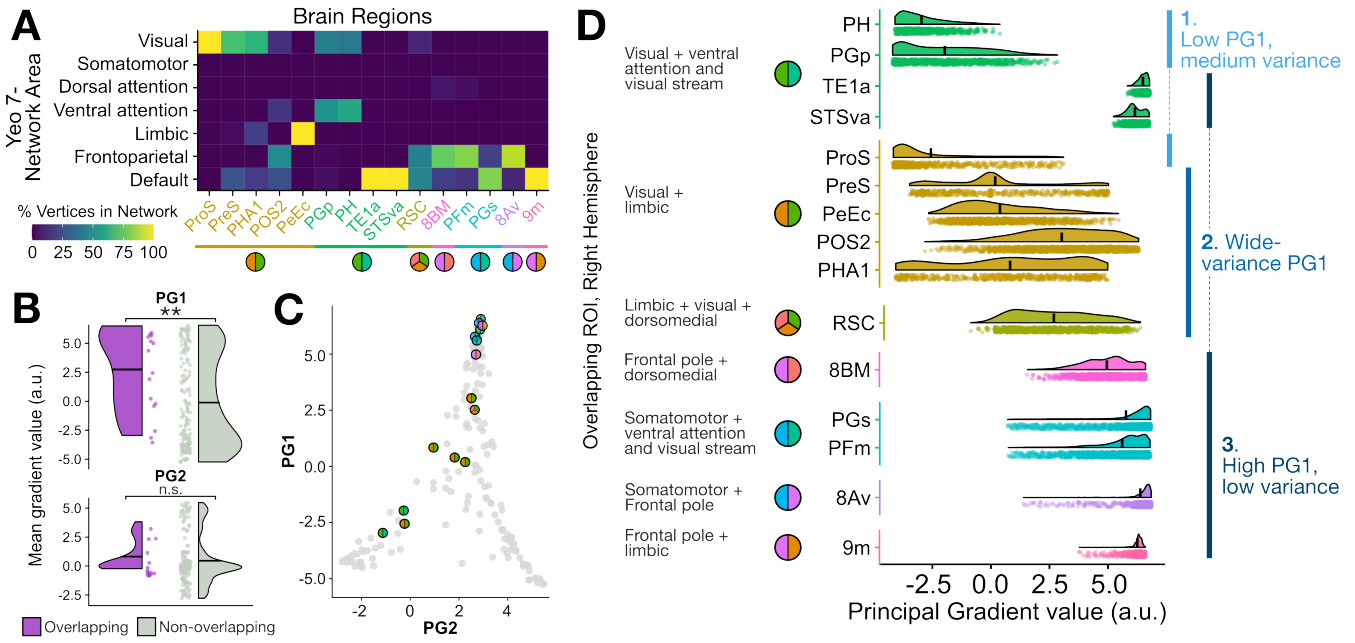


Figure 7.3.4: Regions participating in multiple structural communities generally sit higher in the human cortical hierarchy. A. For the fifteen cortical brain regions assigned to two or more modules, the percentage of vertices corresponding to each of the seven resting-state networks from [185] is shown as a heatmap. **B.** The mean first (PG1) and second (PG2) principal gradient values computed across all vertices per region are shown in raincloud plots for overlapping regions (purple, 15 regions) and non-overlapping regions (gray, 164 regions). **, $P < 0.01$, Wilcox rank-sum test; n.s., $P > 0.05$. **C.** The mean PG1 versus PG2 values are plotted per brain region, with overlapping regions indicated by multi-tone circles as in **A**. **D.** For each group of overlapping nodes (e.g., 'Visual + cingulate and parahippocampal'), the PG1 values are plotted for all vertices as raincloud plots. a.u., arbitrary units. In all plots, the black bar spanning each violin half represents the mean of the corresponding distribution. On the righthand side of raincloud plots, nodes have been grouped into one of three categories based on the patterns of PG1 value distributions.

The spread of PG1 values among overlapping regions is surprising, indicating that the nodes do not all sit narrowly at the top of the cortical hierarchy. Rather, the overlapping regions might be positioned across multiple hierarchical levels, which could allow them to act as intermediaries in integrating across systems that may sit at very different levels of the functional hierarchy. Indeed, Najafi et al. [69] found that overlapping regions exhibit greater functional diversity than those participating in only one network. While the overlapping regions did not exhibit more variance in vertex-wise PG1 values compared to non-overlapping regions (as shown in Fig. S9), we aimed to investigate the overall distribution of PG1 values across vertices in each overlapping region. As shown in Fig. 7.3.4D, we found interesting and distinctive patterns of vertex-wise PG1 values in overlapping regions, which we separated into three distinct patterns: (1) Low PG1 values with moderate variance, corresponding to *ProS*, *PH*, and *PGP*, which map primarily to the visual and ventral attention resting-state networks; (2) Moderate PG1 values with high variance, corresponding to heteromodal integration of domain-general processing, in the tri-modular *RSC* node and in four of the bi-modular 'Visual + limbic' nodes: *PreS*, *PeEc*, *POS2*, and *PHA1*; and (3) High PG1 values with low to moderate variance, corresponding to transmodal association of higher-order functions, in the remaining nodes: *TE1a*, *STSva*, *8BM*, *PGs*, *PFm*, *8Av*, and *9m*. These distinctive patterns in the mean and variance of PG1 values suggest that while overlapping regions generally sit higher in the putative functional connectivity hierarchy, poten-

tially enabling more synergistic systems-level processing—though this pattern is not universal across all overlapping regions. Rather, our findings suggest that the role of a given overlapping node within its complex network topology may depend upon the two or more communities it bridges, along with other structural and functional properties, from cytoarchitecture to neurotransmitter receptor densities [616, 617].

7.3.3 COMPARING OSLOM WITH LOUVAIN DEMONSTRATES THE IMPORTANCE OF ALLOWING NETWORK-BRIDGING OVERLAPS

Having demonstrated that `OSLOM_30` generates biologically plausible intra-hemispheric structural networks bridged by overlapping nodes, we next asked how much additional information this overlapping modular network decomposition provides beyond that of a conventional, non-overlapping partition. In particular, we wanted to understand whether the overlapping nodes assigned by OSLOM exhibit distinctive nodal metrics of network integration relative to those of a non-overlapping partition, such as participation coefficient, P, and within-module strength (z-score), z [61]. As originally described in Guimerà and Nunes Amaral [61] and further extended to brain connectivity subsequently [618–620], comparing P and z can provide insight into the functional role that a given node may fulfill in its community, from peripheral hubs (low z, low P) through global connectors (high z, high P). Since the fifteen `OSLOM-30`-defined overlapping nodes exhibited many connections across communities and between each other (cf. Figs. 7.3.3B,C), we expected that they would sit in the 'global connectors' (high z, high P) space.

For comparison to the `OSLOM_30` overlapping partition, we selected the Louvain clustering method [63] for consistency with the benchmarking analysis. The Louvain method yielded a non-overlapping modular decomposition, from which z and P were derived for each node (as described in Sec. 7.2.7). Using a representative decomposition with the resolution parameter $\gamma = 1$—which is the default in the Brain Connectivity Toolbox [601] and is supported by our robustness analyses in Fig. S4—we identified six non-overlapping Louvain communities in the right hemisphere, shown on the cortical surface in Figs. S4B,C. We also performed a robustness analysis with $\gamma = 1.5$ to enforce a seven-module (non-overlapping) decomposition in the right hemisphere, and the results were comparable to those obtained with $\gamma = 1$ (cf. Fig. S4D-G), so we focus on the $\gamma = 1$ Louvain results hereafter. To characterize nodes according to Louvain versus OSLOM partitioning, we compared P versus z scores for each node. We show each node in P_{Louvain}–z_{Louvain} space in Fig. 7.3.5B, where overlapping nodes (colored circles) are compared with non-overlapping nodes (gray circles). Most of the `OSLOM-30`-defined overlapping nodes exhibit $z_{\text{Louvain}} > 0$ (11 out of 15 nodes), indicating strong within-Louvain-module connectivity relative to other regions. However, only half of the overlapping nodes (8 out of 15) exhibit $P_{\text{Louvain}} > 0.5$, and even those do not stand out as exhibiting particularly high P_{Louvain} relative to non-overlapping nodes. For example, regions *8BM* and *9m*—judged as overlapping by `OSLOM-30`—exhibit strong within-Louvain-module connectivity (high z_{Louvain}) yet sit towards the lower end of the P_{Louvain} distribution. This is surprising, as we expected regions that `OSLOM-30` identified as overlapping to show a greater spread of connectivity to nodes across multiple Louvain-defined modules (i.e., closer to 'global connectors', or at least 'provincial hubs').

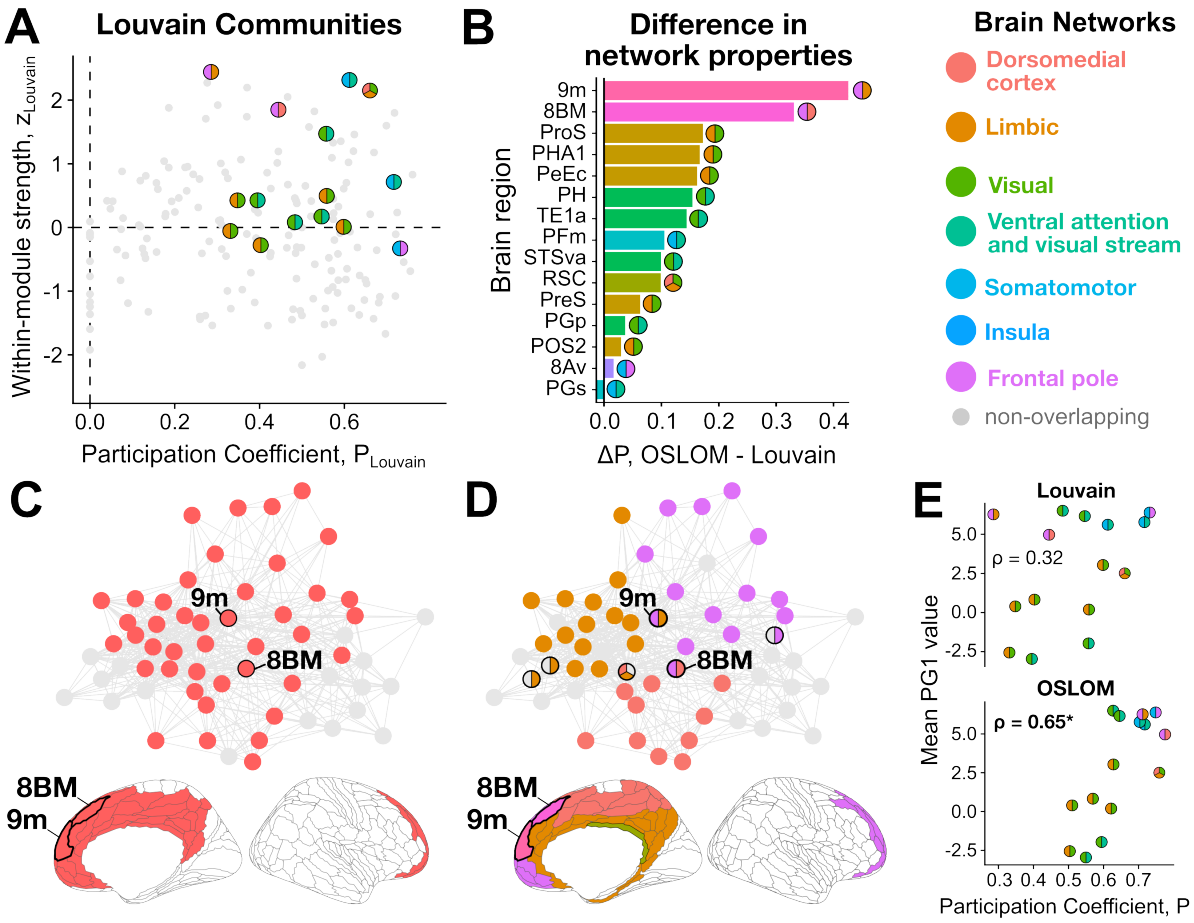


Figure 7.3.5: Overlapping nodes cannot straightforwardly be inferred from nodal metrics of a non-overlapping community decomposition. A. Scatter plots for P versus z for every node in the right hemisphere cortical connectome using the Louvain (non-overlapping) partition. The overlapping nodes (obtained by `OSLOM-30`) are marked in two-tone circles, with the two colors indicating the pair of communities bridged by the node in the OSLOM decomposition. **B.** The difference in P_{OSLOM} versus P_{Louvain} is compared for each OSLOM-identified overlapping region, ranked by the value of ΔP. **C.** Louvain assigns nodes *8BM* and *9m* to a community of 46 total nodes spanning frontal, cingulate, and dorsomedial cortices (n.b., only a subset of these nodes are shown that are structurally connected to *8BM* and/or *9m*), whereas **D.** OSLOM subdivides this set of nodes into three communities, collectively bridged by *8BM* and *9m*. The two-tone circles for *8BM* and *9m* indicate the two structural communities to which they were assigned. Circles with pink-gray or orange-gray coloring indicate overlap between the 'Frontal pole' or 'Limbic and parahippocampal' and other modules, respectively. Completely gray circles represent non-overlapping nodes. Nodes *8BM* and *9m* are annotated with bolded edges on the brain's surface below the network plots in **C** and **D**. **E.** For each overlapping node, its mean PG1 value is compared with its participation coefficient from either Louvain or OSLOM partitioning. Spearman's ρ is shown for each comparison; $^*, P < 0.05$.

Of note, P is a label-invariant measure, as it focuses on the relative frequency of connected neighbor nodes in each module rather than the identity the modules themselves—meaning it is still valid regardless of whether the corresponding community detection algorithm (OSLOM or Louvain, in this case) recovers the ground-truth community structure from the empirical right-hemisphere structural connectome. To compare network properties with the OSLOM partition, we adapted the participation coefficient, P, to allow each overlapping node to be counted as part of its constituent two or three communities (noting that the

specific module assignment label does not affect the value of P). With this adaptation (cf. Sec. 7.2.7), P will increase by definition, so we focus on the relative differences in the magnitude of change (i.e., ΔP) in Fig. 7.3.5B. Regions *9m* and *8BM* stood out as the two regions with the greatest increase in P as measured by OSLOM relative to Louvain—prompting more detailed investigation of these regions. To investigate the structural changes in modular structure that may have underpinned these substantial increases in participation coefficient for 9m and 8BM, we further examined their 'sub-network graphs'—comprised of the cortical regions sharing intra-hemispheric structural connections to one or both of *8BM* and *9m*. These subnetworks that include *9m* and *8BM* are plotted in Figs. 7.3.5C and D, for the Louvain and OSLOM modular decompositions, respectively. As shown in Fig. 7.3.5C for the Louvain decomposition, both *9m* and *8BM* were assigned to the same community, comprised in total of 46 right-hemisphere nodes (of which 40 are included in this subnetwork) that collectively span frontal, cingulate, and dorsomedial cortices. We find that the same nodes depicted in the OSLOM decomposition in Fig. 7.3.5C are partitioned into two distinct communities that are bridged by *9m* (the 'Frontal pole' and 'Limbic' modules) and another two by *8BM* (the 'Frontal pole' and 'dorsomedial cortex').

These distinct modular decompositions demonstrate that OSLOM can disentangle multiple partially overlapping modules, capturing more complex yet subtle structure—which can be lost with a non-overlapping algorithm that cannot accommodate overlaps and therefore merges nodes into one larger module. While OSLOM detects different structures in the network, and thus yields different network nodal statistics like the participation coefficient, we next aimed to investigate whether the participation coefficient derived from Louvain or OSLOM may better capture other independently measured properties of brain organization among overlapping nodes, such position on the topographical cortical hierarchy [621]. As plotted in Fig. 7.3.5E for the fifteen overlapping right-hemisphere nodes identified by OSLOM, the mean PG1 values (cf. Figs. 7.3.4B-D) were significantly correlated with P_{OSLOM} (Spearman's $\rho = 0.65, P = 0.01$) but not with P_{Louvain} ($\rho = 0.32, P = 0.24$). Similarly, we found no association between PG1 values and P_{Louvain} in our robustness analysis with $\gamma = 1.5$ ($\rho = 0.37, P = 0.2$; cf. Fig. S4G). This indicates that the overlapping regions identified by OSLOM exhibit a greater degree of cross-module (intra-hemispheric) connectivity that positively correlates with position in the topographical hierarchy of the cortex. The same relationship holds when examining all 180 regions from the HCP–MMP1 right-hemisphere atlas [564]: as shown in Fig. S10, all mean PG1 values were significantly correlated with P_{OSLOM} (Spearman's $\rho = 0.33, P = 6 \times 10^{-6}$) but not with P_{Louvain} ($\rho = 0.01, P = 0.9$). In other words, testing the hypothesis that the participation coefficient increases along the cortical functional hierarchy yields different conclusions using an overlapping versus non-overlapping decomposition. Only the OSLOM-derived participation coefficient P provides evidence for this hypothesis, with regions broadly increasing in their P from the unimodal end of the hierarchy up to the transmodal, associative end. Collectively, our comparison of metrics derived from OSLOM versus Louvain partitions underscores the utility of flexible overlapping decompositions for interpreting the resulting structural networks and derived nodal statistics, which are widely interpreted across the network neuroscience literature.

7.4 DISCUSSION

This work introduces a general data-driven framework for selecting an overlapping community-detection algorithm (OCDA) for a given application. We further demonstrate how such overlapping modular decompositions can provide novel insights into the structure of the structure of human cortical connectivity. For a given target empirical dataset—here the right-hemisphere structural connectome—the proposed method systematically compares how well each OCDA recapitulates the ground-truth overlapping community structure across benchmark networks generated from properties of the target network. This problem of calibrating methods to data relates to the concept of an algorithmic 'footprint' characterized in related work on other data types [622], that captures the differing regions (or 'footprints') in the data space in which different types of algorithms exhibit strong performance. Acknowledging that each OCDA makes different assumptions about network structure and inter-regional communication—and therefore that no OCDA can exhibit good performance on every type of network (cf. the 'no free lunch' theorem [93, 95])—our method circumvents the typical subjectivity in selecting an OCDA for a given problem, at the expense of constraints imposed on the simulated network ensemble by the benchmark generation process [556]. This demonstrates the importance of empirically tailoring OCDAs to a given network problem at hand, rather than relying on overly general statements about the relative performance of one algorithm over another [94]. We demonstrate an application of our approach to the structural connectome of the human right-hemisphere cortex, where we select an appropriate OCDA algorithm (OSLOM) from an ensemble of brain-like networks. We then show how OSLOM quantifies new types of structure in brain networks, particularly by identifying overlapping regions that bridge across multiple biologically sensible modules and sit higher in the topographical cortical hierarchy. Further, the OSLOM decomposition yields nodal metrics that better distinguish these overlapping regions and align with independent measures of cortical organization, supporting the hypothesized hierarchical variation of nodal participation.

Using OSLOM to infer the overlapping community structure of the right-hemisphere cortical connectome, we obtained a useful and interpretable representation of the network structure with seven biologically plausible processing modules, collectively bridged by fifteen integrative overlapping nodes. While our method cannot directly confirm the accuracy of these identified overlapping nodes in the absence of a ground-truth empirical right-hemisphere community structure, these findings are generally in line with previous studies mapping structural and functional connectomic data. For example, prior work suggests that higher-order integrative regions like *PFm* and *PFs*—which were identified by OSLOM as exhibiting community overlap between the 'Somatomotor' and 'ventral attention and visual stream' networks—serve as important functional and cytoarchitectonic transition zones between neighboring brain areas [623, 624]. In addition to *PFm* and *PFs*, regions *8BM* and *9m* in the dorsomedial prefrontal cortex also exhibited among the highest principal gradient of functional connectivity (PG1) values [414] of the regions we evaluated (cf. Fig. 7.3.4C). For these nodes, the intra-hemispheric structural connectivity overlaps may provide physical substrates for transmodal and/or associative functional integration. Region *8BM*, which showed overlapping community affiliation with the 'frontal pole' and 'dorsomedial cortex' modules here and is a part of the frontoparietal resting-state network [185], has previously been highlighted as a higher-order cognitive control hub [625]. Moreover, Assem et al. [612] showed that *8BM*, *PFm*, and *PGs* are also components of the 'multiple-demand' network that orchestrates a variety of cognitive tasks, underlying flexible organization and cognitive control. Our findings

are therefore consistent with the hypothesis that overlapping regions span multiple structural communities to enable functional integration.

However, some overlapping nodes' PG1 values position them closer to heteromodal limbic regions (e.g., *POS2* and *RSC*) or even lower-order unimodal sensorimotor regions (e.g., *PH* and *PGp*). Despite sitting lower in the first putative topographical hierarchy of the cortex than other overlapping nodes, *POS2* (a component of the parieto-occipital sulcus in the HCP–MMP1 atlas [564]) participates in individual-specific functional network overlaps and is distinct from neighboring regions based on its cortical microstructure, functional connectivity, and task activation [546]. *RSC* (standing for 'retrosplenial cortex' in the HCP–MMP1 atlas [564]) is noteworthy as the only region identified to bridge across *three* structural networks here: 'Limbic', 'visual', and 'dorsomedial cortex'. The RSC exhibits dense inter-network connectivity [626] and plays a critical integrative role in synthesizing viewpoints to process one's environment [627, 628]. Taken together with the literature, our findings suggest that *POS2* and *RSC* serve as integrative transmodal hubs that bridge two or three distinct networks, respectively. For the lower-PG1 regions—*PH* in the temporo-occipital junction and *PGp* in the inferior parietal cortex—the potential role for structural overlap is less clear from the literature. *PGp* is a part of the angular gyrus (AG) believed to mediate semantic processing [629, 630], although evidence suggests it plays less of an 'integrative' role than other AG subregions [631]. Region *PH* contributes to visual processing from the foveal field [632], in line with the shared connectivity between the 'visual' and 'ventral attention and visual stream' networks identified herein.

Taken together, the differing patterns in functional connectivity principal gradient (PG1) values among the overlapping nodes suggest a wide spectrum of functions served by cross-network structural bridges, which likely depend on a variety of factors—including dynamic brain state, cytoarchitecture, receptor densities, and functional properties of the underlying brain systems. In support of this, Najafi et al. [69] described different 'bridging' behaviors exhibited across different overlapping regions based on the functional connectome, and Bijsterbosch et al. [546] found evidence for both inter-community coupling as well as 'interdigitation' (i.e., spatial interweaving of two networks) among areas of overlap between functional modules. Previous work has also suggested that such regions play a key role in information integration across brain modules and dynamical network states [55, 633–636]. Given this heterogeneity, future study is warranted to more fully characterize how the nature of overlapping regions relates to their own intrinsic structural, functional, and/or cytoarchitectonic properties, as well as those of the network communities they span.

The relatively low number of overlapping regions (15/180) is notable, with interesting implications in terms of both methodology and biology that warrant future investigation. While OCDAs have been applied to structural connectomes in mice [537] and macaques [538], to our knowledge, this is the first study to apply an OCDA to the human structural connectome derived from diffusion MRI. In light of this, there is limited precedent for directly comparing our observed overlap ratio to previous work. However, we speculate that the variance-based consensus thresholding method applied here—which was necessary to tune the required parameters to a single empirical adjacency matrix for the benchmark generation algorithm from Lancichinetti and Fortunato [556]—might contribute to the low proportion of overlapping nodes in the right cortical structural connectome, as sparsification reduces the opportunity for nodes to participate in multiple communities. Indeed, Gu et al. [637] applied an OCDA to individual resting-state (functional) connectomes, reporting much higher overlap densities across individuals (80-100 out of 246 nodes) with age-related dif-

ferences. By contrast, Li et al. [539] implemented group-averaging and reported only 16 overlapping nodes out of the 90 regions in the automated anatomical atlas (AAL) parcellation [638]. These differences in overlapping region density are likely attributable to factors including algorithm choice, imaging modality, and whether the analyses are performed on individual or group-level networks. Biologically, the minimal overlap in our results suggests intriguing questions for future work about the efficiency and specificity of structural brain organization. Given the high metabolic cost of maintaining long-range axonal connections [639], it may be that structural modules are relatively segregated, with only a minimal set of overlapping 'bridge' regions facilitating inter-modular communication. Of note, OSLOM identified at least one overlapping node per structural community (with the exception of the insular network), suggesting that while rare, overlaps are distributed across modules.

Comparing intra- and inter-module network statistics, like P and within-module z, highlighted distinct network properties derived from an overlapping versus non-overlapping modular partition. Our results indicate that the inflexibility of the Louvain decomposition—which seeks the optimal non-overlapping partition—can 'force' an overlapping modular structure into a single larger community in attempting to represent an underlying overlapping connectivity structure within the constraints of a non-overlapping partition, consistent with prior work with functional MRI [69]. In one example of this phenomenon (Fig. 7.3.5C), the Louvain partition consolidated regions *9m* and *8BM* into a singular 46-node community, with high z and low P (i.e., high intra-modular connectivity but low inter-modular participation). We also note that enforcing a seven-module Louvain decomposition with $\gamma = 1.5$ still placed *9m* and *8BM* in a singular, 32-node community. More generally, the fact that all the OSLOM-defined overlapping nodes were not distinctively concentrated in a specific area in the $z_{\text{Louvain}} - P_{\text{Louvain}}$ space underscores how non-overlapping community-detection methods can yield suboptimal representations of the integrative, cross-module connectivity within a network structure. In other words, properties of a hard partition (like the within-module degree or participation coefficient in a non-overlapping decomposition) are insufficient to identify overlaps between communities. Since P is agnostic to the specific labels ascribed to each discovered community, results obtained with this measure do not hinge on the accuracy of community assignment by a given community detection algorithm, lending further credence to the performance of OSLOM for the right-hemisphere structural connectome. We note that z and P were both developed for non-overlapping community structure—thus requiring us to make slight modifications to these algorithms to account for overlapping nodes. Future work developing network characteristics that are readily amenable to both overlapping and non-overlapping algorithms would enable more rigorous direct comparison between two resulting decompositions.

Our methodology and results come with assumptions and limitations that should be interpreted in the context of this work. First, we consider limitations in our simulated network benchmark analysis, and proposed future directions to address them. We assume that the generated ensemble of benchmark networks reflects the overlapping community structure of the target empirical network, and in turn that the top-performing OCDA on the benchmark ensemble will also perform well on the target network—such that ensemble properties are crucial to determining the 'best' OCDA. Moreover, we used the generative network model of Lancichinetti and Fortunato [556], which forces a power-law degree distribution that may not be optimally suited for many real-world networks—including the human connectome studied here, in line with previous findings [605–607]. The model also does not incorporate other types of network properties, such as hierarchical community structure, which is a reported feature of brain-network organization [640]. In addi-

tion, given a generative model of overlapping community structure in networks, we had to make subjective choices about how to sample from the corresponding distribution over networks numerically. Here we sampled from sensible ranges of the relevant parameters—including community size distribution τ_2, mixing parameters μ_t and μ_w, and the number of overlapping nodes, O_n—although these remain choices that can affect the assessment of the resulting OCDAs, and could be explored more thoroughly in future work. As more sophisticated generative network models are developed that can incorporate a larger range of topological constraints, they can readily be incorporated into our framework to allow more precise calibration of OCDAs to target empirical data.

Second, we consider ways in which future work could extend the application of OCDAs introduced here beyond the scope of the present study. Future work should investigate the consistency of community detection assignments of a given OCDA and also between different algorithms (including for different parameter configurations). Similarly, we considered a hand-selected range of parameter values for each algorithm, though future extensions of this work could fit parameters for each OCDA using a training ensemble of benchmark networks. Such parameter tuning is likely to improve the ENMI values in general, allowing for more precise calibration of high-performing parameter values for each OCDA to a given application. More broadly, future work could expand beyond the five broad OCDA classes examined here to incorporate alternative methods that warrant investigation [641–643].

Finally, we consider limitations related to our empirical structural intra-hemispheric connectome analysis, which generally reflect issues inherent to structural connectomes estimated from diffusion MRI. Moreover, tractography-based connectivity estimates can be biased by systematic anatomical differences between cortical and subcortical structures, including the small size of nuclei relative to diffusion imaging voxels and different properties of laminar organization and the gray–white matter interface [566, 644]. Limitations in reconstructing long-range fibers also motivated our decision to focus on intra-hemispheric connections within the right hemisphere (noting highly similar connectomic network properties between the two hemispheres, such that this choice was arbitrary and unlikely to affect results). Indeed, it is relatively common practice in structural connectomics to analyze each hemisphere separately—or focus on a single hemisphere altogether—to improve anatomical reliability [28, 567–570]. However, this precluded our ability to examine homotopic and other inter-hemispheric connections, both of which are core components of brain organization [605] (and contribute to dysfunction in diverse pathologies [552, 645, 646]). Therefore, extending these analyses to networks that incorporate inter-hemispheric connections in the cortex, subcortex, and cerebellum represents an important direction for future work to characterize the distributed modular architecture on the scale of the whole brain [573, 647].

Indeed, the seven distinct modules identified by OSLOM in the present study are generally spatially contiguous, with the exception of the ‘Visual’ and ‘Ventral attention and visual stream’ networks. Moreover, all of the overlapping nodes sit at the spatial interface between their constituent and adjacent structural communities, suggesting they form connectivity ‘bridges’. This finding is consistent with similar findings reported in the macaque structural connectome [538]; however, future work should investigate whether this spatial contiguity in overlapping nodes arises from general biological patterns of brain organization and/or from the inherent contiguity of the specific parcellation atlas analyzed. For example, we applied the multimodal parcellation from Glasser et al. [564], motivated by its strong neurobiological validity in integrating properties

of local microstructure and global functional activity to delineate contiguous cortical regions. However, the use of any specific parcellation imposes spatial and topological constraints on network analyses; in particular, Bijsterbosch et al. [648] highlighted how the choice of parcellation scheme can influence resulting patterns of (functional) connectivity and network topology.

On a related note, we found that the average overlapping node occupies significantly more cortical surface area than the average non-overlapping node. A larger surface area can lead to an increased number of streamlines assigned to that region in the tractography estimation process [606, 649], and the overlapping regions do indeed have a higher degree on average. However, both region size and parcellation atlas can affect nodal statistics and the spatial location of overlapping regions within the brain [650, 651]. To more precisely resolve these biological and methodological questions, future work should extend our analyses using alternate parcellation schemes—including those defined specifically based on structural connectivity [29] or the intrinsic geometry of the cortex [3]—to assess the robustness of spatial patterns in overlapping community structure.

The method introduced here can be used to tailor overlapping community detection methods to network data, facilitating the additional insights that OCDAs can provide to diverse application areas. In our application to the human cortical connectome, our results highlight the importance of uncovering meaningful overlapping network partitions that better capture how the brain balances functional integration and segregation on a static backbone of structural connectivity. Future work could map out such a space for empirical networks with overlapping modular structure [652, 653] to more comprehensively characterize which parts of network space are best suited to which OCDAs. With this work, we also provide a comprehensive and extendable code base for overlapping community detection, which we hope will be expanded upon in future work, incorporating improvements in both new types of overlapping community-detection algorithms (including those that incorporate higher-order interactions [613]) and in overlapping benchmark network-generating algorithms.

ACKNOWLEDGMENTS

The authors thank Mac Shine, Isabella Orlando, Joshua Tan, and Christopher Whyte for their valuable input on data visualization and biological interpretation.

8
Conclusions and Future Directions

Tuition is valuable, but you know what's invaluable? Intuition.
Michael Scott, The Office, Season 6 Episode 12

8.1 SUMMARY OF CONTRIBUTIONS

The main content presented in this thesis consists of six chapters, each designed to contribute to the overarching goal of developing accessible and interpretable methods for quantifying dynamical properties of brain activity using interdisciplinary methods from complex systems analysis. Chapter 2 introduced the theoretical motivation and practical implementation of diverse statistical formulations for local dynamics and pairwise interactions that are applied throughout subsequent chapters of this thesis. We expanded upon the conceptual basis for these methods in Chapter 3, where we introduced and applied a systematic framework for characterizing pairwise interactions using hundreds of measures implemented in *pyspi*, a unified and open-source Python package. The benefits of this highly comparative approach were then demonstrated in Chapter 4, where we applied the *pyspi* library to systematically evaluate and interpret candidate neural correlates of conscious visual perception.

In Chapter 5, we extended the comparative ethos of the preceding chapters to integrate representations of local dynamics and pairwise coupling together in one comprehensive yet interpretable framework for resting-state functional magnetic resonance imaging (fMRI) analysis. Our systematic comparison of five different representations of resting-state fMRI dynamics, each with increasing complexity, validated the use of measures sensitive to linear activity and coupling, and also highlighted the benefits of combining functional connectivity with local dynamics—the latter of which performed surprisingly well in distinguishing schizophrenia and bipolar I disorder cases from controls. This notion of comparing across levels of com-ponent granularity carried through to Chapter 6, in which we demonstrated how the strength of functional coupling between a region and its mirror homolog was tightly associated with its overall functional connec-tivity strength within and across hemispheres (as quantified with nodal strength). Finally, in Chapter 7, we established a comprehensive methodology to tailor the optimal overlapping community detection algorithm (OCDA) and parameter configuration for a given empirical network, in this case the structural connectome of the right-hemisphere cortex. We demonstrated how this approach can,

in turn, highlight biologically meaningful communities in the intra-hemispheric structural connectome, bridged by several overlapping regions that generally sit towards the higher-order association end of the principal gradient of functional connectivity.

8.2 UNIFYING COMPLEX METHODS WITH VISUALS AND ACCESSIBLE LANGUAGE: CONTRIBUTIONS OF CHAPTERS 1 AND 2

In the introductory chapter of this thesis (Chapter 1), I provided a general overview of interdisciplinary methods for characterizing dynamical properties of the brain as a complex system, emphasizing the importance of accessible language and visuals to bridge the gap between different theoretical traditions. The aim of this chapter was to establish fundamental concepts and definitions for the remainder of this thesis, providing an integrative complex-systems perspective to enable researchers across disciplines to navigate this vast methodological landscape when interpreting brain dynamics. This motivating aim was carried through to Chapter 2, in which we depicted and discussed the theoretical underpinnings and practical implementation of diverse statistical formulations based on information theory. Specifically, we presented a novel unified framework for information-theoretic time-series measures, discussing both theoretical motivation and practical implementation to capture information storage and flow within neural time series. Together, these first two chapters established a visually-focused and interpretable foundation for distilling salient features of a given multivariate time series (MTS) dataset, which is then directly put into practice throughout the remainder of the thesis.

8.3 SYSTEMATICALLY COMPARING MEASURES OF PAIRWISE INTERACTIONS: CONTRIBUTIONS OF CHAPTERS 3, 4, AND 5

In Chapter 3, we expanded upon the theoretical groundwork built in Chapters 1 and 2 with the *pyspi* library, which provides a unified algorithmic implementation of over 200 measures of pairwise interactions for MTS data (284 as of August 2025, and growing)—including those based on spectral analysis, information theory, and distance-based alignment. Of note, the neuroimaging case studies presented in Chapter 3 were designed to demonstrate the utility of the *pyspi* library in practice, rather than to provide a deep biological interpretation of the results. However, in addition to our own applications in subsequent chapters, the *pyspi* library has already played a vital role in quantifying complex pairwise interactions in the brain among external research groups investigating seizure dynamics [654] and macroscopic properties of cortical organization in health [655] and disease [656]. Clarifying the relative strengths and similarities of different measures of pairwise interactions provides a foundation for future work to extend and develop new theory.

Importantly, the spatial and temporal scales at which neural activity is measured dictate the types of pairwise interactions that are optimally suited for a given application. We found several alternative measures that outperformed the Pearson correlation coefficient in distinguishing naturalistic film viewing from resting-state fMRI; however, more generally, no single measure of pairwise interactions outperformed all others across all three classification case studies presented in Chapter 3 (smartwatch actigraphy data; EEG while modulating cortical ‘up’ or ‘down’ states; and fMRI at rest or during naturalistic film viewing). This finding is reminiscent

of the 'no free lunch' theorem in machine learning [93, 232], which suggests that no single aspect of statistical coupling will be universally optimal across all tasks. However, our results did suggest a middle ground of generalizability: while the number-one performing measure varied across the three classification case studies, relaxing the scope to the top ten measures per problem identified common top-performing measures, such as the cross-distance correlation [657]. Collectively, there are two key benefits to this approach—first, this represents a systematic method to identify *the most* informative type(s) of pairwise interactions for a given dataset, which may be the end goal of a given analysis. Second, this approach can be applied in a meta-comparative way to identify families of measures that perform well across a range of problems, which can then be used to inform future theoretical developments and applications. Altogether, findings from Chapter 3 indicate that highly comparative frameworks like *pyspi* are powerful to distill the most informative type(s) of pairwise interactions based on the specific dataset and research question at hand.

The benefits of this highly comparative approach to pairwise interactions were directly demonstrated in Chapter 4, in which we applied the *pyspi* library to address an open problem in consciousness science: identifying, and theoretically interpreting, candidate neural correlates of conscious visual perception. Our comprehensive comparison in magnetoencephalography (MEG) data, which was the most extensive to date, examined 246 measures of functional connectivity from the *pyspi* library across brain regions hypothesized to drive conscious visual perception by one or both of two theories of consciousness, Global Neuronal Workspace Theory (GNWT) [191, 192] and Integrated Information Theory (IIT) [189, 190]. We found that a family of barycenter-based measures yielded particularly strong stimulus decoding performance in a way that generalized across both theories' predictions, outperforming more conventional measures typically used for MEG analysis, such as the pairwise phase consistency [200] or power envelope correlation [201, 202]. However, even the barycenter-based methods exhibited marked variability in classification performance across different experimental parameters, including stimulus types, relevance types, and region–region pairs. This result connects with the point raised in Chapter 3 that these systematic comparisons can identify general classes of measures that are particularly well suited to a given imaging modality and application, rather than a single measure that is optimal across all contexts.

8.4 EXPANDING ACROSS MULTIPLE SCALES OF DYNAMIC INTERACTIONS: KEY FINDINGS FROM CHAPTERS 5, 6, AND 7

In the first chapter, I introduced the concept of analyzing brain activity at multiple levels of component granularity, from local dynamics to pairwise interactions and distributed network activity. As with any complex system, the brain can be understood as a hierarchy of interacting components, in which each scale exhibits its own unique dynamics that interface in a bidirectional and nonlinear manner. Chapters 3 and 4 infused a highly comparative approach to pairwise coupling, which is a key component of this multi-scale framework. However, one of the core aims of my thesis was to develop and apply a systematic framework for quantifying dynamical properties of the brain at *multiple levels* of component granularity. This goal motivated the work presented in Chapters 5, 6, and 7, each of which bridged across levels of component granularity to deeply characterize brain dynamics in both health and disease.

As described in the introduction to Chapter 5, researchers face a 'statistical smorgasbord' of options when

it comes to extracting feature-based representations of the brain's dynamical structure from a structural and/or functional neuroimaging dataset. This chapter introduced a methodological foundation for systematically invoking informative features from a rich interdisciplinary literature on time-series analysis, built upon five distinct representations of resting-state fMRI dynamics—from focusing on localized dynamics of a single region at a time to combining whole-brain maps of local dynamics and pairwise coupling. One of the most noteworthy findings in this chapter was the relatively strong performance of local (intra-regional) dynamics in distinguishing cases from controls, especially in schizophrenia and bipolar I disorder. This point was surprising since many in the field focus more on functional connectivity, owing in part to the hypothesis that neuropsychiatric disorders arise from distributed alterations rather than localized dysfunction [249, 272, 274, 275]. However, prior characterizations of local dynamics have generally been constrained to a small subset of measures, such as fractional amplitude of low-frequency fluctuations (fALFF) [4, 5], regional homogeneity (ReHo) [6], or voxel-mirrored homotopic connectivity (VMHC) [7], whereas our approach evaluated a broader range of diverse univariate time-series features. Taken together, these findings reaffirm the importance of local dynamics in characterizing the brain's dynamical structure, in both basic and translational neuroscience.

As a first foray into combining these two scales in a systematic way, we used the simplest method of concatenating feature vectors comprising local dynamics and pairwise coupling into a single input vector per classifier—leaving the door open for future work to explore more sophisticated ways of combining these two scales of interaction. For example, novel data fusion techniques have emerged to handle nuanced issues with heterogeneous input data types into one classifier [40, 41]. With our simple concatenation approach, we reported in Chapter 5 that combining local dynamics with pairwise coupling improved the case–control classification performance of all 14 evaluated functional connectivity measures, even elevating the performance of several measures that were not statistically significant when used in isolation. This suggests that a representation sensitive to both local dysfunction *and* spatially distributed changes in functional connectivity networks is particularly informative for characterizing brain dynamics in neuropsychiatric disorders—consistent with our view of the brain as a complex system, in which local dynamics and pairwise interactions comprise related yet fundamentally distinct processes. Beyond the neuropsychiatric disorder case study presented in Chapter 5, this comprehensive yet interpretable approach has broad applicability to a variety of domains, given its generalizability to complex time-varying systems as a whole. For example, in Chapter 1, I introduced several other types of complex systems that share dynamical similarities with the brain: financial systems, ecological food webs, and social networks.

The information-theoretic foundation established in Chapter 2 and empirically implemented in Chapter 3, together with the highly comparative approach to pairwise interactions in Chapter 4 and local dynamics in Chapter 5, set the stage for the work presented in Chapter 6. In this chapter, our aim was to map the directed flow of information in the human cerebral cortex with measures such as directed information or transfer entropy, which capture time-lagged and nonlinear coupling between regions. We constrained our spatial scope to homotopic region–region pairs as a starting point, grounded in a deeply robust biological phenomenon. Interestingly, comparing across over 200 measures from the *pyspi* library (Chapter 3) revealed that these more complex measures of functional connectivity captured highly similar properties of coupling to the simpler Pearson correlation coefficient. Guided by the motivation of methodological parsimony, we therefore progressed with the Pearson correlation coefficient for the remainder of the project presented in Chapter 6,

which focused on mapping biological correlates to the spatial variation in homotopic functional connectivity (HoFC) across the human cerebral cortex. While it was beyond the scope of this chapter to investigate directly, we note that the intrinsic neural timescale co-varies along the cortical sheet with the anatomical hierarchy [9, 658], and that homotopic region–region pairs are more likely to exhibit similar timescales than non-homotopic pairs [41]—which might explain why nonlinear, time-lagged measures like transfer entropy and directed information yielded very similar connectivity values to the Pearson correlation coefficient in this context. It would be interesting to investigate this further in future work, and to quantify the extent to which linear versus nonlinear measures of functional connectivity 'de-couple' as the timescale of neural activity diverges between region–region pairs in general.

Results from this chapter demonstrated that HoFC is a surprisingly robust phenomenon that is conserved across two distinct cohorts, including a subgroup of individuals with schizophrenia, bipolar I disorder, and attention-deficit hyperactivity disorder. This is noteworthy given the extent to which resting-state functional connectivity varies across individuals [490], and even more so in the context of neuropsychiatric disorders [491]. Previous work has shown that HoFC is a temporally stable phenomenon within individuals [404, 405], such that this presents an exciting avenue with which to investigate both similarities and differences in HoFC across individuals in different populations. The lack of an association with structural connectivity (including in a separate robustness analysis) is also surprising, given the hypothesis that HoFC is mechanistically supported by direct monosynaptic connections via the corpus callosum. However, these two points are in fact not mutually exclusive: while the sheer quantity of white-matter tracts does not appear to relate to HoFC magnitude across the cortex, intrinsic *qualities* of the tracts may be more relevant. I would hypothesize that polysynaptic non-callosal connections (possibly via subcortical structures) could also potentially account for the lack of observed association between HoFC and (direct) anatomical connectivity. This is a promising avenue for future work, and accordingly warrants hypothesis-driven follow-up research.

We also found in this chapter that the spatial variation in HoFC is associated with both thalamic functional connectivity and transcriptomic expression of intra-telencephalic layer 5B pyramidal neuron marker genes, the latter of which are developmentally vital for establishing interhemispheric connections [477–479, 482, 483]. It was beyond the scope of this chapter to investigate these associations in detail, but together they suggest that common subcortical drive (especially from the thalamus) might mechanistically support homotopic coupling [51, 404, 472, 486]. In a similar vein, we found that homotopic coupling was closely related to overall functional connectivity strength both within and across hemispheres. This suggests that resting-state HoFC and global functional connectivity may be related, and that the dominance of both HoFC and global synchrony decrease with ascending position along the anatomical hierarchy of the cortex. While beyond the scope of this chapter, I would hypothesize that the spatial variation in both HoFC and global synchrony may arise in part from varying contributions of subcortical drive to regional dynamics—and that the nature of homotopic coupling can take different forms, from direct communication between regions to independent activation from a common external input. For example, Markicevic et al. [43] noted that unimodal regions receive fewer overall inputs (primarily feedforward subcortical projections) and were more sensitive to experimental stimulation of the mouse basal ganglia than transmodal regions, for which thalamocortical projections comprise a smaller proportion of overall inputs. This finding suggests that future work investigating HoFC should endeavor to disentangle direct communication between homotopic region–region pairs from shared subcortical drive propagating in parallel, to clarify the potential thalamic contributions to

resting-state HoFC.

To complete the arc of this thesis, in Chapter 7 we pointed the highly comparative lens at distributed networks constructed from pairwise anatomical connectivity, with the aim of systematically evaluating the ability of different overlapping community detection algorithms (OCDAs) to infer overlapping community structure in the cortical connectome. As noted in the discussion in Chapter 7, since each OCDA inherently makes different assumptions about network organization, no single OCDA is expected to be universally optimal across all empirical networks—which links back to the 'no free lunch' theorem, introduced above in regard to functional connectivity measures compared across different consciously perceived stimulus types in Chapter 4. To our knowledge, our examination of overlapping community detection in the human structural connectome in Chapter 7 is the first of its kind, against a backdrop of mounting interest in the overlapping community structure of *functional* connectomes [69, 539–547]. Indeed, this approach opened up a new avenue for investigating the structural organization of the human connectome, allowing us to quantitatively explore interesting biological hypotheses that can only be supported through an analysis that embraces an overlapping representation.

For example, we hypothesized that the fifteen regions which the top-performing OCDA (OSLOM) assigned to multiple modules (out of seven overall) may serve a more integrative function rather than unimodal specialization. We tested this hypothesis by comparing two independent properties of network structure: (1) the value of the first principal gradient of functional connectivity (PG1) per region, which was introduced in Chapter 6 as a measure from Margulies et al. [414] that organizes regions along an axis of functional connectivity from primary unimodal regions (lower PG1) to higher-order association cortices (higher PG1); and (2) the participation coefficient of each region, which is a network topological measure that summarizes the extent to which that region is anatomically connected across all other communities. We found a positive correlation between the participation coefficient and PG1 for the fifteen overlapping nodes identified by OSLOM, but not for the non-overlapping Louvain decomposition—consistent with the interpretation that allowing flexible overlap among communities can more organically capture how the brain balances functional integration and segregation, anchored in the fixed topology of anatomical wiring. This work therefore offers a novel contribution to the literature on community detection in the human connectome, with promising implications for future work to expand upon some of the key biological insights uncovered in Chapter 7.

8.5 NEW HORIZONS: LIMITATIONS AND NEXT STEPS FROM THIS THESIS

This thesis has made several contributions to the field of computational neuroscience, and in complex systems analysis more broadly, through the development and application of a systematic framework for quantifying the brain's dynamical structure across scales of interaction. As with any scientific endeavor, there are limitations to the work presented here that inspire exciting avenues for future research going forward, which I will now summarize.

8.5.1 FROM THEORETICAL RELATIONSHIPS TO SYSTEMATIC QUANTIFICATION: EXTENDING THE WORK IN CHAPTER 2

Connecting theoretical relationships between neural processes with systematic quantification of empirical data formed a central tenet of this thesis. Indeed, this was a key motivation for the review presented in Chapter 2—as well as our comprehensive examination of functional connectivity measures for visual stimulus decoding in Chapter 4. The review of information-theoretic time-series measures in Chapter 2 introduced the theoretical underpinnings and practical implementation of diverse measures, pointing out similarities between measures (based on underlying theory and/or resulting values in the empirical case study) ad hoc. By including only one participant for our illustrative case study, we were constrained to a qualitative review of the theoretical relationships between measures, such that a natural next step would be to systematically compare the statistical relationship between measures across a broader pool of participants and anatomical connections. More broadly, to complement empirically observed neuroimaging data, future work could also incorporate model-simulated brain networks with parametrically-varying relationships between measures to better characterize the dynamical regimes in which different measures co-vary [204].

8.5.2 COMBINATORIAL EXPLOSIONS: STATISTICAL POWER LIMITATIONS AND FUTURE DIRECTIONS IN CHAPTERS 3, 4, AND 5

With great scope of time-series measures in a given highly comparative library comes the potential for combinatorial explosion, a consistent challenge across multiple chapters of this thesis. The small sample sizes inherent to human neuroimaging studies limits the statistical power, with overfitting a particularly important consideration [659, 660]. To mitigate this, even within a single level of component granularity, the number of measures examined needs to be carefully balanced with the number of brain regions included in a given analysis. These three key axes that each contribute to feature dimensionality—the number of measures, the number of regions, and the component granularity—are depicted schematically in Fig. 8.5.1.

Throughout this thesis, each chapter has found a unique balancing point along these three axes, typically by constraining at least one (if not two) axes in order to more fully explore the third, depending on the specific question at hand. For example, in Chapter 3, we intentionally focused on applied neuroimaging case studies with only 6 (EEG) or 7 (fMRI) brain areas measured, and in Chapter 4, we constrained our analysis to four brain regions collectively hypothesized to drive conscious visual perception by IIT and/or GNWT. These spatial constraints, together with exclusively focusing on pairwise interactions, allowed us to systematically compare across over 200 measures in the *pyspi* library in a computationally tractable and statistically appropriate manner, although this also meant that the majority of cortical activity and interactions were beyond the scope of our analysis.

By contrast, in Chapter 5, we relaxed the spatial constraint to include all cortical regions across the clinical groups (including subcortical regions for the UCLA CNP cohort) and instead focused on a subset of time-series features sensitive to local dynamics (25 features) and pairwise coupling (14 features), allowing us to thoroughly map disease-related changes to regional and distributed brain activity across the whole brain. These feature subsets were both identified based on minimal redundancy across a wide variety of time-series datasets, though neither were specifically tailored to neuroimaging applications; for example, the *catch22* feature set was selected according to overall classification performance across 93 diverse

univariate time series datasets (including yoga pose angles and beef spectrograms) [351]. We were not able to test whether an alternative measure(s) to those included in these two subsets could better capture disease-related perturbations to dynamical structure, which is especially relevant given the known biological heterogeneity in neuropsychiatric disorders [352].

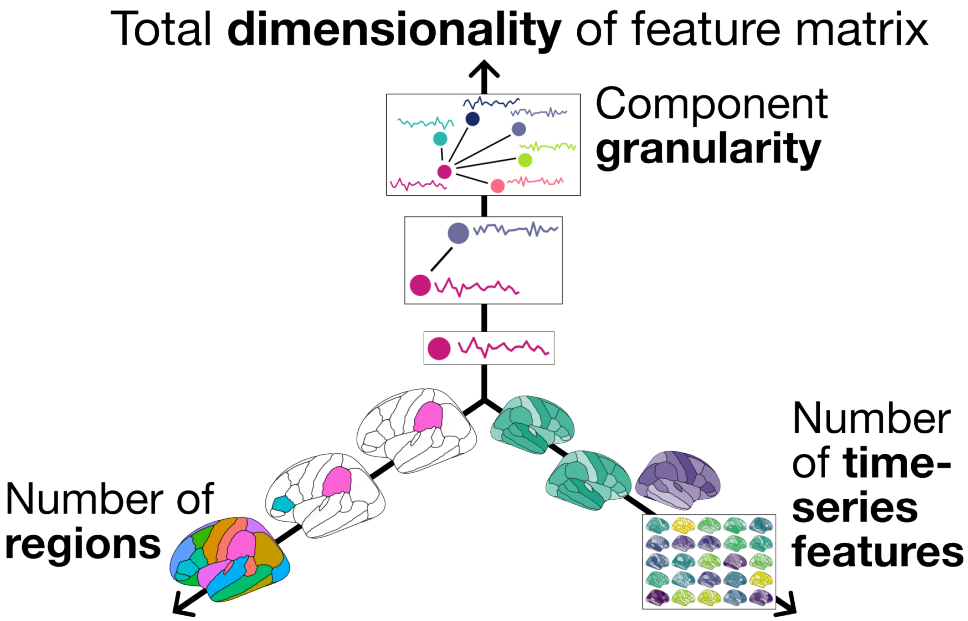


Figure 8.5.1: The total dimensionality of the feature matrix can grow exponentially with the component granularity, number of regions, and number of time-series measures considered. The total number of features in the feature matrix depends on three factors. (1) The component granularity: local activity, pairwise coupling, or distributed networks; (2) The number of regions considered: this could be one region, a handful of regions, or all regions in a given parcellation and/or segmentation; and (3) The number of time-series features considered: this could be one measure, a handful of measures, or all measures in a given feature set—which is in the hundreds for *pyspi* and the thousands for *hctsa*. Here, we use each shaded brain to represent a different feature.

Collectively, this measure–region tradeoff remains an ongoing challenge in highly comparative time-series analysis for neuroimaging and beyond. With increasing position along the continuum of component granularity, introduced in Chapter 1 and depicted schematically in Fig. 8.5.1, the dimensionality of the resulting feature matrix also increases exponentially. Moreover, while beyond the scope of this present thesis, higher-order interactions are increasingly quantified in the computational neuroscience literature, which further exacerbate the potential combinatorial explosion. This is not an insurmountable challenge, though, and there are opportunities for future work to pave the way for principled approaches for combining highly comparative time-series feature libraries with hierarchical representations of the brain's dynamical structure. For example, the same team that developed the *catch22* library recently introduced a new feature subset, *catchaMouse16* [661], which tailored the much broader *hctsa* library specifically to functional neuroimaging data in the mouse brain. Both feature sets were derived in a data-driven manner to minimize redundancy while maximizing the information captured by each feature, with an approach that can be readily extended to other imaging modalities and species in the future. Such an extension could be particularly valuable for translational neuroscience, with the aim of developing a data-driven nosology [354] to identify disorder

subtypes characterized by distinctive dynamical profiles [353].

The *catchaMouse16* method represents one option for feature selection, which is part of a broader class of techniques to reduce the dimensionality of a feature space while retaining the most informative features. Future work could systematically evaluate other dimensionality reduction approaches, including regularization techniques (such as the graphical lasso [344]), similarity network fusion [41, 345], or condensing region pairs into canonical functional networks [185]. Incorporating these approaches into the highly comparative framework presented in this thesis would allow researchers to extend beyond the finite set of representations considered here. For example, in Chapter 4, we focused exclusively on one level of component granularity—pairwise coupling—in part due to aims of the COGITATE challenge, but also because the number of measures in the *pyspi* library is already quite large. However, combining across different levels of component granularity, from local to higher-order, could more fully elucidate properties of neural activity driving a complex biological phenomenon like conscious visual perception. Similarly, in Chapter 5, we defined five distinct representations of resting-state fMRI dynamics, from local to pairwise to distributed network activity, which allowed us to systematically survey across representation types. Within each representation, we only considered a subset of measures in order to mitigate the combinatorial explosion, and there are also many other ways to represent neural activity that could be explored in tandem with appropriate dimensionality reduction techniques. Future work could extend beyond contiguous and localized 'brain regions' to consider properties of other spatiotemporal patterns, including spirals and traveling waves [348, 349] or spatially distributed modes [27, 28, 30, 31].

8.5.3 EXAMINING INDIVIDUAL HETEROGENEITY IN CLINICAL POPULATIONS: NEXT STEPS FROM CHAPTERS 5 AND 6

Chapters 5 and 6 both examined alterations to the brain's dynamical structure in distinct neuropsychiatric disorder populations. For the analysis structure in these two chapters, we chose to focus on group-level differences between cases and controls for each disorder separately—though we did also briefly examine cross-disorder classification with a multi-class support vector machine (SVM) as a supplementary analysis in Chapter 5, which yielded lower overall balanced accuracy but a similar trend of relative time-series feature performance compared to the main analysis. However, mounting evidence suggests that these clinical groups exhibit alterations to neural dynamics that are heterogeneous and patient-specific [319, 339, 340], such that consolidating individuals into a single clinical group may obscure important individual differences. This issue is further compounded by diagnostic variability, arising from a confluence of factors including patient heterogeneity [257, 258] and inter-rater reliability issues [259, 260] that complicate the clinical diagnosis of psychiatric disorders based on behavioral criteria [119]. Collectively, while the approaches and insights introduced in this thesis are a step towards systematically quantifying the brain's dynamical structure in a translational setting, future work could incorporate techniques that preserve individual differences, such as normative modeling [341, 662].

8.5.4 TASK-EVOKED DYNAMICS: FUTURE OUTLOOK FROM CHAPTERS 3, 5, AND 6

Much work in this thesis focused on resting-state brain dynamics with BOLD fMRI, which typically involves several minutes of continuously acquired data in the absence of an explicit task. The brain exhibits a

nonlinear dynamical structure, such that small changes with a given input can give rise to large changes in the output, including chaos, oscillations, and bifurcations [663, 664]. However, prior applications of the *hctsa* library to MEG [239] and fMRI [83] data have shown that linear properties of local dynamics tend to dominate the principal brain-wide axes of temporal variation of the brain at rest. Moreover, findings from this thesis (i.e., Chapters 5 and 6) suggest that the Pearson correlation coefficient can be better suited to capture informative resting-state dynamical structure from fMRI data than more complex measures of pairwise coupling, depending on the specific application. This is consistent with recent comparative modeling work demonstrating that at the macroscale, resting-state brain dynamics are best approximated with a linear autoregressive model [73, 74]—with several potential explanations, including spatial and/or temporal smoothing that are inherent to macroscale neuroimaging methods such as MEG and fMRI.

Collectively, emerging work suggests that in the resting state, without a task-evoking input, the brain's dynamics are closer to an equilibrium, around which the governing dynamics are *approximately* linear [75, 665, 666]. Nozari et al. [74] further posit that task conditions are considerably more likely to yield observable nonlinear dynamics, even in macroscopic neuroimaging modalities like MEG and fMRI. In light of this, a natural extension of the work presented in this thesis would be to apply the systematic framework for quantifying the brain's dynamical structure to task-evoked data—as in Chapter 4, which focused on neural activity while viewing supra-threshold visual stimuli at a relatively short timescale (1 second). MEG is particularly well suited to this type of analysis, given its high temporal resolution and ability to capture fast task-associated neural dynamics. Methodologically, such high temporal resolution is inherently well suited to the current time-series feature computation frameworks embedded in *hctsa* [89, 90] and *pyspi* [92], both of which are designed for continuous time series (ideally with 100 or more timepoints for stable estimation). For example, a 2.5-second MEG epoch, down-sampled to 1 Hz, yields 2,500 timepoints. The sampling frequency of BOLD fMRI is often several orders of magnitude lower; this issue is largely circumvented in resting-state scanning sequences, which typically run for several minutes (or longer).

