DATA SCIENCE APPROACHES ON BRAIN CONNECTIVITY: COMMUNICATION DYNAMICS AND FINGERPRINT GRADIENTS

by

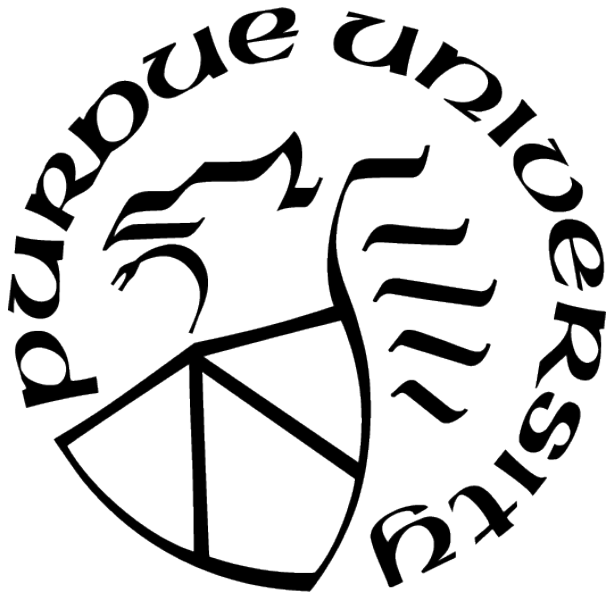
Uttara Tipnis

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



School of Industrial Engineering

West Lafayette, Indiana

May 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Joaquín Goñi, Chair

School of Industrial Engineering, Purdue University

Dr. Alan D. Kaplan

Computational Engineering Division, Lawrence Livermore National Laboratory

Dr. Mario Ventresca

School of Industrial Engineering, Purdue University

Dr. Thomas Talavage

School of Biomedical Engineering, University of Cincinnati

Dr. Jaroslaw Harezlak

Department of Epidemiology and Biostatistics, School of Public Health-Bloomington, Indiana University

Approved by:

Dr. Abhijit Deshmukh

Head of Department, School of Industrial Engineering, Purdue University

I would much rather fail gloriously, than not venture, not try.

-Anthony Bourdain

To Mogambo Vantablack, my beautiful little dumbass with very few brain cells to spare, for keeping me (relatively) sane and constantly entertained through the unending dumpster fire that was 2020

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, mentor, and friend Dr. Joaquín Goñi. Joaquín has always been incredibly supportive throughout this entire journey and has made it very enjoyable indeed. He has set a high standard for what an ideal mentor-mentee relationship should look like and I hope to one day be able to provide the same to early-career researchers I might be mentoring in the future.

I would also like to express my sincere thanks to Lawrence Livermore National Lab for the incredible opportunity to intern there for two summers and to Dr. Alan Kaplan for mentoring me there. I have learned much from Alan about new and exciting research and I look forward to keep growing under his tutelage in the future. The internships also gave me the opportunity to drive across the country through some of the most dramatic landscapes and fueled my love for travelling and photography, for which I will forever be grateful. Also a special shoutout to the people at West gate badge office at LLNL for letting Joaquín go inside the Lab when he visited without any documents, since that visit kickstarted the collaboration.

I also would like to express my sincere gratitude to my committee members – Dr. Mario Ventresca, Dr. Jaroslaw Harezlak, and Dr. Thomas Talavage – for their support and guidance throughout my PhD.

I would like to thank my colleagues and friends from CONN*plexity* Lab: Dr. Enrico Amico, Dr. Kausar Abbas, Dr. Diana O. Svaldi, Federico Ramírez Toraño, Meenusree Rajapandian, Duy Anh Duong-Tran, Mintao Liu, Benjamin Chiêm, and Ángeles Tepper for all the interesting discussions, research related or otherwise. I would especially like to acknowledge Enrico and Kausar for the daily chats over coffee and Federico for his support during the early months of the pandemic.

Next, I would like to acknowledge my friends who became my family away from home and made my time in Indiana a little more enjoyable, in no particular order: Vasundhara Kaul, Dr. Alessandra Costantini, Dima Kondratyev, Dr. Maru Cabreba, Dr. Edgar Rojas-Muñoz, Dr. Mario Tindaro Migliorino, Aparna Pidaparthi, Pallavi Mishra, Dr. Abhishek Ray, Arpan Chakraborty, Dr. Audrey Reinert, Dr. Liz Wachs, and Oscar Rincón-Guevara.