However, task-based fMRI studies often employ a block design or event-related design (or a combination of the two) with a small number of frames per trial, depending on parameters including the repetition time (TR) and spatial coverage. For a block design, the stimulus or task is presented for a fixed duration (usually between 10–60 seconds), followed by a rest period of the same duration [667]. For an event-related de-sign paradigm, stimuli are generally presented as discrete and brief events (e.g., 1–2 seconds each), with either a fixed or variable inter-stimulus interval between events [668, 669]. In either case, the measured activity must be convolved with the hemodynamic response function (HRF) to yield a stable estimate of the BOLD signal. As a result, the highly comparative time-series framework cultivated in this thesis should be adapted to account for the lower temporal resolution and non-continuous nature of task-based fMRI. Future work in this space could open up exciting avenues for systematically quantifying the brain's dynamical structure in task-evoked contexts. For example, it would be very interesting to comprehensively quantify how cortical homotopic functional connectivity evolves with increasing task complexity, as it is possible that regions with relatively low resting-state homotopic coupling (namely, association cortices) may ramp up inter-hemispheric synchrony in response to a more cognitively demanding task [51, 404, 407, 504].

8.5.5 BALANCING INTEGRATION AND SEGREGATION: FUTURE DIRECTIONS FROM CHAPTERS 6 AND 7

Examining properties of distributed functional and structural connectomic networks in Chapters 6 and 7, respectively, raised compelling questions about the balance between integration and segregation in the brain's overall dynamical structure. Shine et al. [670] demonstrated that the human brain's network structure dynamically reorganizes to traverse between states of segregation (i.e., separation into a modular structure of distinct communities) and integration (i.e., communication between otherwise disparate communities). While these states refer to the functional organization of the brain, these dynamics are mechanistically supported by the structural connectome, which comprises the physical architecture of axonal connections between brain regions. As described in Sporns [611], the structural connectome shapes information flow between brain regions and supports an essential balance between functional segregation and integration.

We found in Chapter 6 that in the resting state, the spatial variation in HoFC magnitude was tightly coupled to global synchrony (defined as the average functional connectivity between a given region and all others in the brain). It was beyond the scope of this work to investigate integration specifically, though this preliminary finding suggests that the extent to which a region is synchronized with its mirror homolog may relate to its overall functional integration within and across hemispheres at rest. Future work could investigate this relationship in more detail; for example, it would be interesting to incorporate the comparative OCDA framework from Chapter 7 to examine whether HoFC magnitude relates to the extent of participation across multiple functional modules, in the resting state as well as with increasing task complexity. It would also be interesting to examine whether the relative magnitude of HoFC compared to all other functional connections per region varies across the cortical sheet—if so, that could suggest a relative prioritization of homotopic connections compared to all other intra-hemispheric and heterotopic connections. Additionally, given prior evidence that both the spatial variation in HoFC magnitude [7] and the functional integration–segregation balance [671] both evolve with development, it would be informative to compare these two properties across the lifespan, from early childhood through older adulthood.

Since the structural connectome contributes a physical backbone for functional connectivity, our findings in Chapter 7 invite further investigation into the functional role(s) of a node that is assigned to multiple modules based on anatomical connectivity. Bijsterbosch et al. [546] recently explored a spectrum of plausible roles for overlapping nodes in the functional connectome, including mixing (characterized by a spatially heterogeneous area with 'interdigitation' of neuronal populations belonging to distinct communities), switching (communicating exclusively with one network at a time), and coupling (serving as an integrative hub that exhibits additive or multiplicative coupling). It would be interesting to extend this framework to the structural connectome, particularly in light of our findings in Chapter 7 that the fifteen overlapping nodes were fairly evenly distributed across the right-hemisphere cortical sheet and assigned to two modules each—with the exception of the retrosplenial cortex (*RSC*), which was assigned to three, consistent with its dense inter-network connectivity [626] and integrative sensory roles [627, 628].

One final point to note from Chapter 7 is that the seven structural communities were largely spatially contiguous (which is similar to findings in the macaque structural connectome [538]), and the fifteen overlapping nodes each sat at the spatial interface of the two (or three) modules to which they were assigned. Interest-

ingly, previous work has suggested that higher-order integrative regions like *PFm* and *PFs*, which OSLOM assigned to two structural communities each, act as important functional and cytoarchitectonic transition zones between adjacent brain regions [623, 624]. Moreover, the top-performing OCDA (OSLOM) inferred at least one overlapping region per structural community, barring the insular network, which suggests that overlaps are sparse yet well-distributed across modules. Given the high metabolic cost of developing and maintaining longer inter-modular axonal connections [639], this raises intriguing follow-up questions about how the brain's structural connectome inherently strikes a delicate balance between integration and segregation in its overlapping modular architecture.

8.6 FINAL REMARKS

This thesis has presented a systematic framework for quantifying the multiscale dynamical structure of the human brain using a highly comparative approach, with a focus on unifying interdisciplinary methods through both theoretical and empirical lenses. In treating the brain as a complex system, this thesis has explored multiple levels of component granularity, including local dynamics, pairwise coupling, and distributed network activity. Across diverse applications, from neuropsychiatric disorder classification to analyzing network topology in the human connectome, this thesis has demonstrated the benefits of comprehensively comparing measures within *and across* representations to distill the most salient properties of the brain's dynamical structure for a given application. This work was guided by my overarching motivations for this thesis, which blended rigorous statistical analysis with methodological parsimony and accessible language, with the aim of bridging the gap between interdisciplinary theory and practical implementation in computational neuroscience. Taken together, this thesis suggests exciting new directions for future work to systematically quantify multiscale brain dynamics to address a multitude of biologically compelling questions in both basic research and clinical neuroscience—such as characterizing individualized neural dynamics across disorder subtypes, examining the mechanistic underpinnings of inter-hemispheric coupling, and exploring the evolving balance of integration and segregation in the human connectome across the lifespan.

Thesis References

[1] Luiz Pessoa. Understanding brain networks and brain organization. *Physics of life reviews* 11.3 (2014), pp. 400–435.

[2] Corey Horien et al. Regions and connections: Complementary approaches to characterize brain organization and function. *The Neuroscientist* 26.2 (2020), pp. 117–133.

[3] James C Pang et al. Geometric influences on the regional organization of the mammalian brain. *bioRxiv* (2025), pp. 2025–01.

[4] Qi-Hong Zou et al. An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF. *Journal of neuroscience methods* 172.1 (2008), pp. 137–141.

[5] Xi-Nian Zuo et al. The oscillating brain: Complex and reliable. *Neuroimage* 49.2 (2010), pp. 1432–1445.

[6] Yufeng Zang et al. Regional homogeneity approach to fMRI data analysis. *Neuroimage* 22.1 (2004), pp. 394–400.

[7] Xi-Nian Zuo et al. Growing together and growing apart: Regional and sex differences in the lifespan developmental trajectories of functional homotopy. *Journal of Neuroscience* 30.45 (2010), pp. 15034–15043.

[8] John D Murray et al. A hierarchy of intrinsic timescales across primate cortex. *Nature neuroscience* 17.12 (2014), pp. 1661–1663.

[9] Uri Hasson et al. Hierarchical process memory: Memory as an integral component of information processing. *Trends in cognitive sciences* 19.6 (2015), pp. 304–313.

[10] Stephanie Noble et al. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *Neuroimage* 203 (2019), p. 116157.

[11] Jaroslav Hlinka et al. Functional connectivity in resting-state fMRI: Is linear correlation sufficient? *Neuroimage* 54.3 (2011), pp. 2218–2225.

[12] Robert L Peach et al. HCGA: Highly comparative graph analysis for network phenotyping. *Patterns* 2.4 (2021).

[13] Rosaleena Mohanty et al. Rethinking measures of functional connectivity via feature extraction. *Scientific reports* 10.1 (2020), p. 1298.

[14] Pavel Prado et al. Source space connectomics of neurodegeneration: One-metric approach does not fit all. *Neurobiology of disease* (2023), p. 106047.

[15] Claudio Gentili et al. Not in one metric: Neuroticism modulates different resting state metrics within distinctive brain regions. *Behavioural brain research* 327 (2017), pp. 34–43.

[16] Martijn P van den Heuvel and BT Thomas Yeo. A spotlight on bridging microscale and macroscale human brain architecture. *Neuron* 93.6 (2017), pp. 1248–1251.

[17] Philipp Haueis. Descriptive multiscale modeling in data-driven neuroscience. *Synthese* 200.2 (2022), p. 129.

[18] Richard F Betzel and Danielle S Bassett. Multi-scale brain networks. *Neuroimage* 160 (2017), pp. 73–83.

[19] Mikhail Prokopenko et al. An information-theoretic primer on complexity, self-organization, and emergence. *Complexity* 15.1 (2009), pp. 11–28.

[20] Danielle S Bassett and Michael S Gazzaniga. Understanding complexity in the human brain. *Trends in cognitive sciences* 15.5 (2011), pp. 200–209.

[21] Robert F Engle and Clive WJ Granger. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55.2 (1987), pp. 251–276.

[22] Jakob Runge et al. Inferring causation from time series in Earth system sciences. *Nat Commun* 10.1 (2019), p. 2553.

[23] G. Sugihara et al. Detecting causality in complex ecosystems. *Science* 338.6106 (2012), pp. 496–500.

[24] Yuanzhao Zhang et al. Higher-order interactions shape collective dynamics differently in hypergraphs and simplicial complexes. *Nature communications* 14.1 (2023), p. 1605.

[25] Andrea Santoro et al. Higher-order connectomics of human brain function reveals local topological signatures of task decoding, individual identification, and behavior. *Nature Communications* 15.1 (2024), p. 10244.

[26] Borja Camino-Pontes et al. Brain structural modules associated to functional high-order interactions in the human brain. *bioRxiv* (2025), pp. 2025–03.

[27] Eli J Müller et al. The music of the hemispheres: Cortical eigenmodes as a physical basis for large-scale brain activity and connectivity patterns. *Frontiers in human neuroscience* 16 (2022), p. 1062487.

[28] James C Pang et al. Geometric constraints on human brain function. *Nature* (2023), pp. 1–9.

[29] Adrian Allen et al. CoCoNest: A continuous structural connectivity-based nested family of parcellations of the human cerebral cortex. *Network Neuroscience* 8.4 (2024), pp. 1439–1466.

[30] Neil W Bailey et al. Uncovering a stability signature of brain dynamics associated with meditation experience using massive time-series feature extraction. *Neural networks* 171 (2024), pp. 171–185.

[31] Alex H Williams et al. Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron* 98.6 (2018), pp. 1099–1115.

[32] Andrew Zalesky et al. The relationship between regional and inter-regional functional connectivity deficits in schizophrenia. *Human brain mapping* 33.11 (2012), pp. 2535–2549.

[33] Bratislav Mišić and Olaf Sporns. From regions to connections and networks: New bridges between brain and behavior. *Current opinion in neurobiology* 40 (2016), pp. 1–7.

[34] Lia Papadopoulos et al. Relations between large-scale brain connectivity and effects of regional stimulation depend on collective dynamical state. *PLoS computational biology* 16.9 (2020), e1008144.

[35] Du Lei et al. Integrating machining learning and multimodal neuroimaging to detect schizophrenia at the level of the individual. *Human brain mapping* 41.5 (2020), pp. 1119–1135.

[36] Zhengjia Dai et al. Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). en. *Neuroimage* 59 (Feb. 2012), pp. 2187–2195. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2011.10.003. URL: https://www.sciencedirect.com/science/article/pii/S1053811911011645 (visited on 02/16/2022).

[37] Xiaoyu Ding et al. Combining multiple resting-state fMRI features during classification: Optimized frameworks and their application to nicotine addiction. *Frontiers in human neuroscience* 11 (2017), p. 362.

[38] Likun Xia et al. A physiological signal-based method for early mental-stress detection. In: *Cyber-Enabled Intelligence*. Taylor & Francis, 2019, pp. 259–289.

[39] Adrián Ponce-Alvarez et al. Macroscopic quantities of collective brain activity during wakefulness and anesthesia. *Cerebral cortex* 32.2 (2022), pp. 298–311.

[40] Xiaofeng Zhu et al. Fusing functional connectivity with network nodal information for sparse network pattern learning of functional brain networks. *Information fusion* 75 (2021), pp. 131–139.

[41] Justine Y Hansen et al. Integrating multimodal and multiscale connectivity blueprints of the human cerebral cortex in health and disease. *PLoS biology* 21.9 (2023), e3002314.

[42] Marija Markicevic et al. Cortical excitation:inhibition imbalance causes abnormal brain network dynamics as observed in neurodevelopmental disorders. eng. *Cerebral cortex* 30 (July 2020), pp. 4922–4937. ISSN: 1460-2199. DOI: 10.1093/cercor/bhaa084.

[43] Marija Markicevic et al. Neuromodulation of striatal D1 cells shapes BOLD fluctuations in anatomically connected thalamic and cortical regions. *eLife* 12 (2023), e78620.

[44] Yuling Luo et al. Associating multimodal neuroimaging abnormalities with the transcriptome and neurotransmitter signatures in schizophrenia. *Schizophrenia bulletin* (2023).

[45] Amir Ebneabbasi et al. How Do Neurotransmitter Pathways Contribute to Neuroimaging Phenotypes? *medRxiv* (2024), pp. 2024–04.

[46] Zhiqing Tang et al. Static and temporal dynamic changes in brain activity in patients with post-stroke balance dysfunction: A pilot resting state fMRI. *Frontiers in Neuroscience* 19 (2025), p. 1558069.

[47] Karl J Friston. Functional and effective connectivity: A review. *Brain connectivity* 1.1 (2011), pp. 13–36.

[48] Joanes Grandjean et al. Structural basis of large-scale functional connectivity in the mouse. *Journal of Neuroscience* 37.34 (2017), pp. 8092–8101.

[49] Christopher J Honey et al. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences* 106.6 (2009), pp. 2035–2040.

[50] Yusuke Adachi et al. Functional connectivity between anatomically unconnected areas is shaped by collective network-level effects in the macaque cortex. *Cerebral cortex* 22.7 (2012), pp. 1586–1592.

[51] Lorenzo Mancuso et al. The homotopic connectivity of the functional brain: A meta-analytic approach. *Scientific Reports* 9.1 (2019), p. 3346.

[52] Yao Fei et al. Diverse and asymmetric patterns of single-neuron projectome in regulating interhemispheric connectivity. *Nature Communications* 15.1 (2024), p. 3403.

[53] Martijn P. van den Heuvel and Hilleke E. Hulshoff Pol. Exploring the brain network: A review on resting-state fMRI functional connectivity. en. *Eur Neuropsychopharmacol* 20.8 (2010), pp. 519–534.

[54] Joshua Faskowitz et al. Edges in brain networks: Contributions to models of structure and function. *Network Neuroscience* 6.1 (2022), pp. 1–28.

[55] Edward T Bullmore and Olaf Sporns. The economy of brain network organization. *Nat. Rev. Neurosci.* 13 (May 2012), pp. 337–349.

[56] Alex Fornito et al. *Fundamentals of brain network analysis*. Academic Press, 2016.

[57] Olaf Sporns and Richard F Betzel. Modular brain networks. *Annual review of psychology* 67 (2016), pp. 613–640.

[58] Bratislav Mišić et al. Network-level structure-function relationships in human neocortex. *Cerebral Cortex* 26.7 (2016), pp. 3285–3296.

[59] Joana Cabral et al. Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms. *NeuroImage* 160 (2017), pp. 84–96.

[60] Bertha Vázquez-Rodríguez et al. Gradients of structure–function tethering across neocortex. *Proceedings of the National Academy of Sciences* 116.42 (2019), pp. 21219–21227.

[61] Roger Guimerà and Luís A Nunes Amaral. Functional cartography of complex metabolic networks. *Nature* 433.7028 (Feb. 2005), pp. 895–900. ISSN: 0028-0836.

[62] Richard F Betzel et al. Functional brain modules reconfigure at multiple scales across the human lifespan. *arXiv preprint arXiv:1510.08045* (2015).

[63] Vincent Blondel et al. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment* 2008 (Apr. 2008). DOI: 10.1088/1742-5468/2008/10/P10008.

[64] Richard F Betzel. Community detection in network neuroscience. In: *Connectome Analysis*. Elsevier, 2023, pp. 149–171.

[65] Karl Zilles and Katrin Amunts. Anatomical basis for functional specialization. *fMRI: From nuclear spins to brain functions* (2015), pp. 27–66.

[66] Danhong Wang et al. Functional specialization in the human brain estimated by intrinsic hemispheric interaction. *Journal of Neuroscience* 34.37 (2014), pp. 12341–12352.

[67] Semir Zeki et al. A direct demonstration of functional specialization in human visual cortex. *Journal of neuroscience* 11.3 (1991), pp. 641–649.

[68] Eiji Hoshi and Jun Tanji. Functional specialization in dorsal and ventral premotor areas. *Progress in brain research* 143 (2004), pp. 507–511.

[69] Mahshid Najafi et al. Overlapping communities reveal rich structure in large-scale brain networks during rest and task conditions. *Neuroimage* 135 (2016), pp. 92–106.

[70] Hadi Vafaii et al. Multimodal Measures of Spontaneous Brain Activity Reveal Both Common and Divergent Patterns of Cortical Functional Organization. *Nature Communications* 15.1 (Jan. 2024), p. 229.

[71] M-Marsel Mesulam. From sensation to cognition. *Brain: a journal of neurology* 121.6 (1998), pp. 1013–1052.

[72] Valerie J Sydnor et al. Neurodevelopment of the association cortices: Patterns, mechanisms, and implications for psychopathology. *Neuron* 109.18 (2021), pp. 2820–2846.

[73] Arnaud Messé et al. Predicting functional connectivity from structural connectivity via computational models using MRI: An extensive comparison study. *NeuroImage* 111 (2015), pp. 65–75.

[74] Erfan Nozari et al. Macroscopic resting-state brain dynamics are best described by linear models. *Nature biomedical engineering* 8.1 (2024), pp. 68–84.

[75] Viktor Sip et al. Characterization of regional differences in resting-state fMRI with a data-driven network model of brain dynamics. *Science Advances* 9.11 (2023), eabq7547.

[76] German Rodriguez-Bermudez and Pedro J Garcia-Laencina. Analysis of EEG signals using nonlinear dynamics and chaos: A review. *Applied mathematics & information sciences* 9.5 (2015), p. 2309.

[77] Saime Akdemir Akar et al. Nonlinear analysis of EEGs of patients with major depression during different emotional states. *Computers in biology and medicine* 67 (2015), pp. 49–60.

[78] Yan Ma et al. Nonlinear dynamical analysis of sleep electroencephalography using fractal and entropy approaches. *Sleep medicine reviews* 37 (2018), pp. 85–93.

[79] Annie G Bryant et al. Extracting interpretable signatures of whole-brain dynamics through systematic comparison. *PLOS Computational Biology* 20.12 (2024), e1012692.

[80] Giulio Tani Raffaelli et al. Nonlinear brain connectivity from neurons to networks: Quantification, sources and localization. *bioRxiv* (2024), pp. 2024–11.

[81] Oliver M Cliff et al. Assessing the significance of directed and multivariate measures of linear dependence between time series. *Phys Rev Res* 3.1 (2021), p. 013145.

[82] André M Bastos and Jan-Mathijs Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience* 9 (2016), p. 175.

[83] Golia Shafiei et al. Topographic gradients of intrinsic dynamics across neocortex. *eLife* 9 (Dec. 2020), e62116. ISSN: 2050-084X. (Visited on 01/21/2022).

[84] Zhen-Qi Liu et al. Benchmarking methods for mapping functional connectivity in the brain. *bioRxiv* (2024), pp. 2024–05.

[85] Annika C Linke et al. Dynamic time warping outperforms Pearson correlation in detecting atypical functional connectivity in autism spectrum disorders. *Neuroimage* 223 (2020), p. 117383.

[86] Regina J Meszlényi et al. Resting state fMRI functional connectivity analysis using dynamic time warping. *Frontiers in neuroscience* 11 (2017), p. 75.

[87] Alex Novaes Santana et al. Using deep learning and resting-state fMRI to classify chronic pain conditions. *Frontiers in neuroscience* 13 (2019), p. 1313.

[88] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on signal processing* 26.1 (1978), pp. 43–49.

[89] Ben D. Fulcher et al. Highly comparative time-series analysis: The empirical structure of time series and their methods. *Journal of The Royal Society Interface* 10.83 (June 2013). (Visited on 01/21/2022).

[90] Ben D. Fulcher and Nick S. Jones. hctsa: Computational Framework for Automated Time-Series Phenotyping Using Massive Feature Extraction. en. *Cell systems* 5.5 (Nov. 2017). Number: 5, 527–531.e3. ISSN: 2405-4712. DOI: 10.1016/j.cels.2017.10.001. URL: https://www.sciencedirect.com/science/article/pii/S2405471217304386 (visited on 01/21/2022).

[91] Carl H Lubba et al. catch22: CAnonical Time-series CHaracteristics: Selected through highly comparative time-series analysis. *Data mining and knowledge discovery* 33.6 (2019), pp. 1821–1852.

[92] Oliver M Cliff et al. Unifying pairwise interactions in complex dynamics. *Nature Computational Science* 3.10 (2023), pp. 883–893.

[93] David H Wolpert and William G Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation* 1.1 (1997), pp. 67–82.

[94] Leto Peel et al. The ground truth about metadata and community detection in networks. *Science Advances* 3.5 (2017).

[95] Arya D McCarthy et al. An exact no free lunch theorem for community detection. In: *International Conference on Complex Networks and Their Applications*. Springer. 2019, pp. 176–187.

[96] Mark S Cohen and Susan Y Bookheimer. Localization of brain function using magnetic resonance imaging. *Trends in neurosciences* 17.7 (1994), pp. 268–277.

[97] Stuart Currie et al. Understanding MRI: Basic MR physics for physicians. *Postgraduate medical journal* 89.1050 (2013), pp. 209–223.

[98] Adam J Schwarz. The use, standardization, and interpretation of brain imaging data in clinical trials of neurodegenerative disorders. *Neurotherapeutics* 18.2 (2021), pp. 686–708.

[99] John Mazziotta et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 356.1412 (2001), pp. 1293–1322.

[100] Mara Cercignani and Mark A Horsfield. The physical basis of diffusion-weighted MRI. *Journal of the neurological sciences* 186 (2001), S11–S14.

[101] Vinit Baliyan et al. Diffusion weighted imaging: Technique and applications. *World journal of radiology* 8.9 (2016), p. 785.

[102] Robert E Smith et al. Anatomically-constrained tractography: Improved diffusion MRI streamlines tractography through effective use of anatomical information. *Neuroimage* 62.3 (2012), pp. 1924–1938.

[103] Nikos K Logothetis and Brian A Wandell. Interpreting the BOLD signal. *Annu. Rev. Physiol.* 66 (2004), pp. 735–769.

[104] Kâmil Uludağ and Kâmil Uğurbil. Physiology and physics of the fMRI Signal. *fMRI: From nuclear spins to brain functions* (2015), pp. 163–213.

[105] Costantino Iadecola. Neurovascular regulation in the normal brain and in Alzheimer's disease. *Nature Reviews Neuroscience* 5.5 (2004), pp. 347–360.

[106] Richard B Buxton et al. Modeling the hemodynamic response to brain activation. *Neuroimage* 23 (2004), S220–S233.

[107] Sylvain Baillet. Magnetoencephalography for brain electrophysiology and imaging. *Nature neuroscience* 20.3 (2017), pp. 327–339.

[108] Terrence J Sejnowski et al. Putting big data to good use in neuroscience. *Nat. Neurosci* 17.11 (2014), pp. 1440–1441.

[109] Sylvain Baillet et al. Electromagnetic brain mapping. *IEEE Signal processing magazine* 18.6 (2001), pp. 14–30.

[110] Matti Hämäläinen et al. Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics* 65.2 (1993), p. 413.

[111] Francesca Pizzo et al. Deep brain activities can be detected with magnetoencephalography. *Nature communications* 10.1 (2019), p. 971.

[112] Michal Teplan. Fundamentals of EEG measurement. *Measurement science review* 2.2 (2002), pp. 1–11.

[113] György Buzsáki et al. The origin of extracellular fields and currents–EEG, ECoG, LFP and spikes. *Nature reviews neuroscience* 13.6 (2012), pp. 407–420.

[114] Joachim Gross. Magnetoencephalography in cognitive neuroscience: A primer. *Neuron* 104.2 (2019), pp. 189–204.

[115] Sylvain Baillet and Line Garnero. A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. *IEEE transactions on Biomedical Engineering* 44.5 (2002), pp. 374–385.

[116] John C Mosher et al. EEG and MEG: Forward solutions for inverse methods. *IEEE Transactions on biomedical engineering* 46.3 (2002), pp. 245–259.

[117] Arjan Hillebrand and Gareth R Barnes. Beamformer analysis of MEG data. *International review of neurobiology* 68 (2005), pp. 149–171.

[118] Michael Baumgartner. Parsimony and causality. *Quality & Quantity* 49 (2015), pp. 839–856.

[119] Katherine H Taber et al. Diagnosis and treatment of neuropsychiatric disorders. *Annual review of medicine* 61 (2010), pp. 121–133.

[120] Michael S Breen et al. Systematic review of blood transcriptome profiling in neuropsychiatric disorders: Guidelines for biomarker discovery. *Human psychopharmacology: Clinical and experimental* 31.5 (2016), pp. 373–381.

[121] Walid Yassin et al. Machine-learning classification using neuroimaging data in schizophrenia, autism, ultra-high risk and first-episode psychosis. *Translational psychiatry* 10.1 (2020), p. 278.

[122] Najmeh Khalili-Mahani et al. Biomarkers, designs, and interpretations of resting-state fMRI in translational pharmacological research: A review of state-of-the-Art, challenges, and opportunities for studying brain chemistry. *Human brain mapping* 38.4 (2017), pp. 2276–2325.

[123] Choong-Wan Woo et al. Building better biomarkers: Brain models in translational neuroimaging. *Nature neuroscience* 20.3 (2017), pp. 365–377.

[124] Bingni W Brunton and Michael Beyeler. Data-driven models in human neuroscience and neuroengineering. *Current opinion in neurobiology* 58 (2019), pp. 21–29.

[125] Michel Verleysen and Damien François. The curse of dimensionality in data mining and time series prediction. In: *Computational Intelligence and Bioinspired Systems: 8th International Work-Conference on Artificial Neural Networks, IWANN 2005, Vilanova i la Geltrú, Barcelona, Spain, June 8-10, 2005. Proceedings 8*. Springer. 2005, pp. 758–770.

[126] Trevor Hastie and Robert Tibshirani. Efficient quadratic regularization for expression arrays. *Biostatistics* 5.3 (2004), pp. 329–340.

[127] Majed Ghanbari et al. Diagnosis of schizophrenia using brain resting-state fMRI with activity maps based on deep learning. *Signal, Image and Video Processing* 17.1 (2023), pp. 267–275.

[128] Xiaoyi Chen et al. Classification of schizophrenia patients using a graph convolutional network: A combined functional MRI and connectomics analysis. *Biomedical Signal Processing and Control* 80 (2023), p. 104293.

[129] Vikas Khullar et al. Deep learning-based binary classification of ADHD using resting state MR images. *Augmented Human Research* 6.1 (2021), p. 5.

[130] Faria Zarin Subah et al. A deep learning approach to predict autism spectrum disorder using multisite resting-state fMRI. *Applied Sciences* 11.8 (2021), p. 3636.

[131] Fayyaz Ahmad et al. Classification of schizophrenia-associated brain regions in resting-state fMRI. *The European Physical Journal Plus* 138.1 (2023), pp. 1–12.

[132] Nils R Winter et al. A systematic evaluation of machine learning–based biomarkers for major depressive disorder. *JAMA psychiatry* 81.4 (2024), pp. 386–395.

[133] Vladimir Belov et al. Multi-site benchmark classification of major depressive disorder using machine learning on cortical and subcortical measures. *Scientific reports* 14.1 (2024), p. 1084.

[134] Claas Flint et al. Systematic misestimation of machine learning performance in neuroimaging studies of depression. *Neuropsychopharmacology* 46.8 (2021), pp. 1510–1517.

[135] Ub Waheed et al. Winning with simple learning models: Detecting earthquakes in Groningen, the Netherlands. In: *EAGE 2020 annual conference & exhibition online*. Vol. 2020. 1. European Association of Geoscientists & Engineers. 2020, pp. 1–5.

[136] Arnaud Mignan and Marco Broccardo. One neuron versus deep learning in aftershock prediction. *Nature* 574.7776 (2019), E1–E3.

[137] Jeroen Van Der Donckt et al. Do not sleep on traditional machine learning: Simple and interpretable techniques are competitive to deep learning for sleep scoring. *Biomedical Signal Processing and Control* 81 (2023), p. 104429.

[138] Ralph G Andrzejak et al. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E* 64.6 (2001), p. 061907.

[139] Edward R Tufte and Peter R Graves-Morris. *The visual display of quantitative information*. Vol. 2. 9. Graphics press Cheshire, CT, 1983.

[140] Edward R Tufte. *Visual explanations: Images and quantities, evidence and narrative*. Graphics Press, 1997.

[141] Jill H Larkin and Herbert A Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive science* 11.1 (1987), pp. 65–100.

[142] Arthur M Glenberg and William E Langston. Comprehension of illustrated text: Pictures help to build mental models. *Journal of memory and language* 31.2 (1992), pp. 129–151.

[143] Julie Jebeile. Explaining with simulations: Why visual representations matter. *Perspectives on Science* 26.2 (2018), pp. 213–238.

[144] Eliza Bobek and Barbara Tversky. Creating visual explanations improves learning. *Cognitive research: principles and implications* 1 (2016), pp. 1–14.

[145] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal* 27.3 (1948), pp. 379–423.

[146] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience, 2005.

[147] David JC MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[148] Alexander Borst and Frédéric E Theunissen. Information theory and neural coding. *Nature neuroscience* 2.11 (1999), pp. 947–957.

[149] Nicholas M Timme and Christopher Lapish. A tutorial for information theory in neuroscience. *eneuro* 5.3 (2018).

[150] Cogitate Consortium et al. Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature* (2025), pp. 1–10.

[151] Michael S Harré. Information theory for agents in artificial intelligence, psychology, and economics. *Entropy* 23.3 (2021), p. 310.

[152] Kris V Parag and Christl A Donnelly. Using information theory to optimise epidemic models for real-time prediction and estimation. *PLoS computational biology* 16.7 (2020), e1007990.

[153] Henry Pinkard and Laura Waller. A visual introduction to information theory. *arXiv preprint arXiv:2206.07867* (2022).

[154] Cesar Huerta et al. Comparison of causality determination techniques in studying Hall-effect thrusters. *Journal of Electric Propulsion* 3.1 (2024), p. 23.

[155] Adrián Lozano-Durán and Gonzalo Arranz. Information-theoretic formulation of dynamical systems: Causality, modeling, and control. *Physical Review Research* 4.2 (2022), p. 023195.

[156] Charles Murphy et al. Duality between predictability and reconstructability in complex systems. *Nature Communications* 15.1 (2024), p. 4478.

[157] Hyunju Kim et al. Informational architecture across non-living and living collectives. *Theory in Biosciences* (2021), pp. 1–17.

[158] Conor Finn and Joseph T Lizier. Generalised measures of multivariate information content. *Entropy* 22.2 (2020), p. 216.

[159] Joseph T Lizier. JIDT: An information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI* 1 (2014), p. 11.

[160] Oliver Cliff et al. *DynamicsAndNeuralSystems/pyspi*. Version 0.4.1. URL: https://doi.org/10.5281/zenodo.8223340. Aug. 2023. DOI: 10.5281/zenodo.8223340.

[161] David C Van Essen et al. The WU-Minn Human Connectome Project: An overview. *Neuroimage* 80 (2013), pp. 62–79.

[162] David P Shorten et al. Estimating transfer entropy in continuous time between neural spike trains or other event-based data. *PLoS Comput Biol* 17.4 (2021), e1008054.

[163] Gorana Mijatovic et al. An information-theoretic framework to measure the dynamic interaction between neural spike trains. *IEEE Transactions on Biomedical Engineering* 68.12 (2021), pp. 3471–3481.

[164] Cesare Magri et al. A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. *BMC neuroscience* 10 (2009), pp. 1–24.

[165] Joseph T Lizier et al. Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *Journal of computational neuroscience* 30 (2011), pp. 85–107.

[166] Zhe Wang et al. Causality analysis of fMRI data based on the directed information theory framework. *IEEE Transactions on Biomedical Engineering* 63.5 (2015), pp. 1002–1015.

[167] John Fallon et al. Timescales of spontaneous fMRI fluctuations relate to structural connectivity in the brain. *Network Neuroscience* 4.3 (2020), pp. 788–806.

[168] Linden Parkes et al. Transcriptional signatures of connectomic subregions of the human striatum. *Genes, Brain and Behavior* 16.7 (2017), pp. 647–663.

[169] Rahul S. Desikan et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. en. *Neuroimage* 31.3 (July 2006), pp. 968–980. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2006.01.021. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811906000437 (visited on 11/18/2022).

[170] Hoang Anh Dau et al. The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica* 6.6 (2019), pp. 1293–1305.

[171] Anthony Bagnall et al. The UEA multivariate time series classification archive, 2018. *arXiv* (2018), p. 1811.00075.

[172] Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nat Neurosci* 20.3 (2017), pp. 353–364.

[173] Randy L Buckner et al. Molecular, structural, and functional characterization of Alzheimer's disease: Evidence for a relationship between default activity, amyloid, and memory. *J Neurosci* 25.34 (2005), pp. 7709–7717.

[174] Robert L. Peach et al. HCGA: Highly comparative graph analysis for network phenotyping. en. *Patterns* 2.4 (Apr. 2021). Number: 4, p. 100227. ISSN: 2666-3899. DOI: 10.1016/j.patter.2021.100227. URL: https://www.sciencedirect.com/science/article/pii/S2666389921000416 (visited on 03/02/2022).

[175] Alejandro Pasos Ruiz et al. The great multivariate time series classification bake off: A review and experimental evaluation of recent algorithmic advances. *Data Min Knowl Discov* 35.2 (2021), pp. 401–449.

[176] Niels Birbaumer et al. A spelling device for the paralysed. *Nature* 398.6725 (1999), pp. 297–298.

[177] Lisa Byrge and Daniel P Kennedy. Accurate prediction of individual subject identity and task, but not autism diagnosis, from functional connectomes. *Human brain mapping* 41.9 (2020), pp. 2249–2262.

[178] Ying Liu and Selin Aviyente. Quantification of effective connectivity in the brain using a measure of directed information. *Comput Math Methods Med* 2012 (2012).

[179] Ketan Mehta and Jörg Kliewer. Directional and causal information flow in EEG for assessing perceived audio quality. *IEEE Trans Mol Biol Multiscale Commun* 3.3 (2017), pp. 150–165.

[180] Gang Wang and Morikuni Takigawa. Directed coherence as a measure of interhemispheric correlation of EEG. *Int J Psychophysiol* 13.2 (1992), pp. 119–128.

[181] Björn Schelter et al. Testing for directed influences among neural signals using partial directed coherence. *J Neurosci Methods* 152.1-2 (2006), pp. 210–219.

[182] Stephen M Smith et al. Network modelling methods for FMRI. English. *NeuroImage* 54.2 (2011), pp. 875–891. DOI: 10.1016/j.neuroimage.2010.08.063.

[183] Ben D Fulcher. Feature-based time-series analysis (2018), pp. 87–116.

[184] Richard Betzel et al. *Resting-state and movie-watching data. Figshare dataset: https://doi.org/10.6084/m9.figshare.* Sept. 2020. DOI: 10.6084/m9.figshare.12971162.v2.

[185] BT Thomas Yeo et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology* (2011).

[186] Fabian Pedregosa et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12 (2011), pp. 2825–2830.

[187] Carlo Bonferroni. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze* 8 (1936), pp. 3–62.

[188] Christof Koch et al. Neural correlates of consciousness: Progress and problems. *Nature reviews neuroscience* 17.5 (2016), pp. 307–321.

[189] Giulio Tononi and Gerald M Edelman. Consciousness and complexity. *Science* 282.5395 (1998), pp. 1846–1851.

[190] Larissa Albantakis et al. Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLoS computational biology* 19.10 (2023), e1011465.

[191] Stanislas Dehaene and Lionel Naccache. Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 79.1-2 (2001), pp. 1–37.

[192] George A Mashour et al. Conscious processing and the global neuronal workspace hypothesis. *Neuron* 105.5 (2020), pp. 776–798.

[193] Itay Yaron et al. The ConTraSt database for analysing and comparing empirical studies of consciousness theories. *Nature Human Behaviour* 6.4 (2022), pp. 593–604.

[194] Bernard J Baars. *A cognitive theory of consciousness*. Cambridge University Press, 1993.

[195] Shimon Edelman et al. Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* 26 (1998), pp. 309–321.

[196] Nancy Kanwisher et al. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of neuroscience* 17 (1997), pp. 4302–4311.

[197] Melanie Boly et al. Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *Journal of Neuroscience* 37.40 (2017), pp. 9603–9613.

[198] Lucia Melloni et al. An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory. *Plos one* 18.2 (2023), e0268577.

[199] Daming Wang and Yaoru Sun. The pairwise phase consistency in cortical network and its relationship with neuronal activation. In: *BIO Web of Conferences*. Vol. 8. EDP Sciences. 2017, p. 02006.

[200] Martin Vinck et al. The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization. en. *NeuroImage* 51.1 (May 2010), pp. 112–122. ISSN: 10538119. DOI: 10.1016/j.NeuroImage.2010.01.073. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053811910000959 (visited on 06/03/2021).

[201] Joerg F. Hipp et al. Large-scale cortical correlation structure of spontaneous oscillatory activity. en. *Nature Neuroscience* 15.6 (June 2012). Publisher: Nature Publishing Group, pp. 884–890. ISSN: 1546-1726. DOI: 10.1038/nn.3101. URL: https://www.nature.com/articles/nn.3101 (visited on 06/14/2024).

[202] Sheraz Khan et al. Maturation trajectories of cortical resting-state networks depend on the mediating frequency band. eng. *NeuroImage* 174 (July 2018), pp. 57–68. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2018.02.018.

[203] Masafumi Oizumi et al. Measuring integrated information from the decoding perspective. *PLoS computational biology* 12.1 (2016), e1004654.

[204] Michael Breakspear. Dynamic models of large-scale brain activity. *Nature neuroscience* 20.3 (2017), pp. 340–352.

[205] L Liu et al. *Cogitate M-EEG Data [Dataset]*. 2024. DOI: https://doi.org/10.17617/1.WQA3-WK71.

[206] Alexandre Gramfort et al. MEG and EEG Data Analysis with MNE-Python. *Frontiers in Neuroscience* 7.267 (2013), pp. 1–13. DOI: 10.3389/fnins.2013.00267.

[207] Alexander Schaefer et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral cortex* 28.9 (2018), pp. 3095–3114.

[208] Oscar Ferrante et al. FLUX: A pipeline for MEG analysis. *NeuroImage* 253 (2022), p. 119047.

[209] Samu Taulu et al. Suppression of interference and artifacts by the signal space separation method. en. *Brain Topography* 16.4 (June 2004), pp. 269–275. ISSN: 1573-6792. DOI: 10.1023/B:BRAT.

0000032864.93890.f9. URL: https://doi.org/10.1023/B:BRAT.0000032864.93890.f9 (visited on 06/12/2024).

[210] A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks* 13.4 (June 2000), pp. 411–430. ISSN: 0893-6080. DOI: 10.1016/S0893-6080(00)00026-5. URL: https://www.sciencedirect.com/science/article/pii/S0893608000000265 (visited on 06/12/2024).

[211] Anders M Dale et al. Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26.1 (2000), pp. 55–67.

[212] Tijl Grootswagers et al. Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of cognitive neuroscience* 29.4 (2017), pp. 677–697.

[213] Pauli Virtanen et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[214] Fabian Pedregosa et al. Scikit-learn: Machine learning in Python. *the Journal of machine learning research* 12 (2011). Publisher: JMLR. org, pp. 2825–2830.

[215] Wulfram Gerstner et al. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

[216] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press, 2005.

[217] Hugh Reid Wilson. Spikes, decisions, and actions: the dynamical foundations of neuroscience. *(Oxford University Press)* (1999).

[218] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics* 22.1 (1951), pp. 79–86.

[219] Leonid Nisonovich Vaserstein. Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii* 5.3 (1969), pp. 64–72.

[220] Annie G. Bryant and Christopher Whyte. *anniegbryant/MEG_functional_connectivity*. 2025. DOI: 10.5281/zenodo.15122630. URL: https://doi.org/10.5281/zenodo.15122630.

[221] Oliver Ratcliffe et al. Fronto-medial theta coordinates posterior maintenance of working memory content. *Current Biology* 32.10 (2022), pp. 2121–2129.

[222] Catriona L Scrivener et al. Visual imagery of familiar people and places in category selective cortex. *bioRxiv* (2024), pp. 2024–09.

[223] Victor Brumberg. *Essential Relativistic Celestial Mechanics*. CRC Press, 1991.

[224] François Petitjean et al. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition* 44.3 (2011), pp. 678–693.

[225] David Schultz and Brijnesh Jain. Nonsmooth analysis and subgradient methods for averaging in dynamic time warping spaces. *Pattern Recognition* 74 (2018), pp. 340–358.

[226] Iris Deffke et al. MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus. *Neuroimage* 35.4 (2007), pp. 1495–1501.

[227] Nouchine Hadjikhani et al. Early (M170) activation of face-specific cortex by face-like objects. *Neuroreport* 20.4 (2009), pp. 403–407.

[228] Marco Cuturi and Mathieu Blondel. Soft-dtw: a differentiable loss function for time-series. In: *International Conference on Machine Learning*. PMLR. 2017, pp. 894–903.

[229] Niccolò Negro. (Dis) confirming theories of consciousness and their predictions: towards a Lakatosian consciousness science. *Neuroscience of Consciousness* 2024.1 (2024), niae012.

[230] Andrew W Corcoran et al. Accelerating scientific progress through Bayesian adversarial collaboration. *Neuron* 111.22 (2023), pp. 3505–3516.

[231] Peter Godfrey-Smith. Theory and reality: An introduction to the philosophy of science. In: *Theory and reality*. University of Chicago Press, 2009.

[232] David H. Wolpert. The lack of a priori distinctions between learning algorithms. en. *Neural Computation* 8.7 (Oct. 1996), pp. 1341–1390. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/neco.1996.8.7.1341. URL: https://direct.mit.edu/neco/article/8/7/1341-1390/6016 (visited on 06/14/2024).

[233] Alexandre Gramfort. MEG and EEG data analysis with MNE-Python. en. *Frontiers in Neuroscience* 7 (2013). ISSN: 1662453X. DOI: 10.3389/fnins.2013.00267. URL: http://journal.frontiersin.org/article/10.3389/fnins.2013.00267/abstract (visited on 06/14/2024).

[234] Silvana Silva Pereira et al. Effect of field spread on resting-state magneto-encephalography functional network analysis: A computational modeling study. *Brain Connectivity* 7.9 (Nov. 2017), pp. 541–557. ISSN: 2158-0014. DOI: 10.1089/brain.2017.0525. URL: https://www.liebertpub.com/doi/10.1089/brain.2017.0525 (visited on 06/14/2024).

[235] Aurina Arnatkevičiūtė et al. Hub connectivity, neuronal diversity, and gene expression in the Caenorhabditis elegans connectome. en. *PLOS Computational Biology* 14.2 (Feb. 2018). Number: 2 Publisher: Public Library of Science, e1005989. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005989. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005989 (visited on 01/21/2022).

[236] Jorge F Mejias et al. Feedforward and feedback frequency-dependent interactions in a large-scale laminar network of the primate cortex. *Science advances* 2.11 (2016), e1601335.

[237] Leonardo Novelli et al. Spectral dynamic causal modeling: A didactic introduction and its relationship with functional connectivity. *Network Neuroscience* 8.1 (2024), pp. 178–202.

[238] Peter Zeidman et al. A primer on Variational Laplace (VL). *NeuroImage* 279 (2023), p. 120310.

[239] Golia Shafiei et al. Neurophysiological signatures of cortical micro-architecture. *Nature communications* 14.6000 (2023).

[240] Christopher J Whyte et al. On the minimal theory of consciousness implicit in active inference. *arXiv preprint arXiv:2410.06633* (2024).

[241] Christopher J Whyte et al. A Burst-dependent Thalamocortical Substrate for Visual Rivalry. *bioRxiv* (2023), pp. 2023–07.

[242] Ulysse Klatzmann et al. A dynamic bifurcation mechanism explains cortex-wide neural correlates of conscious access. *Cell Reports* 44.3 (2025).

[243] Stanislas Dehaene et al. A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences* 100.14 (2003), pp. 8520–8525.

[244] Stanislas Dehaene et al. A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the national Academy of Sciences* 95.24 (1998), pp. 14529–14534.

[245] Stephen M Fleming. Awareness as inference in a higher-order state space. *Neuroscience of consciousness* 1 (2020).

[246] Emily S Finn et al. Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature neuroscience* 18.11 (2015), pp. 1664–1671.

[247] Irina Rish and Guillermo A Cecchi. Functional network disruptions in schizophrenia. *Biological networks and pathway analysis* (2017), pp. 479–504.

[248] Klaas E Stephan et al. Dysconnection in schizophrenia: From abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia bulletin* 35.3 (2009), pp. 509–527.

[249] João Ricardo Sato et al. Evaluation of pattern recognition and feature extraction methods in ADHD prediction. *Frontiers in systems neuroscience* 6 (2012), p. 68.

[250] Zhiqiang Sha et al. Meta-connectomic analysis reveals commonly disrupted functional architectures in network modules and connectors across brain disorders. *Cerebral cortex* 28.12 (2018), pp. 4179–4194.

[251] Baila S Hall et al. Glucocorticoid mechanisms of functional connectivity changes in stress-related neuropsychiatric disorders. *Neurobiology of stress* 1 (2015), pp. 174–183.

[252] Liangliang Liu et al. An enhanced multi-modal brain graph network for classifying neuropsychiatric disorders. *Medical image analysis* 81 (2022), p. 102550.

[253] Tumbwene E Mwansisya et al. Task and resting-state fMRI studies in first-episode schizophrenia: A systematic review. *Schizophrenia research* 189 (2017), pp. 9–18.

[254] Nicholas H Barbara et al. Classifying Kepler light curves for 12000 A and F stars using supervised feature-based machine learning. *Monthly notices of the Royal Astronomical Society* 514.2 (2022), pp. 2793–2804.

[255] John Fallon et al. Timescales of spontaneous fMRI fluctuations relate to structural connectivity in the brain. *Network neuroscience* 4.3 (Sept. 2020). Number: 3, pp. 788–806. ISSN: 2472-1751. DOI: 10.1162/netn_a_00151. URL: https://doi.org/10.1162/netn_a_00151 (visited on 01/21/2022).

[256] Li An et al. Local synchronization and amplitude of the fluctuation of spontaneous brain activity in attention-deficit/hyperactivity disorder: a resting-state fMRI study. *Neuroscience bulletin* 29 (2013), pp. 603–613.

[257] Alina Surís et al. The evolution of the classification of psychiatric disorders. *Behavioral sciences* 6.1 (2016), p. 5.

[258] Eric Feczko et al. The heterogeneity problem: Approaches to identify psychiatric subtypes. *Trends in cognitive sciences* 23.7 (2019), pp. 584–601.

[259] Darrel A Regier et al. DSM-5 field trials in the United States and Canada, Part II: Test-retest reliability of selected categorical diagnoses. *American journal of psychiatry* 170.1 (2013), pp. 59–70.

[260] Robert Freedman et al. The initial field trials of DSM-5: New blooms and old thorns. *American journal of psychiatry* 170.1 (2013), pp. 1–5.

[261] R.A. Poldrack et al. A phenome-wide examination of neural and cognitive function. en. *Scientific data* 3.1 (Dec. 2016), p. 160110. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.110. URL: http://www.nature.com/articles/sdata2016110 (visited on 10/29/2022).

[262] A. Di Martino et al. The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. eng. *Molecular psychiatry* 19.6 (June 2014), pp. 659–667. ISSN: 1476-5578. DOI: 10.1038/mp.2013.78.

[263] Adriana Di Martino et al. Enhancing studies of the connectome in autism using the Autism Brain Imaging Data Exchange II. eng. *Scientific data* 4 (Mar. 2017), p. 170010. ISSN: 2052-4463. DOI: 10.1038/sdata.2017.10.

[264] Li Zhang et al. A survey on deep learning for neuroimaging-based brain disorder analysis. *Frontiers in neuroscience* 14 (2020), p. 779.

[265] Wutao Yin et al. Deep learning for brain disorder diagnosis based on fMRI images. *Neurocomputing* 469 (2022), pp. 332–345.

[266] Yanchi Chen. Altered functional dynamics gradient in schizophrenia with cigarette smoking. *Cerebral cortex* 33.11 (2023), pp. 7185–7192.

[267] Karsten Specht. Current challenges in translational and clinical fMRI and future directions. *Frontiers in psychiatry* 10 (2020), p. 924.

[268] Peter J Uhlhaas and Wolf Singer. Neuronal dynamics and neuropsychiatric disorders: Toward a translational paradigm for dysfunctional large-scale networks. *Neuron* 75.6 (2012), pp. 963–980.

[269] Annie G. Bryant et al. *DynamicsAndNeuralSystems/fMRI_FeaturesDisorders. https://doi.org/10.5281/zenodo.* Version v1.0.0. 2023. DOI: 10.5281/zenodo.10467891.

[270] Nicolas Traut et al. *IMaging-PsychiAtry Challenge rfMRI data*. URL: https://doi.org/10.5281/zenodo.3625740. Zenodo, Jan. 2020. DOI: 10.5281/zenodo.3625740. URL: https://doi.org/10.5281/zenodo.3625740.

[271] Trent Henderson et al. Never a Dull Moment: Distributional Properties as a Baseline for Time-Series Classification. *arXiv* (2023). DOI: 10.48550/arXiv.2303.17809. arXiv: 2303.17809 [stat.ML]. URL: https://arxiv.org/abs/2303.17809.

[272] Yuan Zhou et al. Spontaneous brain activity observed with functional magnetic resonance imaging as a potential biomarker in neuropsychiatric disorders. *Cognitive neurodynamics* 4 (2010), pp. 275–294.

[273] Kay Henning Brodersen et al. The balanced accuracy and its posterior distribution. In: *2010 20th International Conference on Pattern Recognition*. IEEE. 2010, pp. 3121–3124.

[274] Akhil Kottaram et al. Brain network dynamics in schizophrenia: Reduced dynamism of the default mode network. *Human brain mapping* 40.7 (2019), pp. 2212–2228.

[275] Maria J Rosa et al. Sparse network-based models for patient classification using fMRI. *Neuroimage* 105 (2015), pp. 493–506.

[276] Xiaozhe Wang et al. Structure-based statistical features and multivariate time series clustering. In: *Seventh IEEE international conference on data mining (ICDM 2007)*. IEEE. 2007, pp. 351–360.

[277] Peter Talkner and Rudolf O Weber. Power spectrum and detrended fluctuation analysis: Application to daily temperatures. *Physical Review E* 62.1 (2000), p. 150.

[278] Giulia Baracchini et al. The biological role of local and global fMRI BOLD signal variability in human brain organization. *bioRxiv* (2023). DOI: 10.1101/2023.10.22.563476. eprint: https://www.biorxiv.org/content/early/2023/10/23/2023.10.22.563476.full.pdf.

[279] Tobias Kaufmann et al. Disintegration of sensorimotor brain networks in schizophrenia. *Schizophrenia bulletin* 41.6 (2015), pp. 1326–1335.

[280] Wanqing Xie et al. Functional brain lateralization in schizophrenia based on the variability of resting-state fMRI signal. *Progress in neuro-psychopharmacology and biological psychiatry* 86 (2018), pp. 114–121.

[281] Amanda K Easson and Anthony R McIntosh. BOLD signal variability and complexity in children and adolescents with and without autism spectrum disorder. *Developmental cognitive neuroscience* 36 (2019), p. 100630.

[282] Jonathan D Power et al. Sources and implications of whole-brain fMRI signals in humans. *Neuroimage* 146 (2017), pp. 609–625.

[283] Jintao Sheng et al. The coupling of BOLD signal variability and degree centrality underlies cognitive functions and psychiatric diseases. *Neuroimage* 237 (2021), p. 118187.

[284] Hamidreza Jamalabadi et al. Classification based hypothesis testing in neuroscience: Below-chance level classification rates and overlooked statistical properties of linear parametric classifiers. *Human brain mapping* 37.5 (2016), pp. 1842–1855.

[285] Daniel Berrar. Cross-Validation. In: *Encyclopedia of Bioinformatics and Computational Biology*. Ed. by Shoba Ranganathan et al. Oxford: Academic Press, 2019, pp. 542–545. ISBN: 978-0-12-811432-2. DOI: https://doi.org/10.1016/B978-0-12-809633-8.20349-X. URL: https://www.sciencedirect.com/science/article/pii/B978012809633820349X.

[286] Ziad S Saad et al. Correcting brain-wide correlation differences in resting-state FMRI. *Brain connectivity* 3.4 (2013), pp. 339–352.

[287] Michael Greicius. Resting-state functional connectivity in neuropsychiatric disorders. *Current opinion in neurology* 21.4 (2008), pp. 424–430.

[288] T Takamura and T Hanakawa. Clinical utility of resting-state functional connectivity magnetic resonance imaging for mood and cognitive disorders. *Journal of neural transmission* 124 (2017), pp. 821–839.

[289] Fumitada Itakura. Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on acoustics, speech, and signal processing* 23.1 (1975), pp. 67–72.

[290] James Massey. Causality, feedback and directed information. In: *Proc. Int. Symp. Inf. Theory Applic. (ISITA-90)*. 1990, pp. 303–305.

[291] Trent Henderson. *hendersontrent/correctR: v0.1.2*. Version v0.1.2. URL: https://doi.org/10.5281/zenodo.7450146. Dec. 2022. DOI: 10.5281/zenodo.7450146.

[292] Patrik Hoyer et al. Nonlinear causal discovery with additive noise models. *Advances in Neural Information Processing Systems* 21 (2008), pp. 689–696.

[293] Akhil Kottaram et al. Spatio-temporal dynamics of resting-state brain networks improve single-subject prediction of schizophrenia diagnosis. en (2018). Publisher: John Wiley & Sons, Inc. DOI: 10.1002/hbm.24202. URL: https://acuresearchbank.acu.edu.au/item/8w098/spatio-temporal-dynamics-of-resting-state-brain-networks-improve-single-subject-prediction-of-schizophrenia-diagnosis (visited on 02/18/2023).

[294] Guang-Zhong Wang et al. Correspondence between resting-state activity and brain gene expression. *Neuron* 88.4 (2015), pp. 659–666.

[295] Matthew J Hoptman et al. Amplitude of low-frequency oscillations in schizophrenia: A resting state fMRI study. *Schizophrenia research* 117.1 (2010), pp. 13–20.

[296] Pengshuo Wang et al. Amplitude of low-frequency fluctuation (ALFF) may be associated with cognitive impairment in schizophrenia: A correlation study. *BMC psychiatry* 19.1 (2019), pp. 1–10.

[297] Yanqing Tang et al. Neural activity changes in unaffected children of patients with schizophrenia: A resting-state fMRI study. *Schizophrenia research* 168.1-2 (2015), pp. 360–365.

[298] Min-Jie Liang et al. Identify changes of brain regional homogeneity in bipolar disorder and unipolar depression using resting-state fMRI. *PLoS one* 8.12 (2013), e79999.

[299] Michael A Cerullo et al. The functional neuroanatomy of bipolar disorder. *International review of psychiatry* 21.4 (2009), pp. 314–322.

[300] Anton Lord et al. Brain parcellation choice affects disease-related topology differences increasingly from global to local network levels. *Psychiatry Research: Neuroimaging* 249 (2016), pp. 12–19.

[301] Kenneth Wengler et al. Distinct hierarchical alterations of intrinsic neural timescales account for different manifestations of psychosis. *eLife* 9 (2020), e56151.

[302] Lavinia Carmen Uscătescu et al. Intrinsic neural timescales in autism spectrum disorder and schizophrenia. A replication and direct comparison study. *Schizophrenia* 9.1 (2023), p. 18.

[303] Takamitsu Watanabe et al. Atypical intrinsic neural timescale in autism. *eLife* 8 (2019), e42256.

[304] Yuki Hosaka et al. Linear models replicate the energy landscape and dynamics of resting-state brain activity. *bioRxiv* (2024), pp. 2024–05.

[305] Maren Wehrheim et al. Reliability of Variability and Complexity Measures for Task and Task-Free BOLD fMRI (2023).

[306] Douglas D Garrett et al. Blood oxygen level-dependent signal variability is more than just noise. *Journal of neuroscience* 30.14 (2010), pp. 4914–4921.

[307] Douglas D Garrett et al. Age differences in brain signal variability are robust to multiple vascular controls. *Scientific reports* 7.1 (2017), pp. 1–13.

[308] Vanessa Scarapicchia et al. Resting state BOLD variability in Alzheimer's disease: A marker of cognitive decline or cerebrovascular status? *Frontiers in aging neuroscience* 10 (2018), p. 39.

[309] Douglas D Garrett et al. Local temporal variability reflects functional integration in the human brain. *Neuroimage* 183 (2018), pp. 776–787.

[310] Danielle S Bassett et al. Understanding the emergence of neuropsychiatric disorders with network neuroscience. *Biological psychiatry: Cognitive neuroscience and neuroimaging* 3.9 (2018), pp. 742–753.

[311] Gordon Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory* 14.1 (1968), pp. 55–63.

[312] Gerard V Trunk. A problem of dimensionality: A simple example. *IEEE transactions on pattern analysis and machine intelligence* 3 (1979), pp. 306–307.

[313] Sarab S. Sethi et al. Structural connectome topology relates to regional BOLD signal dynamics in the mouse brain. *Chaos: An interdisciplinary journal of nonlinear science* 27.4 (Apr. 2017). Number: 4 Publisher: American Institute of Physics, p. 047405. ISSN: 1054-1500. DOI: 10.1063/1.4979281. URL: https://aip.scitation.org/doi/abs/10.1063/1.4979281 (visited on 01/21/2022).

[314] Gerhard Kramer. *PhD Dissertation: Directed information for channels with feedback*. Citeseer, 1998.

[315] Giulio Tononi. An information integration theory of consciousness. *BMC neuroscience* 5 (2004), pp. 1–22.

[316] Bin He and Zhongming Liu. Multimodal functional neuroimaging: Integrating functional MRI and EEG/MEG. *IEEE reviews in biomedical engineering* 1 (2008), pp. 23–40.

[317] Sean Tanabe et al. Cohort study of electroencephalography markers of amyloid-tau-neurodegeneration pathology. en. *Brain Communications* 2.2 (July 2020). Number: 2, fcaa099. ISSN: 2632-1297. DOI: 10.1093/braincomms/fcaa099. URL: https://academic.oup.com/braincomms/article/doi/10.1093/braincomms/fcaa099/5871920 (visited on 08/05/2022).

[318] Ben Goertzel et al. Glocal memory: A critical design principle for artificial brains and minds. *Neurocomputing* 74.1-3 (2010), pp. 84–94.

[319] Alfredo A Pulini et al. Classification accuracy of neuroimaging biomarkers in attention-deficit/hyperactivity disorder: Effects of sample size and circular analysis. *Biological psychiatry: Cognitive neuroscience and neuroimaging* 4.2 (2019), pp. 108–120.

[320] Andrius Vabalas et al. Machine learning algorithm validation with a limited sample size. *PLoS one* 14.11 (2019), e0224365.

[321] Rajamannar Ramasubbu et al. Accuracy of automated classification of major depressive disorder as a function of symptom severity. *Neuroimage: Clinical* 12 (2016), pp. 320–331.

[322] Lourdes García Murillo et al. Locomotor activity measures in the diagnosis of attention deficit hyperactivity disorder: Meta-analyses and new findings. *Journal of neuroscience methods* 252 (2015), pp. 14–26.

[323] Heath R. Pardoe et al. Motion and morphometry in clinical and nonclinical populations. en. *Neuroimage* 135 (July 2016), pp. 177–185. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2016.05.005. URL: https://www.sciencedirect.com/science/article/pii/S1053811916301197 (visited on 02/18/2023).

[324] Xiang-zhen Kong et al. Individual differences in impulsivity predict head motion during magnetic resonance imaging. *PloS one* 9.8 (2014), e104989.

[325] Baptiste Couvy-Duchesne et al. Head motion and inattention/hyperactivity share common genetic influences: implications for fMRI studies of ADHD. *PloS one* 11.1 (2016), e0146271.

[326] Phoebe Thomson et al. Head motion during MRI predicted by out-of-scanner sustained attention performance in attention-deficit/hyperactivity disorder. *Journal of Attention Disorders* 25.10 (2021), pp. 1429–1440.

[327] Yuta Aoki et al. Research Review: Diffusion tensor imaging studies of attention-deficit/hyperactivity disorder: meta-analyses and reflections on head motion. *Journal of Child Psychology and Psychiatry* 59.3 (2018), pp. 193–202.

[328] Jared Nielsen et al. Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in human neuroscience* 7 (2013). ISSN: 1662-5161. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2013.00599 (visited on 11/19/2022).

[329] Xin Yang et al. A Deep Neural Network Study of the ABIDE Repository on Autism Spectrum Classification. *International journal of advanced computer science and applications* 11 (Jan. 2020). DOI: 10.14569/IJACSA.2020.0110401.

[330] Ahmed El-Gazzar et al. A Hybrid 3DCNN and 3DC-LSTM Based Model for 4D Spatio-Temporal fMRI Data: An ABIDE Autism Classification Study. en. In: *OR 2.0 Context-Aware Operating Theaters and Machine Learning in Clinical Neuroimaging*. Ed. by Luping Zhou et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 95–102. ISBN: 978-3-030-32695-1. DOI: 10.1007/978-3-030-32695-1_11.

[331] Marcel Adam Just et al. Identifying Autism from Neural Representations of Social Interactions: Neurocognitive Markers of Autism. en. *PLoS one* 9.12 (Dec. 2014). Publisher: Public Library of Science, e113879. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0113879. (Visited on 11/19/2022).

[332] Xin Yang et al. Functional connectivity magnetic resonance imaging classification of autism spectrum disorder using the multisite ABIDE dataset. In: *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. ISSN: 2641-3604. May 2019, pp. 1–4. DOI: 10.1109/BHI.2019.8834653.

[333] Nicolas Traut et al. Insights from an autism imaging biomarker challenge: Promises and threats to biomarker discovery. en. *Neuroimage* 255 (July 2022), p. 119171. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2022.119171. URL: https://www.sciencedirect.com/science/article/pii/S1053811922002981 (visited on 07/31/2022).

[334] Okito Yamashita et al. Computational Mechanisms of Neuroimaging Biomarkers Uncovered by Multicenter Resting-State fMRI Connectivity Variation Profile. *bioRxiv* (2024), pp. 2024–04.

[335] Madhura Ingalhalikar et al. Functional Connectivity-Based Prediction of Autism on Site Harmonized ABIDE Dataset. *IEEE transactions on biomedical engineering* 68.12 (Dec. 2021). Conference Name: IEEE Transactions on Biomedical Engineering, pp. 3628–3637. ISSN: 1558-2531. DOI: 10.1109/TBME.2021.3080259.

[336] Oswaldo Artiles et al. Confounding Effects on the Performance of Machine Learning Analysis of Static Functional Connectivity Computed from rs-fMRI Multi-site Data. *Neuroinformatics* (2023), pp. 1–18.

[337] Qingqing Li et al. Atypical hemispheric lateralization of brain function and structure in autism: a comprehensive meta-analysis study. *Psychological medicine* (2023), pp. 1–12.

[338] Bin Wan et al. Diverging asymmetry of intrinsic functional organization in autism. *bioRxiv* (2023), pp. 2023–04.

[339] Amanda M Buch et al. Molecular and network-level mechanisms explaining individual differences in autism spectrum disorder. *Nature neuroscience* 26.4 (2023), pp. 650–663.

[340] Maya A Reiter et al. Performance of machine learning classification models of autism using resting-state fMRI is contingent on sample heterogeneity. *Neural computing and applications* 33 (2021), pp. 3299–3310.

[341] Ashlea Segal et al. Regional, circuit and network heterogeneity of brain abnormalities in psychiatric disorders. *Nature Neuroscience* (2023), pp. 1–17.

[342] Takashi Itahashi et al. Generalizable and transportable resting-state neural signatures characterized by functional networks, neurotransmitters, and clinical symptoms in autism. *Molecular Psychiatry* (2024), pp. 1–13.

[343] Iain M Johnstone and D Michael Titterington. Statistical challenges of high-dimensional data. *Philosophical transactions of the Royal Society A: Mathematical, physical and engineering sciences* 367.1906 (2009), pp. 4237–4253.

[344] Kirsten L Peterson et al. Regularized partial correlation provides reliable functional connectivity estimates while correcting for widespread confounding. *bioRxiv* (2023), pp. 2023–09.

[345] Bo Wang et al. Similarity network fusion for aggregating data types on a genomic scale. *Nature methods* 11.3 (2014), pp. 333–337.

[346] Felix Brandl et al. Common and specific large-scale brain changes in major depressive disorder, anxiety disorders, and chronic pain: a transdiagnostic multimodal meta-analysis of structural and functional MRI studies. *Neuropsychopharmacology* 47.5 (2022), pp. 1071–1080.

[347] Yuelu Liu et al. Highly predictive transdiagnostic features shared across schizophrenia, bipolar disorder, and adhd identified using a machine learning based approach. *bioRxiv* (2018), p. 453951.

[348] Yiben Xu et al. Interacting spiral wave patterns underlie complex brain dynamics and are related to cognitive processing. *Nature human behaviour* (2023), pp. 1–20.

[349] Rory G Townsend and Pulin Gong. Detection and analysis of spatiotemporal patterns in brain activity. *PLoS computational biology* 14.12 (2018), e1006643.

[350] Joana Cabral et al. Role of local network oscillations in resting-state functional connectivity. *Neuroimage* 57.1 (2011), pp. 130–139.

[351] Anthony Bagnall et al. The Great Time Series Classification Bake Off: A Review and Experimental Evaluation of Recent Algorithmic Advances. *Data mining and knowledge discovery* 31 (2017), p. 606.

[352] Javier Arnedo et al. Uncovering the hidden risk architecture of the schizophrenias: Confirmation in three independent genome-wide association studies. *American journal of psychiatry* 172.2 (2015), pp. 139–153.

[353] Maya Bleich-Cohen et al. Machine learning fMRI classifier delineates subgroups of schizophrenia patients. en. *Schizophrenia research* 160.1-3 (Dec. 2014). Number: 1-3, pp. 196–200. ISSN: 09209964. DOI: 10.1016/j.schres.2014.10.033. URL: https://linkinghub.elsevier.com/retrieve/pii/S0920996414006021.

[354] Lianjie Niu et al. Resolving heterogeneity in schizophrenia, bipolar I disorder, and attention-deficit/hyperactivity disorder through individualized structural covariance network analysis. *Cerebral Cortex* (2023), bhad391.

[355] Oscar Esteban et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. en. *Nature Methods* 16.1 (Jan. 2019). Number: 1 Publisher: Nature Publishing Group, pp. 111–116. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0235-4. URL: https://www.nature.com/articles/s41592-018-0235-4 (visited on 11/18/2022).

[356] Raimon H. R. Pruim et al. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. eng. *Neuroimage* 112 (May 2015), pp. 267–277. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2015.02.064.

[357] Kevin M. Aquino et al. Identifying and removing widespread signal deflections from fMRI data: Rethinking the global signal regression problem. en. *Neuroimage* 212 (May 2020), p. 116614. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2020.116614. URL: https://www.sciencedirect.com/science/article/pii/S1053811920301014.

[358] Linden Parkes et al. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. en. *Neuroimage* 171 (May 2018), pp. 415–436. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2017.12.073. URL: https://www.sciencedirect.com/science/article/pii/S1053811917310972.

[359] Kevin M. Aquino et al. On the intersection between data quality and dynamical modelling of large-scale fMRI signals. en. *Neuroimage* 256 (Mar. 2022), p. 119051. DOI: 10.1016/j.neuroimage.2022.119051.

[360] Bruce Fischl et al. Whole Brain Segmentation. en. *Neuron* 33.3 (Jan. 2002), pp. 341–355. ISSN: 08966273. DOI: 10.1016/S0896-6273(02)00569-X. URL: https://linkinghub.elsevier.com/retrieve/pii/S089662730200569X (visited on 11/18/2022).

[361] Mark Jenkinson et al. FSL. *Neuroimage* 62.2 (2012), pp. 782–790.

[362] Jonathan D. Power et al. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. en. *Neuroimage* 59.3 (Feb. 2012), pp. 2142–2154. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2011.10.018. URL: https://www.sciencedirect.com/science/article/pii/S1053811911011815 (visited on 12/03/2022).

[363] Xinyu Liu et al. Disentangling age-and disease-related alterations in schizophrenia brain network using structural equation modeling: A graph theoretical study based on minimum spanning tree. *Human brain mapping* 42.10 (2021), pp. 3023–3041.

[364] Willem Huijbers et al. Less head motion during MRI under task than resting-state conditions. *Neuroimage* 147 (2017), pp. 111–120.

[365] Emmanuel Peng Kiat Pua et al. Individualised MRI training for paediatric neuroimaging: A child-focused approach. *Developmental cognitive neuroscience* 41 (2020), p. 100750.

[366] Jessica A Turner et al. A multi-site resting state fMRI study on the amplitude of low frequency fluctuations in schizophrenia. *Frontiers in neuroscience* 7 (2013), p. 137.

[367] Ali M Golestani et al. The effect of low-frequency physiological correction on the reproducibility and specificity of resting-state fMRI metrics: Functional connectivity, ALFF, and ReHo. *Frontiers in neuroscience* 11 (2017), p. 546.

[368] Trent Henderson and Ben D. Fulcher. *Feature-based time-series analysis in R using the theft package*. Aug. 2022. DOI: 10.48550/arXiv.2208.06146. URL: http://arxiv.org/abs/2208.06146 (visited on 09/04/2022).

[369] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning* 20 (1995), pp. 273–297.

[370] Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[371] Stephen R Cole and Miguel A Hernán. Constructing inverse probability weights for marginal structural models. *American journal of epidemiology* 168.6 (2008), pp. 656–664.

[372] Philipp Thölke et al. Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *Neuroimage* 277 (2023), p. 120253.

[373] Nicole Bussola et al. AI slipping on tiles: Data leakage in digital pathology. In: *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part I*. Springer. 2021, pp. 167–182.

[374] Gaël Varoquaux et al. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *Neuroimage* 145 (2017), pp. 166–179.

[375] Jingwei Li et al. Global signal regression strengthens association between resting-state functional connectivity and behavior. *Neuroimage* 196 (2019), pp. 126–141.

[376] Bart Larsen et al. A developmental reduction of the excitation:inhibition ratio in association cortex during adolescence. *Science advances* 8.5 (2022).

[377] Robert Tibshirani et al. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67.1 (2005), pp. 91–108.

[378] F. Pedregosa et al. Scikit-learn: Machine Learning in Python. *Journal of machine learning research* 12 (2011), pp. 2825–2830.

[379] Sébastien Lê et al. FactoMineR: an R package for multivariate analysis. *Journal of statistical software* 25 (2008), pp. 1–18.

[380] Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1 (1995). Publisher: [Royal Statistical Society, Wiley], pp. 289–300. ISSN: 0035-9246. URL: https://www.jstor.org/stable/2346101 (visited on 11/19/2022).

[381] Michael P Fay and Michael A Proschan. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics surveys* 4 (2010), p. 1.

[382] Remco R Bouckaert and Eibe Frank. Evaluating the replicability of significance tests for comparing learning algorithms. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer. 2004, pp. 3–12.

[383] C Nadeau and Y Bengio. Inference for the generalization error machine learning. *Machine learning* 52 (2003), pp. 239–281.

[384] Arno Klein and Jason Tourville. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in neuroscience* 6 (2012), p. 171.

[385] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

[386] Athanasia M Mowinckel and Didac Vidal-Piñeiro. Visualization of brain statistics with R packages ggseg and ggseg3d. *Advances in Methods and Practices in Psychological Science* 3.4 (2020), pp. 466–483.

[387] Athanasia M. Mowinckel and Didac Vidal-Piñeiro. Visualisation of Brain Statistics with R-packages ggseg and ggseg3d (2019). arXiv: 1912.08200 [stat.OT]

[388] Daniel Lüdecke et al. see: An R Package for Visualizing Statistical Models. *Journal of Open Source Software* 6.64 (2021), p. 3393. DOI: 10.21105/joss.03393.

[389] Zuguang Gu et al. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32.18 (2016), pp. 2847–2849.

[390] Zuguang Gu. Complex heatmap visualization. *Imeta* 1.3 (2022), e43.

[391] Hanbo Chen and Paul C Boutros. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC bioinformatics* 12 (2011), pp. 1–7.

[392] Annie G. Bryant et al. *All data to support this manuscript. https://zenodo.org/doi/10.5281/zenodo.10431854*. Version v1. 2023. DOI: 10.5281/zenodo.10431854. URL: https://doi.org/10.5281/zenodo.10431854.

[393] Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* 58.301 (1963), pp. 236–244.

[394] David E Stark et al. Regional variation in interhemispheric coordination of intrinsic hemodynamic fluctuations. *Journal of Neuroscience* 28.51 (2008), pp. 13754–13764.

[395] Jeroen Mollink et al. The spatial correspondence and genetic influence of interhemispheric connectivity with white matter microstructure. *Nature neuroscience* 22.5 (2019), pp. 809–819.

[396] Xuhong Liao et al. Spontaneous functional network dynamics and associated structural substrates in the human brain. *Frontiers in Human Neuroscience* 9 (2015), p. 478.

[397] Elliot A Layden et al. Interhemispheric functional connectivity in the zebra finch brain, absent the corpus callosum in normal ontogeny. *NeuroImage* 195 (2019), pp. 113–127.

[398] MA Urbin et al. Resting-state functional connectivity and its association with multiple domains of upper-extremity function in chronic stroke. *Neurorehabilitation and neural repair* 28.8 (2014), pp. 761–769.

[399] Moriah E Thomason et al. Age-related increases in long-range connectivity in fetal functional neural connectivity networks in utero. *Developmental cognitive neuroscience* 11 (2015), pp. 96–104.

[400] Willem M Otte et al. Altered contralateral sensorimotor system organization after experimental hemispherectomy: A structural and functional connectivity study. *Journal of Cerebral Blood Flow & Metabolism* 35.8 (2015), pp. 1358–1367.

[401] L Federico Rossi et al. Focal cortical seizures start as standing waves and propagate respecting homotopic connectivity. *Nature communications* 8.1 (2017), p. 217.

[402] Adam Q Bauer et al. Effective connectivity measured using optogenetically evoked hemodynamic signals exhibits topography distinct from resting state functional connectivity in the mouse. *Cerebral Cortex* 28.1 (2018), pp. 370–386.

[403] Teppei Matsui et al. Direct comparison of spontaneous functional connectivity and effective connectivity measured by intracortical microstimulation: An fMRI study in macaque monkeys. *Cerebral Cortex* 21.10 (2011), pp. 2348–2356.

[404] Kelly Shen et al. Stable long-range interhemispheric coordination is supported by direct anatomical projections. *Proceedings of the National Academy of Sciences* 112.20 (2015), pp. 6473–6478.

[405] Javier Gonzalez-Castillo et al. The spatial structure of resting state connectivity stability on the scale of minutes. *Frontiers in neuroscience* 8 (2014), p. 138.

[406] Richard F Betzel et al. Generative models of the human connectome. *Neuroimage* 124 (2016), pp. 1054–1064.

[407] Nathalie Tzourio-Mazoyer et al. Variation in homotopic areas' activity and inter-hemispheric intrinsic connectivity with type of language lateralization: An fMRI study of covert sentence generation in 297 healthy volunteers. *Brain Structure and Function* 221 (2016), pp. 2735–2753.

[408] Kevin Jarbo et al. In vivo quantification of global connectivity in the human corpus callosum. *Neuroimage* 59.3 (2012), pp. 1988–1996.

[409] Raymond Salvador et al. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral cortex* 15.9 (2005), pp. 1332–1342.

[410] Sean M Tobyne et al. *A surface-based technique for mapping homotopic interhemispheric connectivity: Development, characterization, and clinical application*. Tech. rep. Wiley Online Library, 2016.

[411] Lucina Q Uddin et al. Residual functional connectivity in the split-brain revealed with resting-state functional MRI. *Neuroreport* 19.7 (2008), pp. 703–709.

[412] Gabrielle Di Virgilio et al. Cortical regions contributing to the anterior commissure in man. *Experimental Brain Research* 124 (1999), pp. 1–7.

[413] Jill X O'reilly et al. Causal effect of disconnection lesions on interhemispheric functional connectivity in rhesus monkeys. *Proceedings of the National Academy of Sciences* 110.34 (2013), pp. 13982–13987.

[414] Daniel S Margulies et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences* 113.44 (2016), pp. 12574–12579.

[415] Julia M Huntenburg et al. Large-scale gradients in human cortical organization. *Trends in cognitive sciences* 22.1 (2018), pp. 21–31.

[416] Claus C Hilgetag and Alexandros Goulas. 'Hierarchy' in the organization of brain networks. *Philosophical Transactions of the Royal Society B* 375.1796 (2020), p. 20190319.

[417] Kathleen S Rockland and Deepak N Pandya. Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain research* 179.1 (1979), pp. 3–20.

[418] Daniel J Felleman and David C Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)* 1.1 (1991), pp. 1–47.

[419] David Mumford. On the computational architecture of the neocortex: The role of cortico-cortical loops. *Biological cybernetics* 66.3 (1992), pp. 241–251.

[420] Rajesh PN Rao and Dana H Ballard. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience* 2.1 (1999), pp. 79–87.

[421] Xinhu Jin et al. The relationship between interhemispheric homotopic functional connectivity and left-right difference of intrahemispheric functional integration in the human brain. *Imaging Neuroscience* 2 (2024), pp. 1–18.

[422] Li-Zhen Chen and Xi-Nian Zuo. Human Brain Mapping of Homotopic Functional Affinity. *bioRxiv* (2024), pp. 2024–01.

[423] Aysenil Belger and Marie T Banich. Interhemispheric interaction affected by computational complexity. *Neuropsychologia* 30.10 (1992), pp. 923–929.

[424] Michelle Hampson et al. Brain connectivity related to working memory performance. *Journal of Neuroscience* 26.51 (2006), pp. 13338–13343.

[425] Qing Chen et al. Abnormal voxel-mirrored homotopic connectivity in first-episode major depressive disorder using fMRI: A machine learning approach. *Frontiers in Psychiatry* 14 (2023), p. 1241670.

[426] Maayan Harel et al. Distinct homotopic functional connectivity patterns of the amygdalar sub-regions as biomarkers in major depressive disorder. *Journal of Affective Disorders* 365 (2024), pp. 285–292.

[427] Hong Yang et al. Aberrant interhemispheric functional connectivity in first-episode, drug-naive major depressive disorder. *Brain imaging and behavior* 13 (2019), pp. 1302–1310.

[428] Jeffrey S Anderson et al. Decreased interhemispheric functional connectivity in autism. *Cerebral cortex* 21.5 (2011), pp. 1134–1146.

[429] Adriana Di Martino et al. The Autism Brain Imaging Data Exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry* 19.6 (2014), pp. 659–667.

[430] Shuxia Yao et al. Decreased homotopic interhemispheric functional connectivity in children with autism spectrum disorder. *Autism Research* 14.8 (2021), pp. 1609–1620.

[431] Ying Wang et al. Classification of unmedicated bipolar disorder using whole-brain functional activity and connectivity: A radiomics analysis. *Cerebral Cortex* 30.3 (2020), pp. 1117–1128.

[432] Hui-Jie Li et al. Homotopic connectivity in drug-naïve, first-episode, early-onset schizophrenia. *Journal of Child Psychology and Psychiatry* 56.4 (2015), pp. 432–443.

[433] Hengyu Zhang et al. Association between homotopic connectivity and clinical symptoms in first-episode schizophrenia. *Heliyon* 10.9 (2024).

[434] Mengjing Cai et al. Homotopic functional connectivity disruptions in schizophrenia and their associated gene expression. *Neuroimage* 289 (2024), p. 120551.

[435] Sara Larivière et al. The ENIGMA Toolbox: Multiscale neural contextualization of multisite neuroimaging datasets. *Nature Methods* 18.7 (2021), pp. 698–700.

[436] Sara Larivière et al. Network-based atrophy modeling in the common epilepsies: A worldwide ENIGMA study. *Science advances* 6.47 (2020), eabc6457.

[437] Ronald A Fisher. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10.4 (1915), pp. 507–521.

[438] Biao Jie et al. Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification. *Human brain mapping* 35.7 (2014), pp. 2876–2897.

[439] Sebastian Markett et al. Assessing the function of the fronto-parietal attention network: Insights from resting-state fMRI and the attentional network test. *Human brain mapping* 35.4 (2014), pp. 1700–1709.

[440] Ross D Markello et al. Neuromaps: structural and functional interpretation of brain maps. *Nature Methods* 19.11 (2022), pp. 1472–1479.

[441] Michael J Hawrylycz et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489.7416 (2012), pp. 391–399.

[442] Aaron F Alexander-Bloch et al. On testing for spatial correspondence between maps of human brain structure and function. *Neuroimage* 178 (2018), pp. 540–551.

[443] Helen Barbas and Nancy Rempel-Clower. Cortical structure predicts the pattern of corticocortical connections. *Cerebral cortex (New York, NY: 1991)* 7.7 (1997), pp. 635–646.

[444] Nikola T Markov et al. Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of comparative neurology* 522.1 (2014), pp. 225–259.

[445] Matthew F Glasser and David C Van Essen. Mapping human cortical areas in vivo based on myelin content as revealed by T1-and T2-weighted MRI. *Journal of neuroscience* 31.32 (2011), pp. 11597–11616.

[446] Joshua B Burt et al. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nature neuroscience* 21.9 (2018), pp. 1251–1259.

[447] Justine Y Hansen et al. Mapping gene transcription and neurocognition across human neocortex. *Nature human behaviour* 5.9 (2021), pp. 1240–1250.

[448] Ben D Fulcher et al. Multimodal gradients across mouse cortex. *Proceedings of the National Academy of Sciences* 116.10 (2019), pp. 4689–4695.

[449] Joanna E Moodie et al. General and specific patterns of cortical gene expression as spatial correlates of complex cognitive functioning. *Human Brain Mapping* 45.4 (2024), e26641.

[450] Shuxia Yao and Keith M Kendrick. Reduced homotopic interhemispheric connectivity in psychiatric disorders: Evidence for both transdiagnostic and disorder specific features. *Psychoradiology* 2.4 (2022), pp. 129–145.

[451] Mária Ercsey-Ravasz et al. A predictive network model of cerebral cortical connectivity based on a distance rule. *Neuron* 80.1 (2013), pp. 184–197.

[452] Szabolcs Horvát et al. Spatial embedding and wiring cost constrain the functional layout of the cortical network of rodents and primates. *PLoS biology* 14.7 (2016), e1002512.

[453] James A Henderson and Peter A Robinson. Relations between the geometry of cortical gyrification and white-matter network architecture. *Brain connectivity* 4.2 (2014), pp. 112–130.

[454] James A Roberts et al. The contribution of geometry to the human connectome. *Neuroimage* 124 (2016), pp. 379–393.

[455] Pierre Bellec et al. Identification of large-scale networks in the brain using fMRI. *Neuroimage* 29.4 (2006), pp. 1231–1243.

[456] Aaron F Alexander-Bloch et al. The anatomical distance of functional connections predicts brain network topology in health and schizophrenia. *Cerebral cortex* 23.1 (2013), pp. 127–138.

[457] Vincent Bazinet et al. Multiscale communication in cortico-cortical networks. *NeuroImage* 243 (2021), p. 118546.

[458] Massimiliano Facca et al. Multiscale and multimodal signatures of structure-function coupling variability across the human neocortex. *NeuroImage* 302 (2024), p. 120902.

[459] Xinhu Jin et al. Functional integration between the two brain hemispheres: Evidence from the homotopic functional connectivity under resting state. *Frontiers in neuroscience* 14 (2020), p. 932.

[460] Michaël Bernier et al. The morphology of the human cerebrovascular system. *Human brain mapping* 39.12 (2018), pp. 4962–4975.

[461] Richard F Betzel et al. Distance-dependent consensus thresholds for generating group-representative structural brain networks. *Network neuroscience* 3.2 (2019), pp. 475–496.

[462] Patrick J Drew. Vascular and neural basis of the BOLD signal. *Current opinion in neurobiology* 58 (2019), pp. 61–69.

[463] Celine Mateo et al. Entrainment of arteriole vasomotor fluctuations by neural activity is a basis of blood-oxygenation-level-dependent "resting-state" connectivity. *Neuron* 96.4 (2017), pp. 936–948.

[464] Patrick J Drew et al. Ultra-slow oscillations in fMRI and resting-state connectivity: Neuronal and vascular contributions and technical confounds. *Neuron* 107.5 (2020), pp. 782–804.

[465] Sebastian C Schneider et al. Resting-state BOLD functional connectivity depends on the heterogeneity of capillary transit times in the human brain: A combined lesion and simulation study about the influence of blood flow response timing. *Neuroimage* 255 (2022), p. 119208.

[466] Nastaran Schwarz et al. Correlates of asymmetric venous drainage in resting state functional magnetic resonance imaging data. *medRxiv* (2025), pp. 2025–01.

[467] Mahmood Mirza et al. Variability in intracranial vessel diameters and considerations for neurovascular models: A systematic review and meta-analysis. *Stroke: Vascular and Interventional Neurology* 4.4 (2024), e001177.

[468] William H Eskew et al. The venous circle of Trolard: An anatomical study with application to approaches to the basal brain. *World Neurosurgery* 175 (2023), e238–e242.

[469] Viktor Weiss et al. Distance from main arteries influences microstructural and functional brain tissue characteristics. *NeuroImage* 285 (2024), p. 120502.

[470] Michael W Miller and Brent A Vogt. Heterotopic and homotopic callosal connections in rat visual cortex. *Brain Research* 297.1 (1984), pp. 75–89.

[471] J Michael Tyszka et al. Intact bilateral resting-state networks in the absence of the corpus callosum. *Journal of Neuroscience* 31.42 (2011), pp. 15154–15162.

[472] Abraham Z Snyder and Adam Q Bauer. Mapping structure-function relationships in the brain. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 4.6 (2019), pp. 510–521.

[473] Xi-Han Zhang et al. The cell-type underpinnings of the human functional cortical connectome. *Nature Neuroscience* 28.1 (2025), pp. 150–160.

[474] Arielle Baker et al. Specialized subpopulations of deep-layer pyramidal neurons in the neocortex: Bridging cellular properties to functional consequences. *Journal of Neuroscience* 38.24 (2018), pp. 5441–5455.

[475] Zoltán Molnár and Amanda FP Cheung. Towards the classification of subpopulations of layer V pyramidal projection neurons. *Neuroscience research* 55.2 (2006), pp. 105–115.

[476] Xiaoyi Mao and Jochen F Staiger. Multimodal cortical neuronal cell type classification. *Pflügers Archiv-European Journal of Physiology* 476.5 (2024), pp. 721–733.

[477] Elizabeth A Alcamo et al. Satb2 regulates callosal projection neuron identity in the developing cerebral cortex. *Neuron* 57.3 (2008), pp. 364–377.

[478] Dino P Leone et al. Satb2 regulates the differentiation of both callosal and subcerebral projection neurons in the developing cerebral cortex. *Cerebral cortex* 25.10 (2015), pp. 3406–3419.

[479] Qiufang Guo et al. In the developing cerebral cortex: Axonogenesis, synapse formation, and synaptic plasticity are regulated by SATB2 target genes. *Pediatric Research* 93.6 (2023), pp. 1519–1527.

[480] Paola Arlotta et al. Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. *Neuron* 45.2 (2005), pp. 207–221.

[481] Yasuhiro Itoh et al. Bcl11b orchestrates subcerebral projection neuron axon development via cell-autonomous, non-cell-autonomous, and subcellular mechanisms. *bioRxiv* (2024), pp. 2024–10.

[482] Tom Lickiss et al. Examining the relationship between early axon growth and transcription factor expression in the developing cerebral cortex. *Journal of anatomy* 220.3 (2012), pp. 201–211.

[483] Swathi Srivatsa et al. Unc5C and DCC act downstream of Ctip2 and Satb2 and contribute to corpus callosum formation. *Nature communications* 5.1 (2014), p. 3708.

[484] Han Zhao et al. Genetic mechanisms underlying brain functional homotopy: A combined transcriptome and resting-state functional MRI study. *Cerebral Cortex* 33.7 (2023), pp. 3387–3400.

[485] S Murray Sherman. Thalamus plays a central role in ongoing cortical functioning. *Nature neuroscience* 19.4 (2016), pp. 533–541.

[486] Giulio Ruffini et al. LSD-induced increase of Ising temperature and algorithmic complexity of brain dynamics. *PLoS computational biology* 19.2 (2023), e1010811.

[487] Friedrich Sanides. Comparative architectonics of the neocortex of mammals and their evolutionary interpretation. *Annals of the New York Academy of Sciences* 167.1 (1969), pp. 404–423.

[488] Zahra Shams et al. A comparison of in vivo MRI based cortical myelin mapping using T1w/T2w and R1 mapping at 3T. *PloS one* 14.7 (2019), e0218089.

[489] Rudolf Nieuwenhuys and Cees AJ Broere. A map of the human neocortex showing the estimated overall myelin content of the individual architectonic areas based on the studies of Adolf Hopf. *Brain Structure and Function* 222.1 (2017), pp. 465–480.

[490] Sophia Mueller et al. Individual variability in functional connectivity architecture of the human brain. *Neuron* 77.3 (2013), pp. 586–595.

[491] Martijn P Van Den Heuvel and Hilleke E Hulshoff Pol. Exploring the brain network: A review on resting-state fMRI functional connectivity. *European neuropsychopharmacology* 20.8 (2010), pp. 519–534.

[492] Cheng Chen et al. Reduced inter-hemispheric auditory and memory-related network interactions in patients with schizophrenia experiencing auditory verbal hallucinations. *Frontiers in Psychiatry* 13 (2022), p. 956895.

[493] Yi Liu et al. Decreased resting-state interhemispheric functional connectivity correlated with neurocognitive deficits in drug-naive first-episode adolescent-onset schizophrenia. *International Journal of Neuropsychopharmacology* 21.1 (2018), pp. 33–41.

[494] Xiaole Z Zhong et al. Predicting the macrovascular contribution to resting-state fMRI functional connectivity at 3 Tesla: A model-informed approach. *Imaging Neuroscience* 2 (2024), pp. 1–22.

[495] Dorothy VM Bishop et al. Approaches to measuring language lateralisation: An exploratory study comparing two fMRI methods and functional transcranial doppler ultrasound. *Neurobiology of Language* 5.2 (2024), pp. 409–431.

[496] Jared A Nielsen et al. An evaluation of the left-brain vs. right-brain hypothesis with resting state functional connectivity magnetic resonance imaging. *PloS one* 8.8 (2013), e71275.

[497] Francisco Aboitiz et al. Fiber composition of the human corpus callosum. *Brain research* 598.1-2 (1992), pp. 143–153.

[498] Lo"ıc Labache et al. The molecular and cellular underpinnings of human brain lateralization. *bioRxiv* (2025), pp. 2025–04.

[499] Eli J Müller et al. Core and matrix thalamic sub-populations relate to spatio-temporal cortical connectivity gradients. *NeuroImage* 222 (2020), p. 117224.

[500] Christina Mo and S Murray Sherman. A sensorimotor pathway via higher-order thalamus. *Journal of Neuroscience* 39.4 (2019), pp. 692–704.

[501] Marc Joliot et al. AICHA: An atlas of intrinsic connectivity of homotopic areas. *Journal of neuroscience methods* 254 (2015), pp. 46–59.

[502] Xiaoxuan Yan et al. Homotopic local-global parcellation of the human cerebral cortex from resting-state functional connectivity. *NeuroImage* 273 (2023), p. 120010.

[503] Xingyu Liu et al. A studyforrest extension, MEG recordings while watching the audio-visual movie "Forrest Gump". *Scientific data* 9.1 (2022), p. 206.

[504] Zeus Gracia-Tabuenca et al. Hemispheric asymmetry and homotopy of resting state functional connectivity correlate with visuospatial abilities in school-age children. *NeuroImage* 174 (2018), pp. 441–448.

[505] Marvin Petersen et al. Brain network architecture constrains age-related cortical thinning. *NeuroImage* 264 (2022), p. 119721.

[506] Matthew F Glasser et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* 80 (2013), pp. 105–124.

[507] Kevin M. Aquino et al. Identifying and removing widespread signal deflections from fMRI data: Rethinking the global signal regression problem. en. *Neuroimage* 212 (May 2020), p. 116614. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2020.116614. URL: https://www.sciencedirect.com/science/article/pii/S1053811920301014 (visited on 01/30/2022).

[508] Linden Parkes et al. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *Neuroimage* 171 (2018), pp. 415–436.

[509] Ross D Markello et al. Standardizing workflows in imaging transcriptomics with the abagen toolbox. *elife* 10 (2021), e72129.

[510] J-Donald Tournier et al. MRtrix: Diffusion tractography in crossing fiber regions. *International journal of imaging systems and technology* 22.1 (2012), pp. 53–66.

[511] Robert E Smith et al. SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *Neuroimage* 119 (2015), pp. 338–351.

[512] Pauli Virtanen et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[513] Alejandro F Frangi et al. Multiscale vessel enhancement filtering. In: *Medical image computing and computer-assisted intervention—MICCAI'98: first international conference cambridge, MA, USA, october 11–13, 1998 proceedings 1*. Springer. 1998, pp. 130–137.

[514] Aurina Arnatkevičiūtė et al. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* 189 (2019), pp. 353–367.

[515] Hadley Wickham and Carson Sievert. *ggplot2: elegant graphics for data analysis*. Vol. 10. springer New York, 2009.

[516] Daniel Lüdecke et al. see: An R Package for Visualizing Statistical Models. *Journal of Open Source Software* 6.64 (2021), p. 3393. DOI: 10.21105/joss.03393.

[517] Martijn P van den Heuvel et al. High-cost, high-capacity backbone for global brain communication. *Proc. Natl. Acad. Sci. USA* 109.28 (July 2012), pp. 11372–11377.

[518] Santo Fortunato. Community detection in graphs. *Physics reports* 486.3-5 (2010), pp. 75–174.

[519] E K Towlson et al. The rich club of the C. elegans neuronal connectome. *J. Neurosci.* 33.15 (2013), pp. 6380–6387.

[520] Chi-Tin Shih et al. Connectomics-Based Analysis of Information Flow in the Drosophila Brain. *Curr. Biol.* (Apr. 2015).

[521] Quanxin Wang et al. Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *J. Neurosci.* 32.13 (Mar. 2012), pp. 4386–4399.

[522] Brian Zingg et al. Neural Networks of the Mouse Neocortex. *Cell* 156.5 (Feb. 2014), pp. 1096–1111.

[523] Mikail Rubinov et al. Wiring cost and topological participation of the mouse brain connectome. *Proc. Natl. Acad. Sci. USA* 112.32 (Aug. 2015), pp. 10032–10037.

[524] Logan Harriger et al. Rich Club Organization of Macaque Cerebral Cortex and Its Role in Network Communication. *PLoS ONE* 7.9 (Sept. 2012), e46497.

[525] M A Bertolero and BTT Yeo. The modular and integrative functional architecture of the human brain. *Proc. Natl. Acad. Sci. USA* (2015).

[526] Richard F Betzel et al. Dynamic fluctuations coincide with periods of high and low modularity in resting-state functional brain networks. *NeuroImage* 127 (2016), pp. 287–297.

[527] Richard F Betzel et al. The modular organization of human anatomical brain networks: Accounting for the cost of wiring. *Network Neuroscience* 1.1 (2017), pp. 42–68.

[528] Nicolas A Crossley et al. Cognitive relevance of the community structure of the human brain functional coactivation network. *Proc. Natl. Acad. Sci. USA* 110.28 (July 2013), pp. 11583–11588.

[529] Ioannis Psorakis et al. Overlapping community detection using Bayesian non-negative matrix factorization. *Phys. Rev. E* 83 (6 June 2011), p. 066114. DOI: 10.1103/PhysRevE.83.066114. URL: https://link.aps.org/doi/10.1103/PhysRevE.83.066114.

[530] Andrea Lancichinetti et al. Finding Statistically Significant Communities in Networks. *PLOS ONE* 6.4 (Apr. 2011), pp. 1–18. DOI: 10.1371/journal.pone.0018961. URL: https://doi.org/10.1371/journal.pone.0018961.

[531] Gergely Palla et al. Uncovering the overlapping community structure of complex networks in nature and society. *nature* 435.7043 (2005), pp. 814–818.

[532] Jierui Xie et al. SLPA: Uncovering Overlapping Communities in Social Networks via A Speaker-listener Interaction Dynamic Process. In: *ICDM 2011 Workshop on DMCCI*. 2011.

[533] Xuan-Chen Liu et al. A generalized stochastic block model for overlapping community detection. *Europhysics Letters* 146.4 (June 2024), p. 41004. DOI: 10.1209/0295-5075/ad4172. URL: https://dx.doi.org/10.1209/0295-5075/ad4172.

[534] Vinícius da Fonseca Vieira et al. A comparative study of overlapping community detection methods from the perspective of the structural properties. *Applied Network Science* 5 (2020), pp. 1–42.

[535] Kai Wu et al. The overlapping community structure of structural brain network in young healthy individuals. *PloS one* 6.5 (2011), e19608.

[536] Hua-Wei Shen et al. Quantifying and identifying the overlapping community structure in networks. *Journal of Statistical Mechanics: Theory and Experiment* 2009.07 (2009), P07042. URL: http://stacks.iop.org/1742-5468/2009/i=07/a=P07042.

[537] Aristeidis Sotiras et al. Finding imaging patterns of structural covariance via non-negative matrix factorization. *Neuroimage* 108 (2015), pp. 1–16.

[538] Jorge Martinez Armas et al. A principled approach to community detection in interareal cortical networks. *bioRxiv* (2024), pp. 2024–08.

[539] Xuan Li et al. Overlapping community structure detection of brain functional network using non-negative matrix factorization. In: *International Conference on Neural Information Processing*. Springer. 2016, pp. 140–147.

[540] Kangjoo Lee et al. SPARK: Sparsity-based analysis of reliable k-hubness and overlapping network structure in brain functional connectivity. *NeuroImage* 134 (2016), pp. 434–449.

[541] Xuan Li et al. Collective sparse symmetric non-negative matrix factorization for identifying overlapping communities in resting-state brain functional networks. *NeuroImage* 166 (2018), pp. 259–275.

[542] S Mirzaei and Hamid Soltanian-Zadeh. Overlapping brain community detection using Bayesian tensor decomposition. *Journal of neuroscience methods* 318 (2019), pp. 47–55.

[543] Alexander J Gates et al. Element-centric clustering comparison unifies overlaps and hierarchy. *Scientific reports* 9.1 (2019), p. 8574.

[544] Joshua Faskowitz et al. Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture. en. *Nature Neuroscience* (Oct. 2020). Publisher: Nature Publishing Group, pp. 1–11. ISSN: 1546-1726. DOI: 10.1038/s41593-020-00719-y. URL: https://www.nature.com/articles/s41593-020-00719-y (visited on 11/04/2020).

[545] Hong Zhu et al. Nodal memberships to communities of functional brain networks reveal functional flexibility and individualized connectome. *Cerebral Cortex* 31.11 (2021), pp. 5090–5106.

[546] Janine D Bijsterbosch et al. Evaluating functional brain organization in individuals and identifying contributions to network overlap. *Imaging Neuroscience* 1 (2023), pp. 1–19.

[547] Tianyuan Lei et al. Functional network modules overlap and are linked to interindividual connectome differences during human brain development. *PLoS biology* 22.9 (2024), e3002653.

[548] Natalie R Smith et al. A guide for choosing community detection algorithms in social network studies: The question alignment approach. *American journal of preventive medicine* 59.4 (2020), pp. 597–605.

[549] Jierui Xie et al. Overlapping Community Detection in Networks: The State-of-the-art and Comparative Study. *ACM Comput. Surv.* 45.4 (Aug. 2013), 43:1–43:35. ISSN: 0360-0300. DOI: 10.1145/2501654.2501657. URL: http://doi.acm.org/10.1145/2501654.2501657.

[550] Nathan Aston et al. Overlapping community detection in dynamic networks. *Journal of Software Engineering and Applications* 7.10 (2014), pp. 872–882.

[551] Darko Hric et al. Community detection in networks: Structural communities versus ground truth. *Physical Review E* 90.6 (2014), p. 062805.

[552] Xuewu Zhang et al. Overlapping community identification approach in online social networks. *Physica A: Statistical Mechanics and its Applications* 421 (2015), pp. 233–248.

[553] Mahdi Hajiabadi et al. IEDC: An integrated approach for overlapping and non-overlapping community detection. *Knowledge-Based Systems* 123 (2017), pp. 188–199.

[554] Vinh Loc Dao et al. Community structure: A comparative evaluation of community detection methods. *Network Science* 8.1 (2020), pp. 1–41.

[555] David C Van Essen et al. The WU-Minn Human Connectome Project: An overview. *NeuroImage* 80 (Oct. 2013), pp. 62–79.

[556] Andrea Lancichinetti and Santo Fortunato. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Physical Review E* 80.1 (2009), p. 016118.

[557] Imre Derényi et al. Clique percolation in random networks. *Physical review letters* 94.16 (2005), p. 160202.

[558] Martin Rosvall and Carl T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* 105.4 (2008), pp. 1118–1123. ISSN: 0027-8424. DOI: 10.1073/pnas.0706851105. eprint: https://www.pnas.org/content/105/4/1118.full.pdf. URL: https://www.pnas.org/content/105/4/1118.

[559] M. Rosvall et al. The map equation. *The European Physical Journal Special Topics* 178.1 (Nov. 2009), pp. 13–23. ISSN: 1951-6401. DOI: 10.1140/epjst/e2010-01179-1. URL: https://doi.org/10.1140/epjst/e2010-01179-1.

[560] H Wu-Minn. 1200 subjects data release reference manual. *URL https://www. humanconnectome. org* 565 (2017), p. 2.

[561] J-Donald Tournier et al. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage* 202 (2019), p. 116137.

[562] J Donald Tournier et al. Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. In: *Proceedings of the international society for magnetic resonance in medicine*. Vol. 18. 2010, p. 1670.

[563] Robert E Smith et al. SIFT: Spherical-deconvolution informed filtering of tractograms. *Neuroimage* 67 (2013), pp. 298–312.

[564] Matthew F. Glasser et al. A multi-modal parcellation of human cerebral cortex. *Nature* 536 (July 2016), 171 EP -. URL: https://doi.org/10.1038/nature18933.

[565] Cibu Thomas et al. Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. *Proceedings of the National Academy of Sciences* 111.46 (2014), pp. 16574–16579.

[566] Stamatios N Sotiropoulos and Andrew Zalesky. Building connectomes using diffusion MRI: why, how and but. *NMR in Biomedicine* 32.4 (2019), e3752.

[567] Richard F Betzel et al. Multi-scale community organization of the human structural connectome and its relationship with resting-state functional connectivity. *Network Science* 1.3 (2013), pp. 353–373.

[568] Aurina Arnatkeviciute et al. Genetic influences on hub connectivity of the human connectome. *Nature communications* 12.1 (2021), p. 4237.

[569] Gustavo Deco and Morten L Kringelbach. Turbulent-like dynamics in the human brain. *Cell reports* 33.10 (2020).

[570] Jakub Vohryzek et al. Human brain dynamics are shaped by rare long-range connections over and above cortical geometry. *Proceedings of the National Academy of Sciences* 122.1 (2025), e2415102122.

[571] Christian Lohse et al. Resolving anatomical and functional structure in human brain organization: Identifying mesoscale organization in weighted network representations. *PLoS computational biology* 10.10 (2014), e1003712.

[572] Joshua Faskowitz et al. Weighted stochastic block models of the human connectome across the life span. *Scientific reports* 8.1 (2018), p. 12997.

[573] Lazaro M Sanchez-Rodriguez et al. Detecting brain network communities: Considering the role of information flow and its different temporal scales. *NeuroImage* 225 (2021), p. 117431.

[574] Casper Schmidt et al. Neural fingerprints of gambling disorder using diffusion tensor imaging. *Psychiatry Research: Neuroimaging* 333 (2023), p. 111657.

[575] Marcel A de Reus and Martijn P Van den Heuvel. The parcellation-based connectome: limitations and extensions. *Neuroimage* 80 (2013), pp. 397–404.

[576] Klaus H Maier-Hein et al. The challenge of mapping the human connectome based on diffusion tractography. *Nature communications* 8.1 (2017), p. 1349.

[577] James A. Roberts et al. Consistency-based thresholding of the human connectome. *NeuroImage* 145 (Jan. 2017), pp. 118–129. DOI: 10.1016/j.neuroimage.2016.09.053.

[578] Fernando Calamante. The seven deadly sins of measuring brain structural connectivity using diffusion MRI streamlines fibre-tracking. *Diagnostics* 9.3 (2019), p. 115.

[579] A. Clauset et al. Power-Law Distributions in Empirical Data. *SIAM Review* 51.4 (2009), pp. 661–703. DOI: 10.1137/070710111. eprint: https://doi.org/10.1137/070710111. URL: https://doi.org/10.1137/070710111.

[580] Illés Farkas et al. Weighted network modules. *New Journal of Physics* 9.6 (2007), p. 180.

[581] Michael Molloy and Bruce Reed. A Critical Point for Random Graphs with a given Degree Sequence. en. *Random Structures & Algorithms* 6.2-3 (1995), pp. 161–180. ISSN: 1098-2418. DOI: 10.1002/rsa.3240060204.

[582] Usha Nandini Raghavan et al. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E* 76.3 (Sept. 2007). Publisher: American Physical Society, p. 036106. DOI: 10.1103/PhysRevE.76.036106. URL: https://link.aps.org/doi/10.1103/PhysRevE.76.036106 (visited on 11/05/2020).

[583] Stephany Rajeh and Hocine Cherifi. A community-aware centrality framework based on overlapping modularity. *Social Network Analysis and Mining* 13.1 (2023), p. 37.

[584] Konstantin Kuzmin et al. Parallelizing SLPA for scalable overlapping community detection. *Scientific Programming* 2015.1 (2015), p. 461362.

[585] Anggita Larasati and Isti Surjandari. Consumer behaviour analysis using speaker-listener label propagation algorithm (SLPA). In: *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*. IEEE. 2017, pp. 155–160.

[586] Mohammad Asadi and Foad Ghaderi. Incremental community detection in social networks using label propagation method. In: *2018 23rd Conference of Open Innovations Association (FRUCT)*. IEEE. 2018, pp. 39–47.

[587] Alcides Viamontes Esquivel and Martin Rosvall. Compression of flow can reveal overlapping-module organization in networks. English. *Physical Review X* 1.2 (Dec. 2011), p. 021025. DOI: 10.1103/PhysRevX.1.021025. URL: http://link.aps.org/doi/10.1103/PhysRevX.1.021025.

[588] Zongqing Lu et al. Algorithms and applications for community detection in weighted networks. *IEEE Transactions on Parallel and Distributed Systems* 26.11 (2014), pp. 2916–2926.

[589] Leon Danon et al. Comparing community structure identification. *Journal of statistical mechanics: Theory and experiment* 2005.09 (2005), P09008.

[590] Mark EJ Newman et al. Improved mutual information measure for clustering, classification, and community detection. *Physical Review E* 101.4 (2020), p. 042304.

[591] Andrea Lancichinetti et al. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics* 11.3 (Mar. 2009), p. 033015. DOI: 10.1088/1367-2630/11/3/033015. URL: https://doi.org/10.1088%2F1367-2630%2F11%2F3%2F033015.

[592] Khawla Asmi et al. The greedy coupled-seeds expansion method for the overlapping community detection in social networks. *Computing* 104.2 (2022), pp. 295–313.

[593] Youngdo Kim and Hawoong Jeong. Map equation for link communities. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 84.2 (2011), p. 026110.

[594] Jierui Xie and Boleslaw K Szymanski. Towards linear time overlapping community detection in social networks. In: *Advances in Knowledge Discovery and Data Mining: 16th Pacific-Asia Conference, PAKDD 2012, Kuala Lumpur, Malaysia, May 29–June 1, 2012, Proceedings, Part II 16*. Springer. 2012, pp. 25–36.

[595] Dongxiao He et al. Identification of hybrid node and link communities in complex networks. *Scientific reports* 5.1 (2015), p. 8638.

[596] Heli Sun et al. LinkLPA: A Link-Based Label Propagation Algorithm for Overlapping Community Detection in Networks. *Computational Intelligence* 33.2 (2017), pp. 308–331.

[597] Borja Esteve-Altava. Challenges in identifying and interpreting organizational modules in morphology. *Journal of Morphology* 278.7 (2017), pp. 960–974.

[598] Simon De Deyne et al. Evidence for widespread thematic structure in the mental lexicon. In: *CogSci*. 2015.

[599] Emiliano Bruner et al. A network approach to brain form, cortical topology and human evolution. *Brain Structure and Function* 224 (2019), pp. 2231–2245.

[600] Mislav Acman et al. Large-scale network analysis captures biological features of bacterial plasmids. *Nature communications* 11.1 (2020), p. 2452.

[601] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* 52.3 (2010), pp. 1059–1069.

[602] Pavle Cajic et al. On the Information-Theoretic Formulation of Network Participation. *Journal of Physics: Complexity* 5.1 (Mar. 2024), p. 015021.

[603] Aric Hagberg et al. *Exploring network structure, dynamics, and function using NetworkX*. Tech. rep. Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), 2008.

[604] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

[605] Patric Hagmann et al. Mapping the structural core of human cerebral cortex. *PLoS biology* 6.7 (2008), e159.

[606] Mehul Gajwani et al. Can hubs of the human connectome be identified consistently with diffusion MRI? *Network Neuroscience* 7.4 (2023), pp. 1326–1350.

[607] Michael T Gastner and Géza Ódor. The topology of large open connectome networks for the human brain. *Scientific reports* 6.1 (2016), p. 27249.

[608] Aaron F Alexander-Bloch et al. Human cortical thickness organized into genetically-determined communities across spatial resolutions. *Cerebral Cortex* 29.1 (2019), pp. 106–118.

[609] Sofie L Valk et al. Shaping brain structure: Genetic and phylogenetic axes of macroscale organization of cortical thickness. *Science advances* 6.39 (2020), eabb3417.

[610] Zhang J Chen et al. Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cerebral cortex* 18.10 (2008), pp. 2374–2381.

[611] Olaf Sporns. Network attributes for segregation and integration in the human brain. *Current opinion in neurobiology* 23.2 (2013), pp. 162–171.

[612] Moataz Assem et al. A domain-general cognitive core defined in multimodally parcellated human cortex. *Cerebral Cortex* 30.8 (2020), pp. 4361–4380.

[613] Maria Grazia Puxeddu et al. Leveraging multivariate information for community detection in functional brain networks. *bioRxiv* (2024), pp. 2024–07.

[614] Gorka Zamora-López et al. Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. *Frontiers in neuroinformatics* 4 (2010), p. 613.

[615] Falko Mecklenbrauck et al. The significance of structural rich club hubs for the processing of hierarchical stimuli. *Human Brain Mapping* 45.4 (2024), e26543.

[616] Timothy Lawn et al. From neurotransmitters to networks: Transcending organisational hierarchies with molecular-informed functional imaging. *Neuroscience & Biobehavioral Reviews* 150 (2023), p. 105193.

[617] Vincent Bazinet et al. Towards a biologically annotated brain connectome. *Nature reviews neuroscience* 24.12 (2023), pp. 747–760.

[618] Olaf Sporns et al. Identification and classification of hubs in brain networks. *PloS one* 2.10 (2007), e1049.

[619] Martijn P Van den Heuvel and Olaf Sporns. Network hubs in the human brain. *Trends in cognitive sciences* 17.12 (2013), pp. 683–696.

[620] Anna Luisa Klahn et al. Provincial and connector qualities of somatosensory brain network hubs in bipolar disorder. *Cerebral Cortex* 34.9 (2024), bhae366.

[621] Jiahe Zhang et al. Topography impacts topology: Anatomically central areas exhibit a “high-level connector” profile in the human cortex. *Cerebral Cortex* 30.3 (2020), pp. 1357–1365.

[622] Kate Smith-Miles et al. Towards objective measures of algorithm performance across instance space. English. *Computers & Operations Research* 45 (2014), p. 12 24. ISSN: 0305-0548. DOI: 10.1016/j.cor.2013.11.015.

[623] Kajsa M Igelström and Michael SA Graziano. The inferior parietal lobule and temporoparietal junction: A network perspective. *Neuropsychologia* 105 (2017), pp. 70–83.

[624] Gagan S Wig et al. An approach for parcellating human cortical areas using resting-state correlations. *Neuroimage* 93 (2014), pp. 276–291.

[625] Zhiguo Luo et al. Frequency-specific segregation and integration of human cerebral cortex: An intrinsic functional atlas. *Iscience* 27.3 (2024).

[626] Andrew S Alexander et al. Rethinking retrosplenial cortex: Perspectives and predictions. *Neuron* 111.2 (2023), pp. 150–175.

[627] Soojin Park and Marvin M Chun. Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage* 47.4 (2009), pp. 1747–1756.

[628] Elizabeth R Chrastil. Heterogeneity in human retrosplenial cortex: A review of function and connectivity. *Behavioral neuroscience* 132.5 (2018), p. 317.

[629] Elizabeth Jefferies. The neural basis of semantic cognition: Converging evidence from neuropsychology, neuroimaging and TMS. *Cortex* 49.3 (2013), pp. 611–625.

[630] Philipp Kuhnke et al. The role of the angular gyrus in semantic cognition: A synthesis of five functional neuroimaging studies. *Brain Structure and Function* 228.1 (2023), pp. 273–291.

[631] Meiqi Niu and Nicola Palomero-Gallagher. Architecture and connectivity of the human angular gyrus and of its homolog region in the macaque brain. *Brain Structure and Function* 228.1 (2023), pp. 47–61.

[632] Noah C Benson et al. The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. *Journal of vision* 18.13 (2018), pp. 23–23.

[633] Martijn P. van den Heuvel and Olaf Sporns. Rich-Club Organization of the Human Connectome. *Journal of Neuroscience* 31.44 (2011), pp. 15775–15786. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.3539-11.2011. eprint: http://www.jneurosci.org/content/31/44/15775.full.pdf. URL: http://www.jneurosci.org/content/31/44/15775.

[634] Martijn P. van den Heuvel and Olaf Sporns. An Anatomical Substrate for Integration among Functional Networks in Human Cortex. *Journal of Neuroscience* 33.36 (2013), pp. 14489–14500. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.2128-13.2013. eprint: http://www.jneurosci.org/content/33/36/14489.full.pdf. URL: http://www.jneurosci.org/content/33/36/14489.

[635] Marcel A. de Reus et al. An edge-centric perspective on the human connectome: Link communities in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1653 (2014), p. 20130527. DOI: 10.1098/rstb.2013.0527. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2013.0527. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2013.0527.

[636] James M. Shine et al. The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron* 92.2 (2016), pp. 544–554. ISSN: 0896-6273. DOI: https://doi.org/10.1016/j.neuron.2016.09.018. URL: http://www.sciencedirect.com/science/article/pii/S0896627316305773.

[637] Yue Gu et al. The overlapping modular organization of human brain functional networks across the adult lifespan. *NeuroImage* 253 (2022), p. 119125.

[638] Nathalie Tzourio-Mazoyer et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15.1 (2002), pp. 273–289.

[639] Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nature reviews neuroscience* 13.5 (2012), pp. 336–349.

[640] David Meunier et al. Modular and Hierarchically Modular Organization of Brain Networks. *Front. Psychiat.* 4 (2010), pp. 1–11.

[641] Prem K Gopalan and David M Blei. Efficient discovery of overlapping communities in massive networks. *Proceedings of the National Academy of Sciences* 110.36 (2013), pp. 14534–14539.

[642] Peiyan Yuan et al. Detecting overlapping community structures with PCA technology and member index. In: *Proceedings of the 9th EAI International Conference on Mobile Multimedia Communications*. 2016, pp. 121–125.

[643] Haiping Ma et al. A local-to-global scheme-based multi-objective evolutionary algorithm for overlapping community detection on large-scale complex networks. *Neural Computing and Applications* 33.10 (2021), pp. 5135–5149.

[644] Burke Q Rosen and Eric Halgren. A whole-cortex probabilistic diffusion tractography connectome. *eneuro* 8.1 (2021).

[645] Xueyan Jiang et al. Connectome analysis of functional and structural hemispheric brain networks in major depressive disorder. *Translational psychiatry* 9.1 (2019), p. 136.

[646] Erlei Wang et al. Convergent reductions in interhemispheric functional, structural and callosal connectivity in Parkinson's disease. *Frontiers in Aging Neuroscience* 17 (2025), p. 1512130.

[647] Stavros I Dimitriadis et al. The impact of graph construction scheme and community detection algorithm on the repeatability of community and hub identification in structural brain networks. *Human brain mapping* 42.13 (2021), pp. 4261–4280.

[648] Janine Bijsterbosch et al. Challenges and future directions for representations of functional brain organization. *Nature neuroscience* 23.12 (2020), pp. 1484–1495.

[649] Hannelore Aerts et al. Modeling brain dynamics in brain tumor patients using the virtual brain. *Eneuro* 5.3 (2018).

[650] Guusje Collin et al. Impaired rich club connectivity in unaffected siblings of schizophrenia patients. *Schizophrenia bulletin* 40.2 (2014), pp. 438–448.

[651] Arnaud Messé. Parcellation influence on the connectivity-based structure–function relationship in the human brain. *Human brain mapping* 41.5 (2020), pp. 1167–1180.

[652] Sumeet Agarwal et al. High throughput network analysis. In: *Proceedings of the workshop on analysis of complex networks (ACNE), European conference on machine learning and principles and practise of knowledge discovery in databases (ECML, PKDD), Barcelona, Spain*. 2010, pp. 13–18.

[653] Andrea Avena-Koenigsberger et al. Network Morphospace. *Journal of The Royal Society Interface* 12.103 (2015), pp. 20140881–20140881. DOI: 10.1098/rsif.2014.0881.