I would like to thank my parents, Sunita and Vinay Tipnis, for making it possible for me to pursue higher education across the globe without letting regressive and patriarchal societal expectations bog me down. I would also like to express my sincere gratitude to my uncle and aunt, Sanjay and Amita Karkhanis, for always supporting me in whatever decisions I have made and my late grandmother, Suman Karkhanis, for being incredibly progressive for her times and leading the latter generations to do better by example.

Last but not least, I would like to thank my partner Greg Henninger for his support and for bringing Mogambo (my cat) into my life when I needed him the most during the pandemic. And also to Mogambo for putting up with me on a daily basis without trying to kill me in my sleep.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

α	Pheromone perception
β	Edge perception
τ_{ij}	Amount of existing pheromone on edge ij
η_{ij}	Fibre density of edge ij
ρ	Pheromone evaporation rate
L_p	Length of path p based on fibre density
$Traffic_p$	Number of times the ants took path p to reach the target
EPL_{ij}	Effective path length between nodes i and j
SPL_{ij}	Number of edges in the shortest path length between nodes i and j
$\mathrm{iter}_{arrival}$	Iteration number when at least 95% of the ants reach the target
num_{Ants}	Number of ants used in the simulation
$Arrivals_{\mathrm{ij}}$	Number of times ants have successfully reached target j from source i by using any path in the path ensemble
AR_{ij}	Arrival rate between nodes i and j
SC^{rand}_{whole}	Whole-brain randomized network
SC^{rand}_{intra}	Intra-hemispheric randomized network
R^2	Explained variance
I_f	Identifiability framework
f_{min}	Minimum frequency for bandpass filter
f_{max}	Maximum frequency for bandpass filter
I_{diff}	Differential identifiability
I_{self}	Self similarity
I_{others}	Similarity between different subjects
I^{twin}_{diff}	Twin differential identifiability
I_{twin}	Similarity between twin subjects
$\delta(I_{diff})$	Difference between original and optimal differential identifiability

ABBREVIATIONS

BOLD signal	Blood oxygen level-dependent signal
MRI	Magnetic resonance imaging
fMRI	Functional magnetic resonance imaging
SC	Structural connectivity
FC	Functional connectivity
EPL	Effective path length
AR	Arrival rate
SI	Search information
TSP	Travelling salesman problem
HCP	Human Connectome Project
DWI	Diffusion weighted imaging
ACT	Anatomically constrained tractography
FA	Fractional anisotropy
rfMRI	Resting-state functional magnetic resonance imaging
tfMRI	Task-based functional magnetic resonance imaging
WM	Working memory
LR	Left-to-right
RL	Right-to-left
TP	Transition probability
MFPT	Mean first passage time
SPL	Shortest path length
MF	Max flow
DMN	Default mode network
TMS	Transcranial magnetic stimulation
HCP-YA	Human Connectome Project – Young Adult
MZ	Monozygotic
DZ	Dizygotic
ADNI	Alzheimer's Disease Neuroimaging Initiative

MEG	Magnetoencephalography
MNI	Montreal Neurological Institute
NIfTI	Neuroimaging Informatics Technology Initiative
GIfTI	Geometry format under the Neuroimaging Informatics Technology Initiative
CIfTI	Connectivity Informatics Technology Initiative
ICA-FIX	Independent component analysis based X-noiseifier
CSF	Cerebro-spinal fluid
GSR	Global signal regression
TR	Repetition time
ROI	Region of interest
PCA	Principal component analysis
RSN	Resting state network
ABCD	Adolescent Brain Cognitive Development

ABSTRACT

The innovations in Magnetic Resonance Imaging (MRI) in the recent decades have given rise to large open-source datasets. MRI affords researchers the ability to look at both structure and function of the human brain. This dissertation will make use of one of these large open-source datasets, the Human Connectome Project (HCP), to study the structural and functional connectivity in the brain.

Communication processes within the human brain at different cognitive states are neither well understood nor completely characterized. We assess communication processes in the human connectome using ant colony-inspired cooperative learning algorithm, starting from a source with no *a priori* information about the network topology, and cooperatively searching for the target through a pheromone-inspired model. This framework relies on two parameters, namely *pheromone* and *edge perception*, to define the cognizance and subsequent behaviour of the ants on the network and the communication processes happening between source and target. Simulations with different configurations allow the identification of path-ensembles that are involved in the communication between node pairs. In order to assess the different communication regimes displayed on the simulations and their associations with functional connectivity, we introduce two network measurements, effective path-length and arrival rate. These measurements are tested as individual and combined descriptors of functional connectivity during different tasks. Finally, different communication regimes are found in different specialized functional networks. This framework may be used as a test-bed for different communication regimes on top of an underlying topology.

The assessment of brain *fingerprints* has emerged in the recent years as an important tool to study individual differences. Studies so far have mainly focused on connectivity fingerprints between different brain scans of the same individual. We extend the concept of brain connectivity fingerprints beyond test/retest and assess *fingerprint gradients* in young adults by developing an extension of the differential identifiability framework. To do so, we look at the similarity between not only the multiple scans of an individual (*subject fingerprint*), but also between the scans of monozygotic and dizygotic twins (*twin fingerprint*). We have carried out this analysis on the 8 fMRI conditions present in the Human Connectome Project –

Young Adult dataset, which we processed into functional connectomes (FCs) and time series parcellated according to the Schaefer Atlas scheme, which has multiple levels of resolution. Our differential identifiability results show that the fingerprint gradients based on genetic and environmental similarities are indeed present when comparing FCs for all parcellations and fMRI conditions. Importantly, only when assessing optimally reconstructed FCs, we fully uncover fingerprints present in higher resolution atlases. We also study the effect of scanning length on subject fingerprint of resting-state FCs to analyze the effect of scanning length and parcellation. In the pursuit of open science, we have also made available the processed and parcellated FCs and time series for all conditions for $\sim$1200 subjects part of the HCP-YA dataset to the scientific community.

Lastly, we have estimated the effect of genetics and environment on the original and optimally reconstructed FC with an ACE model.

1. INTRODUCTION

Magnetic resonance imaging (MRI) is an imaging technique that is used in clinical and research settings to capture images of the anatomy and the physiological processes of the body. This imaging modality is based on the magnetization properties of atomic nuclei. MRI scanners use a powerful and uniform external magnetic field to align the protons that are normally randomly oriented within the water nuclei of the tissue being examined. This alignment of protons is then perturbed by introducing an external radio frequency (RF) energy. The nuclei then return to their resting position through different relaxation processes and emit RF signals themselves while doing so. After a certain time following the initial RF, the emitted signals are measured. MRI is widely used in clinical settings for medical diagnosis and staging and follow-up of disease without exposing the body to radiation. It also has several applications in the research settings to study the properties of different tissues. In this dissertation, we are going to focus on the applications of MRI to study the human brain. Specifically, we are going to focus on the applications of two different modalities – diffusion weighted imaging and functional magnetic resonance imaging – in order to study the structure and function of the human brain, respectively.

1.1 Diffusion weighted imaging

Diffusion weighted imaging (DWI) is an imaging technique that uses specific MRI sequences in order to track the diffusion process of water molecules in the biological tissue along different directions at the voxel (3D pixel) level in a non-invasive manner [1]. Axons, which are extensions of neurons and make up the white matter in the brain, are impermeable due to the myelin sheath surrounding the fibre and they constrain water diffusion to only their longitudinal direction. This is termed as *anisotropic diffusion*. Hence, DWI indirectly assesses white matter microstructure. *Figure 1.1* [2] shows a schematic of the difference between isotropic diffusion taking place as a result of Brownian motion and anisotropic diffusion taking place in the brain white matter. The main bundles of axons can be reconstructed by piecing together local estimates of water diffusion orientation [3]. These diffusion patterns

can tell us a lot about the the tissue architecture. This imaging technique has improved far beyond purely experimental and now finds use also in clinical settings [4][5][6][7][8][9].