[654] Hamid Karimi-Rouzbahani and Aileen McGonigal. Directionality of neural activity in and out of the seizure onset zone in focal epilepsy. *Network Neuroscience* (2025), pp. 1–26.

[655] Zhen-Qi Liu et al. Benchmarking methods for mapping functional connectivity in the brain. *Nature Methods* (2025), pp. 1–10.

[656] Lukas Roell et al. How to measure functional connectivity using resting-state fMRI? A comprehensive empirical exploration of different connectivity metrics. *NeuroImage* 312 (2025), p. 121195.

[657] Sambit Panda et al. hyppo: A Multivariate Hypothesis Testing Python Package. *arXiv preprint arXiv:1907.02088* (2019).

[658] Mehrshad Golesorkhi et al. The brain and its time: Intrinsic neural timescales are key for input processing. *Communications biology* 4.1 (2021), p. 970.

[659] Dustin Scheinost et al. Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage* 193 (2019), pp. 35–45.

[660] Denes Szucs and John PA Ioannidis. Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage* 221 (2020), p. 117164.

[661] Imran Alam et al. Canonical time-series features for characterizing biologically informative dynamical patterns in fMRI. *bioRxiv* (2024), pp. 2024–07.

[662] Andre F Marquand et al. Understanding heterogeneity in clinical cohorts using normative models: Beyond case-control studies. *Biological psychiatry* 80.7 (2016), pp. 552–561.

[663] Henri Korn and Philippe Faure. Is there chaos in the brain? II. Experimental evidence and related models. *Comptes rendus biologies* 326.9 (2003), pp. 787–840.

[664] D Vignesh et al. A review on the complexities of brain activity: Insights from nonlinear dynamics in neuroscience. *Nonlinear Dynamics* 113.5 (2025), pp. 4531–4552.

[665] Gustavo Deco et al. Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. *Journal of Neuroscience* 33.27 (2013), pp. 11239–11252.

[666] Adrián Ponce-Alvarez and Gustavo Deco. The Hopf whole-brain model and its linear approximation. *Scientific reports* 14.1 (2024), p. 2615.

[667] Bärbel Maus et al. Optimization of blocked designs in fMRI studies. *Psychometrika* 75.2 (2010), pp. 373–390.

[668] Anders M Dale. Optimal experimental design for event-related fMRI. *Human brain mapping* 8.2-3 (1999), pp. 109–114.

[669] Rasmus M Birn et al. Detection versus estimation in event-related fMRI: Choosing the optimal stimulus timing. *Neuroimage* 15.1 (2002), pp. 252–264.

[670] James M Shine et al. The dynamics of functional brain networks: Integrated network states during cognitive task performance. *Neuron* 92.2 (2016), pp. 544–554.

[671] Damien A Fair et al. Functional brain networks develop from a “local to distributed” organization. *PLoS computational biology* 5.5 (2009), e1000381.

References to other work co-authored during candidature

2025

Bryant, A. G.*, Malla, S.*, Jayakumar, R., Woost, B., Wolf, N., Li, A., Das, S., Van Veluw, S. J., & Bennett, R. E. (2025). Molecular profiling of frontal and occipital subcortical white matter hyperintensities in Alzheimer's disease. *Frontiers in neurology*, 15. *Co-first authors.
Link to published manuscript: https://doi.org/10.3389/fneur.2024.1470441.

Tan, J. B., Orlando, I., Whyte, C. J., **Bryant, A. G.**, Munn, B. R., Baracchini, G., King, M., O'Callaghan, C., Müller, E. J., & Shine, J. M. (2025). Cerebellar and subcortical contributions to working memory manipulation. *Nature communications biology*, 8.
Link to published manuscript: https://doi.org/10.1038/s42003-025-08467-0.

Munn, B. R., Tan, J., Müller, E. J., **Bryant, A. G.**, Orlando, I. F., Grimm, C., Zerbi, V., & Shine, J. M. Neuromodulation controls adaptive multiscale brain dynamics for cognition. *In submission*.

2024

Serrano-Pozo, A., Li, H., Li, Z., Muñoz-Castro, C., Jaisa-Aad, M., Healey, M. A., Welikovitch, L. A., Jayakumar, R., **Bryant, A. G.**, Noori, A., Connors T. R., Hu, M., Zhao, K., Liao, F., Gen, L., Pastika, T., Tamm, J., Abdourahman, A., Kwon, T., Bennett, R. E., Woodbury, M. E., Wachter, A., Talanian, R. V., Biber, K., Karran, E. H., Hyman, B. T., & Das, S. (2024). Astrocyte transcriptomic changes along the spatiotemporal progression of Alzheimer's disease. *Nature neuroscience*, 1-17.
Link to published manuscript: https://doi.org/10.1038/s41593-024-01791-4.

Lee, H., Fu, J. F., Gaudet, K., **Bryant, A. G.**, Price, J. C., Bennett, R. E., Johnson, K. A., Hyman, B. T., Hedden, T., Salat, D. H., Yen, Y-F., & Huang, S. Y. (2024). Aberrant vascular architecture in the hippocampus correlates with tau burden in mild cognitive impairment and Alzheimer's disease. *Journal of cerebral blood flow & metabolism*, 44(5), 787-800.
Link to published manuscript: https://doi.org/10.1177/0271678X231216144.

2023

Bryant, A. G., Li, Z., Jayakumar, R., Serrano-Pozo, A., Woost, B., Hu, M., Woodbury, M. E., Wachter, A., Lin, G., Kwon, T., Talanian, R. V., Biber, K., Karran, E. H., Hyman, B. T., Das, S., & Bennett, R. E. (2023) Endothelial cells are heterogeneous in different brain regions and are dramatically altered in Alzheimer's disease. *Journal of neuroscience*, 43(24), 4541-4557.
Link to published manuscript: https://doi.org/10.1523/JNEUROSCI.0237-23.2023.

Henderson, T., **Bryant, A. G.**, & Fulcher, B. D. (2023). Never a dull moment: Distributional properties as a baseline for time-series classification. *arXiv* preprint.

Link to *arXiv* preprint: https://doi.org/10.48550/arXiv.2303.17809.

Appendices

Appendix 1: Second half of the information theory review (Chapter 2)

Supplementary material for "Unifying concepts in information-theoretic time-series analysis"

Having introduced our classification of the eleven information-theoretic time-series measures through a schematic depiction (Fig. 3) and tabular overview (Table 1), we now turn to cover each measure in more detail, including a focus on our applied fMRI case study. Subsequent sections are intended to serve as in-depth reference material for the reader interested in a particular measure(s). We begin by examining the first class of measures, which are single-process and order-independent—meaning they capture properties of a signal's distribution that do not depend on temporal structure.

1 Single-process, order-independent measures

In this section, we introduce measures that capture properties within a single time series and are independent of the temporal ordering. In other words, the values of the time series could be shuffled along the time axis, and the values of these measures would not change, as they capture properties of the underlying distribution.

1.1 Entropy

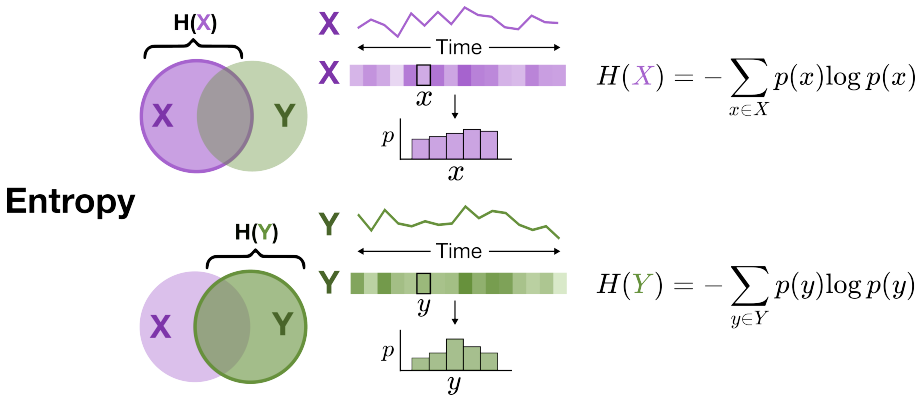


Figure 5. Entropy. Entropy quantifies uncertainty in probability distribution for the univariate process X or Y, with higher values indicating greater unpredictability in the measured time series. The histograms for the probability distributions across realizations (x and y) demonstrate the order-independent nature of entropy computation. The mathematical form for $H(X)$ and $H(Y)$ is provided on the right and discussed in Sec. 1.1.

1.1.1 What (does this measure capture):

Entropy—by which we refer to 'Shannon entropy' [1]—measures the degree to which the observed values from our time series, from which the probability distribution is composed, is surprising. More formally, the entropy of the distribution of a given process is an order-independent measure of the apparent randomness or variability exhibited across time series values, without considering temporal structure or other variables [3]. From another perspective, it can be interpreted as the average uncertainty in predicting the value of this variable, again without taking other information into account.

1.1.2 How (is this measure computed):

Let X be a random variable representing the values taken at individual time points in a time series (i.e., the distribution over time). As schematically depicted in Fig. 5, the entropy—with notation $H(X)$—is then defined

as:

$$H(X) = -\sum_{x \in X} p(x) \log p(x)\,, \tag{1}$$

which takes a weighted sum (over the probability $p(x)$) of the surprise or Shannon information content, $-\log p(x)$ [1], for each possible value x that the random variable X may take over time. Here, the $\log$ may be taken with any base, with the base simply determining the units of the measure. For discrete-valued variables, base 2 is typically used, in which case the units are in 'bits' [1]. Alternatively—by default when no base is indicated, and more often for continuous-valued variables (as defined in the next section)—one may use the natural logarithm with base e (such that the units are therefore in 'nats' [1]). For the remainder of this piece, we will work with the natural logarithm in units of 'nats'. The Shannon entropy, as defined in Eq. (1), is non-negative. When applied to a single variable out of a multivariate set, the measure may be referred to as 'Marginal entropy', indicating the marginalization of the joint PDF applied to obtain each $p(x)$ value.

An important note on computing entropy in continuous space

Here, we briefly introduce the concept of differential entropy for continuous-valued processes, which forms the basis for our simple illustrative fMRI analysis. When there are continuous variables in a given process (or pair of processes), these measures can be derived from 'differential entropy'—a measure generalizing the Shannon entropy to continuous variables—given as [2, Chapter 8]:

$$H(X) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx\,, \tag{2}$$

for a given random variable X, from which individual realizations are denoted with the lower-case x. Within the integral, the probability density function[3] of x ($f(x)$) weights the log-probability ($\log f(x)$), such that the differential entropy is an expectation function (like Shannon entropy). The limits of this integral range from $[-\infty, \infty]$ in general, since each realization corresponds to a continuous random variable. The differential entropy has different properties from the Shannon entropy [2, Chapter 8]; most notably, it can be a negative quantity, which means the continuous distribution is more concentrated than the uniform random distribution on $[0, 1]$.

When we compute information-theoretic measures from empirical data, we obtain an estimate ($\hat{H}$) of an underlying true value (H). This requires us to, implicitly or explicitly, estimate the PDF of the distribution from which the time series data in realizations x of X are drawn. The *JIDT* (version 1.6.1) library for computing information-theoretic measures estimates the PDF and approximates the integral in Eq. (2) above, using one of several possible methods listed below which are applicable to continuous-valued variables:

- Gaussian density estimator (`gaussian`)
- Box kernel density estimator (`kernel`)
- Kozachenko–Leonenko density estimator (`kozachenko`) [25]

Each estimator comes with its own assumptions, benefits, and drawbacks. The simplest Gaussian density estimator models the PDF as, unsurprisingly, a Gaussian distribution with variance σ^2, and under this model the entropy is given by:

$$H(X) = \frac{1}{2} \log 2\pi e \sigma^2, \tag{3}$$

obtained by plugging in the PDF of the Gaussian distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ for $f(x)$ in Eq. (2). The more complex box kernel and Kozachenko–Leonenko estimators are in contrast 'model-free', and approximate probability densities $f(x)$ by counting other samples within various radii of each given sample x, and use these (plus perhaps statistical corrections) to then estimate entropy. We refer the interested reader to Lizier [15] and Cliff et al. [17] for a detailed discussion of the implementation and interpretation of each method.

[3] Notice that here we use $f(x)$ to capture the probability density when X is a *continuous* random variable, while $p(x)$ is the notation for the probability of a specific value from a *discrete* random variable.

1.1.3 Why (is this measure useful):

Having defined the core algorithmic underpinnings of entropy, we continue to explore why it is a useful measure, with applications focused on computational neuroscience. If X represents the activity of a given brain region, a larger $H(X)$ indicates that the activity magnitude values (e.g., BOLD fMRI signal, calcium imaging, etc.) are closer to a more 'spread out' and equiprobable distribution, so each value is more surprising on average. By contrast, a smaller $H(X)$ suggests less diversity in the distribution of values in the given realization x of X (i.e., the observed period of that brain region) in favor of a few dominant values.

Entropy has gained a foothold in computational neuroscience to index the predictability of localized dynamics across spatial and temporal imaging scales [26–28]. We turn to our simple case study example in Fig. 4A, in which we computed the entropy from resting-state BOLD fMRI time series for all brain regions in the left cerebral cortex of one individual (see 'Methods: Examples using empirical neuroimaging data from the Human Connectome Project' for details). As shown on the cortical surface, we observe a narrow range of entropy values across different cortical regions computed with the Kozachenko–Leonenko estimator. This range spans from 1.32 nats in the parahippocampal gyrus to 1.48 nats in the fusiform gyrus, indicating generally homogeneous entropy values across the left cortex in this individual—such that all measured resting-state BOLD time series are comparably concentrated around the mean and contain comparable amounts of surprisal in this individual. The entropy of a single-process time series, which is insensitive to the time ordering, serves as a building block for more complex algorithmic quantities (namely, mutual information and conditional entropy) which are introduced in subsequent sections.

2 Pairwise order-independent measures, undirected

2.1 Joint entropy

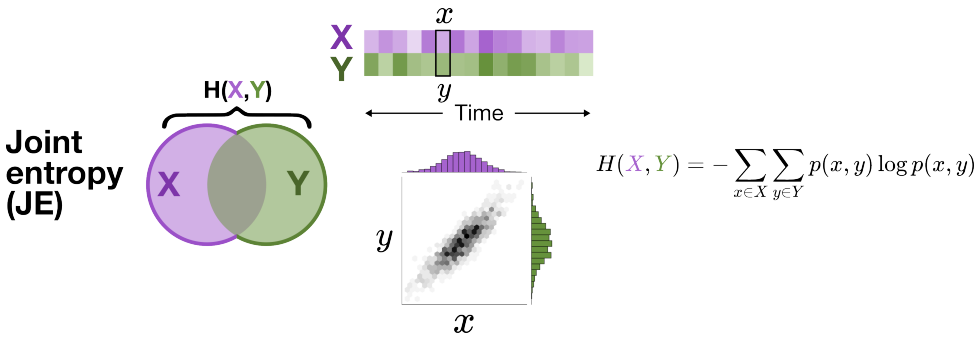


Figure 6. Joint entropy. Joint entropy (JE) quantifies uncertainty in the joint probability distribution for processes X and Y. The hexagonal heatmap of two-dimensional bin counts depicts the density of values in x–y space, demonstrating the invariance to temporal shuffling as a measure that does not depend on temporal order. The mathematical form for $H(X,Y)$ is provided on the right and discussed in Sec. 2.1.

2.1.1 What:

Joint entropy (JE) [3] quantifies the average information content—or 'surprisal'—across all paired observations of two random variables, X and Y, over their full joint distribution. It reflects the total uncertainty associated with jointly observing both X and Y, without assuming any particular relationship between them.

2.1.2 How:

As schematically depicted in Fig. 6, the joint entropy between random variables X and Y is defined as

$$H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x,y)\,, \tag{4}$$

where $x \in X$ and $y \in Y$ indicate the samples spaces of X and Y, and $p(x,y)$ is the joint probability mass function. This definition generalizes the single-process entropy (cf. Sec. 1.1) by considering the joint distribution of *two* variables rather than a single one. Stricktly speaking, joint entropy would belong to our 'single-process' category; however, here we classify it as 'pairwise' in order to focus on how it can be used to specifically capture the relationship *between a pair* of variables. As with single-process entropy, joint entropy is *not* an order-dependent measure. Of course, this can be fully generalized to any number of joint variables for $H(\boldsymbol{X})$; where the boldface $\boldsymbol{X}$ denotes a vector comprising a multivariate time series with $\{X_1,\ldots,X_M\}$ as individual variables corresponding to individual processes, and $H(\boldsymbol{X})$ is defined as a corresponding function of the multidimensional probability distribution function $p(\boldsymbol{x}) = p(x_1,\ldots,x_M)$.

Similarly, we can extend the single-process differential entropy introduced in Sec. 1.1 to the multivariate case here:

$$H(\boldsymbol{X}) = H(X_1,\ldots,X_M) = -\int_{-\infty}^{\infty} f(x_1,\ldots,x_M)\log f(x_1,\ldots,x_M)dx\,, \tag{5}$$

where $f(x_1,\ldots,x_M)$ is the joint probability density function (PDF) of the continuous random vector $\boldsymbol{X} = (X_1,\ldots,X_M)$. The same estimators described in Sec. 1.1 are available for joint entropy—namely, Gaussian, box kernel, and Kozachenko–Leonenko. There is a special case for the Gaussian density estimator: for the multivariate normal distribution $X \sim \mathcal{N}(\mu,\Sigma)$ of dimension M, the differential entropy has the following closed-form expression (i.e., no integral needs to be numerically evaluated):

$$H(\mathbf{X}) = \frac{1}{2}\log(2\pi e)^M|\Omega|, \tag{6}$$

given the covariance matrix Ω (with determinant $|\Omega|$) between its M sub-variables $f(x_1,\ldots,x_M)$. The intuition behind this for Gaussian multivariate distributions is that the larger the covariance among M dimensions, the more 'spread out' the distribution—and therefore, the entropy is greater. If all the M variables in $\boldsymbol{X}$ are independent, Ω is diagonal (i.e., all pairwise covariances are zero), and the joint entropy becomes the sum of the individual entropies.

In empirical settings, such as when working with two observed time series, this theoretical quantity must be estimated from finite samples. $H(X,Y)$ operates in the two dimensions of realizations x and y (with the underlying probability distribution represented as the heatmap, using hexagons to indicate two-dimensional bin counts, in Fig. 6), summing over the product of the joint probability and the log-transformed joint probability for each pair of observations in x and y.

2.1.3 Why:

If X and Y represent the activity of two brain regions, the joint entropy $H(X,Y)$ is a measure of the shared variability and combined uncertainty between the two regions—providing insight into the nature of their functional coupling. A larger $H(X,Y)$ indicates that the two brain regions collectively exhibit greater diversity or variability in their combined states (i.e., across $x-y$ space in the heatmap in Fig. 6). This means that, on average, each pair of x,y activity values is more surprising than if the time series were dominated by a small handful of repeated values, which would result in a more concentrated distribution in $x-y$ space. The inequality $H(X,Y) \leq H(X)+H(Y)$ formalizes that the JE of X and Y is maximized only when X and Y are statistically independent—such that any deviation from this upper bound of $H(X)+H(Y)$ reflects shared structure, or redundancy, between X and Y.

In our simple example in Fig. 4B, we compare the pairwise joint entropy (JE) between an example brain region of focus (lateral occipital cortex, LOC) and all other regions in the left cortex, considered one-by-one, using the Kozachenko–Leonenko estimator. Here, we observe the greatest JE between the caudal anterior cingulate cortex and the LOC, which might suggest that the caudal anterior cingulate adds maximum diversity to the LOC in this example fMRI dataset out of all regions evaluated. This implies that the joint behavior of these two regions is likely the least constrained by a fixed relationship out of all evaluated regions, indicating more independent or less coupled activity between the two regions. By contrast, a smaller $H(X,Y)$ in comparison to $H(X)+H(Y)$ means a stronger indicated dependency between measured activity in regions X and Y. In this case, the caudal middle frontal cortex

shows the lowest JE with the LOC, which could reflect redundancy in the information they encode.

While JE has not been directly applied as extensively in computational neuroscience compared to other measures covered here, JE has yielded insights into inter-areal coupling from both fMRI and electroencephalography (EEG) time series. For example, Baseer et al. [29] computed JE to predict microsleep periods from EEG recordings, positing that its strong performance reflects its ability to thoroughly capture how two regions jointly traverse across states. Martin et al. [30] used JE as part of a maximum entropy estimation framework, analyzing fMRI data from Watanabe et al. [31] to show that functional brain architecture is predominantly shaped by pairwise inter-areal coupling (rather than higher-order interactions). As with single-process entropy, JE is a component of more algorithmically complex measures—namely, mutual information (Sec. 2.2), conditional entropy (Sec. 3.1), and stochastic interaction (Sec. 5.1).

2.2 Mutual information

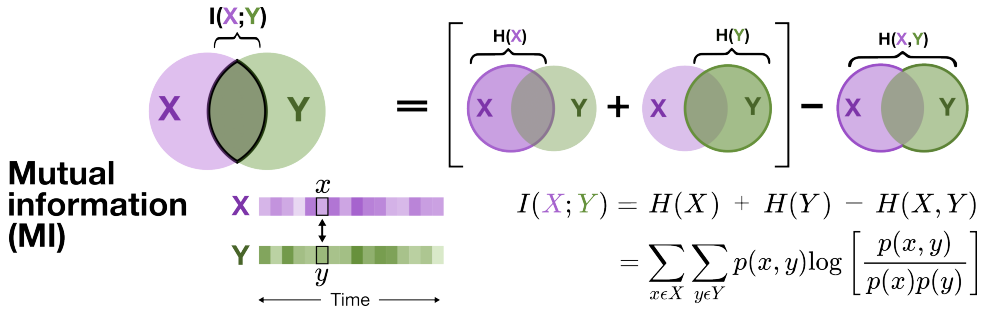


Figure 7. Mutual information. Mutual information (MI) quantifies contemporaneous information shared between processes X and Y—specifically for the contemporaneous bivariate case here. The Venn diagrams illustrate that this quantity is equal to the sum of the individual entropies, $H(X)$ and $H(Y)$, minus the joint entropy, $H(X,Y)$, which is also reflected in the mathematical form below the diagrams and discussed in Sec. 2.2.

2.2.1 What:

Mutual information (MI) [3] is an undirected measure of the average information that the value of one time series (e.g., X) provides about another (e.g., Y), denoted $I(X;Y)$. This can also be phrased in that $I(X;Y)$ tells us how much knowing X reduces uncertainty about Y on average, and vice versa.

2.2.2 How:

MI is schematically depicted in Fig. 7 and given by

$$I(X;Y) = H(X) + H(Y) - H(X,Y), \tag{7}$$

$$= \sum_{x \epsilon X} \sum_{y \epsilon Y} p(x,y) \log \left[\frac{p(x,y)}{p(x)p(y)} \right], \tag{8}$$

which defines $I(X,Y)$ as the difference between the sum of the individual entropies, $H(X)$ and $H(Y)$, minus the joint entropy, $H(X,Y)$ [3] in Eq. (7). Equivalently, Eq. (8) formulates MI directly in terms of probabilities of X and Y, summed across all possible values $x \in X$ and $y \in Y$, where $p(x,y)$ is the probability of the paired observations x and y and $p(x)$ and $p(y)$ are the marginal probabilities for x and y, respectively. $I(X;Y) \geq 0$, with the $I(X;Y) = 0$ case reflecting total independence between X and Y—such that knowing X does not reduce the uncertainty about Y at all, and vice versa. The formulation of MI in Eq. (8) means it is a type of measure known as a Kullback–Leibler (KL) divergence [2, 32]. Consequently, Eq. (8) measures the deviation of the joint distribution $p(x,y)$ from the reference case $p(x)p(y)$, representing what the joint distribution would be if X and Y were independent.

While largely beyond the scope of this review (which focuses on single-process and pairwise measures), we note that MI can be extended to 'conditional mutual information', which conditions the MI between X and Y on knowledge

of a third variable, Z. Put simply, this involves making the relevant probability functions in Eq. (8) conditional on z and then averaging over $p(z)$, or formulating the conditional MI for example as:

$$I(X;Y|Z) = I(X;\{Y,Z\}) - I(X;Z), \tag{9}$$

where $I(X;\{Y,Z\})$ is the MI between variable X and the joint variable $\{Y,Z\}$. We refer the interested reader to Cover and Thomas [2, Chapter 2] for further details and interpretations.

A note on estimating MI for continuous-valued variables

For continuous variables, the MI can be estimated as the sums and differences of the individual entropy terms in Eq. (7). Importantly, unlike differential entropy, MI for continuous variables is non-negative (much like the MI for discrete variables), such that MI has a consistent interpretation regardless of the variable type. While MI is generally a nonlinear measure, when we approximate the PDF with a Gaussian distribution (as with the Gaussian density estimator in *JIDT* [15]), it can be derived directly from the Pearson product-moment correlation between X and Y, r, as

$$\mathrm{MI}_{\mathrm{Gaussian}} = -\frac{1}{2}\log(1-r^2). \tag{10}$$

This MI implementation assumes that the joint distribution $p(x,y)$ is bivariate Gaussian, with linear interactions between X and Y. As such, $\mathrm{MI}_{\mathrm{Gaussian}}$ is only sensitive to linear dependence between X and Y. However, the other estimators introduced thus far (i.e., box kernel and Kozachenko–Leonenko) do not make parametric assumptions about the underlying joint bivariate distributions, and can therefore capture both linear and nonlinear relationships between X and Y. There is an additional estimator available based on the method in Kraskov et al. [33]—referred to as Kraskov–Stögbauer–Grassberger or KSG (and denoted as `kraskov` in *pyspi* and *JIDT*)—building on the Kozachenko–Leonenko estimator with specific statistical corrections tailored for MI [34]. We refer the interested reader to Wibral et al. [35] and Vicente and Wibral [36] for more discussion of the Kraskov–Stögbauer–Grassberger method as a suitable 'hybrid' estimator for MI.

2.2.3 Why:

If X and Y represent the activity of two brain regions, $I(X;Y)$ captures the extent to which knowing the activity of one brain region (X) reduces the uncertainty (i.e., improves the predictability) of the other region (Y), and vice versa. A larger $I(X;Y)$ indicates that the two regions share more information, implying a stronger inter-region dependence. By contrast, a smaller $I(X;Y)$ points to more independent observations in X and Y, with the two regions sharing minimal information with each other. The similarity between MI and the Pearson correlation coefficient (becoming a direct relationship with a Gaussian estimator in Eq. (10)) directly suggests MI as an alternate measure for functional connectivity [37]. In contrast to the Pearson correlation, though, the dependency captured by MI can extend beyond linear relationships, including forms like a function mapping activity of region X to region Y or complex patterns, such that $I(X;Y)$ can capture both linear and nonlinear dependence. This ability to capture both linear and nonlinear relationships, together with the inherently model-free nature, is a strength of MI for measuring inter-areal functional connectivity.

In our simple example in Fig. 4C, we compare the MI between the lateral occipital cortex (LOC) as our region of focus and all other regions in the left cortex using the Kraskov–Stögbauer–Grassberger estimator. Here, we see the greatest MI between the left lingual gyrus and LOC, indicating that knowing the BOLD fMRI time series in the left lingual gyrus maximally reduces uncertainty in that of the LOC out of all regions we evaluated (and vice versa). By contrast, the pars triangularis exhibits the lowest MI with the LOC, which indicates that the LOC shares the least information with the pars triangularis relative to all other regions in this example fMRI dataset. Quantifying inter-regional MI has been used to link microscopic and macroscopic network properties in the macaque functional connectome [38] and to distinguish pediatric epilepsy cases from age-matched control using BOLD fMRI [39]. MI has also proven useful in characterizing region-specific brain dynamics in response to different sensory stimuli across modalities, including EEG, fMRI, and magnetoencelphalography (MEG) [40, 41].

3 Pairwise order-independent measures, directed

3.1 Conditional entropy

3.1.1 What:

Conditional entropy (CE) [3] quantifies the uncertainty in the observations of one process, Y, in the context of simultaneously observing another process, X. CE is a directed measure, with the CE of Y given X—denoted as $H(Y \mid X)$—capturing the amount of remaining uncertainty in Y after conditioning on the observations of X. In other words, $H(Y \mid X)$ tells us how much information is still needed to describe Y even when we know the observed values of the conditional variable X. As with entropy, joint entropy, and mutual information, the conditional entropy is *not* dependent on the order of values in X and Y, making it an order-independent measure.

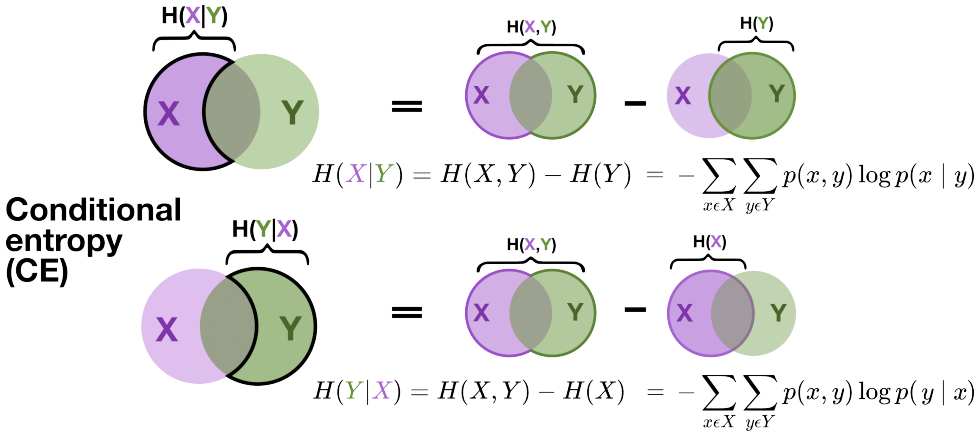


Figure 8. Conditional entropy. Conditional entropy (CE) quantifies uncertainty remaining in the probability distribution of one process (X) after conditioning on that of another process (Y). Both $H(X \mid Y)$ and $H(Y \mid X)$ are shown as this is a directed, and therefore asymmetric, measure. In both cases, the CE (e.g., $H(X \mid Y)$) is equivalent to the joint entropy ($H(X,Y)$) minus the marginal entropy of the process upon which the CE is conditioned ($H(Y)$). The mathematical form is provided for both $H(X \mid Y)$ and $H(Y \mid X)$ and discussed in Sec. 3.1.

3.1.2 How:

As depicted in Fig. 8, $H(Y \mid X)$ is given as

$$H(Y \mid X) = H(X,Y) - H(X), \tag{11}$$

$$= H(Y) - I(X;Y), \tag{12}$$

$$= \sum_{x \epsilon X}\sum_{y \epsilon Y} p(x,y)\log\left[\frac{p(x)}{p(x,y)}\right], \tag{13}$$

$$= -\sum_{x \epsilon X}\sum_{y \epsilon Y} p(x,y)\log p(y \mid x), \tag{14}$$

with Eq. (11) showing that $H(Y \mid X)$ is computed as the joint entropy, $H(X,Y)$, minus the marginal entropy of X, $H(X)$. This is mathematically equivalent to computing the difference between the marginal entropy of Y, $H(Y)$, minus the MI $I(X;Y)$, as in Eq. (12). In practice, for continuous-valued variables, CE is estimated as the difference between two differential entropies for the joint distribution $p(x,y)$, with the differential entropy equation given in Eq. (5), minus the univariate distribution $p(x)$, as given in Eq. (2).

Intuitively, the resulting quantity given by $H(Y|X)$ is the area of the Venn Diagram occupied exclusively by Y in Fig. 8, corresponding to the remaining uncertainty in Y after conditioning on X. It follows that $H(Y \mid X)$ can therefore be equivalently written, as in Eq. (14), as the average uncertainty remaining in the observed realization y after conditioning on the observed realization x, for all $x \in X$ and all $y \in Y$. Here, as in MI (cf. Sec. 2.2), $p(x,y)$ is the probability of the paired observations x and y, and $p(x)$ and $p(y)$ are the marginal distribution probabilities for

x and y, respectively. The form of the conditional entropy as the difference between joint minus marginal entropy in Eq. (11) can be seen as a direct consequence of the expansion of $p(y|x)$ (the probability distribution over which we are computing entropy here) from Eq. (14) using Bayes' rule. Since CE is a directed measure, the amount of uncertainty remaining for Y after simultaneously observing X is not necessarily equivalent to that remaining in X after observing Y. While the Venn diagram circles in Fig. 8 are symmetric (for aesthetic purposes), the two processes can exhibit marginal entropies ($H(X)$ and $H(Y)$) of different magnitudes in general—yielding two different values when subtracting the marginal entropy from the joint entropy.

3.1.3 Why:

If X and Y represent the activity of two brain regions, $H(Y \mid X)$ measures how much uncertainty remains for the distribution of activity values in region Y given knowledge of the activity in region X. Practically speaking, conditional entropy treats each time point in the two brain regions individually without actually capturing any temporal dependence, such that it does not capture any element of forward prediction in time. A larger $H(Y \mid X)$ indicates that simultaneously observing the activity in region X does not reduce the uncertainty in region Y by much. This exhibits a complementary relationship to mutual information, for fixed $H(Y)$, as shown in Eq. (12). By contrast, a smaller $H(Y \mid X)$ means that there is little uncertainty remaining about the observations in region Y after simultaneously observing and having access to the activity values of region X. In the literature, CE has been applied to EEG [42, 43], fMRI [44, 45], and neural spike train time series [46] to evaluate how knowing information about activity in one brain area (or neuronal ensemble) increases or decreases uncertainty about activity in another brain area.

In our simple example in Fig. 4D, we compute the CE for every cortical region in the left hemisphere (our Y in this example) conditioned on the lateral occipital cortex (LOC, our X in this example), using the Kozachenko–Leonenko estimator—in other words, taking the difference between two differential entropies. For the example of the fMRI activity in the entorhinal cortex, this CE would have the notation H(entorhinal|LOC). Here, we observe the highest CE in the left caudal anterior cingulate cortex, which indicates maximal remaining uncertainty in the caudal anterior cingulate even after observing the LOC activity compared to all other target regions. As noted above, this relationship does not automatically hold in the other direction, from caudal anterior cingulate to LOC; indeed, it is still possible that observing activity in the caudal anterior cingulate could reduce the majority of uncertainty remaining for that of the LOC. This asymmetry is important in understanding the relationship between CE and MI: they would be complementary as above for fixed $H(Y)$ if we used the LOC as Y, but not where Y is being changed. This is the reason why CE and MI are not perfectly complementary in this experiment—and therefore why the region with the highest CE after conditioning on the LOC (i.e., the caudal anterior cingulate) is not automatically the region with the lowest MI with the LOC (pars triangularis, cf. Fig. 4C). By contrast, the caudal middle frontal cortex exhibited the lowest CE after conditioning on the LOC, which suggests for this example fMRI dataset that observing the BOLD activity in the LOC maximally reduces the uncertainty remaining in the caudal middle frontal region. We do note that these two regions with the minimum and maximum CE are also the same two for JE (cf. Fig. 4B), which only holds as we are computing (entorhinal|LOC) and (caudal anterior cingulate|LOC) rather than the other way around (i.e., activity in the LOC conditioned on the entorhianl cortex or caudal anterior cingulate). In other words, the direct mapping of the maximum and minimum CE and JE in Eq. (11) is dependent on the conditioning directionality, as observed above for the MI.

4 Single-process, order-dependent measures

4.1 Active information storage

4.1.1 What:

Active information storage (AIS) [47] is a time-dependent measure that captures the amount of information *actively* in use from the past values of a time series to generate the next state. More formally, AIS quantifies information shared between the current time point, X_{t+1}, and previous time points up to memory length k (given as $X_t^{(k)}$) for the time series X with length T. AIS, referred to as $A(X)$, is the first-order component of a broader measure known

as predictive information [48], which quantifies the amount of information about the whole future of a process that can be found in its past—and vice versa.

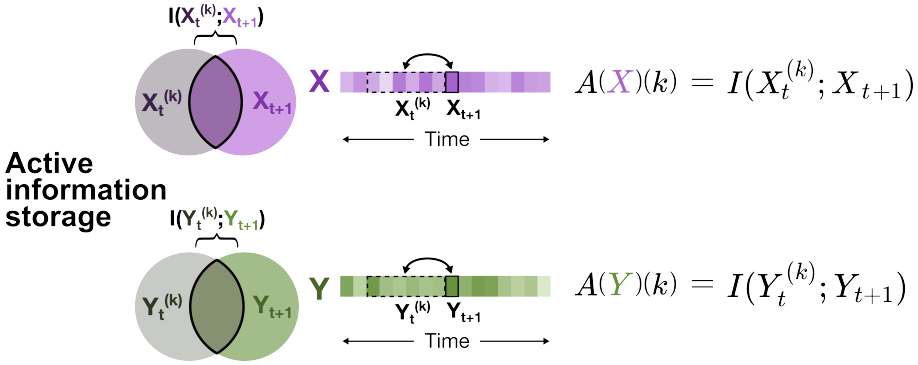


Figure 9. Active information storage. Active information storage (AIS) quantifies the amount of information shared, and actively used, between past and future states of a process (e.g., X) to predict the future of X. This is a single-process, order-dependent measure, equivalent to the mutual information (MI) between the past and present of the process. The mathematical form is provided for both $A(X)(k)$ and $A(Y)(k)$, where k is the memory length, as discussed in Sec. 4.1.

4.1.2 How:

As schematically depicted in Fig. 9, $A(X)$ is given as

$$A(X)(k) = I(X_t^{(k)}; X_{t+1}), \tag{15}$$

where t is a time point in the range $[1, T]$ and T is the total length of time series X. k denotes the memory length (i.e., the number of time points prior to time $t+1$ considered), such that $X_t^{(k)} = \{X_t, X_{t-1}, \ldots, X_{t-k+1}\}$. The expression as $I(X_t^{(k)}; X_{t+1})$ indicates that the AIS is a MI (cf. Sec. 2.2) computed between the present value of X and the past up X up to memory length k. The value chosen for k determines the number of past time points examined, with $A(X)(k)$ measuring the information storage under the assumption that the process X is a k-order Markov chain.

Theoretically, the window length k can be set to the full length of the time series X (T) minus one—meaning that the entire past of X is considered. Empirically, however, evaluating the *entire* past of X can become problematic due to computational burden (which becomes more pronounced in the case of pairwise order-dependent measures, discussed in subsequent sections). Moreover, when we have only a singular time series realization for X, we would only have single samples for the vectors $X_t^{(t)}$ at each time t. One approach to address these issues is for samples $(x_t^{(k)}, x_{t+1})$ across the whole time series to be utilized via an ergodic assumption to make one AIS calculation, $A(X)(k)$—rather than at each time point, t. Optimizing selection of the parameter k usually involves a trade-off between wanting longer k to capture more memory, while avoiding bias due to a too highly-dimensional probability space (see strategies detailed in [49, 50]). In contrast to a linear autocorrelation profile, $A(X)$ captures both linear and nonlinear dependence, while also capturing the full multivariate dependence of the variable X on its whole past—rather than characterizing the pairwise dependence at each lag separately.

4.1.3 Why:

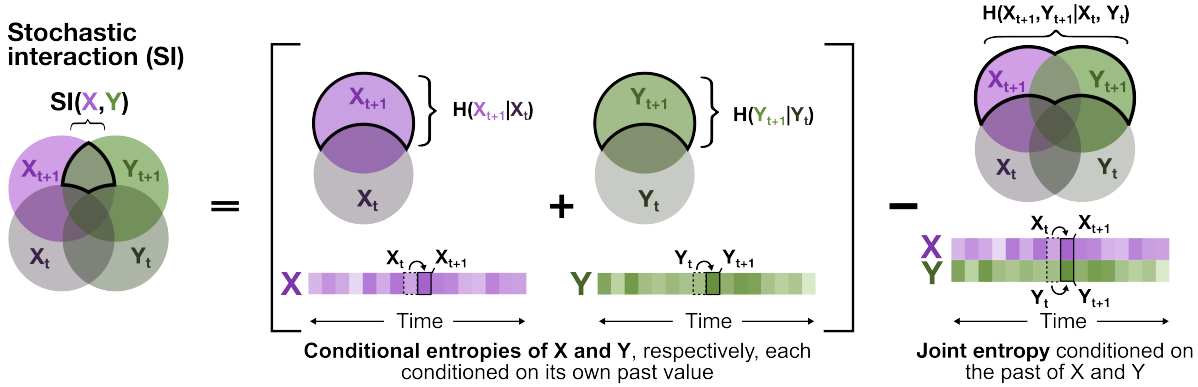
AIS is particularly useful in computational neuroscience to quantify the complexity and predictability of a neural process as it evolves through time, with higher values requiring both dynamical 'richness' (i.e., complex and informative patterns over time) and temporal predictability [51, 52]. If X represents the activity of a given brain region, $A(X)$ captures how much the current activity can be predicted from that region's past activity states. A larger value of $A(X)$ means there is a stronger dependence between the region's past and present activity values—indicating the activity is dynamically rich and highly predictable [52]. By contrast, a smaller $A(X)$ suggests there is weak or no dependence between the region's past and present activity values.

In Fig. 4E, we compare the AIS (using $k = 1$) from BOLD fMRI time series in all regions in the left cortex from one

example HCP participant using the Kraskov–Stögbauer–Grassberger estimator. Here, we observe the highest AIS in the lateral occipital cortex (LOC), suggesting that this region exhibits richer dynamics with more information from past activity values used in future states. By contrast, the parahippocampal gyrus exhibits the lowest overall AIS, meaning its activity is less dependent on its own past states—possibly indicating a greater relative influence of other regions' prior activity on future activity in the parahippocampal gyrus. AIS has been applied to characterize neural dynamics around a critical point [53, 54], learning and memory in neuromorphic computing [55], frequency-specific rhythms during anesthesia [56], and distributed processing of visual stimuli [57, 58]. Prior clinically grounded work has shown that AIS is reduced in the hippocampus of individuals with autism spectrum disorder [52, 59] and in the motor cortex of individuals with Parkinson's disease [60].

5 Pairwise order-dependent measures, undirected

5.1 Stochastic interaction



$$SI(X,Y) = H(X_{t+1}|X_t) + H(Y_{t+1}|Y_t) - H(X_{t+1},Y_{t+1}|X_t,Y_t)$$

Figure 10. Stochastic interaction. Stochastic interaction (SI) measures the integrated information between the two processes X and Y that cannot be explained by considering the two time series separately. It is an order-dependent pairwise measure that considers the past and present of both X and Y. As depicted by the Venn diagrams, $SI(X,Y)$ is equivalent to the sum of the conditional entropy of X and Y conditioned on their own past values, respectively, minus the joint entropy of X and Y both conditioned on their past values together. The mathematical form is provided below the diagram schematics and discussed in Sec. 5.1.

5.1.1 What:

Stochastic interaction (SI) [61, 62], denoted $\mathrm{SI}(X,Y)$, quantifies the integrated information between two processes, X and Y, by capturing the joint dynamical dependence between the two that cannot be entirely explained by their marginal dynamics. As described in Cohen et al. [63], SI captures the amount of information lost in (modeling) a 'disconnected' system compared to (modeling) an 'integrated' system. Specifically, SI measures the extent to which 'disconnecting' X and Y—in other words, modeling them as independent processes—increases the uncertainty about the future values of X and Y based on their respective past values, compared to the uncertainty about future values in the fully 'interconnected' (or jointly modeled) paired system. We previously reported that SI behaves very similarly across data types to measures derived from the integrated information literature [17], such as geometric integrated information and ϕ^* [64]. For an in-depth discussion of how SI relates to the broader field of integrated information (particularly geometric integrated information [64]), we refer the interested reader to Mediano et al. [65].

5.1.2 How:

As schematically depicted in Fig. 10, $SI(X,Y)$ is formulated based on Equation #26 in Ay [62]:

$$SI(X,Y) = H(X_{t+1}|X_t) + H(Y_{t+1}|Y_t) - H(X_{t+1},Y_{t+1}|X_t,Y_t)\,, \tag{16}$$

where $H(X_{t+1}|X_t)$ indicates the entropy of X at time point $t+1$ conditioned on the previous value at time point t—in other words, the amount of uncertainty remaining in the value of X at time $t+1$ after knowing the previous value of X at time t—as is the case for $H(Y_{t+1}|Y_t)$. The subtracted term in Eq. (16), $H(X_{t+1}, Y_{t+1}|X_t, Y_t)$, captures the joint entropy of X and Y at time $t+1$, both conditioned on their previous values at time t—in other words, the amount of uncertainty remaining in the paired observation X_{t+1}, Y_{t+1} after conditioning on the past paired observation X_t, Y_t. We note that the definition of SI here considers only one past value of X and Y, akin to using $k = 1$ for the AIS from the previous section. This effectively assumes a first-order joint Markov process, although in principle, one could formulate it based on longer memory lengths. Taken together, SI captures the difference between the summed entropy of X conditioned on its own past and Y conditioned on its own past (i.e., the two disconnected components) minus the joint entropy of X and Y conditioned on both of their own pasts. Equivalently, SI can be thought of as the average Kullback–Leibler divergence [32] between (i) the joint probability distribution for the present of both X and Y given the past of the connected system, and (ii) the product of the present value probability distributions for X and Y conditioned separately on their own pasts (i.e., 'disconnected' or modeled independently), akin to considerations in [66, 67].

5.1.3 Why:

SI can be applied in computational neuroscience contexts to understand how the activity of two neural processes—from single neurons to ensembles to macroscale regions—evolve through different states together, in a way where the future value of activity in region X depends on its own past and present along with those of region Y, and vice versa. A larger value of $SI(X, Y)$ indicates more shared information between the present and past of the joint distribution of X and Y (i.e., the fully interconnected paired case) compared to information contained in X and Y separately. Since the joint entropy conditioned on the past of X and Y is the subtracted term in Eq. (16) (cf. Fig. 10), a smaller value of $SI(X, Y)$ suggests that this joint entropy term may be comparatively larger—and therefore that there is not so much information shared in the fully connected paired system.

We depict the SI values between the LOC and all other regions in the left cortex from an example HCP participant in Fig. 4F, computed using the Kozachenko–Leonenko estimator. Here, we observe that the LOC exhibits maximal SI with the caudal anterior cingulate cortex, indicating that these two regions share the most information between their past and present respective BOLD fMRI magnitudes in this example fMRI dataset. By contrast, the LOC shows the lowest SI with the caudal middle frontal cortex, indicating minimal difference in the amount of information contained in these two regions separately than combined. We note that these same regions exhibited corresponding trends with CE (cf. Fig. 4D), consistent with the notion that BOLD magnitude in the LOC maximally reduces uncertainty in the caudal anterior cingulate and minimally reduces uncertainty in the caudal middle frontal cortex in this particular individual. SI has largely been used to quantify the spatiotemporal sharing of information among neural ensembles [68] or sensory processing in the feline auditory cortex [69] and macaque cortex [70]. In a comprehensive functional connectivity benchmarking analysis, Liu et al. [71] reported that SI exhibited among the greatest structure–functional coupling out of hundreds of evaluated measures.

6 Pairwise order-dependent measures, directed

6.1 Time-lagged mutual information

6.1.1 What:

Time-lagged mutual information (TLMI) [72, 73] quantifies the statistical dependence between two time series—specifically, how much information a past (time-lagged) value of one process, X, provides about a present or future value of another (Y). TLMI is computed using the mutual information framework (cf. Sec. 2.2), but applied across temporal offsets to capture time-lagged interactions between X and Y, without assuming any specific functional form of the relationship.

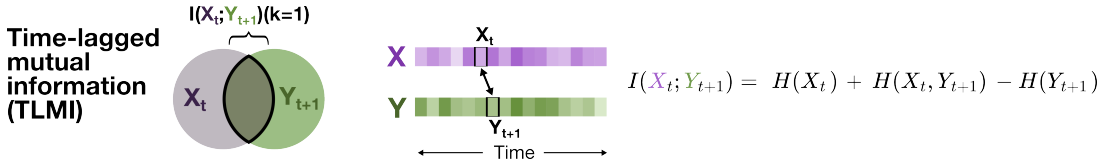


Figure 11. Time-lagged mutual information. Time-lagged mutual information (TLMI) quantifies the information shared between the past of one process (X) and the present of another (Y). While MI itself is not a directed measure, the directionality of this measure refers to the sensitivity to which time series (X or Y) is lagged. The mathematical form is provided to the right of the schematic diagram and discussed in Sec. 6.1.

6.1.2 How:

As depicted in Fig. 11, TLMI is computed as

$$I(X_t;Y_{t+1}) = H(X_t) + H(Y_{t+1}) - H(X_t,Y_{t+1}), \quad \textbf{(17)}$$
$$= H(X_t) - H(X_t|Y_{t+1}), \quad \textbf{(18)}$$
$$= H(Y_t) - H(Y_{t+1}|X_t), \quad \textbf{(19)}$$

indicating that $I(X_t;Y_{t+1})$ is equivalent to the difference between the sum of the marginal entropies, $H(X_t)$ and $H(Y_{t+1})$, minus the joint entropy, $H(X_t,Y_{t+1})$ (cf. 'Joint entropy', Sec. 2.1). We classify this as a *directed* pairwise coupling measure due to the asymmetry that arises depending on which of the two time series processes is treated with the lag. In other words, MI itself is not directed, since $I(X_t;Y_{t+1}) = I(Y_{t+1};X_t)$; however, this equivalence does not hold if the lag is reversed, meaning that $I(X_t;Y_{t+1}) \neq I(X_{t+1};Y_t)$.

6.1.3 Why:

If X and Y represent the activity in two brain regions (or neuronal ensembles), Greenblatt et al. [74] note that the shift(s) incorporated in TLMI can capture the temporally-lagged influence of activity measured in region X on that in region Y and vice versa. As such, a larger value of $I(X_t;Y_{t+1})$ indicates that the past states of region X share more information with the current state of region Y, while a smaller value indicates minimal cross-region dependence. Von Wegner et al. [75] demonstrated that TLMI can better distinguish short- versus long-range memory in neurophysiological time series from both EEG and fMRI compared to the Hurst exponent (which measures variance across timescales [76]). Other applications of TLMI to EEG have yielded connectivity networks with consistent delays across timescales [77] and highlighted the contribution of phase signals to microstate sequences [78, 79].

In our simple empirical example in Fig. 4G, we depict the TLMI values from the past (or lagged) value of the LOC BOLD fMRI time series (realization x_t corresponding to X_t in this example) to the present value of each other target region in the left cortex (realization y_{t+1} corresponding to Y_{t+1} in this example), computed for an example HCP participant using the Kraskov–Stögbauer–Grassberger estimator. This shows that the past of the LOC shares the most information with the current value in the caudal middle frontal cortex. Since we specifically implement this with a time lag of 1 BOLD frame, this suggests that the past state of the LOC is maximally predictive of the next state of the caudal middle frontal cortex in this example fMRI dataset. Of note, the caudal middle frontal cortex exhibited a relatively large MI and the smallest-magnitude CE and JE with the LOC, all of which focus on contemporaneous interactions between the two regions—suggesting that the time-lagged relationship incorporated in TLMI remains relevant to the (potentially nonlinear) contemporaneous interaction between these two regions. Moreover, the maximal TLMI here is a contributing factor to the minimal SI with the caudal middle frontal cortex. By contrast, the LOC exhibits the lowest TLMI with the lateral orbitofrontal cortex as the target, suggesting minimal time-lagged dependence between these two brain regions in this participant.

6.2 Causally conditioned entropy

6.2.1 What:

Causally conditioned entropy (CCE) [80, 81] is a directed measure, denoted $H(X\|Y)(k)$, that quantifies the uncertainty remaining in one process, Y, in the context of its own past up to memory length k (Y_{t+1}) and the past *and present* of another process, X (up to memory length $k+1$). In other words, $H(X\|Y)(k)$ captures how much uncertainty remains for the present value of Y when we know its past k values alongside the past and present of X up to time $k+1$. One could set a single value for k *a priori* or evaluate different values up to a maximum window length of $K \ll T$, where T is the length of time series X and Y (see 'Active information storage', Sec. 4.1, for a discussion on choosing k and ergodicity assumptions). Note that the term 'causal' here refers to the conditioning on terms in the past, which *may* be causal to Y. However, as an observational rather than interventional measure, CCE does not directly measure causality [82].

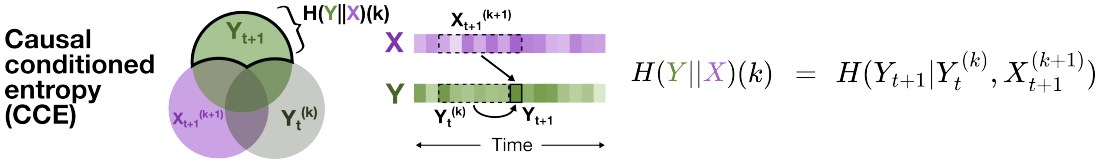


Figure 12. Causally conditioned entropy. Causally conditioned entropy (CCE) quantifies the uncertainty remaining in the present of one process (Y) after observing its own past along with the past and present of another process (X). Note that in this implementation, the memory length k that defines the 'past' is consistent between X and Y. The mathematical form is provided to the right of the schematic diagram and discussed in Sec. 6.2.

6.2.2 How:

As schematically depicted in Fig. 12, the CCE is defined as the entropy of a target process (Y) conditioned on both its own past and the past and present of a source process (X). More formally, for the stochastic processes X and Y, CCE quantifies the uncertainty in the future of Y given its own history and the history of X.

Following Equation 5 in Ziebart et al. [81], we define the time-indexed formulation of CCE, $H(X\|Y)(k)$, as Ziebart et al. [81] as

$$H(Y||X)(k) = H(Y_{t+1}|Y_t^{(k)}, X_{t+1}^{(k+1)}), \tag{20}$$

where $Y_t^{(k)}$ denotes the past k values of Y, and $X_{t+1}^{(k+1)}$ denotes the past k values plus the present value of X up to time $t+1$. This formulation captures the conditional entropy of Y at the next time point ($t+1$), given its own past *and* the past and present of X.

6.2.3 Why:

If X and Y represent the activity of two brain regions, a larger $H(X\|Y)$ means that observing past values of both regions does not reduce the uncertainty in the present value of Y by so much—suggesting a weaker dependency between the present value of Y and its own past, along with the past values of X. Figure 4H depicts the CCE for all brain regions in the left cortex conditioned on our region of focus, the LOC, in one example HCP participant—computed using the Kozachenko–Leonenko estimator. Here, we observe the maximum CCE in the parahippocampal gyrus, which therefore exhibits the greatest remaining uncertainty in its present value after conditioning on its own past and the past of the LOC in this example fMRI dataset. Our observation that the parahippocampal gyrus exhibits minimal self-predictive information as defined by the AIS (cf. Fig. 4E) would be a key contributing factor to this. We might also expect the parahippocampal gyrus to exhibit relatively low MI and TLMI with the time-lagged LOC, although this is not the case in Fig. 4G. Note that MI and TLMI each measure the influence of only one value of the time series we condition on, while our implementation of CCE (using the *pyspi* library [16, 17]) uses $k = 5$ past values jointly, such that values are not directly comparable anyway. Altogether, the high CCE indicates that the current state of the parahippocampal gyrus depends relatively weakly on its own past and the past BOLD activity in the LOC (when considered jointly) for this participant. The LOC exhibited the lowest CCE with the supramamarginal gyrus, which similarly reflects the supramamarginal gyrus ranking among the top of

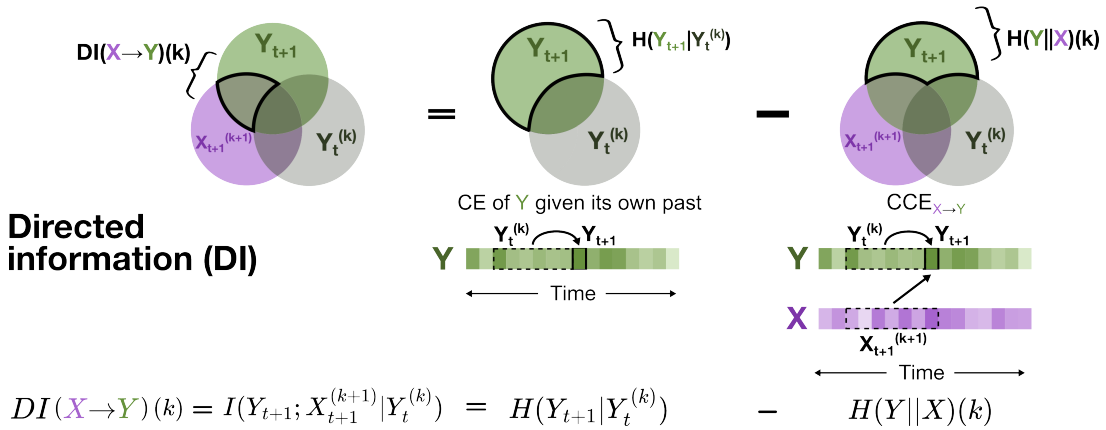
all evaluated left cortical regions for AIS values (cf. Fig. 4E)—such that it exhibits high self-predictive information, and the past BOLD activity in the LOC further reduces uncertainty in the current state of the supramarginal gyrus in this individual.

Prior applications in computational neuroscience have predominantly incorporated CCE as a building block for the directed information measure (cf. Sec. 6.3) [83–85]. Through a comprehensive evaluation of over 200 pairwise time-series measures (including several information-theoretic measures) [17], CCE emerged as a top performer in two classification case studies using EEG (to distinguish cortical 'up' or 'down' states, as in Birbaumer et al. [86]) and fMRI (to distinguish resting versus film-viewing, as in Byrge and Kennedy [87]). By contrast, Karimi-Rouzbahani and McGonigal [88] evaluated CCE along with several other directed connectivity measures to characterize neural dynamics in the seizure onset zone during and outside of epileptogenic activity, finding minimal support for predicted patterns with CCE compared to other methods.

6.3 Directed information

6.3.1 What:

Directed information (DI) is a measure of information flow introduced by Massey [89] to quantify how much information is shared between the present value of one process and the past *and present* of another process. In other words, $\mathrm{DI}(X \to Y)$ captures the total directional influence of X on the present value of Y beyond what can be explained by the past of Y on its own.



$$DI(X \to Y)(k) = I(Y_{t+1}; X_{t+1}^{(k+1)} | Y_t^{(k)}) \quad = \quad H(Y_{t+1} | Y_t^{(k)}) \quad - \quad H(Y||X)(k)$$

Figure 13. Directed information. The sub-component $DI(X \to Y)(k)$ of directed information (DI) quantifies the information shared between the present of the target time series process (Y) and the past and present of another source process (X) after conditioning on the probability distribution of Y's past. Note that in this implementation, the memory length (k) that defines the 'past' is consistent between X and Y. The mathematical form is provided below the schematic diagram and discussed in Sec. 6.3.

6.3.2 How:

For each of the T samples of the present of Y ($Y_{t+1}, t \in [0, T-1]$, where T is the length of the processes X and Y and the first sample is $t = 1$), DI sums the mutual information from the entire past and presence of X (denoted as $X_{t+1}^{(t+1)}$) after conditioning on the entire past of Y ($Y_t^{(t)}$):

$$= \sum_{t=0}^{T-1} I(Y_{t+1}; X_{t+1}^{(t+1)} | Y_t^{(t)}), \tag{21}$$

$$= \sum_{t=0}^{T-1} H(Y_{t+1}|Y_t^{(t)}) - H(Y_{t+1}|Y_t^{(t)}, X_{t+1}^{(t+1)}), \tag{22}$$

$$= \sum_{t=0}^{T-1} H(Y_{t+1}|Y_t^{(t)}) - H(Y||X)(t), \tag{23}$$

where Eq. (21) formulates DI in terms of the conditional mutual information between the present of Y (Y_{t+1}) and the entire past and presence of X ($X_{t+1}^{(t+1)}$), conditioned on the past of Y up to time t ($Y_t^{(t)}$, with $Y_t^{(0)}$ as a null variable by convention). This can be decomposed, as in Eq. (22), into the conditional entropy of Y conditioned on its own past ($H(Y_{t+1}|Y_t^{(t)})$)—which is a form of entropy rate [2, Section 4.2])—minus the conditional entropy of Y conditioned on the past of both X and Y ($H(Y_{t+1}|Y_t^{(t)}, X_{t+1}^{(t+1)})$). The subtracted term in Eq. (22) was introduced in 'Causally conditioned entropy' (cf. Sec. 6.2 [90]) as, equivalently, the CCE with Y conditioned on X. The CCE notation, $H(Y\|X)$, is included in Eq. (23), with the addition of (t) to denote the memory length.

As discussed in 'Active information storage' (cf. Sec. 4.1)—and particularly relevant in the bivariate case—estimating this quantity can be empirically challenging due to the high dimensionality of joint past vectors and the fact that only a single realization of the time series exists per time point t. To address this, the implementation in *pyspi* [16, 17] imposes a maximum past window length $K \ll T$, where the default is $K = 5$. This constraint allows the algorithm to pool information across the full time series using an ergodic assumption, enabling the computation of a single conditional MI estimate $\mathrm{DI}(X{\to}Y)(k)$ for each window length k (up to K), rather than estimating it separately for each time point. This is given below in Eq. (24), and carried throughout subsequent equations:[4]:

$$DI(X{\to}Y)(k < K) = \sum_{k=0}^{K-1} DI(X{\to}Y)(k), \tag{24}$$

$$= \sum_{k=0}^{K-1} I(Y_{t+1}; X_{t+1}^{(k+1)} | Y_t^{(k)}), \tag{25}$$

$$= \sum_{k=0}^{K-1} H(Y_{t+1}|Y_t^{(k)}) - H(Y_{t+1}|Y_t^{(k)}, X_{t+1}^{(k+1)}), \tag{26}$$

$$= \sum_{k=0}^{K-1} H(Y_{t+1}|Y_t^{(k)}) - H(Y||X)(k). \tag{27}$$

The component $\mathrm{DI}(X{\to}Y)(k)$ of $\mathrm{DI}(X{\to}Y)$ for a given past window length k is depicted schematically in Fig. 13. Note that the inclusion of the present state of the source (X, in this case) is one difference between DI and transfer entropy (discussed in the next section, cf. Sec. 6.4), whilst another is the sum over time points in DI [91, App. C].

6.3.3 Why:

DI provides a perspective of information flow that encompasses both lagged and contemporaneous dependence between the source and target time series. For computational neuroscience, this is particularly useful in mapping how information is stored and propagated across spatial and temporal scales [92]. Malladi et al. [83] used DI to track seizure spread from the epileptogenic zone outwards using electrocorticography (ECoG) recordings in a patient with epilepsy. Other work has demonstrated that DI can model information flows from neural spike train recordings [90, 93], fMRI time series [22], and EEG time series [94, 95]. If X and Y represent the activity of two

[4]The quantity $H(Y_{t+1}|Y_t^{(k)})$ is approximated as $H(Y_{t+1}^{(k+1)})/(k+1)$ in *pyspi* [17], which becomes correct asymptotically as $k \to \infty$ for stationary stochastic processes [2, Theorem 4.2.1].

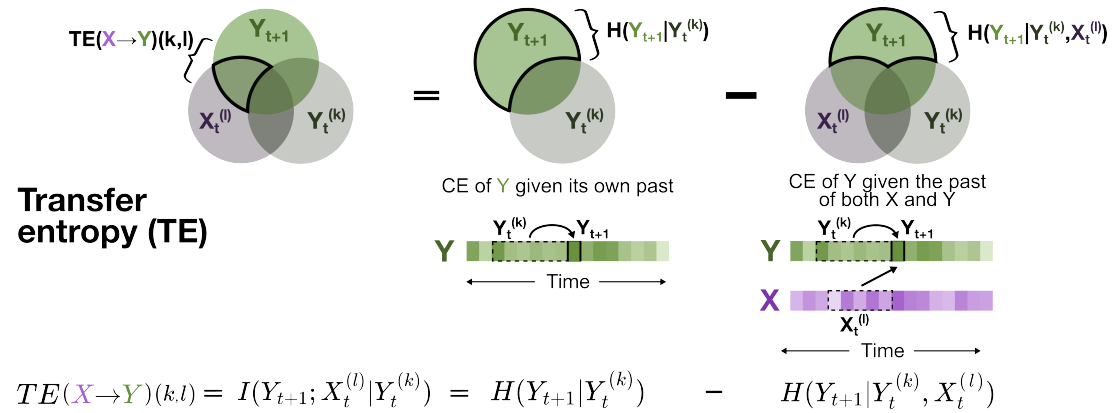
brain regions, then $\mathrm{DI}(X \to Y)$ captures the information shared between the present of Y and the past *and present* of X, after conditioning on the past of Y. A larger value of $\mathrm{DI}(X \to Y)$ indicates that knowing the past and present values of activity in region X reduces substantial uncertainty about the present value of region Y beyond that which is reduced by Y's own past. By contrast, a smaller value means that the past and present of activity in region X does not reduce additional uncertainty in Y beyond its own past—which could indicate minimal directed dependence from X to Y (i.e., a larger CCE $X \to Y$), and/or that the past of Y itself maximally reduces uncertainty about the next state of Y (i.e., a smaller CE of Y with its own past)—as schematically depicted in Fig. 13.

In Fig. 4I, we examine the DI from the LOC (source) to all other target regions in the left cortex for BOLD fMRI time series from an example HCP participant, computed with the Kozachenko–Leonenko estimator. This reveals that the maximum DI occurs from the LOC to the inferior parietal cortex, suggesting that the past and present values of the LOC maximally reduce uncertainty remaining in the inferior parietal cortex activity after conditioning on its own past in this example fMRI dataset. In other words, the LOC past and present BOLD activity is most informative of the present inferior parietal cortex activity after accounting for its own prior activity. By contrast, the LOC exhibits minimal DI to the entorhinal cortex, which is consistent with high CCE from the LOC to the entorhinal cortex shown in Fig. 4H—meaning that the LOC does not reduce much of the uncertainty remaining in the entorhinal cortex after conditioning on its own past, indicating minimal directed dependence between the two regions in this individual.

6.4 Transfer entropy

6.4.1 What:

Transfer entropy is a measure of information flow introduced by Schreiber [72] (see also Bossomaier et al. [96]) to capture the directed dependence from one process (X) to another (Y) in the context of the past of Y. Specifically, TE measures the amount of information shared between the *past values* of X (up to history length l) and the present of Y after conditioning on the past k values of Y. As mentioned above (see 'Directed information', Sec. 6.3), DI and TE are disambiguated by their inclusion (DI) or exclusion (TE) of the present value of X. By not including the present value of the source X in the model for Y, TE alone is akin to building a model of how the target process Y is generated or computed from the past of both [91]. Note that unlike previous order-dependent pairwise features (for which the history length k is consistent between both X and Y), TE allows the history length to differ between the two time series. In other words, $\mathrm{TE}(X \to Y)$ captures the total directional influence of the past l values of X on the present value of Y beyond what can be explained by Y's own past (up to length k). As with k, the memory length l for X can be set to a single value or a range of values can be compared.



$$TE(X \to Y)(k,l) = I(Y_{t+1}; X_t^{(l)} | Y_t^{(k)}) = H(Y_{t+1} | Y_t^{(k)}) - H(Y_{t+1} | Y_t^{(k)}, X_t^{(l)})$$

Figure 14. Transfer entropy. Transfer entropy (TE) quantifies the information shared between the present of one target process (X) and the past of another (X) after conditioning on the probability distribution over the past of Y. Note that in this implementation, X and Y can be assigned different memory lengths (l and k, respectively) that define the past period. The mathematical form is provided below the schematic diagram and discussed in Sec. 6.4.

6.4.2 How:

As depicted schematically in Fig. 14, TE is given as

$$\mathrm{TE}(X \to Y)(k,l) = I(Y_{t+1}; X_t^{(l)} | Y_t^{(k)}) \quad \textbf{(28)}$$

$$= H(Y_{t+1}|Y_t^{(k)}) - H(Y_{t+1}|Y_t^{(k)}, X_t^{(l)}), \quad \textbf{(29)}$$

which presents TE as the mutual information (cf. Sec. 2.2) between the present of Y (Y_{t+1}) and past of X ($X_t^{(l)}$) after conditioning on the past of Y ($Y_t^{(k)}$) in Eq. (28). Equivalently, as in Eq. (29), TE can be formulated as the CE (cf. Sec. 3.1) of Y given its own past ($H(Y_{t+1}|Y_t^{(k)})$) minus the CE of Y given the past of both itself ($Y_t^{(k)}$) and X ($X_t^{(l)}$). As per the other measures here, TE may be extended for multivariate sources and targets, and may be made conditional on other source variables [22, 97, 98]. Unlike with the DI implementation, the TE implementation in *pyspi* [16, 17] (in utilizing *JIDT* [15]) allows for both fixed default history lengths for X (l) and Y (k), as well as different delay lag (τ) values for both X and Y (i.e., with the past vector $Y_t^{(k)}$ sampling every τ-th value, from Y_t backwards), as well as flexible optimization based on active information storage values (following [50, 99]). We have assumed $\tau = 1$ for both X and Y in this work for simplicity.

6.4.3 Why:

TE has become an increasingly popular measure of directed information flow for neural time series data in the last fifteen years. If X and Y represent activity of two brain regions, a larger $\mathrm{TE}(X \to Y)$ indicates that knowledge of the past activity of the source region (X) reduces more uncertainty about the present value of the target region (Y) after already knowing the past of Y, suggesting a stronger (potentially nonlinear) influence of the past of X on the present of Y. As a measure based on observations, transfer entropy measures contribution to predictability, rather than causal effect [82]. Wibral et al. [35] frame the 'causal' nature often attributed to TE (from this predictive effect) as the following question: Given we already know the past state of region Y, how much additional information does the past state of region X provide about the next state of Y?

As above, by not considering any contemporaneous dependence between X and Y, TE isolates influences of the past values of both X and Y on the future values of X or Y. This is in contrast to DI, which also includes contemporaneous dependence between X and Y. While TE is a principled choice for measuring how the target is computed or generated from the past of both time series, the time lag has an impact with functional modalities with a lower sampling frequency, such as BOLD fMRI (sampled on the order of seconds, compared to EEG or neural spike recordings with millisecond resolution), as it may be difficult to identify source–target relationships over such a large timescale in comparison to between contemporaneous samples. Given this, TE may be more practically suited to modalities with higher sampling rates, as the dynamic relationship between the past of X and the present of Y may be more fully captured by the data. TE has been applied largely to simulated data from neural mass models [38, 54, 100, 101] and biophysical synapse models and to empirical data with high temporal resolution, like EEG [102, 103], MEG [104] and neural spike trains [105, 106]. However, successful application to BOLD fMRI time series is indeed possible; for example, Lizier et al. [22] applied TE to measure information flows in visuomotor tracking tasks and how they changed with task difficulty, while Taylor et al. [107] applied TE to demonstrate functional attenuation of cholinergic region communication in the macaque brain after locally inactivating the nucleus basalis of Meynert.

In Fig. 4J, we depict TE values from the LOC as the source to all other target regions in the left cortex of an example HCP participant, computed using the Kraskov–Stögbauer–Grassberger estimator with $k = l = 1$. Here, we observe the highest TE from the LOC to the medial orbitofrontal cortex in this example fMRI dataset, suggesting that the past activity in the LOC maximally reduces uncertainty in future activity of the medial orbitofrontal cortex (after conditioning on its own past BOLD fMRI activity). By contrast, the LOC exhibited the lowest overall TE to the superior parietal cortex, indicating minimal time-lagged dependence between these two regions beyond information shared between the past and present activity in the superior parietal cortex itself (which was marked with a relatively high AIS) in this participant.

6.5 Granger causality

6.5.1 What:

Granger causality (GC) is a measure of information flow introduced by Granger [108] that quantifies the predictive power of the past values of one process (X) adds to an autoregressive model fit to another process (Y). In other words, GC captures how much better a model fit to the past of Y can predict the future of Y when it also has access to the past of X. While not necessarily a pure information-theoretic measure by traditional definition, GC is equivalent to TE (see 'Transfer entropy', Sec. 6.4) with a Gaussian density estimator up to a multiplicative factor [109] and is often applied to characterize brain dynamics and connectivity [110], motivating its inclusion here.

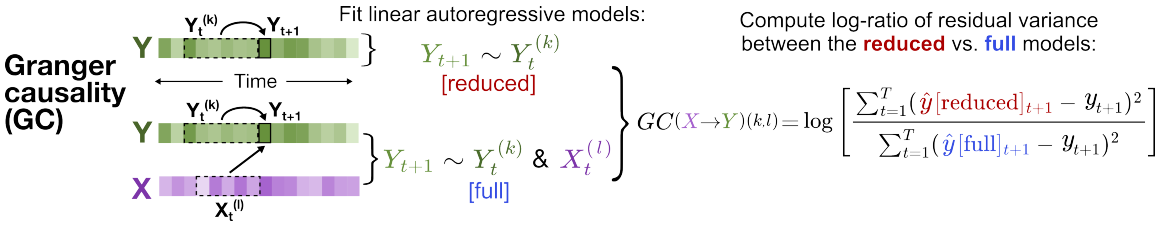


Figure 15. Granger causality. Granger causality (GC) quantifies the predictive information that the past of one time series (e.g., X) adds to an autoregressive model for the other time series (e.g., Y) beyond that of Y's own past. Note that in this implementation, X and Y can be assigned different memory lengths (l and k, respectively) that define the past period. The mathematical form is provided to the right of the schematic diagram and discussed in Sec. 6.5.

6.5.2 How:

As with TE, GC allows for the history length to differ between X (denoted as l) and Y (denoted as k). Specifically, GC measures the log-ratio of the residual variance for the predicted value of Y_{t+1} based solely on the past values of Y up to window length k ($Y_t^{(k)}$; 'reduced' model) versus that of the predicted value of Y_{t+1} based on the past l values of X ($X_t^{(l)}$) as well the past k values of Y ('full' model). As schematically depicted in Fig. 15, GC is given as

$$\mathrm{GC}(X \to Y)(k,l) = \log \left[\frac{\sum_{t=1}^{T} (\hat{y}[\text{reduced}]_{t+1} - y_{t+1})^2}{\sum_{t=1}^{T} (\hat{y}[\text{full}]_{t+1} - y_{t+1})^2} \right], \tag{30}$$

where $\hat{y}$ is the vector of values that the autoregressive model predicted for the realization y at each time point $t+1$, in which $\hat{y}[\text{reduced}]$ corresponds to the 'reduced' model (i.e., only the past of y) and $\hat{y}[\text{full}]$ corresponds to the 'full' model (i.e., the past of both realizations y and x). The numerator in the logarithmic fraction in Eq. (30) corresponds to the sum of squared differences between each $\hat{y}_{t+1}$ (predicted) minus y_{t+1} (real) value—that is, the squared residuals of the 'reduced' model. Similarly, the denominator contains the sum of the squared residuals from the 'full' model.

In *pyspi* [16, 17], GC is computed from the same code as TE with Gaussian density estimator due to their equivalence as above.[5] For an in-depth discussion about the similarities between GC and TE, we direct the interested reader to Milinkovic et al. [111]. Of note, Granger causality can also be computed in the frequency domain, as spectral Granger causality [112]. This allows the contribution of different frequencies to the dependence to be inferred.

6.5.3 Why:

GC is very widely used as an index of 'causal' interactions between brain areas in the computational neuroscience literature [110, 113, 114]. Like transfer entropy, 'causal' here is with reference to a *predictive* model rather than considering interventional notions. If X and Y represent the activity of two brain regions, then $\mathrm{GC}(X \to Y)$ captures the additional predictive information that the past of region X provides for the future of region Y beyond the activity of Y alone. Multivariate extensions have also been developed in the pursuit of 'full-brain connectivity' representations [115, 116].

[5] In principle, Granger causality is a unitless measure; however, given the known equivalence to TE above, it is often expressed in nats when computed in that way.

In Fig. 4K, we examine the GC values from the LOC as the source to all other (target) regions in the left cortex (using $k = l = 1$), noting the maximum value detected in the fusiform gyrus. This indicates that incorporating information about the past activity of the LOC contributes more predictive information to the future of the fusiform gyrus than an autoregressive model fit to the past activity of the fusiform gyrus alone. Of note, the fusiform gyrus exhibited the maximal entropy of all evaluated left cortical regions in this participant (cf. Fig. 4A), suggesting that activity in this region is more unpredictable over time—which provides plenty of uncertainty for the past activity of the LOC to potentially reduce when it is included. We also note that Granger causality assumes a joint Gaussian bivariate distribution, unlike the pure (conventional) information-theoretic measures covered here, especially the TE, that do not inherently assume the form of underlying data (see 'A note on estimating MI in continuous space', Sec. 2.2). By contrast, we observe the lowest GC from the LOC to the lingual gyrus, indicating that the past activity in the LOC does not substantially reduce uncertainty in the future of the lingual gyrus beyond its own past activity. This is particularly interesting because the LOC and lingual gyrus exhibited the maximal MI in Fig. 4C, which indicates that this inter-regional relationship is stronger with a contemporaneous measure that does not condition out the past of the target, and does not assume a joint bivariate Gaussian distribution. Moreover, we note the contrast in results between GC in Fig. 4K and TE in Fig. 4J—with the two measures only differing in their estimators, this contrast reflects the difference observed in information flow when only modeling linear interactions (GC) versus non-linear (TE).

References

1. C. E. Shannon. A mathematical theory of communication. *The Bell system technical journal* **27**, 379 (1948).
2. T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience (2005).
3. D. J. MacKay. *Information theory, inference and learning algorithms*. Cambridge University Press (2003).
4. M. S. Harré. Information theory for agents in artificial intelligence, psychology, and economics. *Entropy* **23**, 310 (2021).
5. K. V. Parag and C. A. Donnelly. Using information theory to optimise epidemic models for real-time prediction and estimation. *PLoS computational biology* **16**, e1007990 (2020).
6. A. Borst and F. E. Theunissen. Information theory and neural coding. *Nature neuroscience* **2**, 947 (1999).
7. N. M. Timme and C. Lapish. A tutorial for information theory in neuroscience. *eneuro* **5** (2018).
8. H. Pinkard and L. Waller. A visual introduction to information theory. *arXiv preprint arXiv:2206.07867* (2022).
9. C. Huerta, C. Greve, and A. Wong. Comparison of causality determination techniques in studying Hall-effect thrusters. *Journal of Electric Propulsion* **3**, 23 (2024).
10. A. Lozano-Durán and G. Arranz. Information-theoretic formulation of dynamical systems: Causality, modeling, and control. *Physical Review Research* **4**, 023195 (2022).
11. C. Murphy, V. Thibeault, A. Allard, and P. Desrosiers. Duality between predictability and reconstructability in complex systems. *Nature Communications* **15**, 4478 (2024).
12. H. Kim, G. Valentini, J. Hanson, and S. I. Walker. Informational architecture across non-living and living collectives. *Theory in Biosciences* pp. 1–17 (2021).
13. E. Bobek and B. Tversky. Creating visual explanations improves learning. *Cognitive research: principles and implications* **1**, 1 (2016).
14. C. Finn and J. T. Lizier. Generalised measures of multivariate information content. *Entropy* **22**, 216 (2020).
15. J. T. Lizier. JIDT: An information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI* **1**, 11 (2014).

16. O. Cliff, A. G. Bryant, B. Fulcher, and A. Nguyen. DynamicsAndNeuralSystems/pyspi (2023). URL: https://doi.org/10.5281/zenodo.8223340.

17. O. M. Cliff, A. G. Bryant, J. T. Lizier, N. Tsuchiya, and B. D. Fulcher. Unifying pairwise interactions in complex dynamics. *Nature Computational Science* pp. 1–11 (2023).

18. D. C. Van Essen, S. M. Smith, D. M. Barch, et al. The WU-Minn Human Connectome Project: An overview. *Neuroimage* **80**, 62 (2013).

19. D. P. Shorten, R. E. Spinney, and J. T. Lizier. Estimating transfer entropy in continuous time between neural spike trains or other event-based data. *PLoS computational biology* **17**, e1008054 (2021).

20. G. Mijatovic, Y. Antonacci, T. Loncar-Turukalo, L. Minati, and L. Faes. An information-theoretic framework to measure the dynamic interaction between neural spike trains. *IEEE Transactions on Biomedical Engineering* **68**, 3471 (2021).

21. C. Magri, K. Whittingstall, V. Singh, N. K. Logothetis, and S. Panzeri. A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. *BMC neuroscience* **10**, 1 (2009).

22. J. T. Lizier, J. Heinzle, A. Horstmann, J.-D. Haynes, and M. Prokopenko. Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *Journal of computational neuroscience* **30**, 85 (2011).

23. Z. Wang, A. Alahmadi, D. C. Zhu, and T. Li. Causality analysis of fMRI data based on the directed information theory framework. *IEEE Transactions on Biomedical Engineering* **63**, 1002 (2015).

24. J. Fallon, P. G. Ward, L. Parkes, et al. Timescales of spontaneous fMRI fluctuations relate to structural connectivity in the brain. *Network Neuroscience* **4**, 788 (2020).

25. L. F. Kozachenko and N. N. Leonenko. Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii* **23**, 9 (1987).

26. R. Cofré and A. Destexhe. Entropy and complexity tools across scales in neuroscience: A review. *Entropy* **27**, 115 (2025).

27. A. Ponce-Alvarez, L. Uhrig, N. Deco, et al. Macroscopic quantities of collective brain activity during wakefulness and anesthesia. *Cerebral cortex* **32**, 298 (2022).

28. F. Cieri, X. Zhuang, J. Z. Caldwell, and D. Cordes. Brain entropy during aging through a free energy principle approach. *Frontiers in Human Neuroscience* **15**, 647513 (2021).

29. A. Baseer, S. J. Weddell, and R. D. Jones. Prediction of microsleeps using pairwise joint entropy and mutual information between EEG channels. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4495–4498. IEEE (2017).

30. E. A. Martin, J. Hlinka, A. Meinke, et al. Network inference and maximum entropy estimation on information diagrams. *Scientific reports* **7**, 7062 (2017).

31. T. Watanabe, S. Hirose, H. Wada, et al. A pairwise maximum entropy model accurately describes resting-state human brain networks. *Nature communications* **4**, 1370 (2013).

32. S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics* **22**, 79 (1951).

33. A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* **69**, 066138 (2004).

34. S. Frenzel and B. Pompe. Partial mutual information for coupling analysis of multivariate time series. *Physical review letters* **99**, 204101 (2007).

35. M. Wibral, R. Vicente, and M. Lindner. Transfer entropy in neuroscience. *Directed information measures in neuroscience* pp. 3–36 (2014).

36. R. Vicente and M. Wibral. Efficient estimation of information transfer. In *Directed information measures in neuroscience*, pp. 37–58. Springer (2014).

37. G. T. Raffaelli, S. Jiříček, and J. Hlinka. Nonlinear brain connectivity from neurons to networks: Quantification, sources and localization. *bioRxiv* pp. 2024–11 (2024).

38. L. Novelli and J. T. Lizier. Inferring network properties from time series using transfer entropy and mutual information: Validation of multivariate versus bivariate approaches. *Network Neuroscience* **5**, 373 (2021).

39. W. Zhang, V. Muravina, R. Azencott, Z. D. Chu, and M. J. Paldino. Mutual information better quantifies brain network architecture in children with epilepsy. *Computational and mathematical methods in medicine* **2018**, 6142898 (2018).

40. R. A. Ince, B. L. Giordano, C. Kayser, et al. A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula. *Human brain mapping* **38**, 1541 (2017).

41. V. Gómez-Verdejo, M. Martínez-Ramón, J. Florensa-Vila, and A. Oliviero. Analysis of fMRI time series with mutual information. *Medical image analysis* **16**, 451 (2012).

42. T. Wadhera and D. Kakkar. Conditional entropy approach to analyze cognitive dynamics in autism spectrum disorder. *Neurological Research* **42**, 869 (2020).

43. X. Liu and Z. Fu. A novel recognition strategy for epilepsy EEG signals based on conditional entropy of ordinal patterns. *Entropy* **22**, 1092 (2020).

44. L. M. Harrison, A. Duggins, and K. J. Friston. Encoding uncertainty in the hippocampus. *Neural Networks* **19**, 535 (2006).

45. J. Liu, J. Ji, G. Xun, and A. Zhang. Inferring effective connectivity networks from fMRI time series with a temporal entropy-score. *IEEE Transactions on Neural Networks and Learning Systems* **33**, 5993 (2021).

46. G. Vivekanandhan, M. Mehrabbeik, K. Rajagopal, et al. Higuchi fractal dimension is a unique indicator of working memory content represented in spiking activity of visual neurons in extrastriate cortex. *Mathematical Biosciences and Engineering: MBE* **20**, 3749 (2022).

47. J. T. Lizier, M. Prokopenko, and A. Y. Zomaya. Local measures of information storage in complex distributed computation. *Information Sciences* **208**, 39 (2012).

48. W. Bialek, I. Nemenman, and N. Tishby. Complexity through nonextensivity. *Physica A* **302**, 89 (2001).

49. M. Lindner, R. Vicente, V. Priesemann, and M. Wibral. TRENTOOL: A Matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC Neuroscience* **12**, 119 (2011).

50. E. Y. Erten, J. T. Lizier, M. Piraveenan, and M. Prokopenko. Criticality and Information Dynamics in Epidemiological Models. *Entropy* **19**, 194 (2017).

51. J. T. Lizier, S. Pritam, and M. Prokopenko. Information dynamics in small-world boolean networks. *Artificial life* **17**, 293 (2011).

52. C. Gómez, J. T. Lizier, M. Schaum, et al. Reduced predictable information in brain signals in autism spectrum disorder. *Frontiers in neuroinformatics* **8**, 9 (2014).

53. P. A. Mediano and M. Shanahan. Balanced information storage and transfer in modular spiking neural networks. *arXiv preprint arXiv:1708.04392* (2017).

54. M. Li, Y. Han, M. J. Aburn, et al. Transitions in information processing dynamics at the whole-brain network level are driven by alterations in neural gain. *PLoS computational biology* **15**, e1006957 (2019).

55. R. Zhu, J. Hochstetter, A. Loeffler, et al. Information dynamics in neuromorphic nanowire networks. *Scientific Reports* **11**, 13047 (2021).

56. E. Pinzuti, P. Wollstadt, O. Tüscher, and M. Wibral. Information theoretic evidence for layer- and frequency-specific changes in cortical information processing under anesthesia. *PLOS Computational Biology* **19**, e1010380 (2023).

57. M. Wibral, J. T. Lizier, S. Vögler, V. Priesemann, and R. Galuske. Local active information storage as a tool to understand distributed neural information processing. *Frontiers in neuroinformatics* **8**, 1 (2014).

58. A. Brodski-Guerniero, G.-F. Paasch, P. Wollstadt, et al. Information-theoretic evidence for predictive coding in the face-processing system. *Journal of Neuroscience* **37**, 8273 (2017).

59. A. Brodski-Guerniero, M. J. Naumer, V. Moliadze, et al. Predictable information in neural signals during resting state is reduced in autism spectrum disorder. *Human Brain Mapping* **39**, 3227 (2018).

60. A. C. Puche Sarmiento, Y. Bocanegra García, and J. F. Ochoa Gómez. Active information storage in Parkinson's disease: a resting state fMRI study over the sensorimotor cortex. *Brain imaging and behavior* **14**, 1143 (2020).

61. N. Ay and T. Wennekers. Temporal infomax leads to almost deterministic dynamical systems. *Neurocomputing* **52**, 461 (2003).

62. N. Ay. Information geometry on complexity and stochastic interaction. *Entropy* **17**, 2432 (2015).

63. D. Cohen, S. Sasai, N. Tsuchiya, and M. Oizumi. A general spectral decomposition of causal influences applied to integrated information. *Journal of neuroscience methods* **330**, 108443 (2020).

64. M. Oizumi, N. Tsuchiya, and S.-i. Amari. Unified framework for information integration based on information geometry. *Proceedings of the National Academy of Sciences* **113**, 14817 (2016).

65. P. A. Mediano, A. K. Seth, and A. B. Barrett. Measuring integrated information: Comparison of candidate measures in theory and simulation. *Entropy* **21**, 17 (2018).

66. T. Wennekers and N. Ay. Finite state automata resulting from temporal information maximization and a temporal learning rule. *Neural computation* **17**, 2258 (2005).

67. A. B. Barrett and A. K. Seth. Practical measures of integrated information for time-series data. *PLoS computational biology* **7**, e1001052 (2011).

68. T. Wennekers and N. Ay. Stochastic interaction in associative nets. *Neurocomputing* **65**, 387 (2005).

69. T. Wennekers, N. Ay, and P. Andras. High-resolution multiple-unit EEG in cat auditory cortex reveals large spatio-temporal stochastic interactions. *BioSystems* **89**, 190 (2007).

70. J. Kitazono, R. Kanai, and M. Oizumi. Efficient algorithms for searching the minimum information partition in integrated information theory. *Entropy* **20**, 173 (2018).

71. Z.-Q. Liu, A. I. Luppi, J. Y. Hansen, et al. Benchmarking methods for mapping functional connectivity in the brain. *bioRxiv* pp. 2024–05 (2024).

72. T. Schreiber. Measuring information transfer. *Physical Review Letters* **85**, 461 (2000). Publisher: APS.

73. V. Chaitankar, P. Ghosh, E. J. Perkins, P. Gong, and C. Zhang. Time lagged information theoretic approaches to the reverse engineering of gene regulatory networks. In *BMC bioinformatics*, volume 11, pp. 1–14. Springer (2010).

74. R. E. Greenblatt, M. Pflieger, and A. Ossadtchi. Connectivity measures applied to human brain electrophysiological data. *Journal of neuroscience methods* **207**, 1 (2012).

75. F. Von Wegner, H. Laufs, and E. Tagliazucchi. Mutual information identifies spurious Hurst phenomena in resting state EEG and fMRI data. *Physical Review E* **97**, 022415 (2018).

76. H. E. Hurst. Long-term storage capacity of reservoirs. *Transactions of the American society of civil engineers* **116**, 770 (1951).

77. J. I. Chapeton, S. K. Inati, and K. A. Zaghloul. Stable functional networks exhibit consistent timing in the human brain. *Brain* **140**, 628 (2017).

78. F. von Wegner, S. Bauer, F. Rosenow, J. Triesch, and H. Laufs. Eeg microstate periodicity explained by rotating phase patterns of resting-state alpha oscillations. *Neuroimage* **224**, 117372 (2021).

79. G. Hermann, I. Tödt, E. Tagliazucchi, et al. Propofol reversibly attenuates short-range microstate ordering and 20 Hz microstate oscillations. *Brain Topography* **37**, 329 (2024).

80. P.-O. Amblard and O. J. J. Michel. On directed information theory and Granger causality graphs. *Journal of Computational Neuroscience* **30**, 7 (2011).

81. B. D. Ziebart, J. A. Bagnell, and A. K. Dey. The principle of maximum causal entropy for estimating interacting processes. *IEEE Transactions on Information Theory* **59**, 1966 (2013).

82. J. T. Lizier and M. Prokopenko. Differentiating information transfer and causal effect. *European Physical Journal B* **73**, 605 (2010).

83. R. Malladi, G. P. Kalamangalam, N. Tandon, and B. Aazhang. Inferring causal connectivity in epileptogenic zone using directed information. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 822–826. IEEE (2015).

84. N. Soltani and A. J. Goldsmith. Directed information between connected leaky integrate-and-fire neurons. *IEEE Transactions on Information Theory* **63**, 5954 (2017).

85. Z. Cai, C. L. Neveu, D. A. Baxter, J. H. Byrne, and B. Aazhang. Inferring neuronal network functional connectivity with directed information. *Journal of neurophysiology* **118**, 1055 (2017).

86. N. Birbaumer, N. Ghanayim, T. Hinterberger, et al. A spelling device for the paralysed. *Nature* **398**, 297 (1999).

87. L. Byrge and D. P. Kennedy. Accurate prediction of individual subject identity and task, but not autism diagnosis, from functional connectomes. *Human brain mapping* **41**, 2249 (2020).

88. H. Karimi-Rouzbahani and A. McGonigal. Directionality of neural activity in and out of the seizure onset zone in focal epilepsy. *Network Neuroscience* pp. 1–26 (2025).

89. J. Massey. Causality, feedback and directed information. In *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*, pp. 303–305. Citeseer (1990).

90. C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos. Estimating the directed information to infer causal relationships in ensemble neural spike train recordings. *Journal of computational neuroscience* **30**, 17 (2011).

91. J. T. Lizier. *The Local Information Dynamics of Distributed Computation in Complex Systems*. Springer Theses. Springer, Berlin / Heidelberg (2013).

92. M. Wibral, R. Vicente, and J. T. Lizier. *Directed information measures in neuroscience*, volume 724. Springer (2014).

93. K. So, A. C. Koralek, K. Ganguly, M. C. Gastpar, and J. M. Carmena. Assessing functional connectivity of neural ensembles using directed information. *Journal of neural engineering* **9**, 026004 (2012).

94. Y. Liu and S. Aviyente. Quantification of effective connectivity in the brain using a measure of directed information. *Computational and mathematical methods in medicine* **2012**, 635103 (2012).

95. K. Mehta and J. Kliewer. Directional and causal information flow in EEG for assessing perceived audio quality. *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications* **3**, 150 (2017).

96. T. Bossomaier, L. Barnett, M. Harré, and J. T. Lizier. *An Introduction to Transfer Entropy: Information Flow in Complex Systems*. Springer, Cham, Switzerland (2016).

97. J. T. Lizier, M. Prokopenko, and A. Y. Zomaya. Local information transfer as a spatiotemporal filter for complex systems. *Physical Review E* **77**, 026110 (2008).

98. J. T. Lizier, M. Prokopenko, and A. Y. Zomaya. Information modification and particle collisions in distributed computation. *Chaos* **20**, 037109 (2010).

99. J. Garland, R. G. James, and E. Bradley. Leveraging information storage to select forecast-optimal parameters for delay-coordinate reconstructions. *Physical Review E* **93**, 022221 (2016).

100. M. Ursino, G. Ricci, and E. Magosso. Transfer entropy as a measure of brain connectivity: A critical analysis with the help of neural mass models. *Frontiers in computational neuroscience* **14**, 45 (2020).

101. M. Conrad and R. B. Jolivet. Comparative performance of mutual information and transfer entropy for analyzing the balance of information flow and energy consumption at synapses. *bioRxiv* pp. 2020–06 (2020).

102. M. H. I. Shovon, N. Nandagopal, R. Vijayalakshmi, J. T. Du, and B. Cocks. Directed connectivity analysis of functional brain networks during cognitive activity using transfer entropy. *Neural Processing Letters* **45**, 807 (2017).

103. N. Yin, H. Wang, Z. Wang, et al. A study of brain networks associated with freezing of gait in Parkinson's disease using transfer entropy analysis. *Brain Research* **1821**, 148610 (2023).

104. J. S. Chan, M. Wibral, C. Stawowsky, et al. Predictive coding over the lifespan: increased reliance on perceptual priors in older adults—a magnetoencephalography and dynamic causal modeling study. *Frontiers in aging neuroscience* **13**, 631599 (2021).

105. S. P. Faber, N. M. Timme, J. M. Beggs, and E. L. Newman. Computation is concentrated in rich clubs of local cortical networks. *Network Neuroscience* **3**, 384 (2019).

106. D. P. Shorten, V. Priesemann, M. Wibral, and J. T. Lizier. Early lock-in of structured and specialised information flows during neural development. *eLife* **11**, e74651 (2022).

107. N. L. Taylor, C. J. Whyte, B. R. Munn, et al. Causal evidence for cholinergic stabilization of attractor landscape dynamics. *Cell reports* **43** (2024).

108. C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* pp. 424–438 (1969).

109. L. Barnett, A. B. Barrett, and A. K. Seth. Granger causality and transfer entropy are equivalent for Gaussian variables. *Physical Review Letters* **103**, 238701 (2009). Publisher: APS.

110. A. K. Seth, A. B. Barrett, and L. Barnett. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience* **35**, 3293 (2015).

111. B. Milinkovic, L. Barnett, O. Carter, A. K. Seth, and T. Andrillon. Capturing the emergent dynamical structure in biophysical neural models. *PLOS Computational Biology* **21**, e1012572 (2025).

112. J. F. Geweke. Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association* **79**, 907 (1984).

113. S. L. Bressler and A. K. Seth. Wiener–Granger causality: A well established methodology. *Neuroimage* **58**, 323 (2011).

114. K. Friston, R. Moran, and A. K. Seth. Analysing connectivity with Granger causality and dynamic causal modelling. *Current opinion in neurobiology* **23**, 172 (2013).

115. W. Liao, J. Ding, D. Marinazzo, et al. Small-world directed networks in the human brain: Multivariate granger causality analysis of resting-state fMRI. *Neuroimage* **54**, 2683 (2011).

116. C. Schmidt, B. Pester, N. Schmid-Hertel, et al. A multivariate Granger causality concept towards full brain functional connectivity. *PloS one* **11**, e0153105 (2016).

117. H. A. Dau, A. Bagnall, K. Kamgar, et al. The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica* **6**, 1293 (2019).

118. A. Bagnall, H. A. Dau, J. Lines, et al. The UEA multivariate time series classification archive, 2018. *arXiv preprint arXiv:1811.00075* (2018).

119. L. Parkes, B. Fulcher, M. Yücel, and A. Fornito. Transcriptional signatures of connectomic subregions of the human striatum. *Genes, Brain and Behavior* **16**, 647 (2017).

120. R. S. Desikan, F. Ségonne, B. Fischl, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968 (2006).

Appendix 2: Full manuscript of the *pyspi* paper (Chapter 3)

nature computational science

Article

https://doi.org/10.1038/s43588-023-00519-x

Unifying pairwise interactions in complex dynamics

Received: 1 July 2022

Accepted: 14 August 2023

Published online: 25 September 2023

Check for updates

Oliver M. Cliff[1,2], Annie G. Bryant[1,2], Joseph T. Lizier[2,3], Naotsugu Tsuchiya[4,5,6] & Ben D. Fulcher[1,2]

Scientists have developed hundreds of techniques to measure the interactions between pairs of processes in complex systems, but these computational methods—from contemporaneous correlation coefficients to causal inference methods—define and formulate interactions differently, using distinct quantitative theories that remain largely disconnected. Here we introduce a large assembled library of 237 statistics of pairwise interactions, and assess their behavior on 1,053 multivariate time series from a wide range of real-world and model-generated systems. Our analysis highlights commonalities between disparate mathematical formulations of interactions, providing a unified picture of a rich interdisciplinary literature. Using three real-world case studies, we then show that simultaneously leveraging diverse methods can uncover those most suitable for addressing a given problem, facilitating interpretable understanding of the quantitative formulation of pairwise dependencies that drive successful performance. Our results and accompanying software enable comprehensive analysis of time-series interactions by drawing on decades of diverse methodological contributions.

A fundamental question in science is how complex dynamics can be characterized by measuring the interactions within a distributed system. To address this question, many approaches have been developed to measure different types of pairwise interactions from dynamical data. For example, in neuroimaging, functional connections between pairs of brain regions are quantified through statistical correlations, which mark changes in human behaviors[1] and differ in neurological diseases[2]. In Earth system science, pairwise causal models have been used to infer mechanistic drivers of natural processes, from the influence of sea-surface temperature on sardine and anchovy populations[3], to the atmospheric drivers of air circulation[4]. Furthermore, economic analysts have studied the co-integration of paired non-stationary time series—including stock-market indices and their associated future contracts—to infer a statistically significant coupling for building econometric models[5].

As illustrated schematically in Fig. 1a, the common goal of these studies is to extract meaningful pairwise relationships from multivariate time series (MTS): sets of observations taken regularly over time[6]. In the age of big data, the scientific problems that are studied in diverse disciplinary contexts—from genomics to astronomy[7]—require novel ways to extract information from MTS data; however, despite the myriad ways to quantify a statistical dependency between two time series, it remains common practice to manually select a single method with minimal comparisons against alternatives. For instance, Pearson correlation remains the most commonly used tool for measuring pairwise relationships in neuroimaging[8] and Earth system science[4], despite rather restrictive (and often unsatisfied) assumptions that the data are serially independent and normally distributed[9]. Fortunately, many more sophisticated and powerful algorithms have been developed, including those where dependencies are lagged in time (for example,

[1]School of Physics, The University of Sydney, Camperdown, New South Wales, Australia. [2]Centre for Complex Systems, The University of Sydney, Camperdown, New South Wales, Australia. [3]School of Computer Science, The University of Sydney, Camperdown, New South Wales, Australia. [4]Turner Institute for Brain and Mental Health & School of Psychological Sciences, Faculty of Medicine, Nursing, and Health Sciences, Monash University, Melbourne, Victoria, Australia. [5]Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), Suita-shi, Japan. [6]Advanced Telecommunications Research Computational Neuroscience Laboratories, Seika-cho, Japan. e-mail: ben.fulcher@sydney.edu.au

https://doi.org/10.1038/s43588-023-00519-x

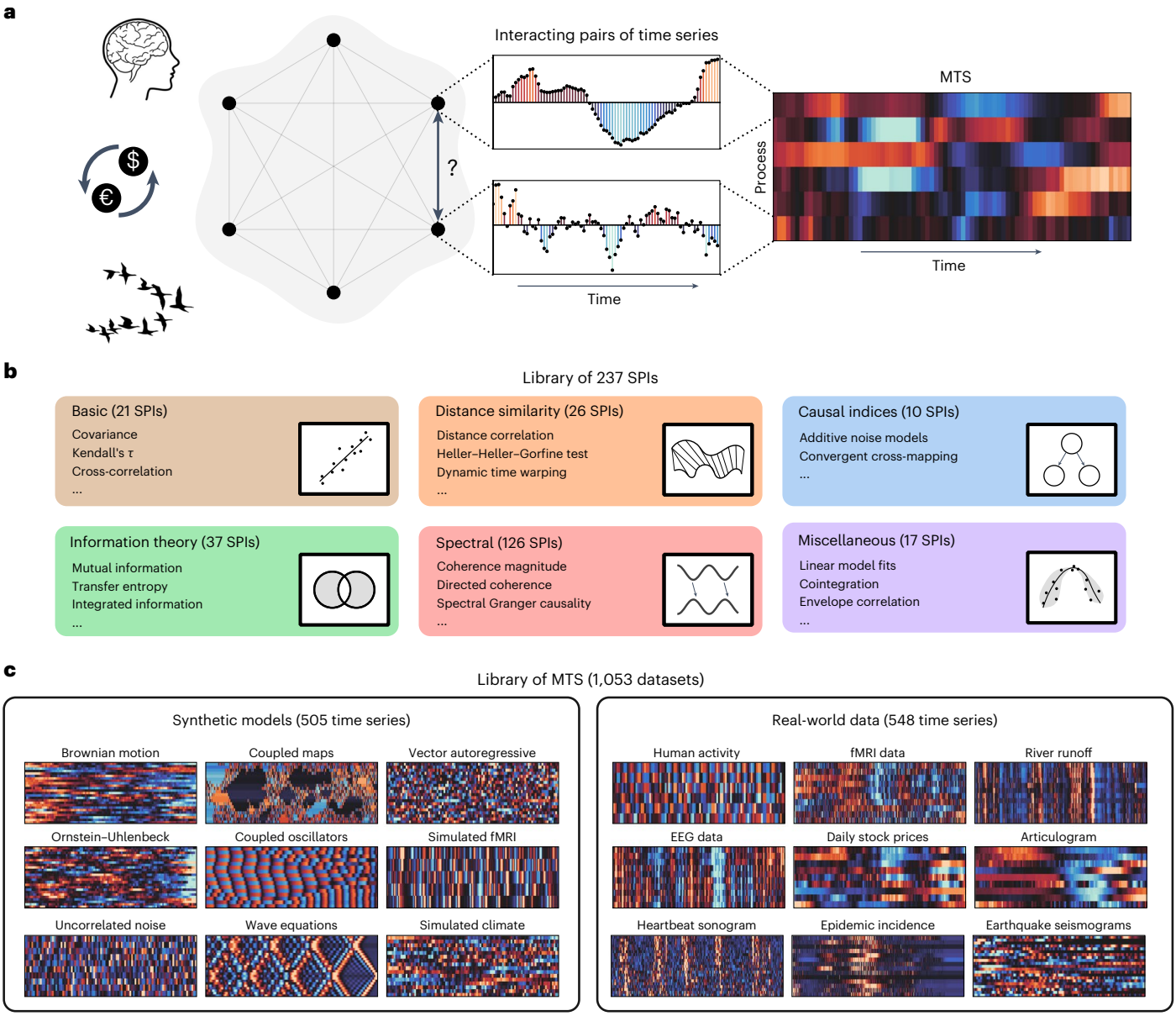


Fig. 1 | The behavior of a scientific library of 237 SPIs was evaluated using a collection of 1,053 MTS. a, In many disciplines, including neuroscience, economics and biology, scientists analyze interactions between pairs of processes in a MTS. For illustration, two time series are plotted, using colour to indicate time-series values on a scale from low (light blue), to the mean (black), to high (light orange). An MTS can be visualized as a heat map using the same coloring, in which individual processes (rows) evolve over time (columns). **b**, The 237 SPIs analyzed here derive from six broad disciplinary categories: 'basic', 'distance similarity', 'causal indices', 'information theory', 'spectral' and 'miscellaneous', as described in detail in Supplementary Note 1; selected examples of SPIs within each category are listed here, along with a schematic depiction. **c**, Selected examples from our diverse collection of 1,053 MTS—which are used to evaluate the empirical behavior of the SPIs—are plotted as process × time heat maps (using the same colors as in **a**). Our library includes data generated from synthetic models (left) and measured from real-world systems (right), as described in detail in Supplementary Note 2.

cross-correlation[6]), may be misaligned (for instance, dynamic time warping[10]), or where the knowledge of one variable improves the predictability of another (for example, Granger causality[11]).

In this Article, we represent algorithms that measure interactions between pairs of time series as real-valued summary statistics, that is, statistics of pairwise interactions (SPIs). Figure 1b illustrates the diverse theoretical tools and types of interactions covered in the scientific literature on SPIs, from covariance (the foundation of statistics and machine learning) to convergent cross-mapping[3] (developed to infer causal effect in complex ecosystems); however, as the theory underlying statistical interactions between pairs of time series has been developed largely independently across disciplinary contexts, these methods remain disconnected from one another. In this work, we unify this wealth of interdisciplinary scientific knowledge, empirically connecting previously disjoint methodological traditions to yield a unified set of tools for quantifying interactions in complex dynamics.

As diverse scientific methods for quantifying pairwise interactions have never been compared at scale, there remain many unanswered questions. Are all of these methods capturing unique information on the interactions occurring within a time-varying system? Is there synergy between complementary approaches that, when combined, tells us more about the underlying system than any single method can? Or, do interesting redundancies exist between techniques used across disciplines that hint at a common theoretical underpinning

that has not previously been identified? Following previous highly comparative studies of univariate time series[12] and graphs[13], this work addresses these questions by simultaneously evaluating hundreds of different SPIs directly from data. Our empirical approach to this problem involves first assembling a comprehensive annotated library of computational methods for quantifying pairwise interactions from data (summarized in Fig. 1b), and then analyzing their behavior across a large and diverse library of MTS (Fig. 1c).

As shown in Fig. 1b, our annotated library of SPIs was organized into six broad categories on the basis of their underlying theory: 'basic', 'distance similarity', 'causal indices', 'information theory', 'spectral' and 'miscellaneous' (see 'Library of 237 SPIs' section in the Methods for more details, and Supplementary Note 1 for a full list of SPIs with descriptions and references). To understand the behavior of these SPIs on data, we constructed a library of 1,053 diverse MTS, curated with the aim of capturing the main classes of systems and dynamics that are studied across scientific disciplines, including synchronization, spatiotemporal chaos, wave propagation, criticality and phase transitions. As depicted in Fig. 1c, our MTS library contains 505 synthetic MTS generated from a range of mathematical models, including coupled maps, ordinary differential equations and partial differential equations, and 548 diverse real-world MTS assembled from public databases such as geophysical, medical and financial data (see 'Library of 1,053 MTS' section in the Methods for further details, and Supplementary Note 2 for a full list of included MTS).

Results

Organizing pairwise interactions by their empirical behavior

Having assembled diverse libraries of methods (237 SPIs) and data (1,053 MTS), we aimed to analyze how similarly the different SPIs behave on the data. To achieve this, we developed an empirical similarity index, R ($0 \leq R \leq 1$), that captures the relationship between any two SPIs by comparing their output when applied to all 195,112 pairwise interactions present across all 1,053 MTS. This index is derived from the average absolute Spearman correlation between a pair of SPIs when applied to all pairs of processes in all datasets (see 'Quantifying similarity between SPIs using the empirical similarity index' section of the Methods for details). The minimum value, $R = 0$, indicates a pair of maximally distinct SPIs (with uncorrelated behavior on all datasets), whereas the maximum, $R = 1$, indicates a pair of SPIs that are perfectly correlated on all datasets (that is, they behave as simple monotonic transformations of one another); thus, a pair of SPIs with a high R reflects broadly similar behavior across MTS containing very different types of structures, and therefore acts as a suitable candidate index of empirical similarity.

Using the dissimilarity measure, $D = 1 - R$, we organized all 237 SPIs using hierarchical clustering, yielding the dendrogram shown in Fig. 2b. This presents a data-driven, structured representation of a diverse literature that allows us to probe and interpret relationships between scientific methods at multiple levels. We focus our analysis here on a 14-module resolution (with modules labeled M1–M14), which captures important methodological connections between groups of SPIs with similar behavior on data. As summarized in Fig. 2a, these fourteen modules group common conceptual and theoretical approaches to measuring interactions between pairs of time series, demonstrating the ability of our empirical approach to meaningfully organize the interdisciplinary literature.

In addition to grouping similar types of methods into modules, we found that different high-level conceptual formulations of dynamical interactions were recapitulated in the relationships between modules. For example, modules M3–M6 contain distinct types of SPIs (including Granger causality[11], directed information[14] and integrated information[15,16]), all of which capture statistical dependencies between two time series by considering the context of their past. This idea that observable interactions are predicated (or, from a statistical standpoint, conditioned) on the history of a process was first proposed by the Wiener–Granger theories of causality and feedback[17], specifically by measuring how one time series might improve the self-predictability of another. Our results group SPIs on the basis of this underlying theoretical formulation, due to their characteristic behavior on data. Other types of SPIs (that do not predicate on the self-predictability of a process) also display distinctive behavior, including measures of contemporaneous relationships (the correlation coefficients of M14), dependencies that account for temporal lags (the coherence measures of M12), or temporal dilation and shifts (dynamic time warping and related methods in M10).

Ten of the fourteen modules are homogeneous, containing methods that are derived from similar underlying theories (as indicated by the color of the category labels in Fig. 2bii). Of these ten homogeneous modules, six of them (M4, M7, M8, M9, M11 and M13) comprise SPIs for measuring one specific type of pairwise interaction, differing in either their: specific algorithmic implementation (for example, M13 contains both the Engle–Granger[5] and the Johansen[18] tests for measuring co-integration); extracted summary statistics (for example, M8 contains both the mean and maximum of the wavelet-based phase lag index[19]); or parameter settings (for example, M4 contains SPIs that use five different estimation techniques for transfer entropy[20]). The remaining four homogeneous modules (M1, M2, M6 and M12) comprise methods with very similar theoretical underpinnings; for example, M12 contains many SPIs for measuring undirected interactions via Fourier transformations, such as the magnitude and the imaginary part of the coherence[21]. Of particular interest are the four heterogeneous modules (M3, M5, M10 and M14), which mix SPIs from different literature categories, revealing interesting connections between different theoretical bases for quantifying pairwise dependencies between time series. Although M3 contains a mix of SPIs based on information theory (six labeled 'information theory' measures and one, information-geometric conditional independence, labeled as a 'causal index'), the remaining three modules establish interesting connections between the behavior of seemingly disparate SPIs on MTS data. Three networks that are derived from these modules are plotted in Fig. 2c–e and are investigated in detail below.

As illustrated in Fig. 2c, module M5 contains a mix of two different types of methods: the first includes five linear estimators for integrated information (such as geometric integrated information, Φ_G (ref. 22), phi star, Φ^* (ref. 16) and stochastic interaction[23]), whereas the second includes 16 estimators for Granger causality in both the time and frequency domains[24]. Although Granger causality and integrated information theory were developed in very different contexts (for example, Granger's investigations into causality between economic time series in 1969[11], versus Tononi's recent integrated information theory (Φ) of consciousness[15,16,22]), our analysis reveals that all SPIs in this module nevertheless behave similarly on data (with an empirical similarity index, $R = 0.52$, which lies in the 95th percentile for all R values; see Supplementary Fig. 1c). Recent results have indeed shown that Granger causality can be formulated as the information-theoretic measure, transfer entropy[25], and can thus be grouped under the same information-geometric framework as integrated information theory[22,26]. However, it was not known whether these information-theoretic SPIs behave similarly in practice and, as such, their relationship was not widely recognized. Module M5 thus demonstrates an important confirmation of our empirical approach in being able to recapitulate emerging theory and unify scientific tools for understanding interacting processes.

Module M10, shown in Fig. 2d, highlights striking connections between three conceptually distinct types of methods: (1) dynamic time warping (DTW), which was developed in the data-mining community to quantify the similarity between two (potentially shifted and dilated) audio signals[10]; (2) cross-spectral phase-based measures—the maximum phase coherence[21], and the mean and maximum phase-locking value[27]—which were developed to examine frequency-specific

https://doi.org/10.1038/s43588-023-00519-x

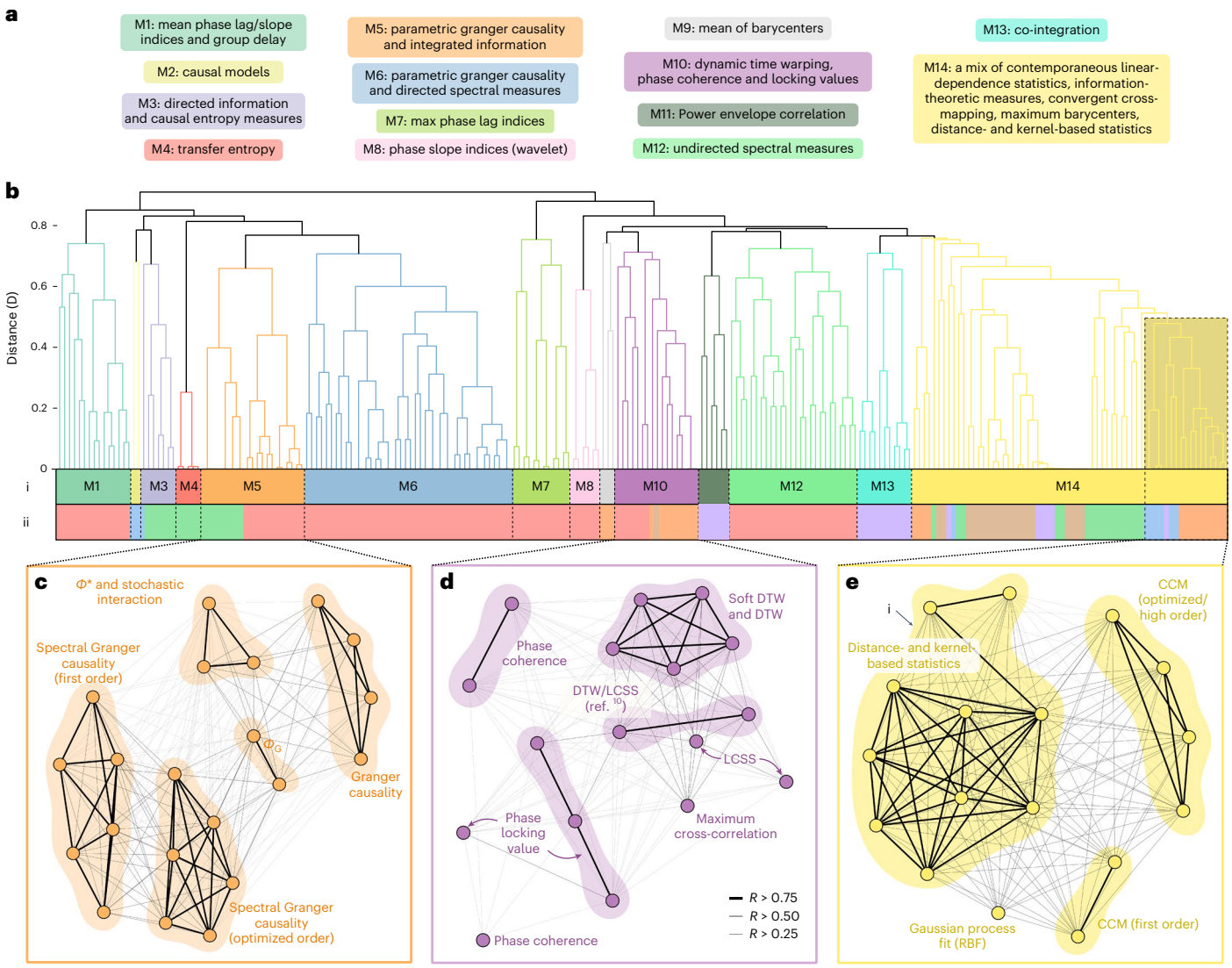


Fig. 2 | Statistics for measuring pairwise interactions between time series can be organized into 14 modules on the basis of their behavior on over 1,000 MTS, providing an intuitive, data-driven organization of interdisciplinary scientific literature. a, A brief summary of the main types of methods in each of the modules (see Supplementary Note 1 for a full list with descriptions). **b**, The dendrogram used to infer the modules, produced by hierarchical clustering using a dissimilarity measure, $D = 1 - R$, that is based on the empirical similarity index, R (see Methods). SPIs are colored according to: (i) their module label (upper row) and (ii) their literature categorization (lower row), displayed using the color scheme defined in Fig. 1b. A high-resolution version of this dendrogram, including each of the SPIs (leaf nodes) that form the modules, is presented in Supplementary Fig. 2 (and the assignment of SPIs to modules is in Supplementary Data 1). **c**–**e**, Three selected modules that include a mixture of SPIs developed in different disciplinary contexts are shown as network plots in M5 (**c**), M10 (**d**) and a subset of M14 (**e**). In these plots, SPIs are represented as nodes, three different edge weight thresholds are displayed (corresponding to $R > 0.25$, $R > 0.5$ and $R > 0.75$), and connected components with high similarities, $R > 0.75$, are shaded yellow. DTW, dynamic time warping; LCSS, longest continuous subsequence; CCM, convergent cross-mapping; RBF, radial basis function.

synchronization in neuroimaging data[21]; and (3) the maximum cross-correlation[6], a classical statistical technique for correlating two time series at different temporal lags. All of these SPIs capture time-lagged interactions between two processes, but in slightly different ways: the maximum cross-correlation finds the highest fixed-lag match, DTW extends this idea by optimizing the distortion of the time axis to best match potentially misaligned time series, and the cross-spectral measures account for time lags in terms of phase differences. Module M10 thus reveals previously unreported connections between diverse approaches to capturing associations between pairs of potentially unaligned time series, indicating a common conceptual basis for methods developed and applied across disciplines—whether they are measuring synchronization between neuroelectric recordings or recognizing speech from audio signals.

Finally, we discuss module M14, which groups 66 SPIs from all literature categories except for 'spectral' (see Fig. 1b for categories). This module recapitulates some theoretical relationships that have already been established, such as the equivalence between linear-Gaussian mutual information and absolute correlation[28] (with a maximum similarity of $R = 1$). To highlight some previously unreported relationships, we focus on a demonstrative submodule (shown in Fig. 2e) comprising 17 SPIs from the 'causal indices', 'distance similarity' and 'miscellaneous' literature categories. We first note the tight cluster of SPIs, labeled 'i' in Fig. 2e, that were developed independently in two different domains: distance correlation-based methods[29] from the statistics community, and kernel-based methods from the machine-learning community. This cluster first highlights a recent finding that distance correlation and the Hilbert–Schmidt independence criteria

https://doi.org/10.1038/s43588-023-00519-x

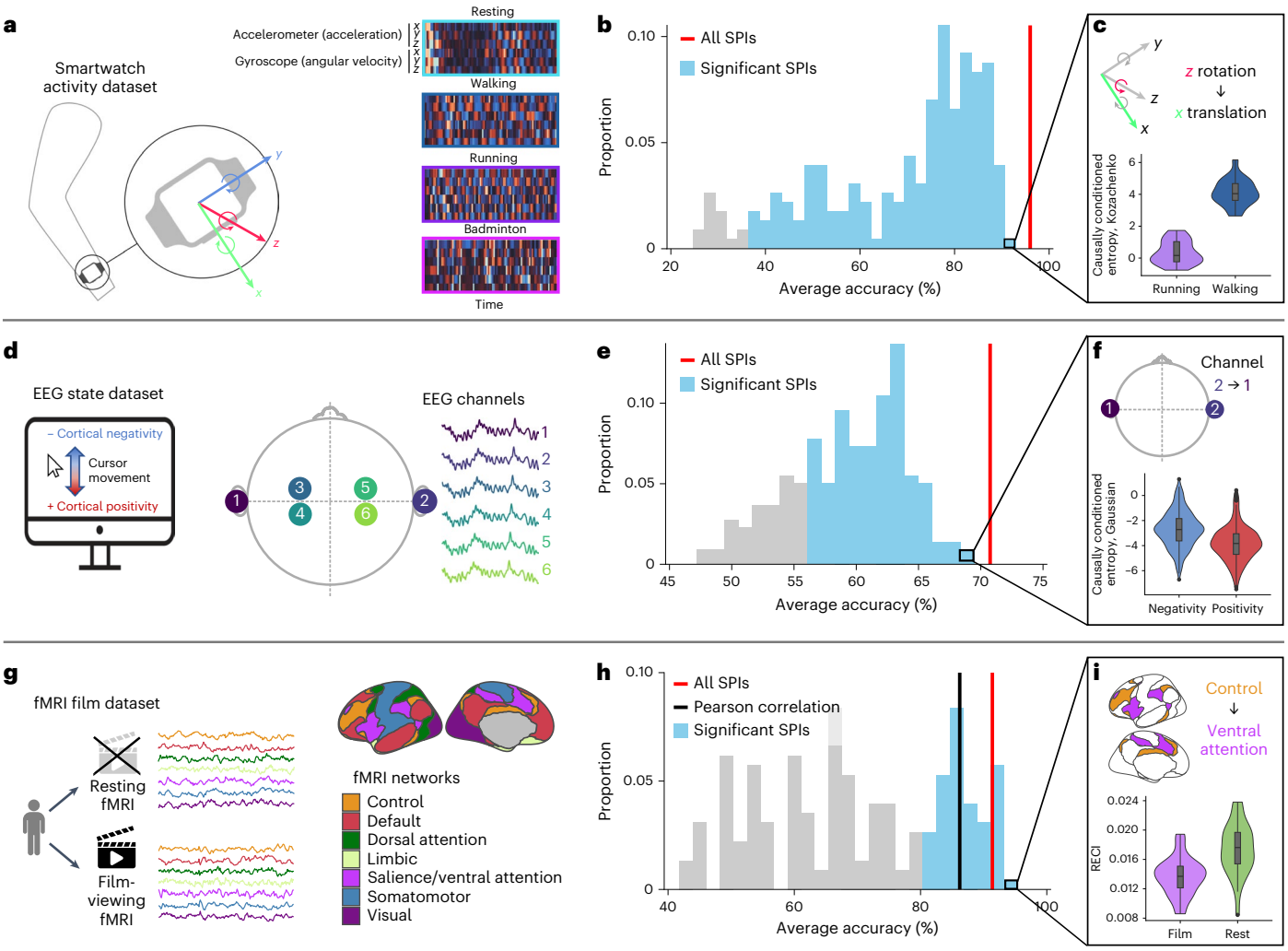


Fig. 3 | A comprehensive library of SPIs can be used to accurately classify and understand differences in human movement and neural activity datasets. **a**, In the smartwatch activity dataset, we aimed to determine which of four activities a participant is performing ('resting', 'walking', 'running' or 'badminton') from smartwatch accelerometer recordings. An example MTS from each of the four classes is shown as a process × time heat map (colored as in Fig. 1a). **b**, Distribution of average classification accuracy (over train–test resamples) across 228 SPIs, where all pairwise interactions are used as the basis for classification for each SPI. The 213 SPIs with significant accuracy are shaded blue (one-sided permutation test, Bonferroni-corrected $P < 0.05$). Concatenating feature vectors from all SPIs into a single classification model yields an average accuracy of 96% (red). **c**, Violin plots showing the distribution of the top-performing SPI (causally conditioned entropy with a Kozachenko–Leonenko density estimator) between the wrist's x-axis translation and z-axis rotation, indicating a stronger interaction during walking (mean ± s.d. = 4.1 ± 0.8, $N = 20$) than running (0.4 ± 0.7, $N = 20$). In each violin plot, the annotated box plot shows the box centre and extent as the median, 25th and 75th percentiles; whiskers extend for 1.5× the interquartile range and outliers are shown as black dots. **d**, In the EEG state dataset, we aimed to classify positive versus negative cortical activity states using data from six EEG channels, for which the electrode locations and example time series are depicted. **e**, Distribution of average classification accuracy across all 219 SPIs, including the 165 significant SPIs, and a combination of all SPIs (71% accuracy). **f**, The top-performing SPI (causally conditioned entropy with a Gaussian density estimator) is visualized from EEG channel 2 to 1 in cortical negativity (−2.7 ± 1.4, $N = 282$) versus positivity (−3.8 ± 1.4, $N = 279$). **g**, In the fMRI film dataset, we aimed to classify rest versus film-viewing using fMRI time series from seven brain networks. **h**, Distribution of classification accuracy across all 227 SPIs, including the 67 significant SPIs (shaded blue) and annotating the performance of Pearson correlation coefficient (86%, black) and the combined set of all 227 SPIs (91%, red) for comparison. **i**, The top-performing SPI (regression error-based causal inference, RECI) is shown from the control to ventral attention networks in film-viewing (0.014 ± 0.003, $N = 29$) and rest (0.018 ± 0.004, $N = 29$) conditions.

(HSIC, a kernel-based method)[30] are equivalent when computed using certain distance kernels[31]; our results suggest that similar theoretical connections can be established between the other SPIs of the cluster (including the Heller–Heller–Gorfine test[32] and multiscale graph correlation[29]). Second, we find that the distance- and kernel-based statistics display strikingly similar behavior as common implementations of the convergent cross-mapping (CCM) algorithm, which was originally developed for inferring causality in complex ecosystems[3]. Convergent cross-mapping aims to measure the causal effect of one time series on another by the ability of the second to reconstruct the first using a nearest-neighbor approach. The fact that these methods behave so similarly on MTS data indicates that the well-studied techniques of phase-space reconstruction (used in the CCM algorithm) have a correspondence to the nonlinear kernel-estimation techniques from the statistics and machine learning communities. These observed connections also have important practical ramifications; for example, our results suggest candidate proxy algorithms to substitute for the computationally expensive CCM, which could yield major computational efficiencies and enable new applications on larger datasets.

In summary, our empirical approach to unifying a large and disjoint literature of methods for characterizing pairwise interactions in time series has allowed us to: (1) capture the commonalities in

the conceptualizations of dynamical interactions (such as whether methods predicate a statistical dependence on self-predictability) and (2) understand empirical and theoretical connections between diverse methods, with implications for shaping theory and practice.

Leveraging diverse methods to address scientific problems

Our results above illustrate the rich diversity of scientific methods for quantifying pairwise interactions. This diversity suggests that, when quantifying pairwise interactions for a given application, there is potential to make comparisons across the scientific literature of SPIs to: (1) select the best-performing SPI in an unbiased, data-driven way; and (2) leverage a synergistic combination of multiple complementary SPIs to better capture complex underlying interactions in MTS. Here we provide a simple demonstration of this strategy (referred to as 'highly comparative' due to the broad methodological comparison involved[12,13]) on three MTS classification problems using three open datasets: (1) the smartwatch activity dataset (Fig. 3a), where the aim is to classify one of four behavioral states (walking, running, resting, or playing badminton) from six-sensor smartwatch MTS (comprising a three-axis accelerometer and three-axis gyroscope)[33]; (2) the electroencephalogram (EEG) state dataset (Fig. 3d), where the aim is to distinguish positive versus negative slow cortical potentials from single-subject EEG data (originally used to move a cursor up or down on a computer screen)[34,35]; and (3) the functional magnetic resonance imaging (fMRI) film dataset (Fig. 3g), where the aim is to classify resting and film-watching conditions from fMRI data[36] (see 'Classification case studies' section in the Methods for additional details on each dataset). To investigate the performance of different SPIs on these tasks, we represented each MTS as a set of features corresponding to all pairwise interactions between its constituent processes and compared their classification performance using a linear support vector machine (SVM) with cross-validation (see 'Classification' section in the Methods for details).

We observed a wide range of SPI performances in each case study, ranging from null performance up to high and statistically significant performance: 27–92% accuracy on the smartwatch activity dataset (Fig. 3b), 48–69% accuracy on the EEG state dataset (Fig. 3e) and 41–95% accuracy on the fMRI film dataset (Fig. 3h). Many SPIs displayed significant classification performance on each dataset (permutation test, Bonferroni-corrected $P < 0.05$): 213 SPIs for the smartwatch activity dataset, 165 SPIs for the EEG state dataset, and 67 SPIs for the fMRI film dataset. This wide range of observed SPI performance on all three datasets demonstrates the crucial importance of selecting an SPI that is able to capture the relevant types of interactions underlying a given dataset.

To understand the types of interactions that characterize the labeled classes of MTS, we analyzed and interpreted the highest-performing SPIs on each dataset (see Supplementary Data 2 for the full results). To provide a simple demonstration of this process, here we focus on the top-performing individual SPI in each case study. In the smartwatch activity dataset, the top-performing SPI was causally conditioned entropy (CCE) using a Kozachenko–Leonenko estimator (92% accuracy; SPI label 'cce_kozachenko'; see Supplementary Note 1.4.7 for details). Its high performance is driven in part by its ability to capture the coupling from the wrist's z-rotation to its x-translation, which differs strongly between running and walking, as shown in Fig. 3c. This tells us that wrist rotation in the z-direction is more informative of subsequent elbow movement in the x-direction (as the elbow moves side to side) in walking than in running.

The top-performing SPI for the EEG state dataset was, as above, a causally conditioned entropy, but using a Gaussian density estimator (69%, SPI label 'cce_gaussian'; see Supplementary Note 1.4.7 for details). Its performance was driven in part by its ability to capture the increased coupling from EEG channel 2 to 1 (from near the right ear to near the left ear) in cortical negativity versus positivity states, as shown in Fig. 3f. Other statistics designed to capture directed information flow (in a way that includes instantaneous interactions in the presence of feedback) also performed well on this task, including directed information with a Gaussian density estimator (69% accuracy, SPI label 'di_gaussian'; see Supplementary Note 1.4.8). Although past applications of SPIs derived from directed information theory to neuroimaging data are limited[37,38], our highly comparative analysis suggests them as high-performing methods for measuring EEG coupling alongside other novel candidates for further investigation, including the direct directed transfer function evaluated over high frequencies (67%, SPI label 'ddtf_multitaper_mean_fs-1_fmin-0-25_fmax-0-5'; see Supplementary Note 1.5.9) and the Hilbert–Schmidt Independence Criterion (66%, SPI label 'hsic'; see Supplementary Note 1.2.5). Of the classical methods for quantifying EEG connectivity, some are recapitulated as high performers by our data-driven analysis, including mean directed coherence across various frequency bands[39] (all 67% accuracy; see Supplementary Note 1.5.10), whereas others exhibited surprisingly low accuracy, such as algorithmic variants of partial directed coherence[40] (between 55% and 64%).

Finally, for the fMRI film dataset, the top-performing SPI was regression error-based causal inference (95% accuracy; SPI label 'reci'; see Supplementary Note 1.3.4). Its high performance is driven in part by its ability to capture the stronger coupling from the control network to the ventral attention network during film-watching compared to rest (Fig. 3i). The dominant way of measuring coupling in whole-brain fMRI is to use the Pearson correlation coefficient[8,41] (annotated in Fig. 3h), which exhibits strong and statistically significant classification accuracy on this problem (86%); however, our data-driven approach highlights 30 alternative SPIs with higher performance (88% to 95%; compare with Supplementary Data 2). These high-performing methods include alternative types of covariance (for example, minimum covariance determinant, 93% accuracy, SPI label 'cov_MinCovDet'; see Supplementary Note 1.1.1) and precision estimates (for example, using Ledoit–Wolf shrinkage, 91%, SPI label 'prec_LedoitWolf'; see Supplementary Note 1.1.2) that better deal with non-Gaussian bivariate distributions. Other high performers include information-theoretic SPIs (such as conditional entropy, joint entropy and mutual information using Gaussian density estimators, all of which exhibited 91% accuracy; see Supplementary Note 1.4.1–1.4.3) and directed SPIs which distinguish asymmetric coupling (for example, the top-performing regression error-based causal inference, 'reci'). Compared with the typically subjective process of selecting an appropriate method to analyze a given dataset, the highly comparative approach demonstrated here highlights the most useful scientific methods automatically, facilitating interpretable understanding of the conceptual formulations of pairwise dependencies that drive successful performance.

As different types of systems involve different types of interactions between measured processes, we expected different SPIs to perform well across the three datasets. Indeed, we found that an SPI with high performance on one problem does not imply its high performance on other problems. In particular, some SPIs performed well on only a single dataset; for example, dynamic time warping with an Itakura parallelogram (SPI label 'dtw_constraint-itakura'; see Supplementary Note 1.2.7), which was a top performer on the fMRI film dataset (91%), but showed weaker performance on the smartwatch activity dataset (78%) and null performance on the EEG state dataset (53%). Yet some SPIs did perform strongly across all three datasets, such as the cross distance correlation (SPI label 'dcorrx_maxlag-10'; see Supplementary Note 1.2.2), which ranked among the top 10 SPIs for all three problems (90%, 66% and 91%, for the smartwatch activity, EEG state and fMRI film datasets, respectively). Moreover, different algorithmic variants of causally conditioned entropy were top performers for the smartwatch activity ('cce_kozachenko', 92%), EEG state ('cee_gaussian', 69%) and fMRI film ('cce_kernel_W-0.5', 90%) datasets (although we note a strong dependence of the density estimation approach on CCE performance). We also found that grouping SPIs based on the fourteen data-driven modules (identified by their similarity of behavior on data, Fig. 2)

better captured their relative performance on these tasks than the six literature categories (from Fig. 1b), as shown in Supplementary Fig. 3, suggesting our modular representation as a useful one for understanding differential SPI performance on a given task.

Relative to investigating individual SPIs one at a time, we finally aimed to investigate the value of drawing on multiple SPIs simultaneously. We developed a combined representation of the pairwise dependence structures captured by all SPIs, allowing us to simultaneously represent each MTS using a large and diverse set of pairwise dependency structures (through feature concatenation, as described in the 'Classification' section of the Methods). Although this approach represents each MTS in a much higher-dimensional space than the individual SPI representation analyzed above (with associated challenges for robust classifier fitting), we expected it to outperform the individual best SPI on datasets involving multiple types of interactions, such that simultaneously leveraging multiple SPIs provides complementary and useful information about class differences. Relative to the top-performing individual SPI, our combined SPI approach improved classification accuracy on the smartwatch activity dataset (to 96%) and the EEG state dataset (to 71%), shown as vertical red lines in Fig. 3b,e. On the fMRI film dataset, it yielded slightly lower performance (91%) than the top individual SPI, 'reci' (95%), suggesting that the associated interactions are well-captured by a single, well-chosen SPI on this dataset (that is, multiple SPIs do not provide an advantage sufficient to overcome the challenges of fitting a classifier in a higher-dimensional space). By simultaneously drawing on a wide range of SPIs, the simple statistical approach demonstrated here can quantify multiple complementary types of interactions from MTS data (and is likely to yield improved accuracy through optimization; see 'Discussion' section).

Discussion

Our empirical organization of SPIs (Fig. 2b) identifies connections between a rich literature of diverse methods for quantifying interactions between time series, and thereby highlights fruitful directions for future research to consolidate, extend and develop new theory. These results were obtained using a similarity index, R, that averages across a large and diverse range of 1,053 MTS but—as the behavior of an SPI depends on the types of interactions present in a dataset—the resulting relationships between SPIs depend on the MTS we chose to include. As a simple illustrative example, consider the case that our MTS library only contained data generated by lag-1 vector autoregressive models, VAR(1). Then the unique behavior of more sophisticated SPIs (such as those capturing nonlinear dependence or causal coupling on longer time lags) would not be observed, as data containing those types of interactions would be absent. Consequently, the variety of behaviors exhibited by our library of SPIs would be reduced, and the modular structure evident in the dendrogram of Fig. 2b would become far more homogeneous. In constructing the MTS data library used in this work, we thus aimed to be as even-handed and comprehensive as possible in sampling from diverse systems, which has been sufficient to yield a meaningful and useful representation of the interdisciplinary literature on SPIs. However, as any such finite sample is incomplete and relies on subjective decisions in its construction, future work may explore the dataset-dependent similarity of SPI pairs in detail to construct more nuanced organizations of the literature. This would provide new understanding of how different mechanisms (and hence empirical dependency structures) play out in different classes of complex systems. Future work may also revisit the simple methodological choices made here, including our decisions to: (1) quantify the similarity of a pair of SPIs using a single number (R); (2) represent the resulting relationships in one dimension (as a dendrogram); and (3) analyze the resulting dendrogram at a single selected resolution (a 14-module decomposition). For example, more complex unsupervised methods, such as overlapping community-detection methods[42,43], could reveal interesting new relationships between SPIs at different resolution scales.

As different time-varying systems contain different statistical relationships between their elements (for example, those that capture instantaneous or time-delayed responses, linear or nonlinear interactions, are conditioned on the past, allow for non-constant delays, or infer directional coupling), our results demonstrate that a highly comparative approach can be used to detect the most informative types of SPIs for capturing the relevant types of interactions underlying a given dataset. Unlike many machine-learning approaches to MTS classification, which can be challenging to interpret[33], this approach connects scientists to interpretable theory that shapes understanding of the most important types of pairwise interactions in a dataset, following recent undertakings to unify interdisciplinary literatures on summary statistics ('features') for univariate time series[12,44,45] and complex networks[13]. This broad, comprehensive methodological comparison stands in contrast to a more conventional approach in which the data analyst manually selects a method—a practice that leaves open the possibility that alternative methods may provide clearer interpretation, better performance, or computational efficiencies. The three classification case studies analyzed here provide a simple demonstration of the procedure, automatically highlighting high-performing SPIs from across the literature and providing interpretable understanding of the relevant interactions in each dataset. We observed a wide range of performance in all cases, highlighting the importance of careful selection of SPIs for a given application. In the EEG state dataset, for example, our analysis flagged high-performing SPIs consistent with common methodological practice in EEG analysis (thus recapitulating existing domain knowledge), and others with surprisingly poor performance, as well as flagging novel high-performing SPIs as promising candidates for methodological innovation in the field. Future work investigating which types of SPIs are best suited to which types of problems will yield new insights into the interactions underlying different complex systems, and is likely to uncover additional novel applications of SPIs to new problems.

Although the ability to compare across the full library of SPIs is powerful, it comes at a computational cost, particularly for larger MTS datasets in which each MTS is sampled over many time points and processes. The high redundancy between many groups of SPIs (Fig. 2), together with the preliminary indications from case studies that some SPIs may exhibit generally high performance, suggest the potential for major efficiency gains by constructing a high-performing and minimally redundant reduced set of SPIs. Beyond a simple approach of selecting a representative SPI from each of the 14 modules, a more systematic approach could compare SPI performance across a wide range of representative MTS tasks, for example, following recent work reducing a large pool of thousands of interdependent univariate time-series features to just 22 representatives[46]. There is also much scope for statistical optimization in applying any given set of SPIs to real-world problems, beyond the simple choices made here. In particular, performance could be optimized by tackling the challenges of high-dimensionality and strong interdependence of SPIs using dimensionality reduction, regularization or feature-selection methods to find small but highly explanatory combinations of complementary SPIs tailored to a given dataset[47]. Given the methodological freedoms involved in highly comparative time-series analysis, pre-registration should be considered for transparency and to avoid bias[48] (for example, ref. 49).

Beyond the applications to MTS classification presented here, the ability to represent the coupling structure in a MTS using a large and diverse set of pairwise interactions could form the foundation for tackling myriad statistical inference problems, including regression, clustering, anomaly detection, and causal network inference. For example, current methods for inferring networks of causal interactions from MTS data[50,51] range from heuristic approaches based on thresholding a set of pairwise dependencies defined by a given SPI through to full statistical inference[50,52–54]. Rather than manually selecting an SPI for this purpose, the ability to compare diverse SPIs provides the flexibility to

capture different types of underlying interactions that could form the basis of improved network-inference algorithms. Such an extension may also go beyond point estimates of pairwise dependence strengths (considered here) towards assessing statistical significance through comparison to an appropriate null model[9,51,55]. Future work could also explore comprehensive statistical representations of MTS using more diverse properties than just the set of pairwise interactions, by also incorporating properties of univariate dynamics of individual system components[45], system-level structure in the full set of pairwise interactions represented as a network[13] and higher-order interactions[56].

The flexible and extendable software accompanying this work, pyspi[57] (see 'Code Availability' section), facilitates the application of our highly comparative approach to quantifying pairwise interactions to myriad applications. Through careful framing of systems-level inference[55], it has the potential to highlight unexpected and powerful methodological approaches to quantifying interaction patterns in time-varying systems. The data library of over 1,000 MTS provided with this work[58] (see 'Data Availability' section) is another resource that will allow researchers to characterize the behavior of their computational methods on a comprehensive range of real-world and simulated systems, addressing issues associated with only testing new complex systems methods on idealized datasets[55]. In summary, this work demonstrates the utility of an empirical approach to unifying diverse complex dynamical systems and their methods of analysis, providing insights and tools for scientific discovery.

Methods

Comprehensive scientific libraries of methods and data

Library of 237 SPIs. This work takes an empirical approach to understanding the relationships between diverse scientific methods for quantifying pairwise interactions between time series, formulated as an SPI. The first step in this approach was to construct a comprehensive, annotated library of SPIs. Here we introduce a library of 237 SPIs, organized into six broad categories based on their underlying theory: 'basic' (such as co-variance, Kendall's τ (ref. 59) and cross-correlation); 'distance similarity' (for example, distance correlations[29], kernel-based independence tests[30,32] and dynamic time warping[10]); 'causal indices' (including additive noise models and convergent cross-mapping[3]); 'information theory' (such as Granger causality[11,25] (note that Granger causality is included in the information theory category as it is equivalent to transfer entropy evaluated with a Gaussian estimator[25]), transfer entropy[20] and integrated information[16]); 'spectral' (derived from Fourier or wavelet transformations, for example, coherence magnitude, phase-locking value[27] and spectral Granger causality[60]); and 'miscellaneous' (for example, co-integration[5] and model fits). A full list of SPIs, along with descriptions and references, is presented in Supplementary Note 1.

Library of 1,053 MTS. To understand how each SPI behaves on different types of multivariate time-series data, we constructed a library of 1,053 diverse model-generated and real-world MTS. Our aim in assembling these data was to capture the main classes of systems and dynamics that are studied across scientific disciplines, including synchronization, spatiotemporal chaos, wave propagation, criticality and phase transitions. Our library contains 505 synthetic MTS generated from mathematical models, including: uncorrelated and correlated noise (for example, Cauchy and normally distributed noise and Brownian motion); coupled maps (such as vector autoregression[6] and coupled map lattices[61]); coupled ordinary differential equations (including Kuramoto oscillators[62], Hodgkin–Huxley and Wilson–Cowan networks); and partial differential equations (namely, wave equations, in which processes are embedded in physical space). It also contains 548 diverse real-world MTS from public databases across: geophysics (such as earthquake seismograms and atmospheric processes); medicine (for example, heartbeat sonograms, fMRI data and EEGs); physiology (including accelerometer and gyroscope readings for sports and basic motions); and finance (for instance, exchange rates and stock prices), among others. Each MTS comprises between 5 and 40 processes, and between 100 and 2,000 observations—characteristics that were designed to match many real-world datasets. Across all MTS in our library, we have a total of 195,112 pairwise interactions that we used to evaluate the SPIs. Descriptions of the full set of MTS analyzed here are presented in Supplementary Note 2.

Quantifying similarity between SPIs using the empirical similarity index

Our main aim in this work was to assess the similarity of behavior of any two SPIs across a diverse range of MTS. In this section we describe the empirical similarity index, R, used to capture the similarity between the behavior of all pairs of SPIs across our diverse library of 1,053 MTS. This index measures the monotonic relationship between a pair of SPIs across a large number of datasets that have different numbers of processes (and thus a different number of pairwise interactions). For an M-variate time series, $\mathbf{z} = (z_1, \ldots, z_M)$, we first consider the $M \times M$ matrix of pairwise interactions (MPI), constructed by evaluating a given SPI for all pairs of M processes within the MTS. The scalar values of the MPI, $S = (s_{ij}) \in \mathbb{R}^{M\times M}$, where the (i, j) entries of the matrix, $s_{ij} = s(z_i, z_j)$, denotes an SPI evaluated on the ith and the jth time series, z_i and z_j, respectively. In general, undirected statistics (for which $s_{ij} = s_{ji}$, such as Kendall's τ) yield symmetric MPIs, whereas directed statistics (such as transfer entropy, which have $s_{ij} \neq s_{ji}$) yield asymmetric MPIs. Furthermore, some SPIs are signed (for example, correlation coefficients are within $[-1, 1]$), whereas others are unsigned (for example, the distance correlation is within $[0, 1]$). To compare all methods appropriately, we converted signed SPIs to their absolute value, such that all statistics increase with the strength of dependency between z_i and z_j (for example, we analyze the magnitude of the covariance rather than its sign). Some example MPIs are shown in Supplementary Fig. 1a. Having computed MPIs for all SPIs on all 1,053 MTS, we computed the empirical similarity index between each pair of SPIs via a two-step process (depicted in Supplementary Fig. 1b). First, we defined the similarity of a pair of SPIs, k and l, on a given dataset, d. We did this by computing MPIs for k and l (yielding S_{kd} and S_{ld}, respectively) and then computing the absolute value of the Spearman's rank correlation coefficient between each of their (off-diagonal) entries, $|r_d|$. The resulting correlation value, $|r_d|$, thus captures the strength of a monotonic relationship between the output of the two SPIs on the dataset d. After repeating this computation for all MTS in our library, we calculated the empirical similarity index, R, as the average of $|r_d|$ across all datasets. Although R provides a useful scalar summary of the similarity between a pair of SPIs, it is important to note that some pairs of SPIs have quite a wide distribution of scores, $|r_d|$, across datasets, indicating that they yield highly correlated outputs on some MTS (high $|r_d|$) but not on others (low $|r_d|$). See Supplementary Note 3.1 for more details, including a detailed visual breakdown of $|r_d|$ distributions for some selected pairs of SPIs. We found a wide range of similarity indices across pairs of SPIs, from $R \approx 0$ to $R \approx 1$, illustrated by the cumulative distribution function in Supplementary Fig. 1c.

Classification case studies

Datasets. Smartwatch activity dataset. The smartwatch activity dataset is derived from the BasicMotions problem in the University of East Anglia (UEA) MTS classification repository[34]. Each MTS includes six sensors—a three-axis accelerometer and a three-axis gyroscope—recorded for 10 s at 10 Hz, yielding 1,000 time points. There are 20 MTS in each of the four classes (resting, walking, running or playing badminton) for a total of 80 MTS in the dataset. This dataset was recently analyzed in a large MTS classification challenge[33] in which the data was split 50%–50% into training and test subsets with 30 stratified repeats. The baseline classifier (based on dynamic time warping) achieved 95.25% accuracy; the best algorithm (HIVE–COTE) achieved 100% accuracy[33].

EEG state dataset. The EEG state dataset corresponds to the SelfRegulationSCP1 problem in the UEA MTS classification repository[34] and was originally published in work by Birbaumer and colleagues[35]. Electrical activity was measured from six EEG channels in one participant as they were instructed to move a cursor up or down on a computer screen by generating negative or positive slow cortical potentials, respectively. The physical placement of these EEG electrodes is depicted in Fig. 3d. Cortical activity (measured in microvolts) was recorded for 3.5 s at 256 Hz, yielding 896 time points in each channel per MTS, with 282 negativity trials and 279 positivity trials. This dataset was analyzed in work by Ruiz et al.[33], assessing accuracy using 30 stratified train–test splits of 268 training and 293 test samples. In the original comparison, the baseline DTW algorithm achieved an accuracy of 81.81% and the best algorithm (TapNet) achieved 95.68% (ref. 33).

fMRI film dataset. The fMRI film dataset is derived from a functional connectome fingerprinting study examining individual signatures of cortical activity in $N = 29$ individuals at rest or while watching a film[36]. In this dataset, blood-oxygen-level-dependent signals were recorded in 114 parcellated cortical regions with a repetition time of 813 ms (a sampling rate of 1.23 Hz) for either 1,200 frames (resting), or between 952 and 1,000 frames (film-watching). Each trial type (rest versus film-watching) was repeated four times per participant. Here we analyzed pre-processed data obtained from work by Betzel and co-workers[63]. For consistency and simplicity, we examined the first rest and first film-watching session per participant; across all participants and trials, only the first 947 frames contain real data, and thus we restricted our analysis to this time range. To retain a comparable number of processes as the first two classification case studies, we averaged blood-oxygen-level-dependent signals from the 114 original brain regions into the seven functional networks from Yeo et al.[64], as depicted in Fig. 3g. We compared the performance of the Pearson correlation coefficient, used to construct functional connectivity matrices in the original publication (ref. 36) to our library of SPIs.

Classification. For all three case studies, our simple approach to SPI-based classification involved computing the MPIs (6×6 for the smartwatch activity and EEG state datasets, and 7×7 for the fMRI film dataset) for each z-scored MTS and repeating for each SPI. We then used the elements of these matrices as features for a linear SVM classifier. Note that we used the most recent version of pyspi (v.0.4.0) to compute SPIs for the classification case studies, which included some improved implementations of some SPIs. Features were extracted from each MPI differently for directed and undirected SPIs (see Supplementary Data 1): we used the upper triangular entries as features for undirected SPIs (for which the corresponding MPI is symmetric), whereas we used all non-diagonal elements as features for directed SPIs. As a preprocessing step, for each case study, we removed any SPI that had invalid entries (due to numerical issues) in any of the MPIs, or gave constant results across all MTS (see Supplementary Data 3 for a list of omitted SPIs). This yielded a set of 228 SPIs for the smartwatch activity problem, 219 SPIs for the EEG state problem and 227 SPIs for the fMRI film problem. For the analysis involving combining all SPIs into a single classifier, this yielded a total of 4,755 features for the smartwatch activity dataset, 4,659 features for the EEG state dataset and 6,743 features for the fMRI film dataset.

The linear SVM was implemented using default settings from Python's scikit-learn[65] as part of a classification pipeline that involved z-score feature normalization (fitted on training data and applied to unseen test data). The very simple methodological choices made here allowed us to focus on demonstrating the key conceptual types of analyses made possible by drawing on a diverse set of SPIs, aiding transparency while acknowledging that more complicated statistical methodologies are likely to improve the classification performance quoted here. We implemented 30 class-stratified train–test splits for cross-validation with the same proportions implemented in ref. 33 using the StratifiedShuffleSplit function for the smartwatch activity and EEG state datasets. As there are $N = 29$ individuals in the fMRI film dataset, we implemented leave-one-individual-out cross-validation, such that each classifier was trained with the rest and film scans of $N = 28$ participants, and tested on the rest and film fMRI scans of the left-out participant.

We measured classification performance using total accuracy for all three case studies. Statistical significance was estimated using a permutation testing approach whereby 100 null models were fitted (using randomly shuffled class labels) and evaluated using the same cross-validation classification procedure described above. The observed classification performance for each SPI was then compared with the pooled null distribution of all SPIs, yielding P-values that were then adjusted for multiple comparisons by controlling the family-wise error at 0.05 using the method of Bonferroni[66]. The performance metric for the union of all SPIs was similarly compared with its corresponding null permutation distribution to yield a single P-value per classification problem.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The full database of 1,053 diverse real-world and simulated MTS analyzed here is available on the Zenodo repository at: https://doi.org/10.5281/zenodo.7118947 (ref. 58). This resource could be used to test scientific methods on a diverse sample of MTS. Time-series data used in the three case studies are from open sources, as described in the 'Classification case studies' section in the Methods. Source data are provided with this paper.

Code availability

Accompanying this paper is an extendable Python-based package called pyspi[57], which includes implementations of all 237 SPIs. This software allows users to compare the behavior of an interdisciplinary literature on methods for quantifying interactions between pairs of time series. Furthermore, code and instructions for reproducing the results and figures presented in this work can be found on the Zenodo repository at: https://doi.org/10.5281/zenodo.8027702 (ref. 67).

References

1. Bassett, D. S. & Sporns, O. Network neuroscience. *Nat. Neurosci.* **20**, 353–364 (2017).
2. Buckner, R. L. et al. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *J. Neurosci.* **25**, 7709 (2005).
3. Sugihara, G. et al. Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
4. Runge, J. et al. Inferring causation from time series in Earth system sciences. *Nat. Commun.* **10**, 2553 (2019).
5. Engle, R. F. & Granger, C. W. Co-integration and error correction: representation, estimation, and testing. *Econometrica* **55**, 251–276 (1987).
6. Reinsel, G. C. *Elements of Multivariate Time Series Analysis* (Springer, 2003).
7. Stephens, Z. D. *PLoS Biol.* **13**, e1002195 (2015).
8. van den Heuvel, M. P. & Hulshoff Pol, H. E. Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* **20**, 519 (2010).
9. Cliff, O. M., Novelli, L., Fulcher, B. D., Shine, J. M. & Lizier, J. T. Assessing the significance of directed and multivariate measures of linear dependence between time series. *Phys. Rev. Res.* **3**, 013145 (2021).

10. Sakoe, H. & Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Signal Process* **26**, 43 (1978).
11. Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424 (1969).
12. Fulcher, B. D., Little, M. A. & Jones, N. S. Highly comparative time-series analysis: the empirical structure of time series and their methods. *J. R. Soc. Interface* **10**, 20130048 (2013).
13. Peach, R. L. et al. HCGA: Highly comparative graph analysis for network phenotyping. *Patterns* **2**, 100227 (2021).
14. Massey, J. Causality, feedback and directed information. In *International Symposium on Information Theory and its Applications (ISITA-90)* 303–305 (IEEE, 1990).
15. Tononi, G. An information integration theory of consciousness. *BMC Neurosci.* **5**, 42 (2004).
16. Oizumi, M., Amari, S.-i, Yanagawa, T., Fujii, N. & Tsuchiya, N. Measuring integrated information from the decoding perspective. *PLoS Comput. Biol.* **12**, e1004654 (2016).
17. Wiener, N. in *Modern Mathematics for Engineers* (ed. Beckenback, E.) (McGraw-Hill, 1956).
18. Pesaran, M. H. & Shin, Y. An autoregressive distributed lag modelling approach to cointegration analysis. In *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium* (ed. Strøm, S.) (Cambridge Univ. Press, 1999).
19. Nolte, G. et al. Robustly estimating the flow direction of information in complex physical systems. *Phys. Rev. Lett.* **100**, 234101 (2008).
20. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **85**, 461 (2000).
21. Bastos, A. M. & Schoffelen, J.-M. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.* https://doi.org/10.3389/fnsys.2015.00175 (2016).
22. Oizumi, M., Tsuchiya, N. & Amari, S.-I. Unified framework for information integration based on information geometry. *Proc. Natl Acad. Sci. USA* **113**, 14817 (2016).
23. Ay, N. Information geometry on complexity and stochastic interaction. *Entropy* **17**, 2432 (2015).
24. Geweke, J. Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat Assoc.* **77**, 304 (1982).
25. Barnett, L., Barrett, A. B. & Seth, A. K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **103**, 238701 (2009).
26. Cliff, O. M., Prokopenko, M. & Fitch, R. Minimising the Kullback–Leibler divergence for model selection in distributed nonlinear systems. *Entropy* **20**, 51 (2018).
27. Lachaux, J.-P., Rodriguez, E., Martinerie, J. & Varela, F. J. Measuring phase synchrony in brain signals. *Hum Brain Mapp* **8**, 194–208 (1999).
28. MacKay, D. J. in *Information Theory, Inference and Learning Algorithms* (Cambridge Univ. Press, 2003).
29. Shen, C., Priebe, C. E. & Vogelstein, J. T. From distance correlation to multiscale graph correlation. *J. Am. Stat Assoc.* **115**, 280 (2020).
30. Gretton, A. et al. A kernel statistical test of independence. In *Advances in Neural Information Processing Systems* (eds Roweis, S. et al.) Vol. 20 (Curran Associates, 2008).
31. Sejdinovic, D., Sriperumbudur, B., Gretton, A. & Fukumizu, K. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Annals Stat.* **41**, 2263–2291 (2013).
32. Heller, R., Heller, Y. & Gorfine, M. A consistent multivariate test of association based on ranks of distances. *Biometrika* **100**, 503–510 (2013).
33. Ruiz, A. P., Flynn, M., Large, J., Middlehurst, M. & Bagnall, A. The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min. Knowl. Discov.* **35**, 401–449 (2021).
34. Bagnall, A. et al. The UEA multivariate time series classification archive. Preprint at https://arxiv.org/abs/1811.00075 (2018).
35. Birbaumer, N. et al. A spelling device for the paralysed. *Nature* **398**, 297–298 (1999).
36. Byrge, L. & Kennedy, D. P. Accurate prediction of individual subject identity and task, but not autism diagnosis, from functional connectomes. *Hum. Brain Mapp.* **41**, 2249 (2020).
37. Liu, Y. & Aviyente, S. Quantification of effective connectivity in the brain using a measure of directed information. *Comput. Math Methods Med.* **2012**, 635103 (2012).
38. Mehta, K. & Kliewer, J. Directional and causal information flow in EEG for assessing perceived audio quality. *IEEE Trans. Mol. Biol. Multiscale Commun.* **3**, 150–165 (2017).
39. Wang, G. & Takigawa, M. Directed coherence as a measure of interhemispheric correlation of EEG. *Int J. Psychophysiol.* **13**, 119–128 (1992).
40. Schelter, B. et al. Testing for directed influences among neural signals using partial directed coherence. *J. Neurosci. Methods* **152**, 210–219 (2006).
41. Smith, S. M. et al. Network modelling methods for FMRI. *NeuroImage* **54**, 875–891 (2011).
42. Evans, T. & Lambiotte, R. Line graphs, link partitions, and overlapping communities. *Phys. Rev. E* **80**, 016105 (2009).
43. Ahn, Y.-Y., Bagrow, J. P. & Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature* **466**, 761–764 (2010).
44. Fulcher, B. D. Feature-based time-series analysis. In *Feature Engineering for Machine Learning and Data Analytics* 87–116 (CRC, 2018).
45. Fulcher, B. D. & Jones, N. S. *hctsa*: a computational framework for automated time-series phenotyping using massive feature extraction. *Cell Syst.* **5**, 527–531.e3 (2017).
46. Lubba, C. H. et al. *catch22*: CAnonical Time-series CHaracteristics. *Data Min. Knowl. Discov.* **33**, 1821–1852 (2019).
47. Phinyomark, A. et al. Navigating features: A topologically informed chart of electromyographic features space. *J. R. Soc.* **14**, 20170734 (2017).
48. Nosek, B. A., Ebersole, C. R., DeHaven, A. C. & Mellor, D. T. The preregistration revolution. *Proc. Natl Acad. Sci. USA* **115**, 2600 (2018).
49. Leung, A. et al. Towards blinded classification of loss of consciousness: distinguishing wakefulness from general anesthesia and sleep in flies using a massive library of univariate time series analyses. Preprint at https://psyarxiv.com/rmsv8/ (2023).
50. Peixoto, T. P. Network reconstruction and community detection from dynamics. *Phys. Rev. Lett.* **123**, 128301 (2019).
51. Banerjee, A., Chandra, S. & Ott, E. Network inference from short, noisy, low time-resolution, partial measurements: application to *C. elegans* neuronal calcium dynamics. *Proc Natl Acad. Sci. USA* **120**, e2216030120 (2023).
52. Hoffmann, T., Peel, L., Lambiotte, R. & Jones, N. S. Community detection in networks without observing edges. *Sci Adv* **6**, eaav1478 (2020).
53. McCabe, S. et al. Netrd: A library for network reconstruction and graph distances. *J. Open Source Softw.* **6**, 2990 (2021).
54. Wollstadt, P. et al. IDTxl: The Information Dynamics Toolkit xl: a Python package for the efficient analysis of multivariate information dynamics in networks. *J. Open Source Softw.* **4**, 1081 (2019).
55. Peel, L., Peixoto, T. P. & De Domenico, M. Statistical inference links data and theory in network science. *Nat. Commun.* **13**, 6794 (2022).
56. Battiston, F. et al. The physics of higher-order interactions in complex systems. *Nat. Phys.* **17**, 1093–1098 (2021).
57. Cliff, O. M. *DynamicsAndNeuralSystems/pyspi* (Zenodo, 2023); https://doi.org/10.5281/zenodo.8223340

58. Cliff, O. M. *Library of Multivariate Time Series* (Zenodo, 2022); https://doi.org/10.5281/zenodo.7118947
59. Kendall, M. G. A new measure of rank correlation. *Biometrika* **30**, 81 (1938).
60. Dhamala, M., Rangarajan, G. & Ding, M. Estimating Granger causality from Fourier and wavelet transforms of time series data. *Phys. Rev. Lett.* **100**, 018701 (2008).
61. Kaneko, K. & Tsuda, I. *Complex systems: Chaos and Beyond* (Springer, 2011).
62. Strogatz, S. H. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D* **143**, 1–20 (2000).
63. Betzel, R., Kennedy, D. & Byrge, L. *Resting-State and Movie-Watching Data* (Figshare, 2020); https://doi.org/10.6084/m9.figshare.12971162.v2
64. Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
65. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825 (2011).
66. Bonferroni, C. *Teoria Statistica Delle Classi e Calcolo Delle Probabilita* Vol. 8, 3 (Seeber, 1936).
67. Bryant, A. G. *DynamicsAndNeuralSystems/pyspi_paper_classification: Documentation ft. readme and code comments* (Zenodo, 2023); https://doi.org/10.5281/zenodo.8027702

Acknowledgements

O.M.C., N.T. and B.D.F. were supported by NHMRC Ideas Grant 1183280. N.T. was supported by by Japan Society for the Promotion of Science, Grant-in-Aid for Transformative Research Areas (20H05710, 23H04830, 23H04829). A.G.B. was supported by an Australian Government Research Training Program Scholarship, an American Australian Association Graduate Education Fund Scholarship, and the University of Sydney Physics Foundation. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper. High-performance computing facilities provided by the School of Physics, the University of Sydney contributed to our results.

Author contributions

N.T. and B.D.F. conceived of the project. O.M.C. developed the software and large data repositories and performed the main analysis. A.G.B. performed the classification case study analyses. B.D.F. supervised the project with input from J.T.L. and N.T. O.M.C. and B.D.F. wrote the paper, with input from all other co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43588-023-00519-x.

Correspondence and requests for materials should be addressed to Ben D. Fulcher.

Peer review information *Nature Computational Science* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Jie Pan, in collaboration with the *Nature Computational Science* team. Peer reviewer reports are available.

2.4.5 Others

- The UEA `LSST` dataset is from the Photometric LSST Astronomical Time Series Classification Challenge to classify astronomical time-series data. The dataset contains simulations of the light curves of 14 astronomical objects across six different filters (passbands). We include 10 instances from each of the thirteen classes of this dataset, giving a total of 130 MTS for this datatype.

- The UEA `EigenWorms` dataset is intended to classify *C. Elegans* as either wild types or (one of four) mutant types based on time series recordings of their movements (which is translated via a projection onto a base shape known as eigenworms). We truncate the dataset to have $T = 3000$ observations. We include 10 instances from each of the five classes of this dataset, giving a total of 50 MTS for this datatype.

- The UEA `PEMS-SF` data describes the occupancy rate (between 0 and 1) of different car lanes of San Francisco bay area freeways from Jan 1, 2008 to Mar 30, 2009. Each time series represents a single day sampled every 10 minutes (giving 144 observations) and the goal was to classify the correct day of the week. We subsample the channels by randomly choosing a subsequence of lanes (processes) with a random length $M \in [10, 30]$. We include 10 instances from each of the seven classes of this dataset, giving a total of 70 MTS for this datatype.

- We obtained county-level COVID-19 cases in the USA, including records from 3006 counties over $T = 700$ days of the pandemic [102]. From this database, we generated 10 instances of both the incidence (number of daily cases) and the cumulative incidence (cumulative sum of daily cases), where, in each instance, we randomly chose $M \in [7, 15]$ counties.

3 Supplementary Methods and Results

3.1 The empirical similarity index

Some example MPIs are shown in Supplementary Fig. 1A for SPIs based on transfer entropy (SPI 73), coherence magnitude (SPI 95), Kendall's τ (SPI 18), and convergent cross-mapping (SPI 56) applied to three different MTS. Different SPIs can yield quite different structures in their corresponding MPIs, and the level of similarity can vary across different datasets. For example, in Supplementary Fig. 1A, Kendall's τ and convergent cross-mapping exhibit markedly different behavior for the currency pair but qualitatively similar behavior on the coupled map lattice and human fMRI data.

The 1053 MTS have an average of approximately 11.89 processes, yielding 195 112 pairwise interactions in total with which to compute the empirical similarity index per SPI pair. Note, however, that undirected statistics have repeated entries in the upper and lower diagonals of the MPIs, and so the number of effective samples (in computing the correlation coefficient) is halved, which may induce a higher variance for undirected SPIs across all datasets when compared to directed SPIs [10].

The calculation of r_d is depicted as ranked scatter plots in Supplementary Fig. 1B for the similarity on human fMRI data between: *(i)* Kendall's τ [1] and convergent cross-mapping (CCM) [29], $r_d = 0.57$; and *(ii)* transfer entropy [35] and coherence, $r_d = -0.34$. For both of these pairs of SPIs, the distribution of $|r_d|$ across all 1053 MTS is also shown in Supplementary Fig. 1B: the empirical similarity index between two SPIs, R, is then the average of this distribution, as $\langle |r_d| \rangle_d$. While it was important to have a single summary of this similarity for a pair of SPIs, as the mean, it is important to note that some pairs of SPIs have quite a wide distribution of scores, $|r_d|$, across datasets, d, indicating that they can give highly correlated outputs on some MTS, but not on others. An example of a wide distribution is shown in Suppelementary Fig. 1B, where the distribution of Kendall's τ and CCM is approximately uniformly

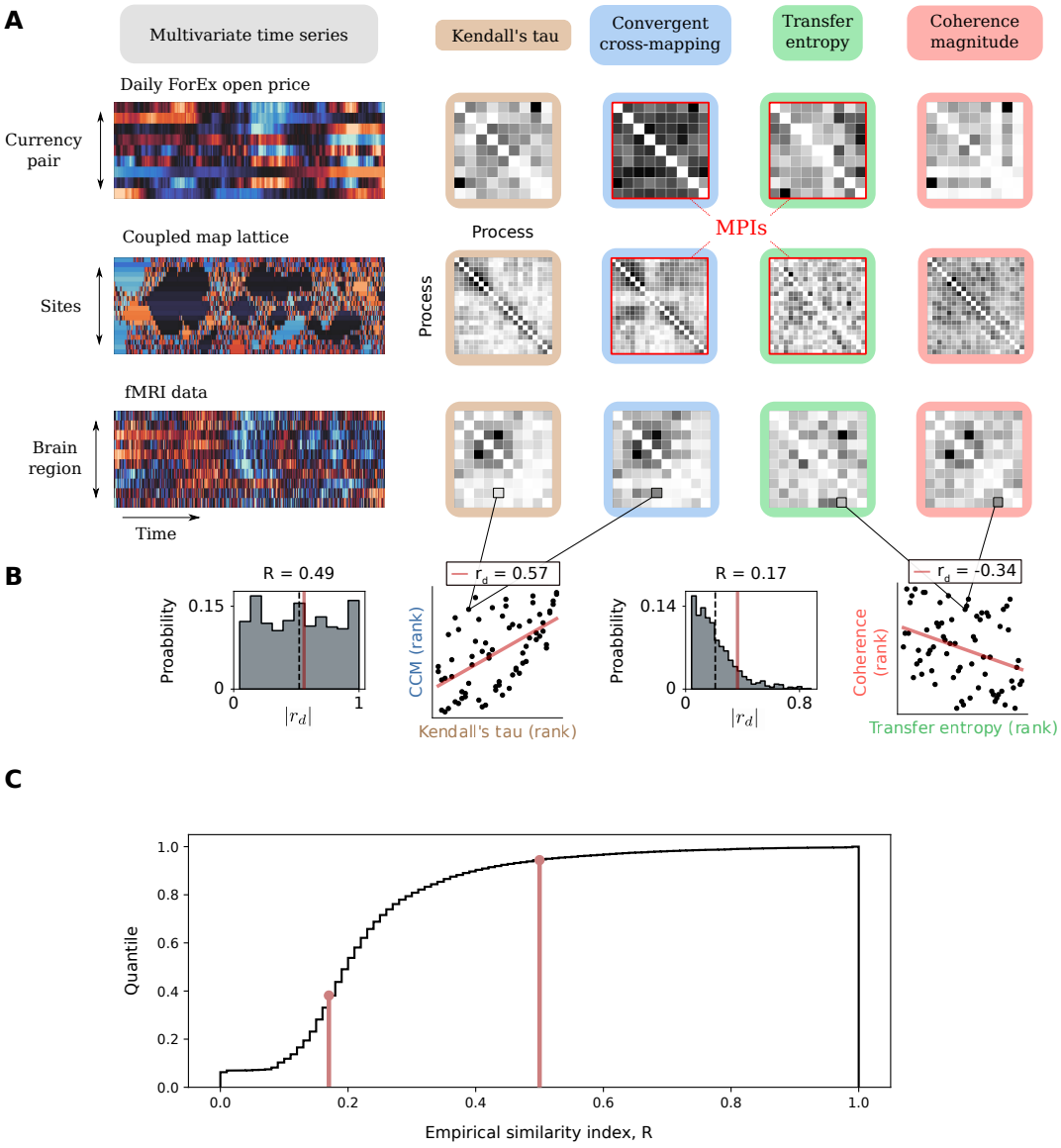
distributed across MTS, ranging from some MTS for which the two SPIs are uncorrelated, $|r_d| \approx 0$, to other MTS for which there is perfect correlation, $|r_d| \approx 1$.

When using this approach to compare each pair of SPIs across the MTS library, we found a wide range of similarity indices, from $R \approx 0$ to $R \approx 1$, illustrated by the cumulative distribution function in Supplementary Fig. 1C. Each index was then used to construct the dendrogram (illustrated in Supplementary Fig. 2) that gave us our findings in the main text. While we were able to use our index to capture important empirical relationships using hierarchical clustering, it is difficult to interpret the raw index value, R, in terms of a statistical model (for, e.g., hypothesis testing). This is because of two reasons: *(i)* many MPIs contain values that are not completely independent of one another (due to higher-order dependencies), and *(ii)* the set of all MPIs are themselves not completely independent (having occasionally overlapping dynamics). As discussed in Sec. 2, the fact that the data used in computing the index are not independent may result in fewer effective samples than expected, which could bias the estimate of an absolute Spearman correlation and make many SPI pairs appear to have a higher correlation than they actually do. Nevertheless, the effective sample size of the dataset does not affect empirical clustering.

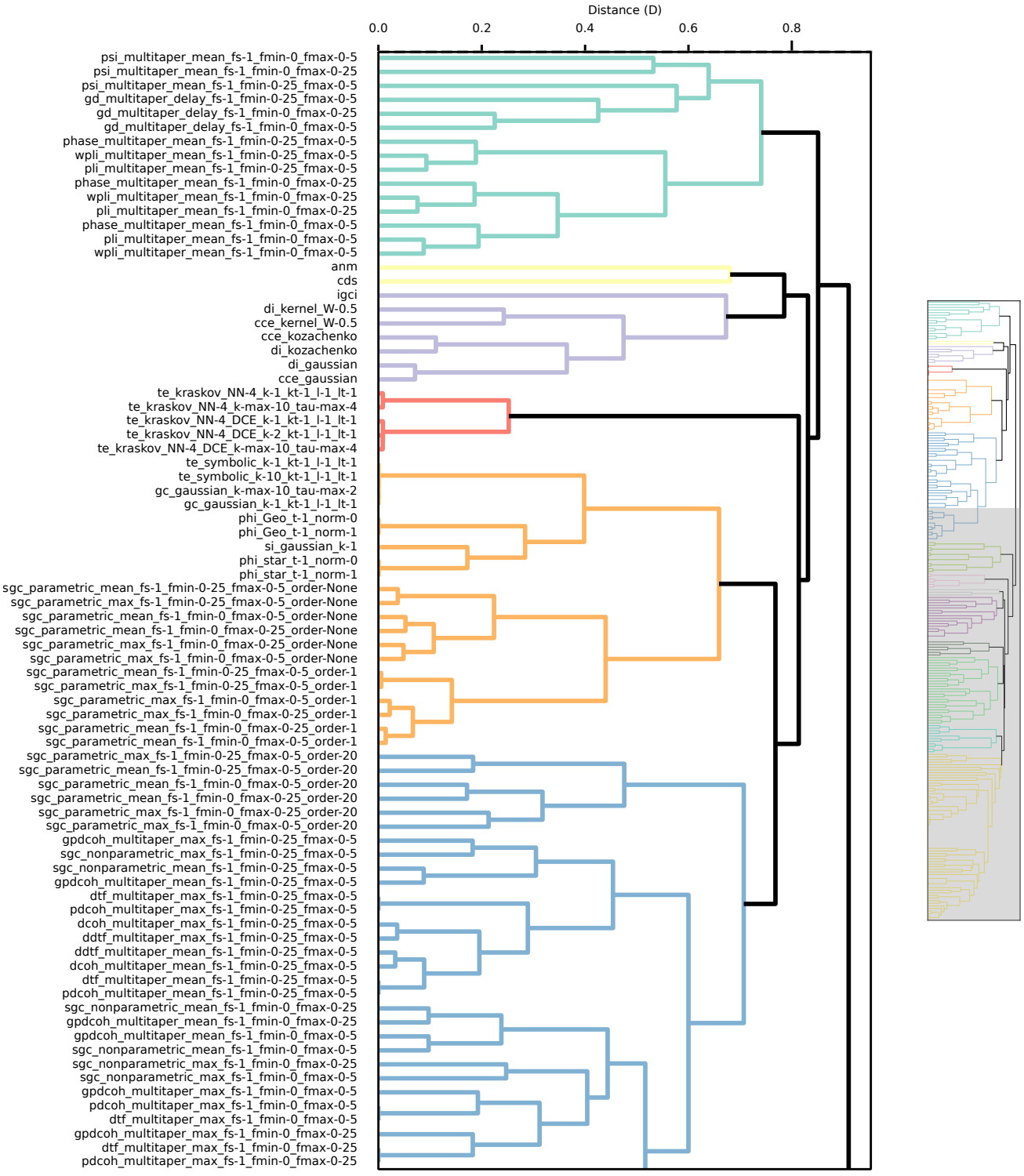
Note that our main results are not highly dependent on the choice to summarize the distribution of $|r_d|$ scores as a mean rather than a median. The two choices yield a set of SPI–SPI similarity scores (that form the basis of the clustering shown here) that are highly correlated to each other, $\rho = 0.99$, and result in a modular decomposition that is highly overlapping (a Rand index of 0.95 [103], using the same threshold, $R = 0.76$, to form clusters).

3.2 Modular versus literature categorization of SPIs

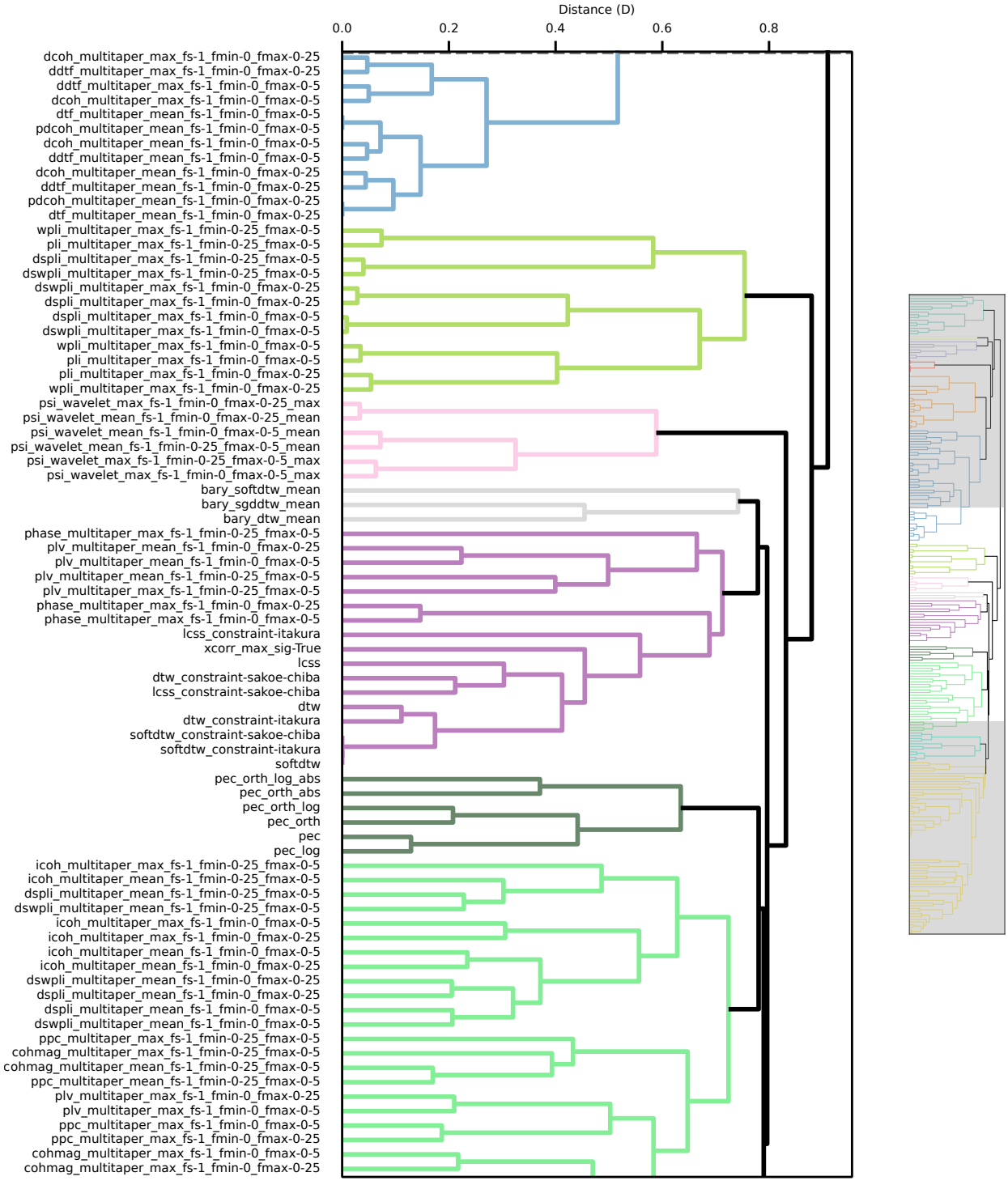
We aimed to explore whether the literature categories or data-driven modules provided a more succinct representation of algorithm performance on the three MTS classification datasets studied in this paper. Grouping the SPIs by the six literature categories (of main text Fig. 1B) revealed a wide variability of classification performance within each category. By contrast, as shown in Supplementary Fig. 3, grouping SPIs by the fourteen data-driven modules (from main text Fig. 2A) results in far more characteristic performance levels within modules, suggesting our data-driven modular representation as a useful way to capture differential performance of SPIs on a given task. For example, on the smartwatch activity dataset, SPIs from each literature category display a wide range of performance, including the information-theoretic SPIs (green in Supplementary Fig. 3A), which vary in accuracy from 45% (for the undirected SPI stochastic interaction with a Kozechenko–Leonenko estimator, `si_kozachenko_k-1`) to 92% (for the directed SPI causally conditioned entropy with a Kozachenko–Leonenko estimator, `cce_kozachenko`). By contrast, we identify two modules containing consistently high-performing SPIs: M4 (transfer entropy) and M5 (parametric Granger causality and integrated information), in Supplementary Fig. 3B. These methods relate to measuring the communication and integration of information in a multidimensional process, adopting the Wiener–Granger theory of 'causality' and 'feedback' discussed earlier, suggesting that this way of conceptualizing and quantifying dependencies (predicated on self-predictability) is a useful one for analyzing these human behavioral recordings. We find qualitatively similar results in the EEG state dataset (Suppelementary Figs 3C,D) and fMRI film dataset (Supplementary Figs 3E,F).



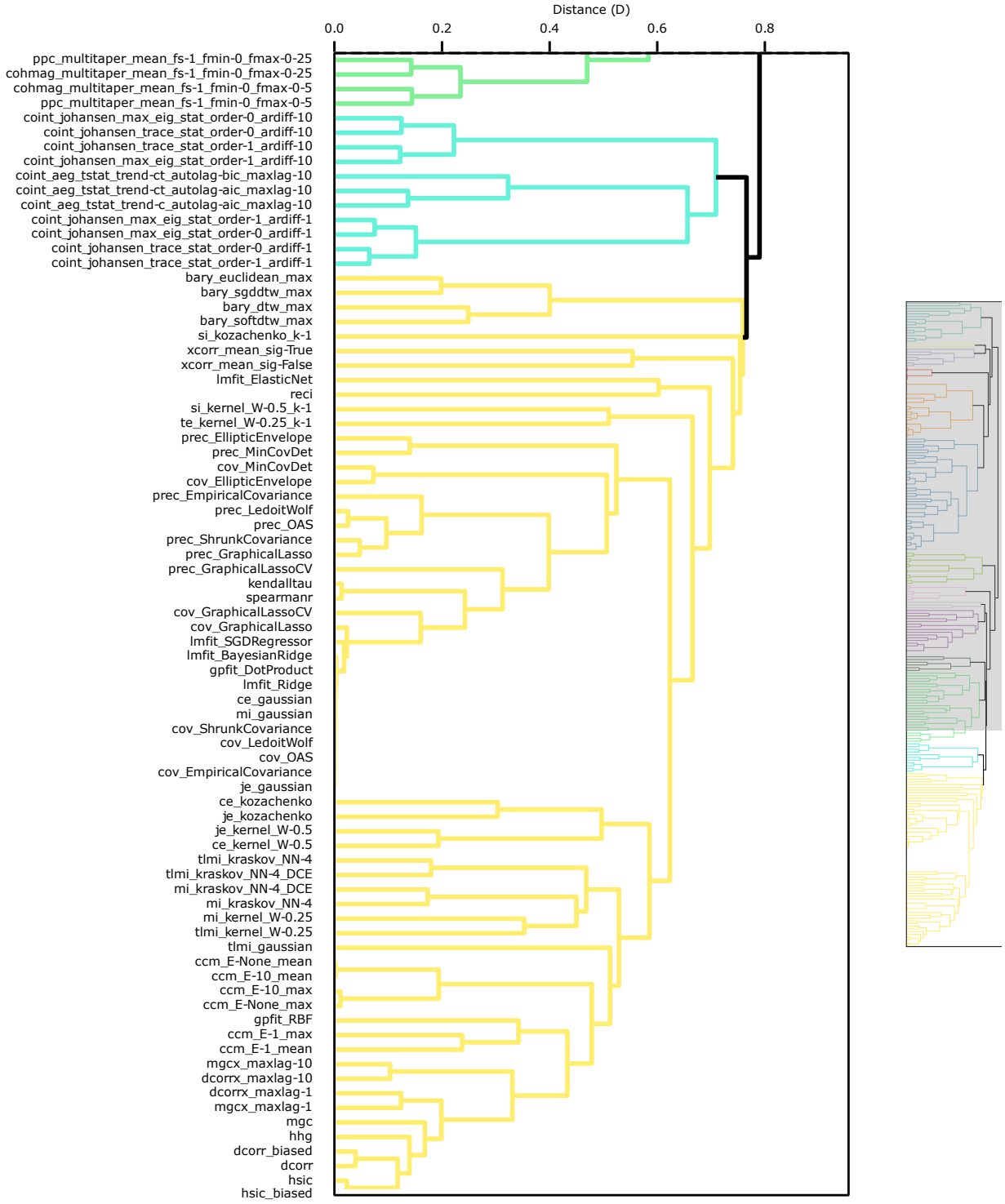
Supplementary Figure 1: **We use the empirical similarity index, R, to measure the similarity between two SPIs, based on the Spearman's correlation coefficient between their pairwise matrices, averaged across all MTS. A.** From our MTS library, here we show three selected examples: *(i)* exchange rates; *(ii)* coupled maps; and *(iii)* a functional Magnetic Resonance Imaging (fMRI) dataset. For each dataset, we also show how pairwise dependencies between all elements of the system can be captured as a matrix of pairwise interactions (MPI). The MPIs are normalized and shaded in grayscale, with white being the lowest value, and black being the highest. We show four selected examples from our library of SPIs: *(i)* Kendall's τ; *(ii)* convergent cross-mapping; *(iii)* transfer entropy; and *(iv)* coherence magnitude. **B.** On a given dataset, we quantified the similarity between each pair of SPIs as the absolute Spearman's correlation coefficient of (non-diagonal) elements of the resulting pairwise matrices, as $|r_d|$, and averaged this quantity across all MTS to compute our empirical similarity score, R. Probability distributions of $|r_d|$ are shown for two SPI pairs: *(i)* Kendall's τ and convergent cross-mapping, and *(ii)* transfer entropy and coherence magnitude. The resulting R, the average of this distribution, is shown as a vertical dashed line. For each of these SPI pairs we also plot an example ranked scatter plot for the fMRI dataset (yielding an $|r_d|$ value annotated red in the distribution plot). **C.** The cumulative distribution function for the empirical similarity index across all pairs of SPIs, with annotations for: *(i)* transfer entropy and coherence magnitude ($R = 0.17$); and *(ii)* Kendall's τ and convergent cross-mapping ($R = 0.49$).



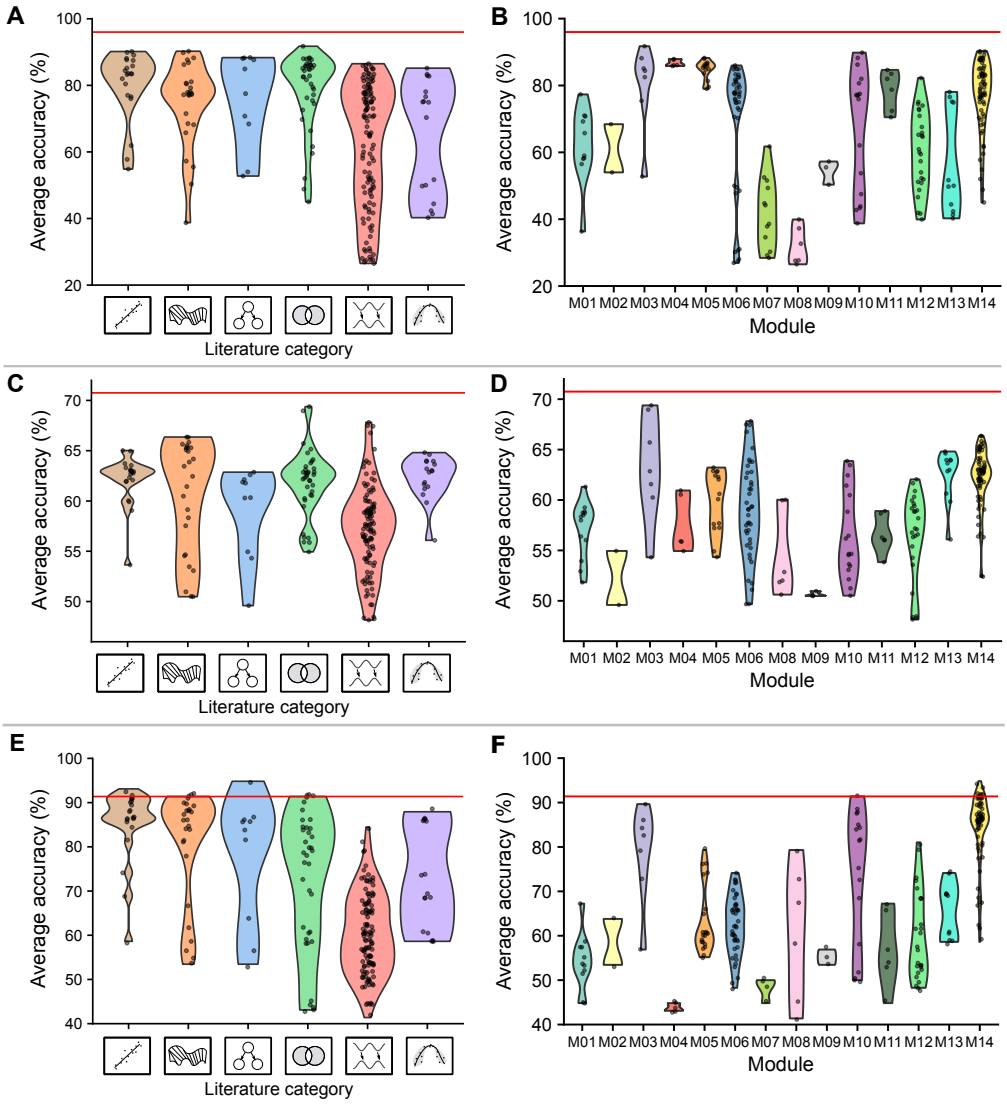
Supplementary Figure 2: **A high-resolution dendrogram illustrates the empirical organization of SPIs at a fine-grained scale.** The color scheme illustrates clusters of SPIs below a cut-off of 0.76, and leaf nodes are labeled according to the SPI identifiers given in Supplementary Sec. 1. See Supplementary Data 1 for a list of SPIs assigned to modules.



(continuation of previous figure)



(continuation of previous figure)



Supplementary Figure 3: **Grouping SPIs by literature category can result in greater performance variability on a given MTS classification problem than grouping SPIs by modules.** Classification accuracy distributions over SPIs are shown as a violin plot when grouping SPIs as literature-derived categories, see main text Fig. 1B (left panels **A**, **C**, **E**) and by modular categories, see main text Fig. 2A (right panels, **B**, **D**, **F**). Results are shown for each datasets analyzed: smartwatch movement (**A**, **B**), EEG state (**C**, **D**), and fMRI film (**E**, **F**). Classification accuracy using the combination of all SPIs is shown as a horizontal red line.

Appendix 3: Supplementary material for Chapter 4

Supporting Information Appendix for 'A data-driven approach to identifying and evaluating connectivity-based neural correlates of consciousness'

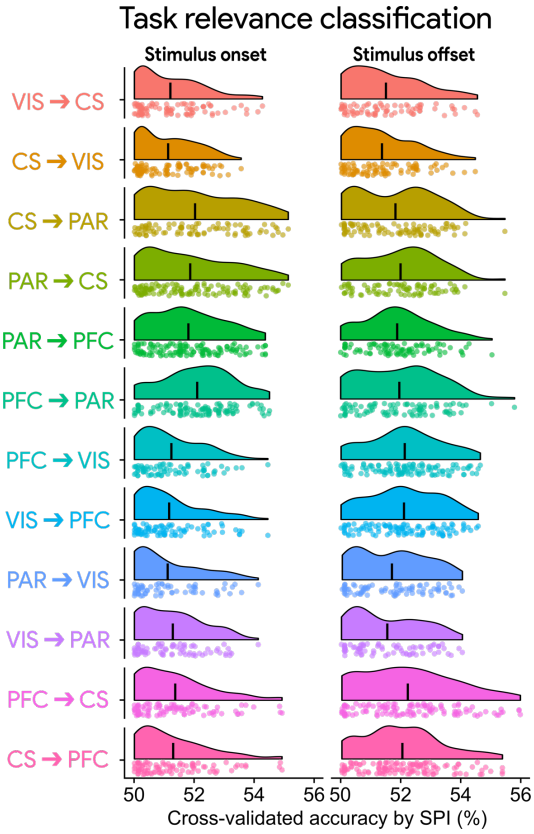


Figure S1. Domain-independent task relevance is not well classified across FC measures and region–region pairs. Each raincloud plot shows the distribution of classification accuracy values for each FC measure in the corresponding region–region pair, partitioned by stimulus onset or stimulus offset period. For visualization purposes, the x-axis was truncated at 50% accuracy for both stimulus presentation periods, such that SPIs yielding lower than 50% accuracy are not shown here.

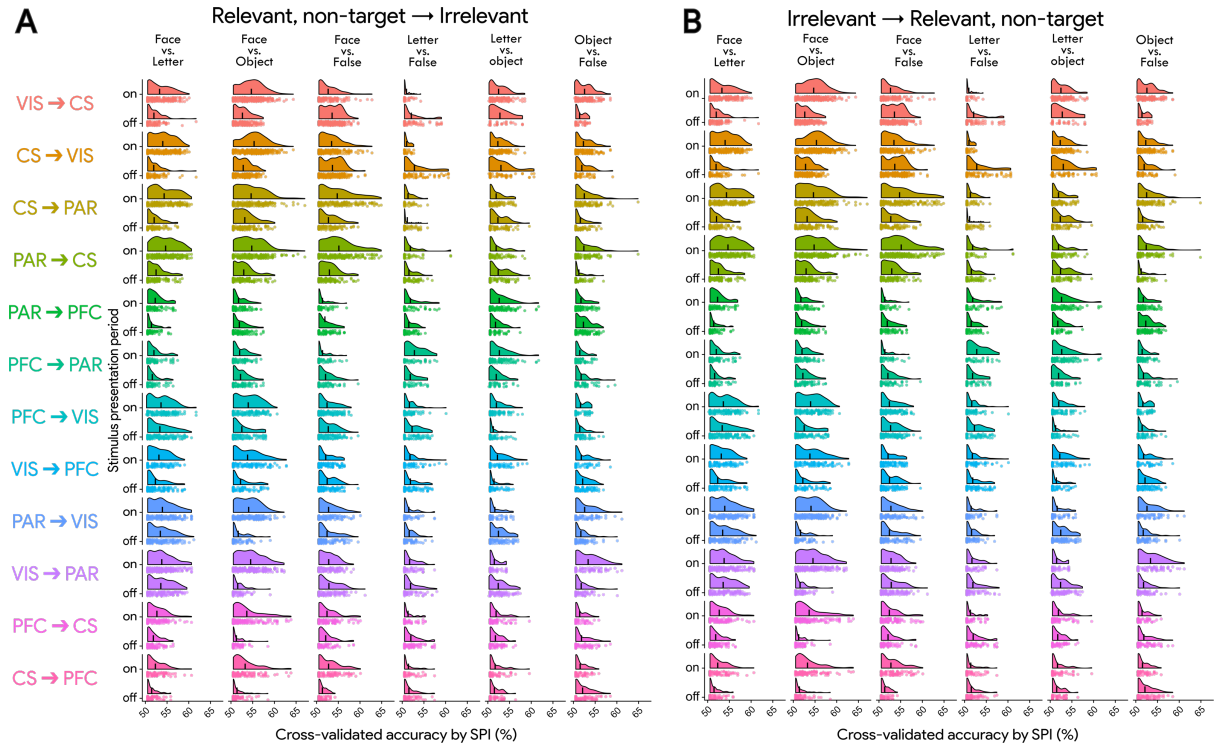


Figure S2. Cross-task classification between stimulus pairs shows comparable performance to that seen in the main stimulus type classification. Each raincloud plot shows the distribution of classification accuracy values for each FC measure in the corresponding region–region pair, resulting from classifiers trained on stimulus pairs in one relevance setting (e.g., 'Relevant non-target') and tested on the same stimulus pairs in the other relevance setting (e.g., 'Irrelevant'). For visualization purposes, the x-axis was truncated at 50% accuracy for both stimulus presentation periods, such that SPIs yielding lower than 50% accuracy are not shown here.

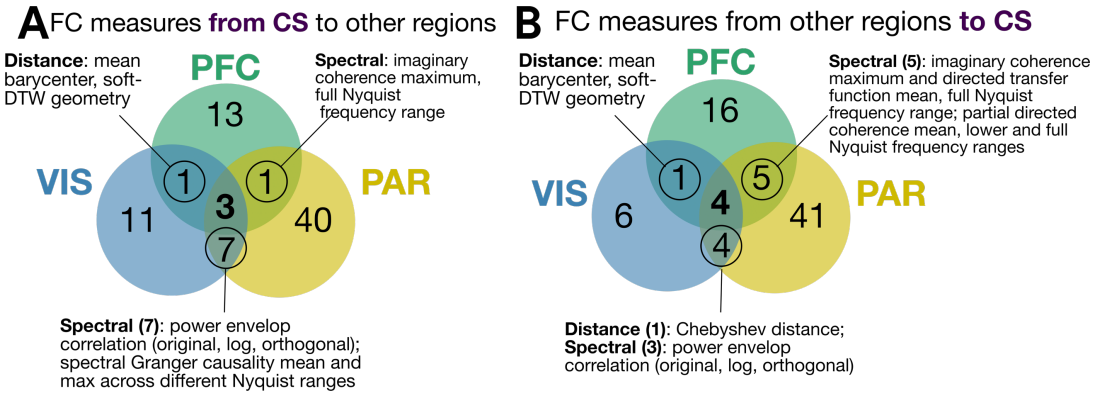


Figure S3. Detailed breakdown of the top-performing FC measures shared between two functional pathways. A The same Venn diagram from Fig. **4**A is reproduced, with detailed annotation of the FC measures comprising the intersection between two functional pathways projecting **from** CS to VIS, PFC, or PAR. **A** The same Venn diagram from Fig. **4**B is reproduced, with detailed annotation of the FC measures comprising the intersection between two functional pathways projecting **to** CS from VIS, PFC, or PAR.

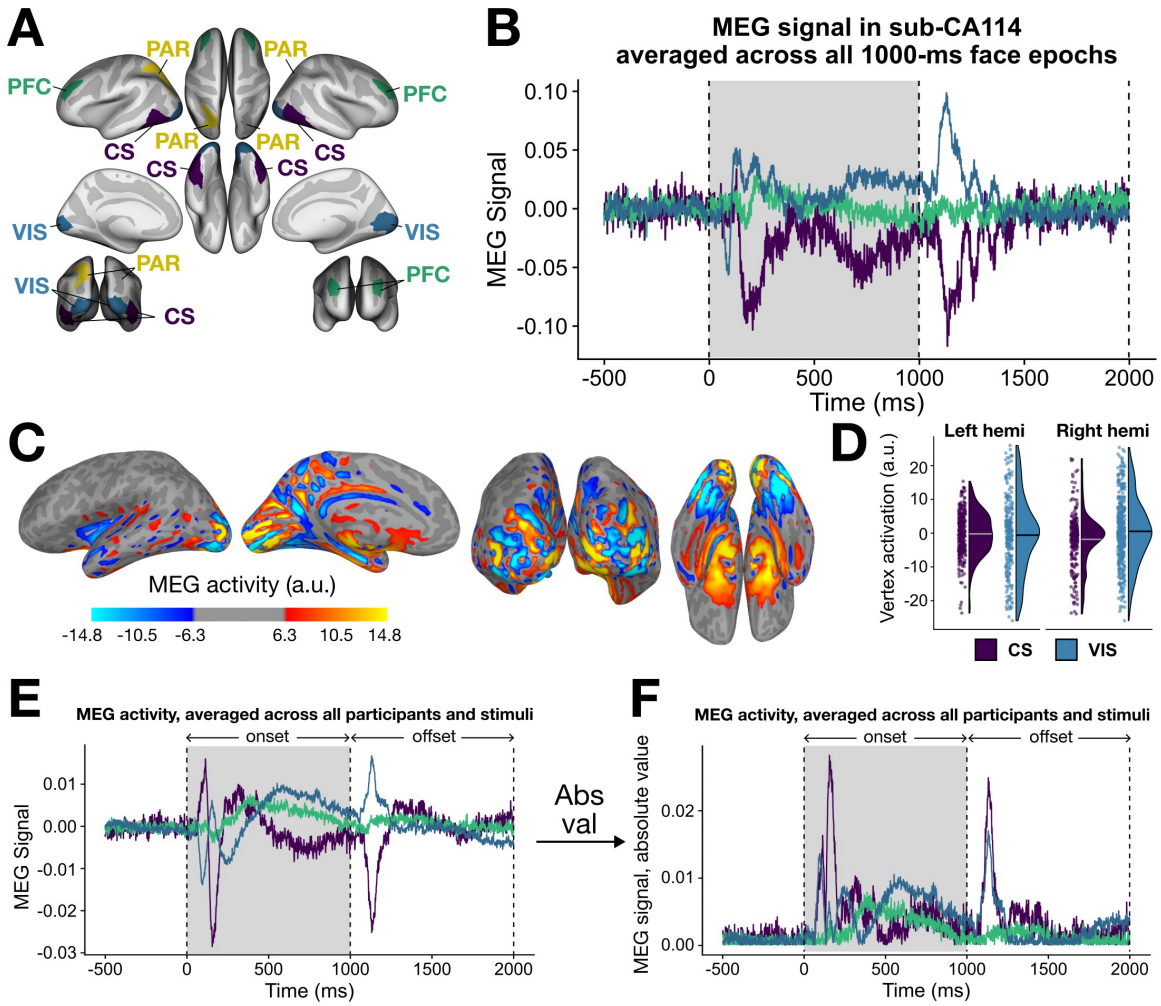


Figure S4. Analyzing regional polarity in MEG signal supports computing the absolute value of MEG amplitude. A. The four regions of interest examined in this study are shown on the cortical surface in `fsaverage` space. **B.** The average MEG signal is shown for one participant (sub-CA114), with activity averaged across all epochs per region. The regions are colored as in **A**. Stimulus onset is denoted with a shaded gray rectangle (occurring between 0 to 1000 ms), followed by the 1000-ms offset period. **C.** The raw MEG signal is plotted at the point of peak brain-wide intensity, corresponding to 153 ms post-onset. Larger activity magnitudes indicate greater electrical current, arising primarily from postsynaptic activity in pyramidal neurons in the cortex, with the sign indicating the orientation relative to the MEG sensors. **D.** Vertex-wise activation values are summarized in raincloud plots for the CS and VIS regions in the left and right hemispheres, respectively. Horizontal bars in each half-violin represent the mean activation value in artificial units (a.u.). **E.** The average MEG signal is shown, computed across all participants (N=94) and all epoch types, per brain region. The regions are colored as in **A**. Stimulus onset is denoted with a shaded gray rectangle (occurring between 0 to 1000ms), followed by the 1000-ms offset period. **F.** The absolute value is obtained from the average MEG time series shown in **E**.

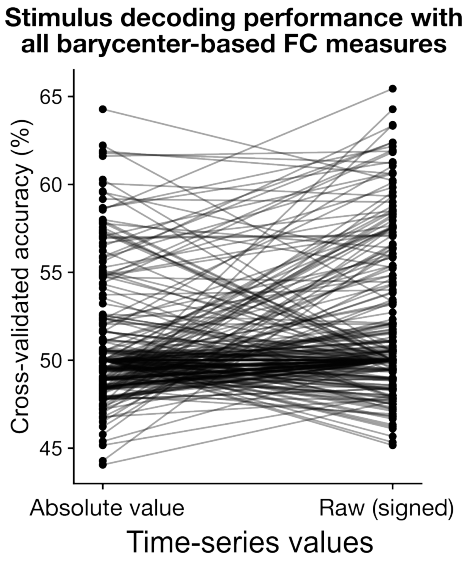


Figure S5. Stimulus decoding performance with barycenter-based FC measures does not systematically vary between absolute value versus raw (signed) empirical MEG time series.

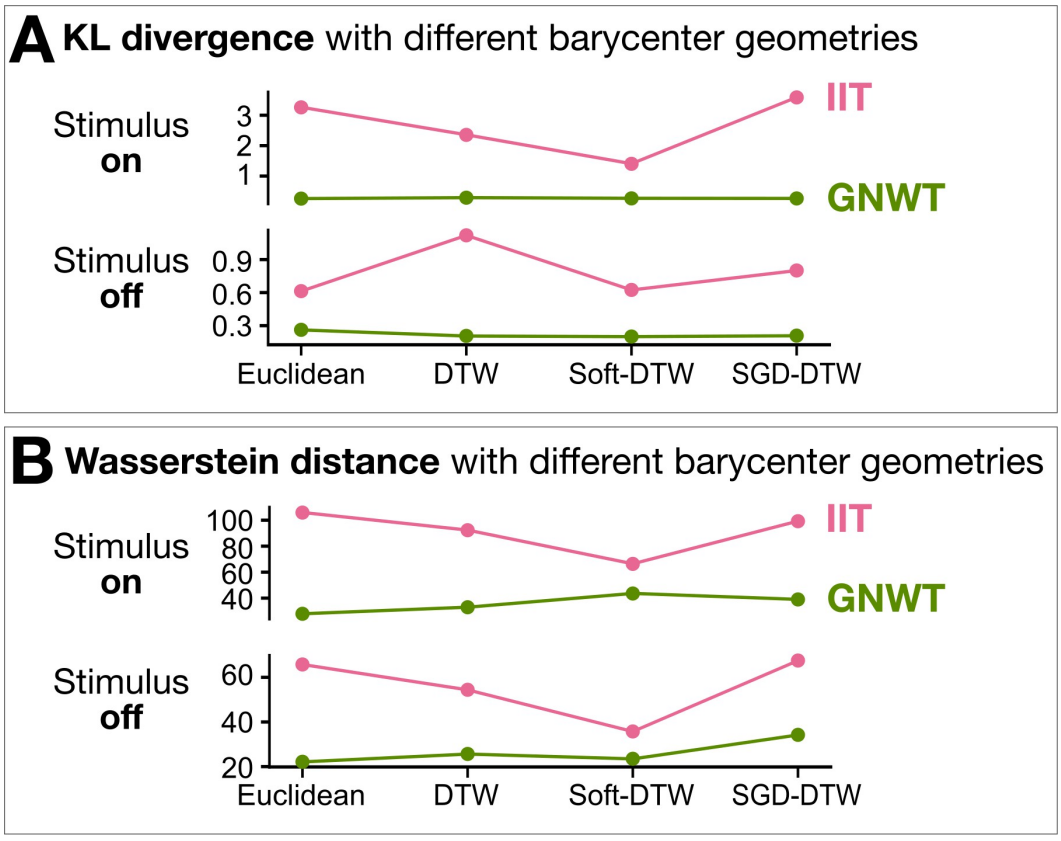


Figure S6. K-L divergence and Wasserstein distance between empirical vs. model-based maximum squared barycenter timepoints show consistent trends across different Euclidean and non-Euclidean barycenter geometries.

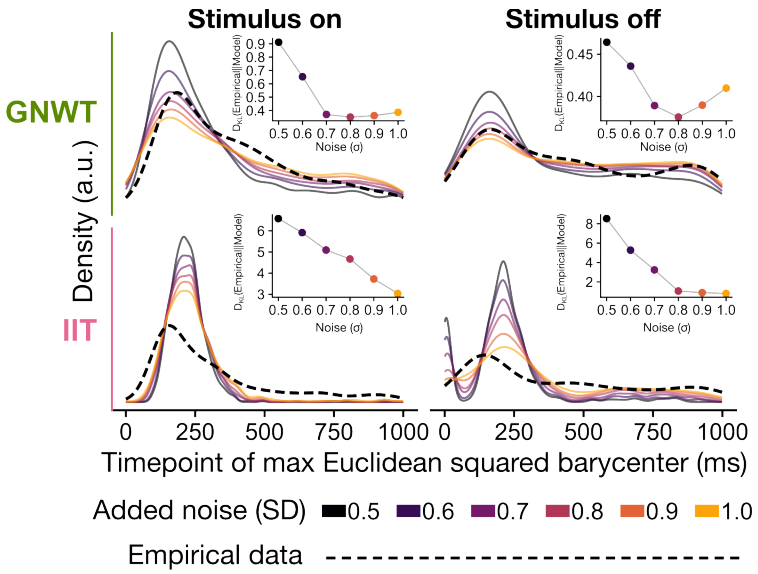


Figure S7. K-L divergence patterns are robust across added noise levels for simulated time series between GNWT and IIT. In each stimulus presentation period ('on' vs. 'off') and model context ('GNWT' vs. 'IIT'), the distribution of timepoints for the maximum squared Euclidean barycenter is plotted for the empirical MEG data (thicker dashed black line) as well as for model-simulated data with noise thresholds ranging from $0.5 \leq \sigma \leq 1.0$. The inset in each graph depicts the K-L divergence between the empirical versus model-simulated time series at the corresponding noise level.

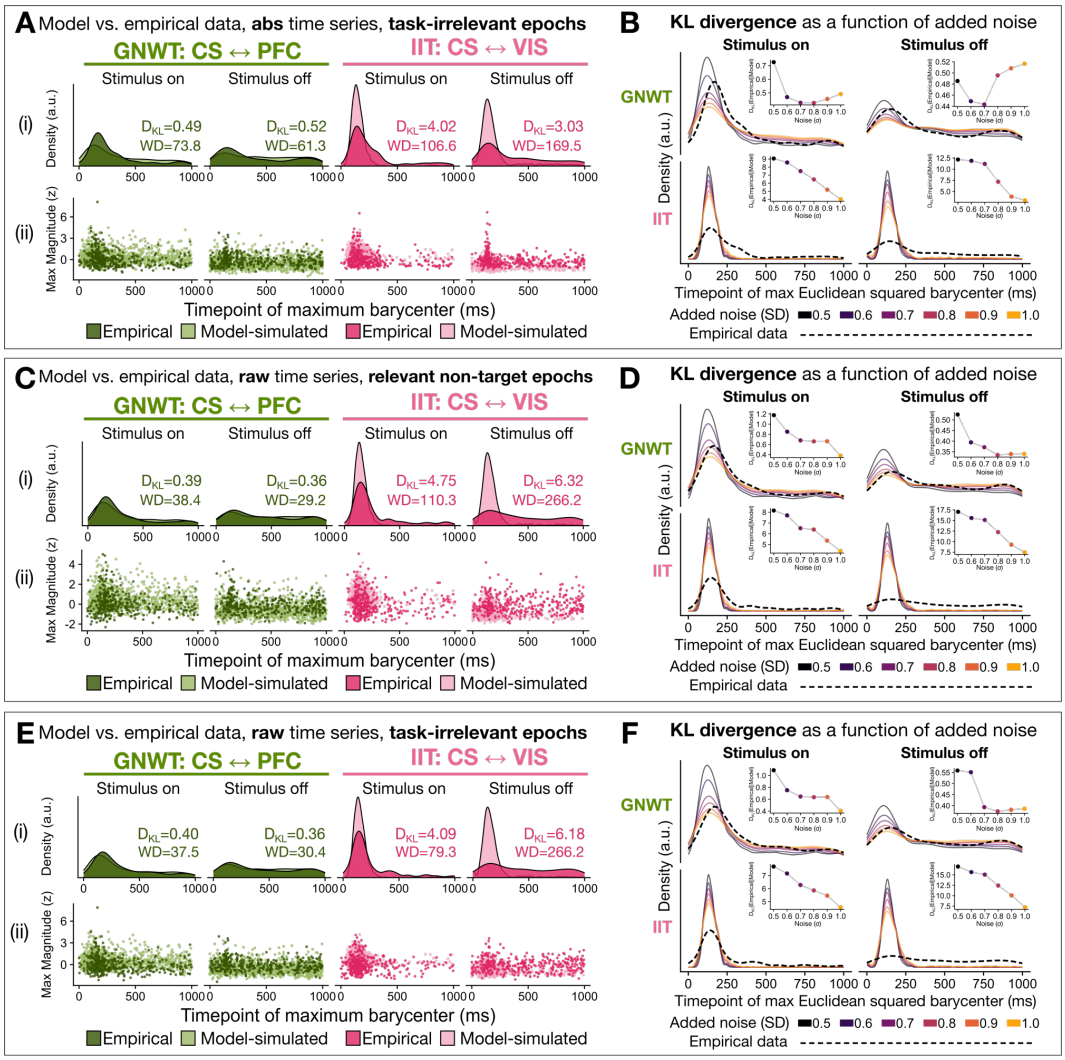


Figure S8. Timing of the maximum squared Euclidean center is better approximated by the GNWT-based than IIT-based model across stimulus relevance settings, with either raw or absolute value time series. A, C, E. In each setting, we compare the distribution of timepoints corresponding to the maximum squared Euclidean barycenter for the empirical data (darker shading) versus model-simulated time series (lighter shading) for GNWT in green, corresponding to CS ↔ PFC, and for IIT in pink, corresponding to CS ↔ VIS in pink. The kernel-estimated densities are shown in each panel in (i), with the K-L divergence (D_{KL}) and Wasserstein distance (WD) given for each empirical vs. model-simulated comparison. The timepoints and corresponding maximum barycenter magnitudes are shown in each (ii) panel. Here, we compare (**A**) absolute time series in task-irrelevant epochs; (**C**) raw time series in relevant non-target epochs; and (**E**) raw time series in task-irrelevent epochs. **B, D, F.** In each stimulus presentation period ('on' vs. 'off') and model context ('GNWT' vs. 'IIT'), the distribution of timepoints for the maximum squared Euclidean barycenter is plotted for the empirical MEG data (thicker dashed black line) as well as for model-simulated data with noise thresholds ranging from $0.5 \leq \sigma \leq 1.0$. The inset in each graph depicts the K-L divergence between the empirical versus model-simulated time series at the corresponding noise level.

Appendix 4: Supplementary material for Chapter 5

Supplementary figures for "Extracting interpretable signatures of whole-brain dynamics through systematic comparison"

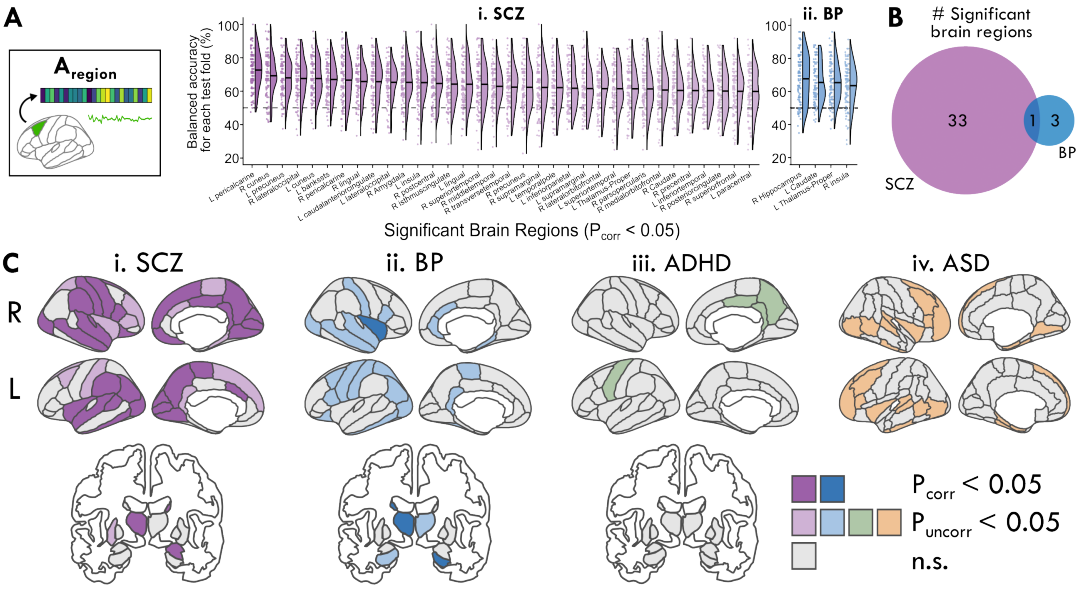


Figure S1. While several brain regions exhibited altered dynamics in each case–control comparison, only a subset were significant in SCZ and BP after multiple comparisons. A. The distribution of balanced accuracy values across test folds is shown as a raincloud plot for each significant brain region ($P_{adj} < 0.05$, corrected across 82 regions) in SCZ (**i**) and BP (**ii**). The horizontal line within each half-violin indicates the mean balanced accuracy for the corresponding brain region. B. The Venn diagram illustrates the number of significant brain regions for each of SCZ (purple) and BP (blue), indicating that one brain region (left thalamus) is shared between the two disorders. **C.** For each of the four disorders, regions are shaded dark to indicate $P_{adj} < 0.05$ (corrected across 82 regions for SCZ, BP, and ADHD; across 48 regions for ASD). Additionally, regions are shaded light to indicate that the nominal uncorrected $P < 0.05$. Gray shading indicates that the uncorrected $P > 0.05$ for the balanced accuracy in the given region.

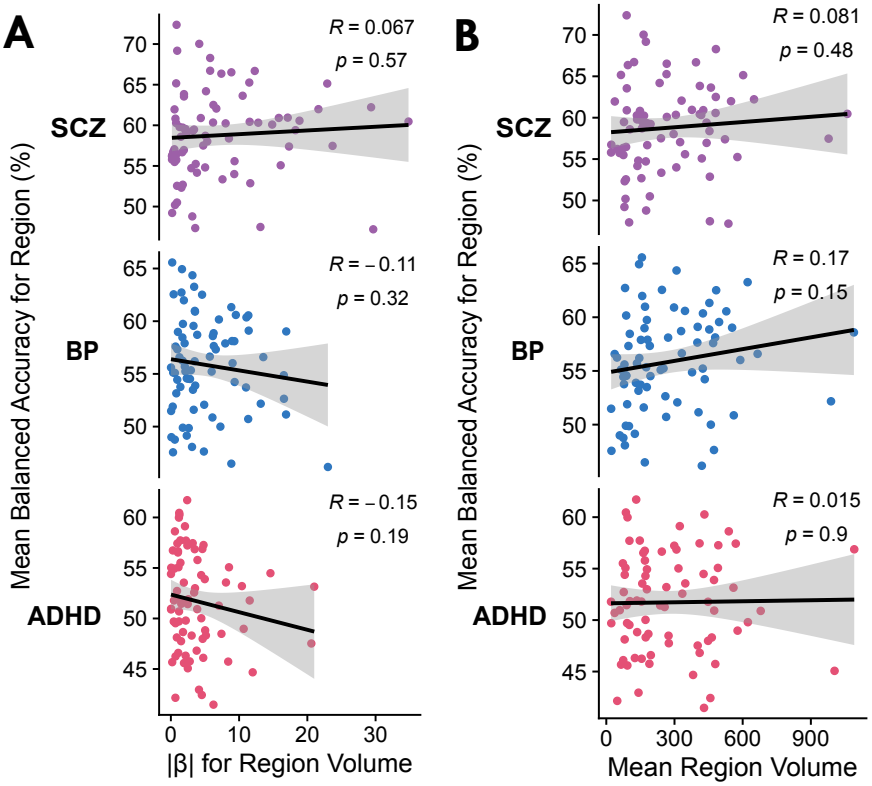


Figure S2. Region-wise classification balanced accuracy is not associated with volumetric differences across clinical groups. A. For each brain region in the UCLA CNP dataset, the mean balanced accuracy is plotted relative to the absolute β coefficient estimated from ordinary least squares regression of region volume on diagnosis per clinical group. Pearson correlation estimates (R) and corresponding P-values are annotated in the top right corners. **B.** As in A, for each brain region, the mean balanced accuracy is plotted relative to the average region volume (measured in number of voxels) across all participants in the UCLA CNP cohort. Pearson correlation estimates, R, and corresponding P-values are shown in the top right corners.

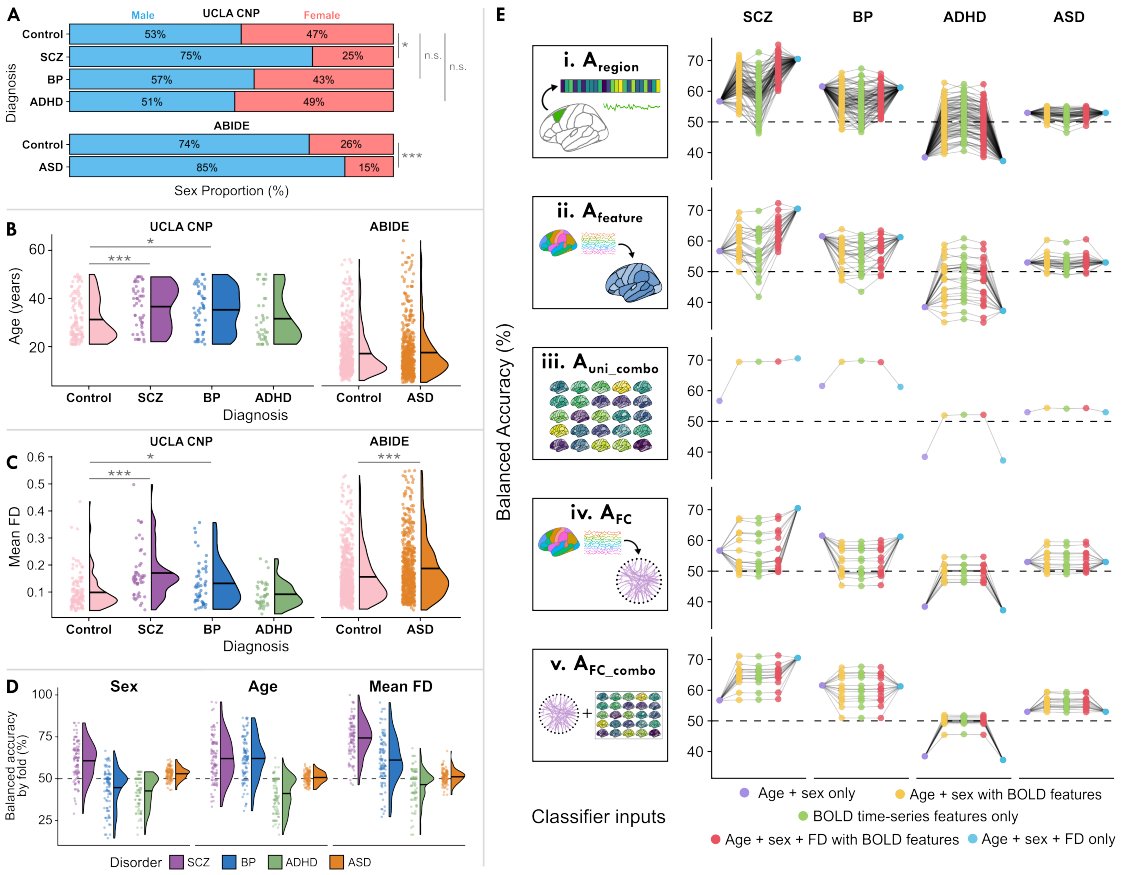


Figure S3. BOLD fMRI feature-based representations generally improve diagnostic classification beyond age, sex, and mean FD. A. The percentage of males (blue) and females (red) per diagnostic group is shown for the UCLA CNP cohort (upper) and the ABIDE cohort (lower). Sex proportions were compared for each disorder relative to the corresponding control group using a chi-square test, with significance level indicated as $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$, n.s. $P > 0.05$. **B.** The distribution of participant ages is shown as a raincloud plot per diagnostic group for the UCLA CNP cohort (left) and the ABIDE cohort (right). The horizontal line within each half-violin indicates the mean age for the corresponding distribution. Age distributions were compared for each disorder relative to the corresponding control group using a Wilcoxon rank-sum test, with significance level indicated as $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$, n.s. $P > 0.05$. **C.** The mean framewise displacement (FD) computed with the method from Power et al. [176] is shown with raincloud plots for all participants in the UCLA CNP (left) and ABIDE cohorts (right). The horizontal line within each half-violin indicates the mean FD for the corresponding group. Mean FD distributions were compared for each disorder relative to the corresponding control group using a Wilcoxon rank-sum test, with significance level indicated as $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$. **D.** Case–control classification balanced accuracy is shown for all 100 test folds per each disorder based on participant sex, age, or mean FD. The horizontal line within each half-violin indicates the mean balanced accuracy for the corresponding distribution. The dashed horizontal line in the background denotes 50% balanced accuracy. **E.** For each representation type (rows) and neuropsychiatric disorder (columns), the classification accuracy of each model (e.g., left pericalcarine cortex in A_{region}) is compared using five distinct classier inputs: (1) age + sex only, in purple; (2) age + sex with BOLD time-series features from the given model, in yellow; (3) just BOLD time-series features, in green; (4) age + sex + mean FD with BOLD time-series features, in red; and (5) age + sex + mean FD only. Each dot corresponds to one model (e.g., the left pericalcarine cortex in [i] A_{region}) and lines connect model types to guide visual comparison. The horizontal dashed line in each plot is included to show the chance baseline performance of 50% balanced accuracy.

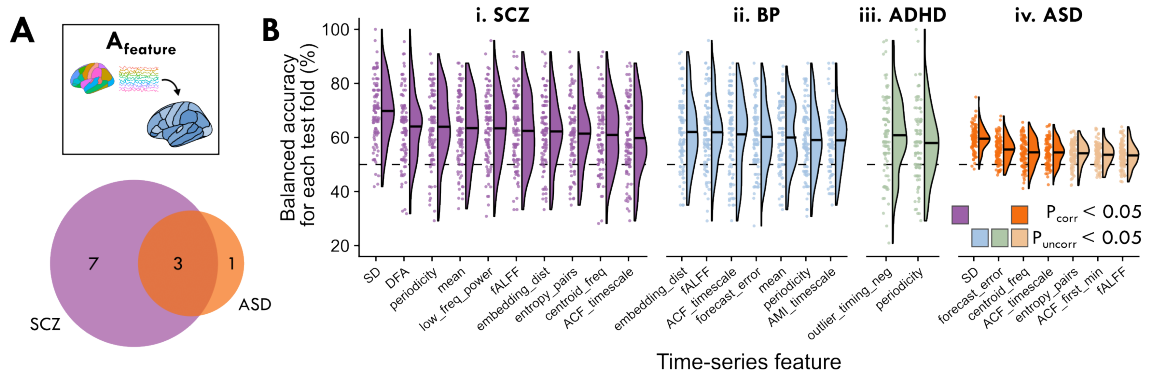


Figure S4. While several brain regions exhibited altered dynamics in each case–control comparison, only a subset were significant in SCZ and BP after multiple comparisons. A. The Venn diagram illustrates the number of significant intra-regional time-series features for each of SCZ (purple) and ASD (orange), indicating that three features are shared between the two disorders. B. The distribution of balanced accuracy values across test folds is shown as a raincloud plot for each intra-regional time-series feature that yielded a balanced accuracy with either $P_{\text{corr}} < 0.05$ (darker) or $P_{\text{uncorr}} < 0.05$ (lighter). The horizontal line within each half-violin indicates the mean balanced accuracy for the intra-regional time-series feature.

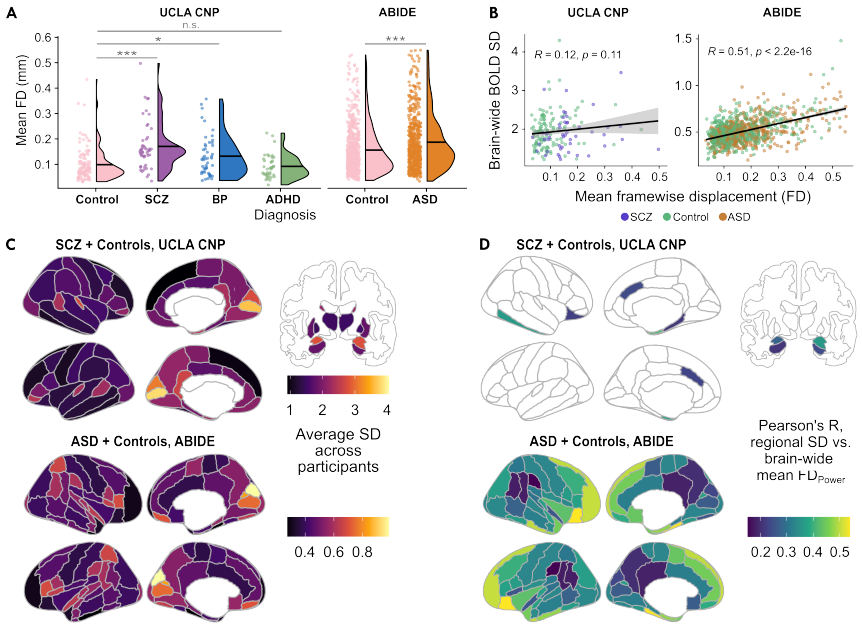


Figure S5. Head motion is generally higher in cases than in controls, but is only associated with brain-wide BOLD SD in the ABIDE cohort. **A.** The mean framewise displacement (FD) computed with the method from Power et al. [176] is shown with raincloud plots for all participants in the UCLA CNP (left) and ABIDE cohorts (right). The horizontal line within each half-violin indicates the mean FD for the corresponding group. Mean FD distributions were compared for each disorder relative to the corresponding control group using a Wilcoxon rank-sum test, with significance level indicated as $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$, n.s. $P > 0.05$. **B.** The brain-wide average BOLD SD is plotted against the mean FD across SCZ and control participants in the UCLA CNP cohort (upper) as well as ASD and control participants in the ABIDE cohort (lower), with the Pearson correlation estimates (R) and corresponding p-values shown in each plot. **B.** For each brain region, the average BOLD SD is shown across all SCZ an control participants in the UCLA CNP cohort (upper) as well as ASD an control participants in the ABIDE cohort (lower). Note that different color scales are used for the two cohorts, respectively. **D.** For each brain region, we computed the Pearson correlation between the region-wise BOLD SD and whole-brain mean FD values in the UCLA CNP cohort (upper) as well as ASD and control participants in the ABIDE cohort (lower). Pearson correlation estimates (R) are shown in brain maps, in which only brain regions for which Benjamini–Hochberg corrected $P < 0.05$ are shaded (corrected across 82 regions for UCLA CNP and 48 regions for ABIDE). Note that the same color scale is used for both cohorts.

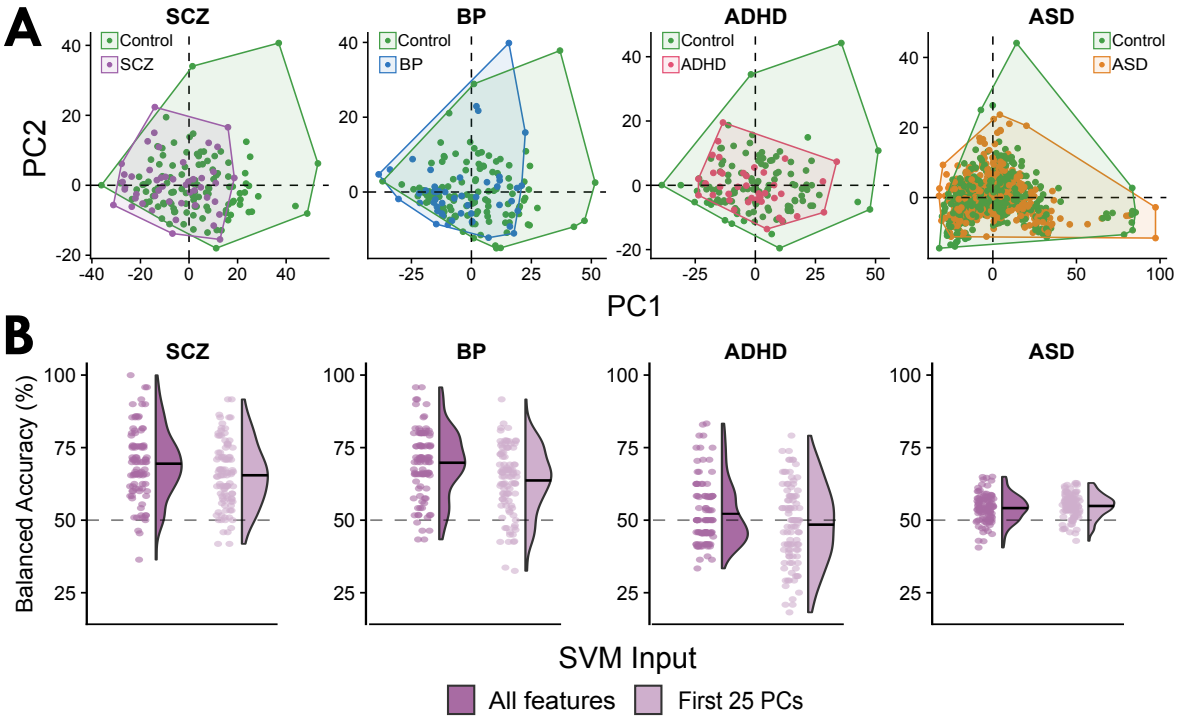


Figure S6. Linear dimensionality reduction and regularization approaches did not improve out-of-sample classification for the univariate region × feature classifiers, A_{uni_combo}. **A.** For each disorder, individual scores for the first two PCs are plotted, with points colored according to diagnosis. Shaded areas reflect convex hulls encapsulating all points for each diagnostic group. Note that each PCA was computed separately for each case–control comparison, so PC1 and PC2 scores are not directly comparable across clinical groups. **B.** For each case–control comparison, we compare the out-of-sample balanced accuracy across the 100 repeats × folds using all region × feature variables (left, dark purple) versus using only scores for the first 25 PCs (right, light purple). Points are randomly jittered along the horizontal axis in each raincloud plot to aid visualization. The horizontal line within each half-violin indicates the mean balanced accuracy for the corresponding distribution. **C.** For each case–control comparison, we compare the out-of-sample balanced accuracy across the 100 repeats×folds using default regularization (left, dark green) versus L1 ('LASSO' [192]) regularization (right, light green).

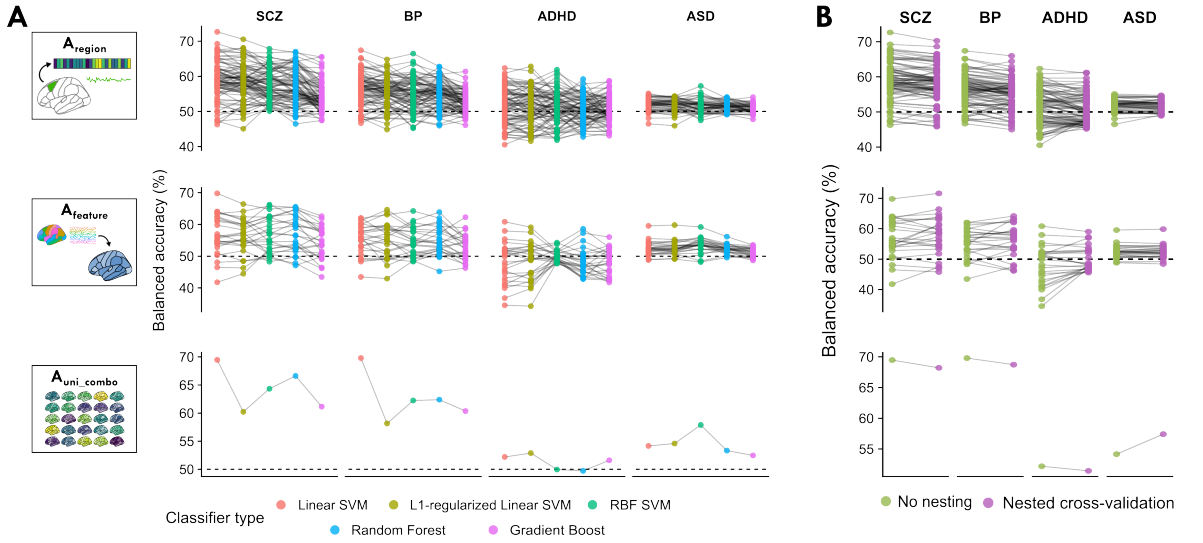


Figure S7. Comprehensive comparison of classifier types and hyperparameter optimization supports the use of linear SVM. A. For each of the three univariate representations—A_{region}, $A_{feature}$, and A_{uni_combo}—the mean cross-validated balanced accuracy is shown per disorder using each of five different classifier types. Each dot corresponds to one model input type (e.g., left pericalcarine cortex in A_{region}) and lines connect model input types across classifiers per disorder to guide visual interpretation. The dashed horizontal lines indicate 50% balanced accuracy in all plots. **B.** For the same univariate representations as in **A**, the mean cross-validated balanced accuracy is shown for the linear SVM classifier without hyperparameter optimization (i.e., explicitly setting $C = 1$ and applying inverse probability weighting; green) or with hyperparameter optimization for the C parameter and sample weighting type in purple. The dashed horizontal lines indicate 50% balanced accuracy in all plots.

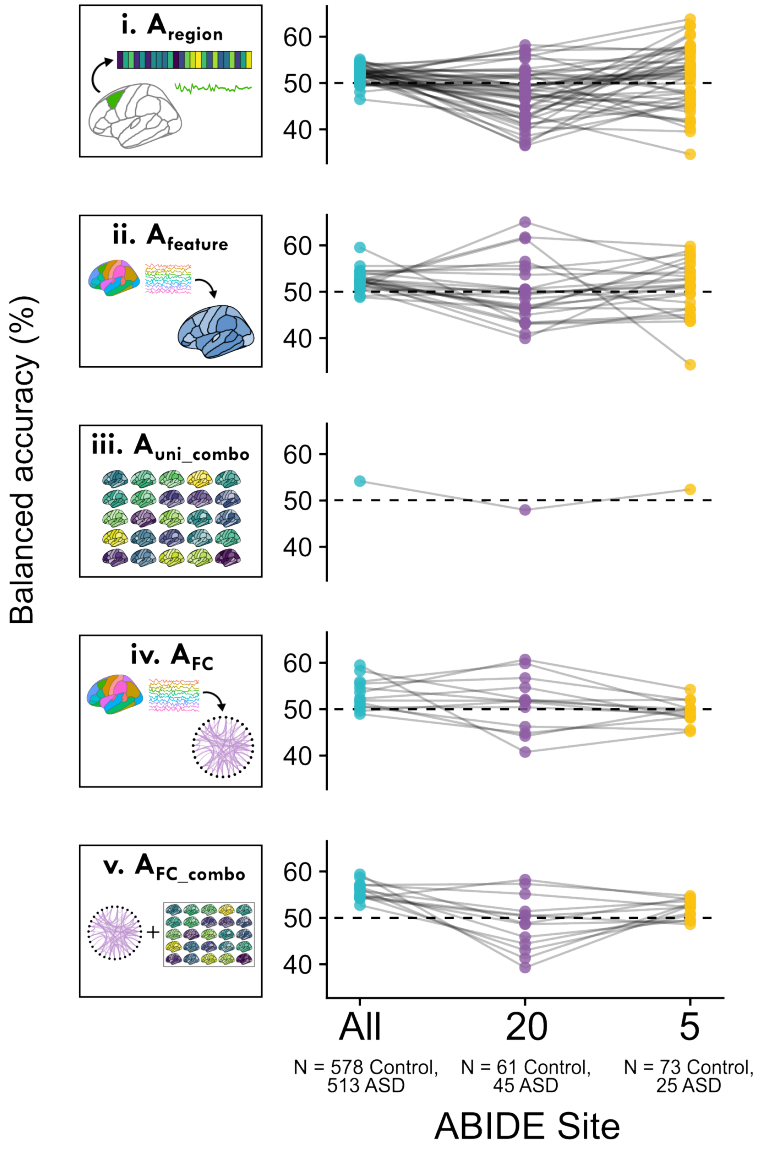


Figure S8. Classification performance is comparable within individual ABIDE consortium imaging sites. The mean cross-validated balanced accuracy is shown with the inclusion of participants from all ABIDE sites together (blue, $N = 1091$ participants) for each of **(i)** A_{region}, **(ii)** $A_{feature}$, **(iii)** A_{uni_combo}, **(iv)** A_{FC}, and **(v)** A_{FC_combo}. The mean cross-validated balanced accuracy is also shown when we restricted classification analyses to each of the two largest ABIDE imaging sites: Site #20 (purple, $N = 106$ participants) and Site #5 (yellow, $N = 98$ participants). Each dot corresponds to one individual model (e.g., the Superior Frontal Gyrus in A_{region}) and lines connect models across ABIDE site analyses to guide visual interpretation. The dashed horizontal line marks 50% balanced accuracy in all plots.

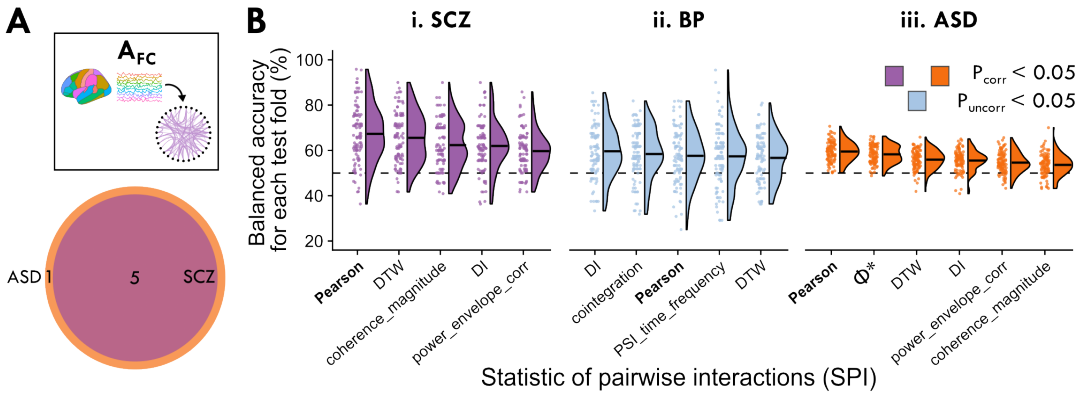


Figure S9. Representing brain activity as the set of all pairwise functional connectivity strengths, $\mathbf{A_{FC}}$, can significantly distinguish cases from controls, with the classical Pearson correlation coefficient (capturing linear contemporaneous coupling) a top performing metric. **A.** We compared 14 statistics of pairwise interactions (SPIs) (from *pyspi* [25], cf. Sec. 4.2.2), as different ways of quantifying functional connectivity (FC) between pairs of brain regions. For a given SPI, each participant was represented by the set of corresponding FC values (calculated for each pair of brain regions), yielding a set of region–region pair values that can be stored as a one-dimensional vector per participant. These vectors were concatenated to yield a participant × region–pair matrix that formed the basis for case–control classification using a linear SVM. **B.** Classification results are shown as a heatmap, with rows representing SPIs that yielded significant balanced accuracy ($p_{\text{corr}} < 0.05$, corrected across 14 SPIs) in at least one disorder, and columns representing each of the four disorders. Of the 14 SPIs we evaluated, eleven significantly distinguished cases from controls in at least one disorder, and are plotted here. The Pearson correlation coefficient is annotated in boldface for easier reference. **C.** The SPI similarity score, $|\rho_{\text{SPI}}|$, is visualized between each pair of the eleven SPIs from B as a heatmap, revealing six clusters of SPIs with similar behavior on the dataset (based on their outputs across all region–pairs and all disorders). As in B, the Pearson correlation coefficient annotation is shown boldface. **D.** The disorder similarity score, ρ_{disorder}, is depicted to compare the balanced accuracy values among all 14 SPIs between each pair of neuropsychiatric disorders; a large positive ρ_{disorder} indicates a strong positive Spearman correlation in case–control classification performance across the 14 SPIs in the given pair of disorders.

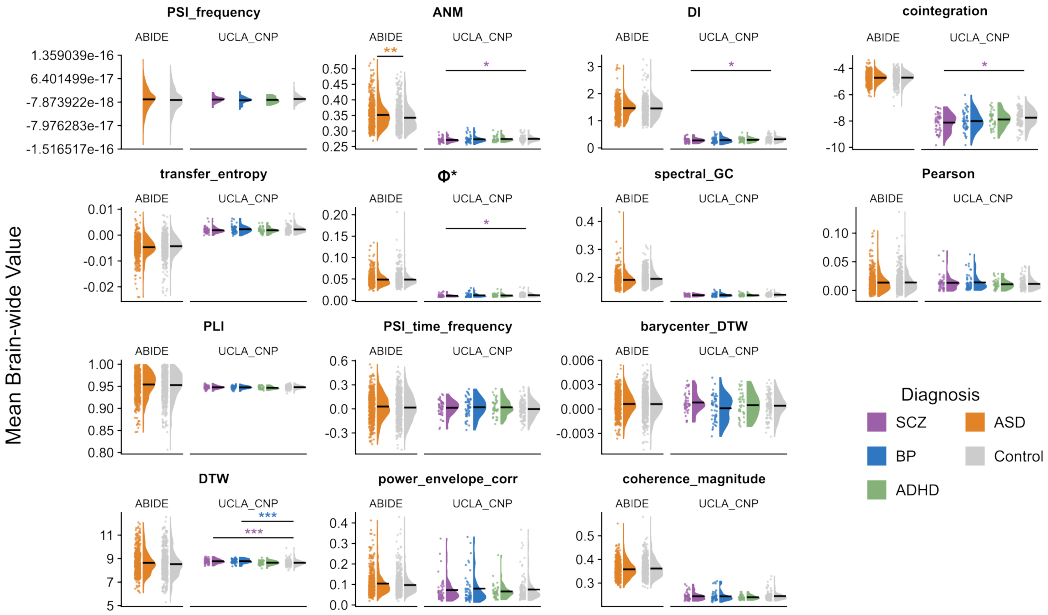


Figure S10. The mean FC value across all region–region pairs by SPI. For each participant in the UCLA CNP and ABIDE cohorts, we calculated the mean FC value across all region–region pairs per SPI and show the distributions across participants as raincloud plots. The solid horizontal line in each half-violin indicates the mean balanced accuracy and the dashed horizontal axis line denotes 50% balanced accuracy. Wilcoxon rank-sum test results are indicated with ***, $P_{\text{corr}} < 0.001$; **, $P_{\text{corr}} < 0.01$; *, $P_{\text{corr}} < 0.05$.

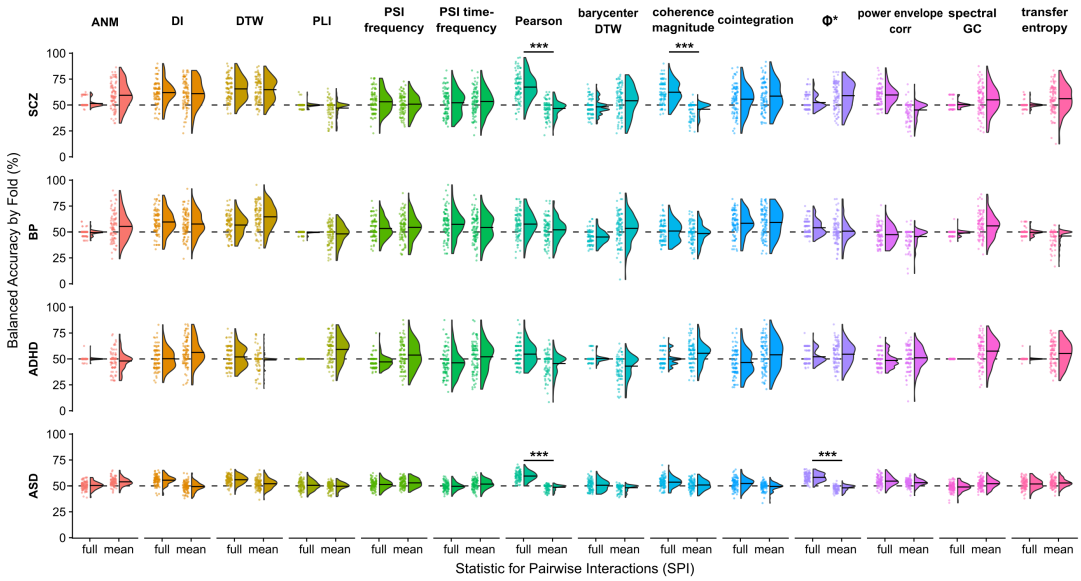


Figure S11. Comparing classification performance with the full region–region feature space or the brain-wide mean FC for each SPI. For each SPI, we compared case–control classification performance in each of the four disorders using either the full region–region pair input matrix (left) or the brain-wide average across all region–region pairs (right). Here, we show the distribution of all 100 test folds (10 repeats × 10 folds) as raincloud plots, where each dot represents one test fold. The solid horizontal line in each half-violin indicates the mean balanced accuracy and the dashed horizontal axis line denotes 50% balanced accuracy. Corrected resampled T-test results are indicated with ***, $P_{\text{corr}} < 0.001$.

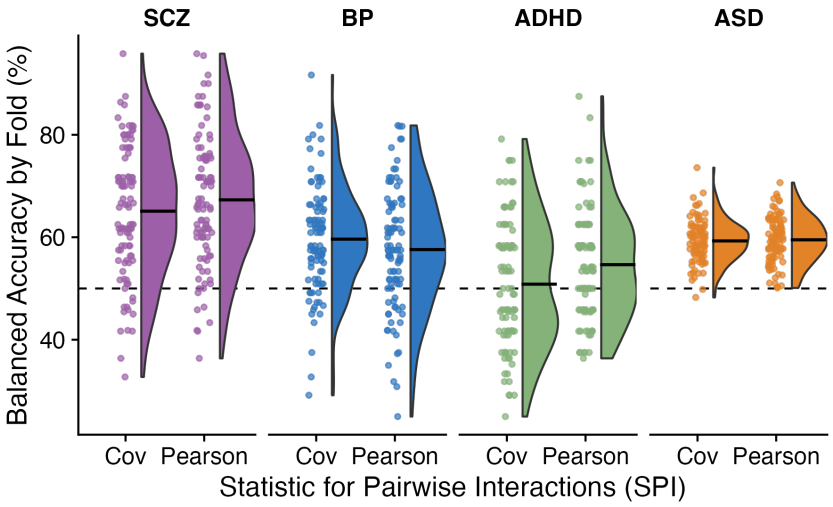


Figure S12. Raw covariance and normalized Pearson correlation coefficient show comparable case–control classification performance. For each disorder (SCZ, BP, ADHD, and ASD), we plot the distribution of test balanced accuracy values across 100 folds (10 repeats × 10 folds) with the same `scikit-learn` pipeline for the raw covariance matrix across all region–pairs ('Cov', left) versus the normalized Pearson correlation matrix ('Pearson', right). The solid horizontal line in each half-violin indicates the mean balanced accuracy and the dashed horizontal axis line denotes 50% balanced accuracy.

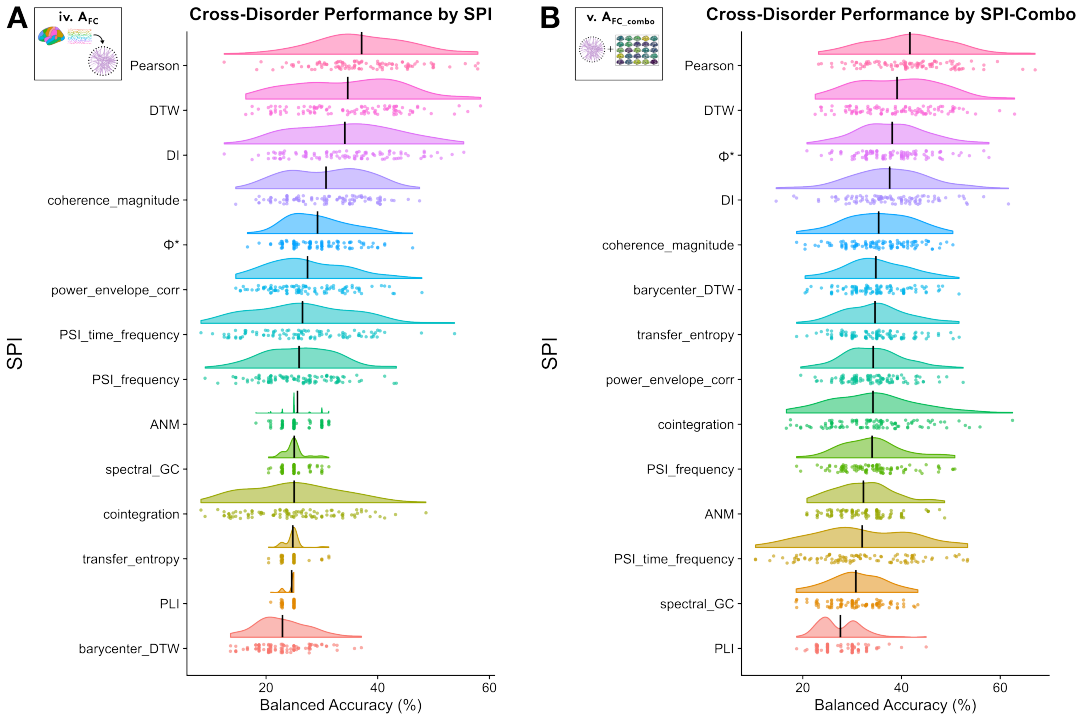


Figure S13. Cross-disorder classification analysis in the UCLA CNP cohort based on pairwise functional connectivity. The out-of-sample balanced accuracy distribution (averaged across SCZ, BP, ADHD, and Control participants in the UCLA CNP cohort) is shown for each SPI on its own (**A**, A_{FC}) or with the inclusion of whole-brain local dynamics (**B**, A_{FC_combo}). Each dot corresponds to one test fold (for a total of 100 data points, across 10 repeats × 10 folds), and the vertical black line in each half-violin corresponds to the mean balanced accuracy across test folds.

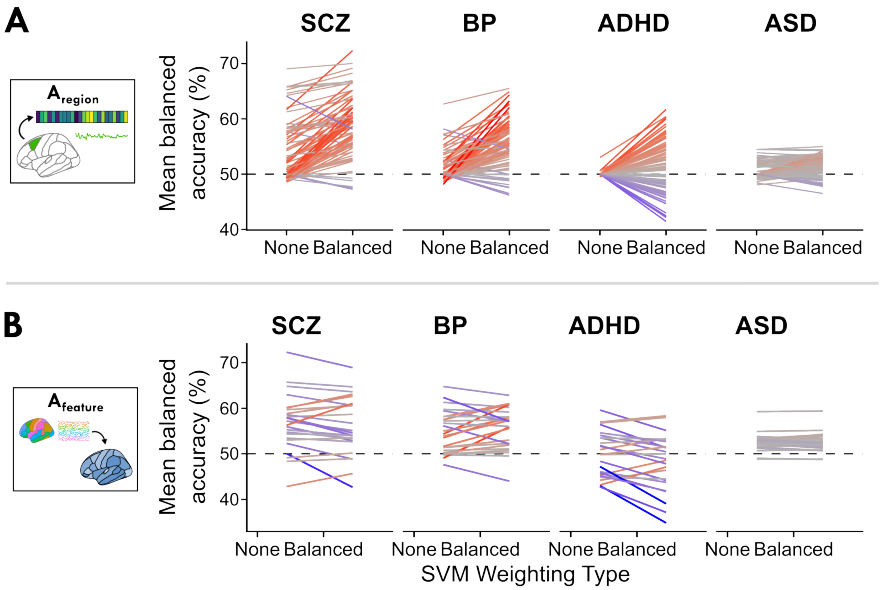


Figure S14. Comparing classification performance with versus without inverse probability weighting.
A. For each disorder, the mean balanced accuracy per brain region is shown with no weighting ('None') or inverse probability weighting ('Balanced'). Lines correspond to each of 82 brain regions for SCZ, BP, and ADHD and each of 48 brain regions for ASD. Colors are included as a visual aid to highlight the difference in performance between the two weighting types, with red corresponding to higher balanced accuracy with inverse probability weighting and blue corresponding to lower balanced accuracy with inverse probability weighting. **B.** For each disorder, the mean balanced accuracy per univariate time-series feature is shown with no weighting ('None') or inverse probability weighting ('Balanced'), as in A.

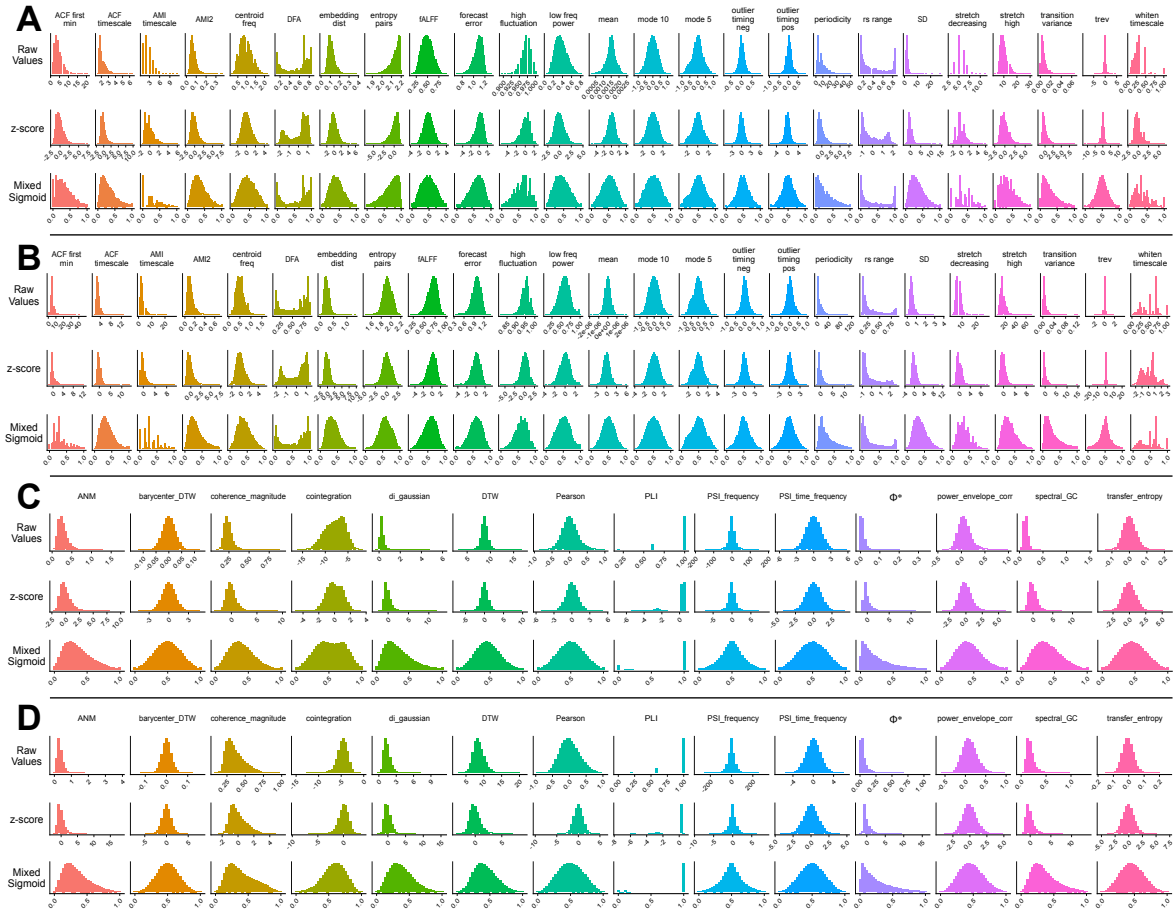


Figure S15. Comparing normalization methods supports the use of the outlier-robust mixed sigmoid method. Univariate time-series feature values were concatenated from all brain regions, with the resulting distributions depicted for all participants in the UCLA CNP cohort (**A**) or ABIDE cohort (**B**) with no normalization (upper row), z-score normalization (middle row), and outlier-robust mixed sigmoid normalization (bottom row; see Methods Sec. 4.3.1 for description). Pairwise SPI feature values were concatenated from all region–region pairs, with the resulting distributions depicted for all participants in the UCLA CNP cohort (**C**) or ABIDE cohort (**D**) with no normalization (upper row), z-score normalization (middle row), and outlier-robust mixed sigmoid normalization (bottom row).

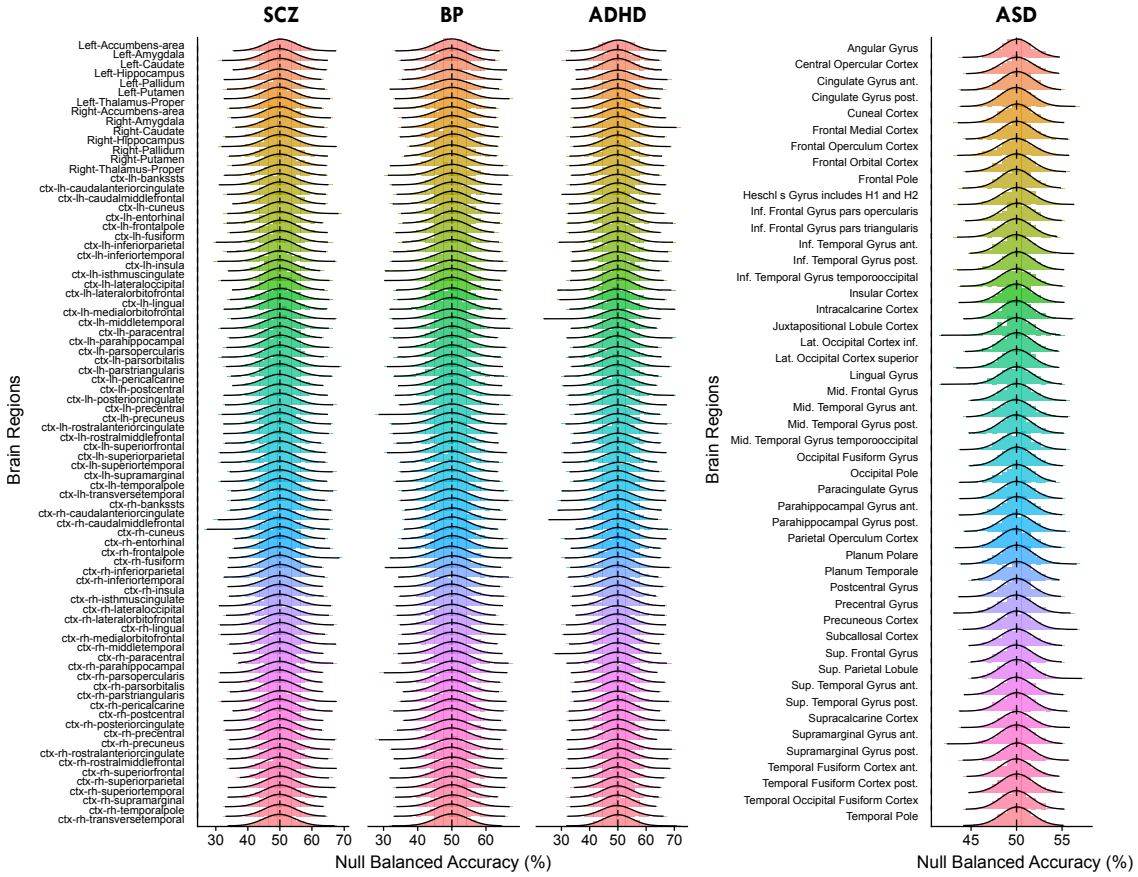


Figure S16. Null balanced accuracy metrics are approximately normally distributed across brain regions in all clinical cohorts. The distribution of 1000 null balanced accuracy values (cross-validated over 10 repeats of 10 folds each) is shown for each brain region by disorder. Empirical null values are visualized as histograms, with the probability density overlaid as black lines for the normal distribution based on the mean and standard deviation for each null distribution. These normal distributions were used to compute p-values for each brain region based on the corresponding probability density function; the same procedure was also applied for univariate and pairwise time-series features. ASD brain regions are plotted separately from the SCZ, BP, and ADHD brain regions as different parcellation atlases were analysed for UCLA CNP (Desikan-Killiany atlas) and ABIDE cohorts (Harvard-Oxford cortical altas). The dashed vertical line marks 50% balanced accuracy in all plots.

Appendix 5: Supplementary material for Chapter 6

Supporting Information for 'Mapping functional, structural, and transcriptomic correlates of homotopic connectivity'

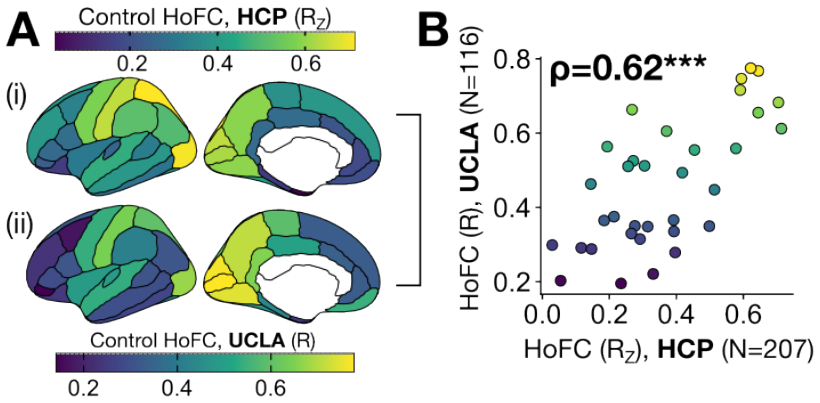


Figure S1. Neurotypical control cohorts show general concordance in homotopic functional connectivity (HoFC) spatial variation. A. The regional homotopic functional connectivity (HoFC) values are projected onto the cortical surface for the group averages representing neurotypical control participants from the (i) HCP cohort (N=207) and (ii) UCLA cohort (N=116). Note that the correlation coefficients were z-transformed only for the HCP cohort as part of the ENIGMA preprocessing pipeline [48]. **B**. The HoFC values are plotted for each region in the HCP cohort (x-axis) and UCLA cohort (y-axis), with the Spearman rank correlation (ρ) annotated. ***, $P < 0.001$, single-hemisphere spin test with 10,000 permutations [56].

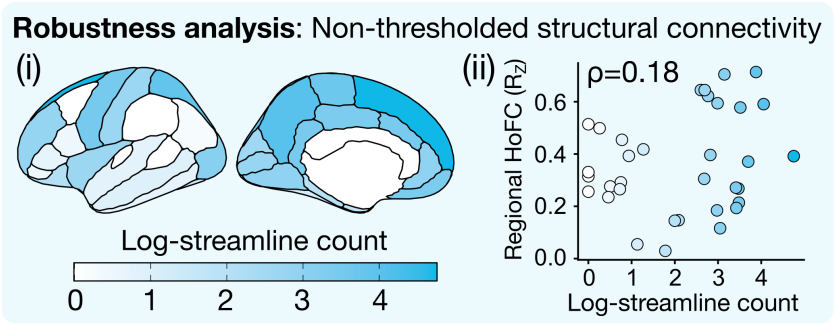


Figure S2. The spatial variation in cortical HoFC is not associated with non-thresholded structural connectivity in a similar HCP cohort. (i) As a robustness analysis, we additionally examined a similar subset of 100 unrelated individuals from the HCP S1200 release as in previous work [82] with a non-thresholded group consensus connectome, derived by computing the mean streamline count per region–region pair across all individuals. Streamline counts are $\log_{10}$-transformed; any edges with zero streamlines, or with negative counts after log-transformation, were then set to zero. (ii) The mean $\log_{10}$-streamline count connecting each homotopic region pair (x-axis) is plotted against the HoFC for each corresponding region (y-axis), with the Spearman rank correlation (ρ) annotated.

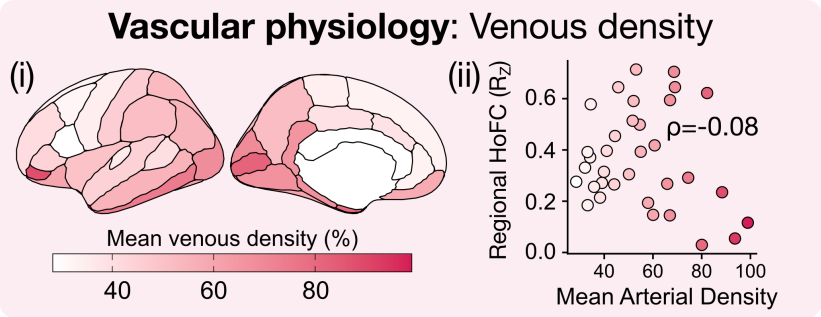


Figure S3. Spatial variation in homotopic functional connectivity (HoFC) is not associated with venous density. (i) The mean venous density in each region is plotted on the cortical surface, averaged between the two hemispheres per region. (ii) The mean hemisphere-averaged arterial density values from the same regions shown in (i) are plotted on the x-axis against the HoFC in each corresponding region on the y-axis. The Spearman rank correlation (ρ) is annotated.

Appendix 6: Supplementary material for Chapter 7

Supplementary material for 'Benchmarking overlapping community detection methods for applications in human connectomics'

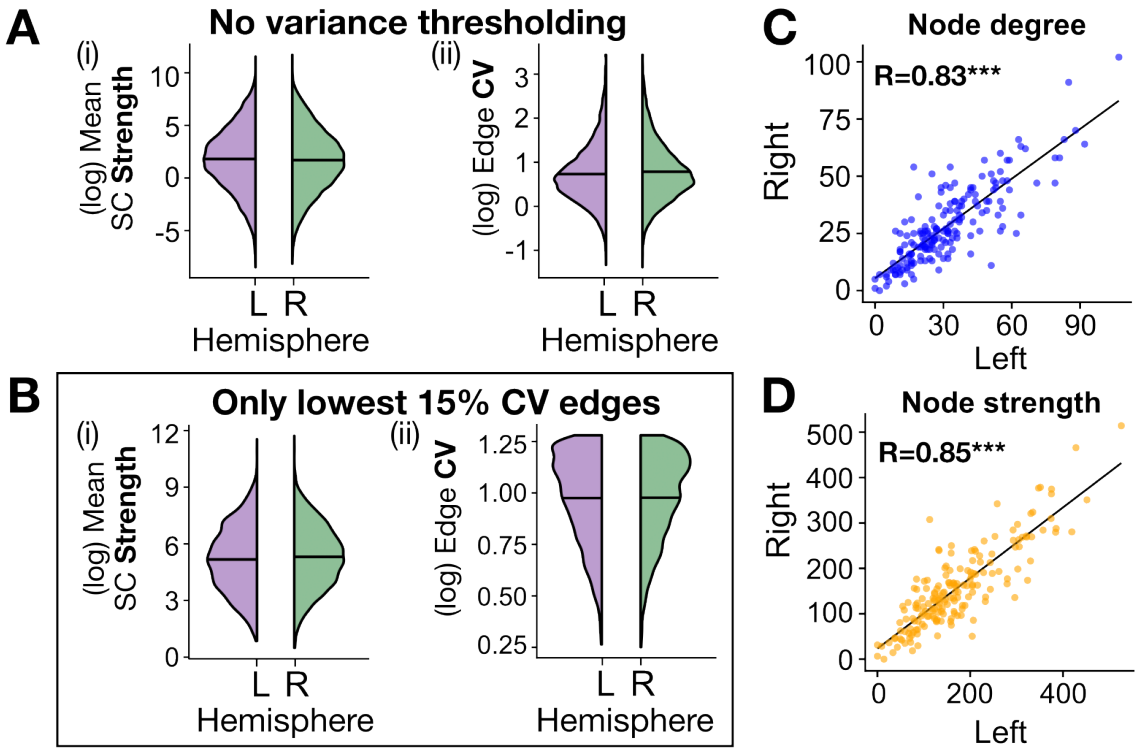


Figure S1: The left and right cerebral hemispheres exhibit highly similar intra-hemispheric structural connectivity profiles. A. For the left (purple) and right (green) hemispheres, the distribution of (i) log-transformed mean edge strengths, averaged across the group of N=973 participants, and (ii) the edge-wise CV values without any thresholding. **B.** As in **A**, after filtering to edges ranking in the lowest 15% of CV values (as in the main text).

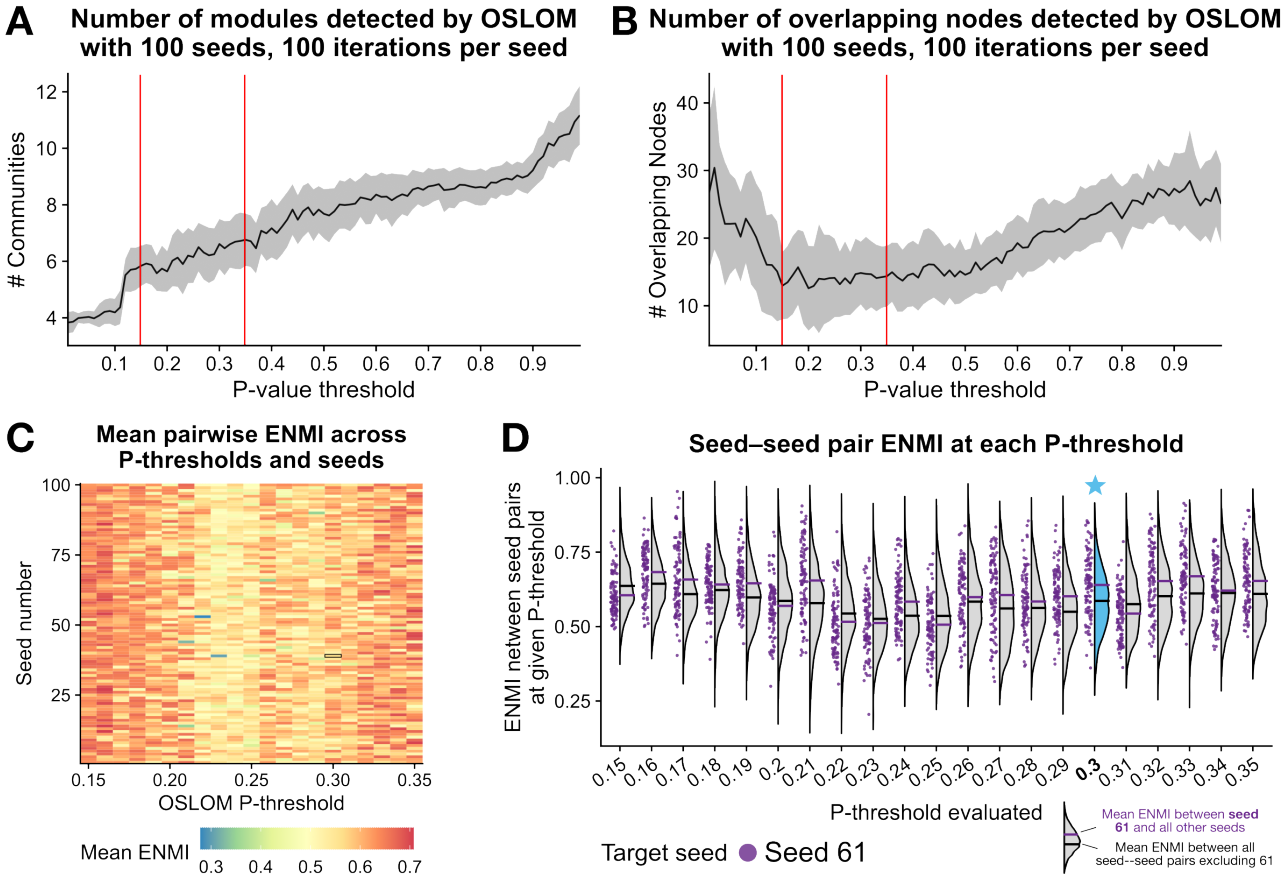


Figure S2: Parameter sweeps support the selection of $P = 0.3$ and seed number 61 for OSLOM. A. For each P-value threshold from 0.01 to 1, the mean number of communities that OSLOM identified from the empirical human cortical connectome is shown as a line plot. The shaded gray ribbon indices $\pm$ 1 SD and the red lines highlight the $0.15 \leq P \leq 0.35$ range. **B.** For each P-value threshold from 0.01 to 1, the mean number of nodes that OSLOM identified as overlapping across two or more communities is shown as a line plot. The shaded gray ribbon indices $\pm$ 1 SD and the red lines highlight the $0.15 \leq P \leq 0.35$ range. **C.** For the $0.15 \leq P \leq 0.35$ range, the mean pairwise ENMI between each seed and all 99 other evaluated seeds is plotted as a heatmap. The color range indicates that the mean ENMI values range between $[0.27, 0.71]$, with an average of 0.58 ± 0.06 (arbitrary units). **D.** For the final seed selection (61), the pairwise ENMI between seed 61 and each of the 99 other seeds is depicted as a raincloud plot. Specifically, for each P-threshold from $0.15 \leq P \leq 0.35$, there are 99 purple points corresponding to the ENMI between seed 61 and each of the 99 other seeds. The gray violins correspond to the ENMI distributions across all 4950 pairs of seeds from 1 to 100 (excluding self–pairs). Within each violin, the black bar measures the mean ENMI between all seed–seed pairs excluding seed 61, while the purple bar indicates the mean ENMI between seed 61 and all other seeds. The violin at $P = 0.3$ is highlighted in blue with a star as this is the final P-threshold choice in combination with seed 61, which yielded a mean ENMI of 0.64 ± 0.11, compared to the average across all other seeds of 0.59 ± 0.10.

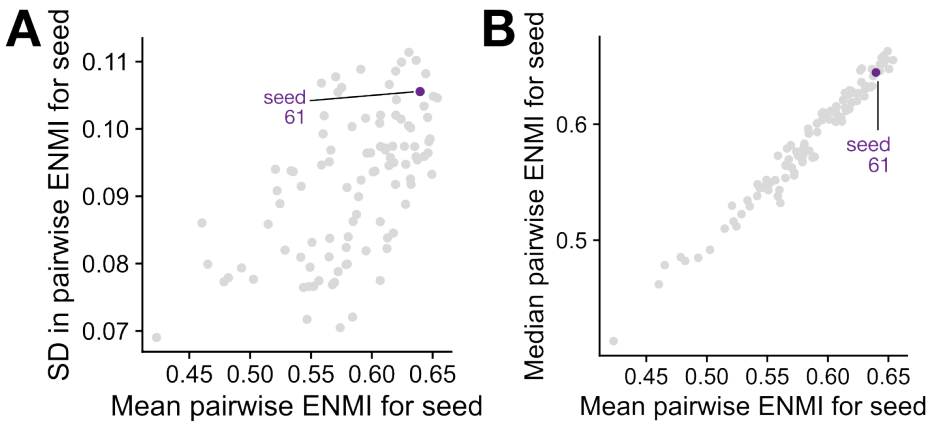


Figure S3: ENMI exhibits a heteroskedastic relationship for OSLOM applied to the right-hemisphere structural connectome, but the mean and median are very closely related. A. The mean (x-axis) and standard deviation (y-axis) of ENMI values for each of 100 initialization seeds supplied to OSLOM with the empirical right-hemisphere cortical connectome. **B.** The mean (x-axis) and median (y-axis) of ENMI values, as in **A**.

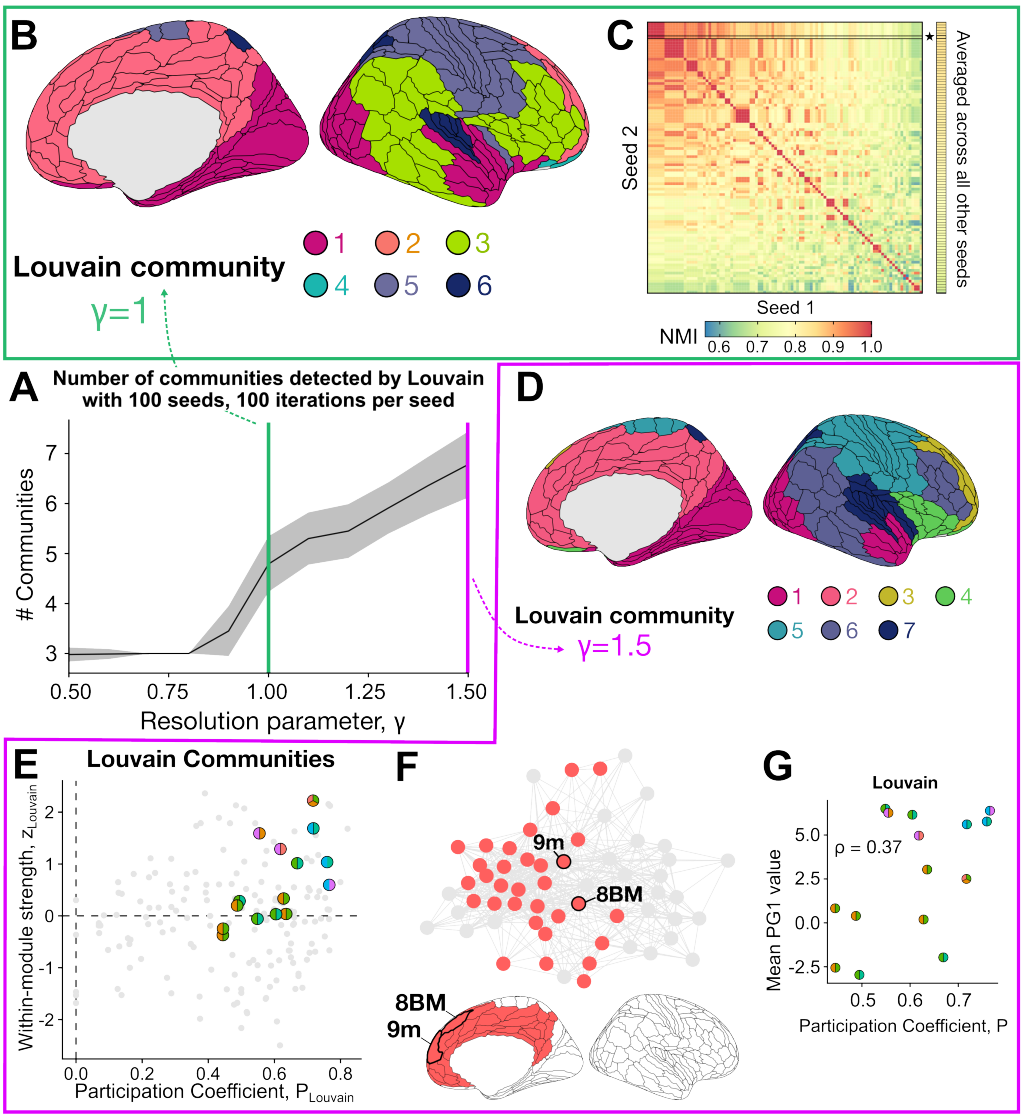


Figure S4: Louvain partitioning with $\gamma = 1$ yields a representative six-community decomposition, with comparable properties to that of a seven-community decomposition with $\gamma = 1.5$. A. For the resolution parameters $\gamma = [0.5, 0.6, \ldots 1.5]$, Louvain partitioning was performed using 100 initialization seeds each. The mean number of detected communities per γ value is plotted as a black line, with the shaded ribbon indicating ± 1 standard deviation. The green line at $\gamma = 1$ is included to guide visual interpretation for this threshold selection parameter, with a representative six-community decomposition projected onto the cortical surface in **B**. **C.** The result in **B** corresponds to the seed number 98, which was selected based on its maximal normalized mutual information (NMI) value to all other evaluated seeds, as shown in the heatmap. The pairwise NMI values are depicted as a heatmap, with the selected seed (98) outlined in bold with a star. **D.** To compare our seven-community OSLOM decomposition with a seven-community Louvain decomposition specifically, we also evaluated a maximally representative seed (seed=91) with the threshold $\gamma = 1.5$. These seven communities are projected onto the right cortical surface here. **E.** Scatter plots for P_{Louvain} versus z_{Louvain} for every node in the right hemisphere cortical connectome, using $\gamma = 1.5$. The overlapping nodes (obtained by `OSLOM-30`) are marked in two-tone circles, with the two colors indicating the pair of communities bridged by the node. **F.** Louvain (with $\gamma = 1.5$) assigns nodes *8BM* and *9m* to a community of 32 total nodes spanning frontal, cingulate, and retrosplenial/dorsomedial cortices (n.b., only a subset of these nodes are shown that are structurally connected to *8BM* and/or *9m*). Gray circles in the network plot represent nodes that are structurally connected to *8BM* and *9m* but not assigned to the same Louvain module. **G.** For each overlapping node identified by `OSLOM-30`, its mean PG1 value is compared with its P_{Louvain}. Spearman's $\rho = 0.37, P = 0.17$.

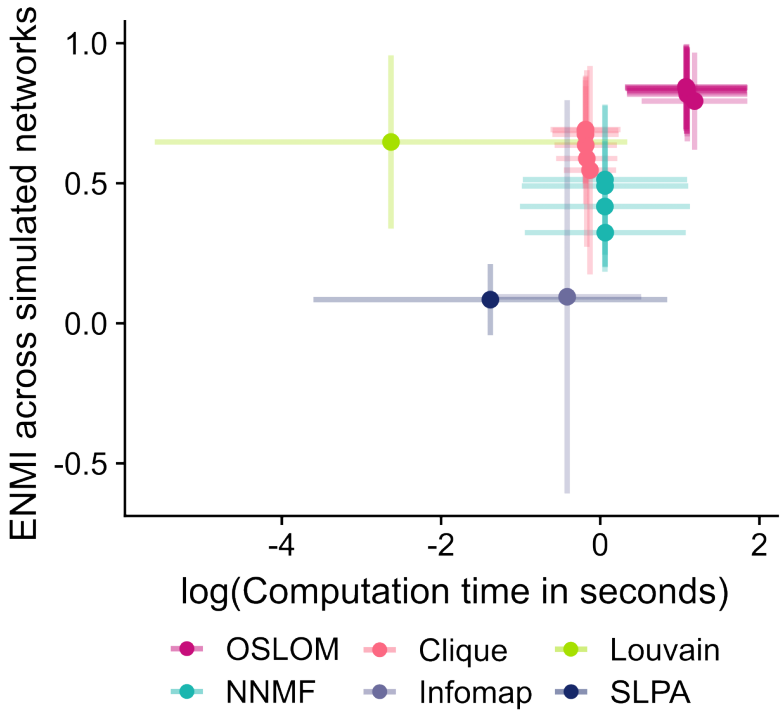


Figure S5: OSLOM with different thresholds consistently require the longest time to run, but yield the highest overall ENMI in the benchmark network ensemble. The mean computation time (in log(10)-transformed number of seconds) is plotted on the x-axis against the mean ENMI on the y-axis, each averaged across the 1,000 simulated networks in the benchmark ensemble. The cross-bars indicate ± 1 standard deviation for computation time (x-axis) or ENMI (y-axis).

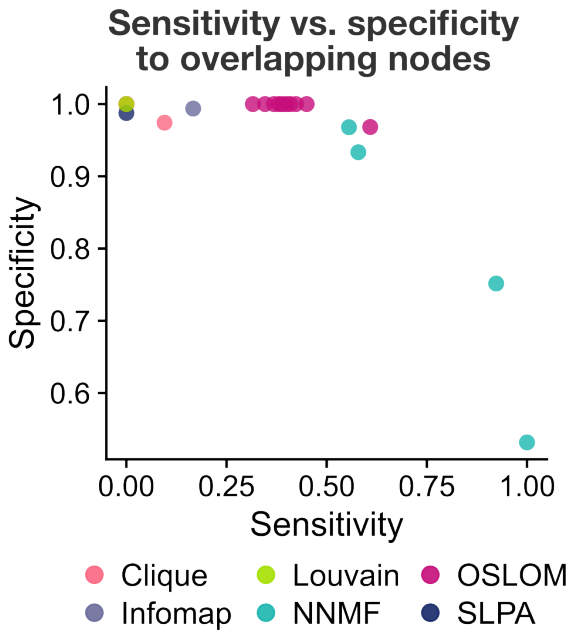


Figure S6: OSLOM yields a good balance between sensitivity and specificity in identifying ground-truth overlapping nodes across the benchmark ensemble. The mean sensitivity is plotted on the x-axis against the specificity on the y-axis, each averaged across the 1,000 simulated networks in the benchmark ensemble.

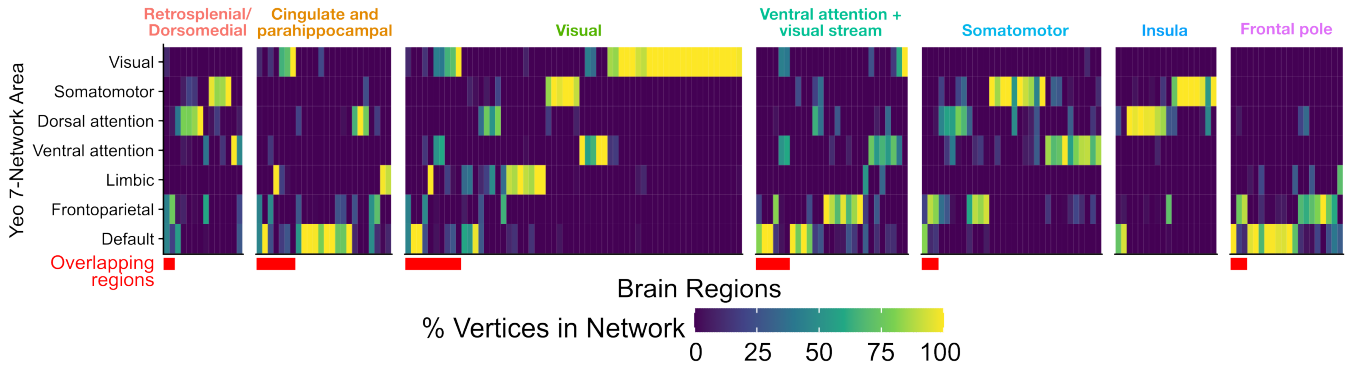


Figure S7: For each brain region, the proportion of vertices mapping to each of the 7-Network areas from Yeo et al. [8] was computed, with the percentages indicated as heatmap tile values. Brain regions are grouped along the x-axis by OSLOM-identified structural module, with the 'overlapping' regions to the left of each cluster (annotated by the red bar below the heatmap). Note that overlapping regions were included as columns in each of the corresponding OSLOM modules, such that each overlapping region appears two or three times in the heatmap.

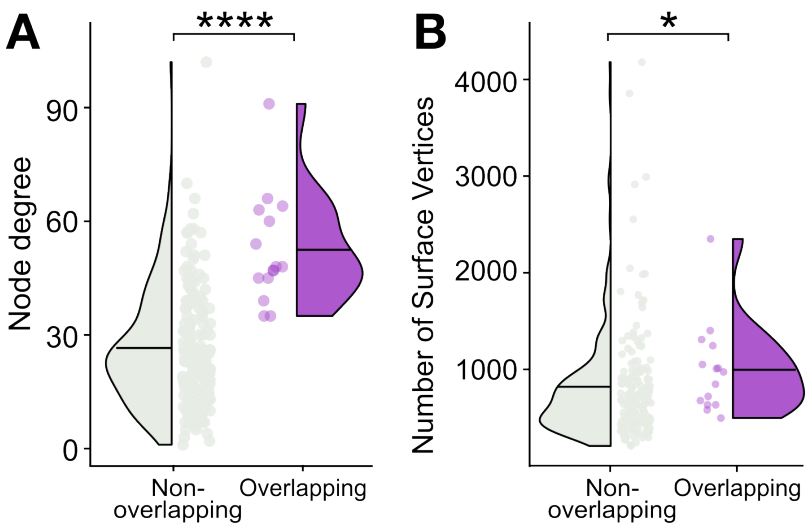


Figure S8: Overlapping regions identified by OSLOM_30 have a greater degree and are larger than non-overlapping regions on average. A. The distributions of nodal degree are shown for overlapping regions (purple, 15 regions) and non-overlapping regions (grey, 164 regions) as raincloud plots. ****, P = 3×10^{-7}, unpaired Wilcox rank-sum test between overlapping vs. non-overlapping degree (i.e., sum of log-transformed streamline counts per region). **B.** The distributions of cortical surface vertices contained within each overlapping region (purple, 15 regions) or non-overlapping region (grey, 164 regions) are shown as raincloud plots. *, P = 0.03, unpaired Wilcox rank-sum test between overlapping vs. non-overlapping vertex counts.

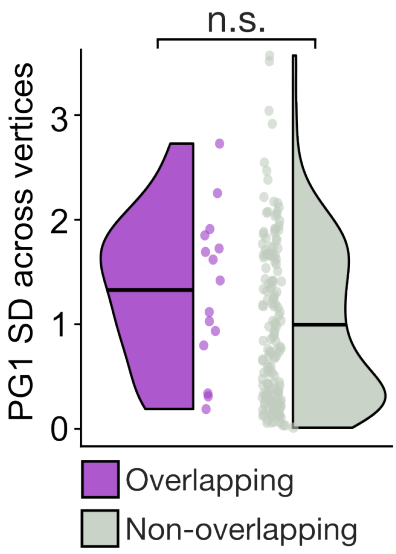


Figure S9: There is no difference in the vertex-wise SD of first principal gradient (PG1) values between overlapping vs. non-overlapping regions. The distributions of PG1 SD (across vertices) within each region are shown as raincloud plots for overlapping regions (purple, 15 regions) or non-overlapping regions (grey, 164 regions). n.s., $P > 0.05$, unpaired Wilcox rank-sum test between overlapping vs. non-overlapping PG1 SD values.

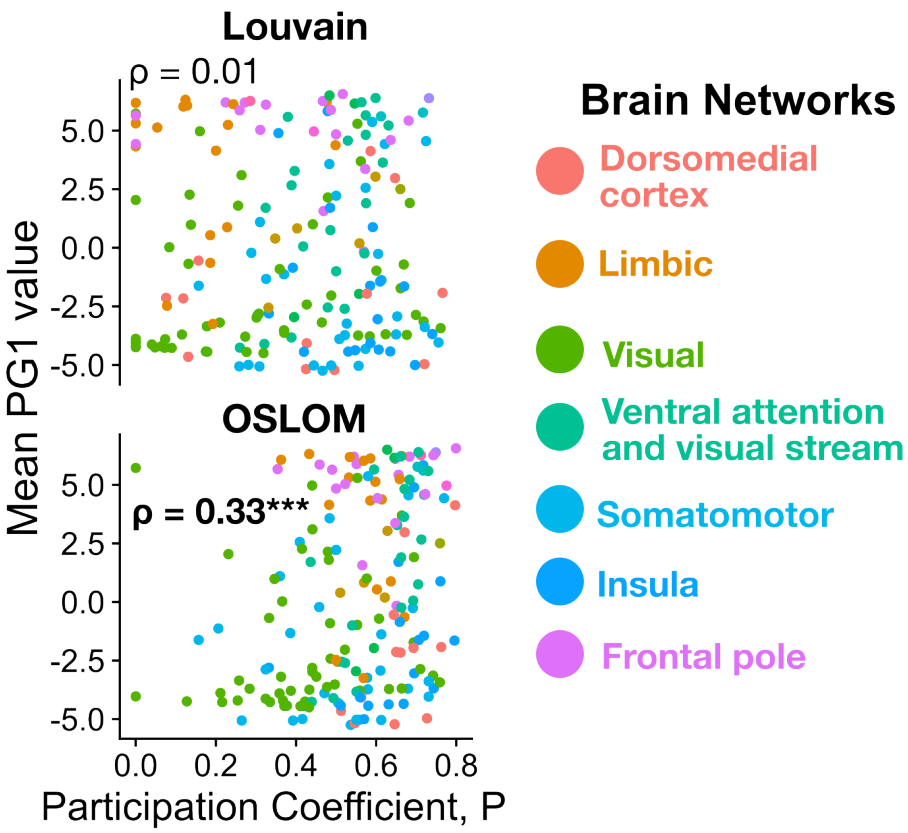


Figure S10: The participation coefficient from the OSLOM decomposition, but not the Louvain decomposition, is significantly correlated with the first principal gradient of functional connectivity (PG1) across the right hemisphere cortex. For each node in the 180-region HCP–MMP1 atlas, its mean PG1 value is compared with its participation coefficient from either Louvain (P_{Louvain}) or OSLOM (P_{OSLOM}) partitioning. Spearman's ρ is shown for each comparison; ***, $P < 0.001$.