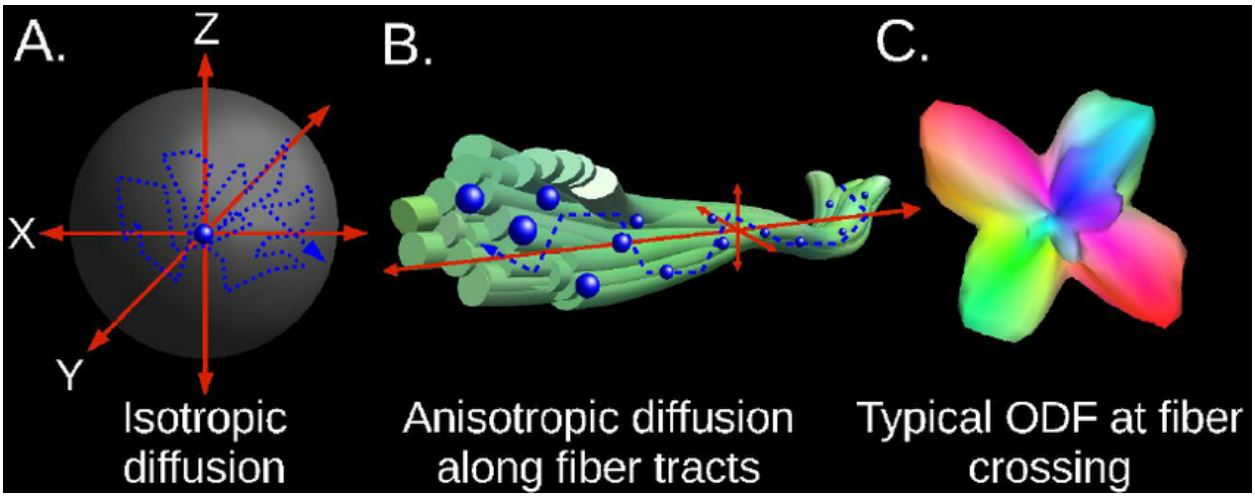


Figure 1.1. A: Isotropic diffusion taking place in a homogeneous medium as a result of Brownian motion. The diffusion pattern is spherical. **B:** Anisotropic diffusion of water molecule within axon and in the interstices of the axon bundles. The diffusion is constrained by axonal walls and myelin sheaths. **C:** Orientation distribution function (ODF) used to model the anisotropic diffusion in the white matter tracts. ODFs can distinguish separate diffusion pathways for multiple fibre tracts crossing at a single location, which is common in the brain. (Figure with permission from Phillips et al., 2012 [2])

Diffusion tensor imaging (DTI) is a special type of DWI that is extensively used to map white matter tracts in the brain through measurement of their location, orientation, and anisotropy. The parallel bundles of axons with myelin sheaths around each one of them facilitates the diffusion water molecules along their main direction, called *anisotropic diffusion*. Thus, DTI measures this restricted diffusion of water molecules in the white matter fibers in order to generate neural tract images. Each voxel in a DTI image contains the information on the rate and three-dimensional directionality of the diffusion process. These parameters are computed for each DTI voxel by tensor mathematics from a number of acquisitions, each having a different orientation of the diffusion gradients. The fibre direction is indicated by the main eigenvector of the tensor computed for each voxel. This information on the directionality can then be used in order to follow white matter fibers in the brain through tractography [10]. Each tract in a tractograph is colour-coded to indicate

its position and direction. The brightness of the tracts indicates their fractional anisotropy, which is a scalar measure of the degree of anisotropy in a voxel. DTI finds uses in clinical settings to localize white matter lesions due to different reasons, such as traumatic brain injury and brain tumors, that do not typically manifest on other forms of clinical MRI [11]. It is also used to assess the development of neurodegenerative diseases such as Alzheimer's and multiple sclerosis [12][13].

1.1.1 Structural connectivity

The human brain has a dense neural architecture made up of billions of neurons that form a highly complex network. There have been numerous studies that have obtain the connectivity patterns of this dense network at the microscale (single neuron) [14][15][16], mesoscale (bundles of neurons) [17][18], or macroscale (distinct gray matter regions) [16][19][20] using various neuroimaging techniques. In this dissertation, we are going to solely focus on the macroscale connectivity patterns that look at the anatomical (or structural) connections between different gray matter regions in the brain.

As mentioned in *Section 1.1*, DTI-based tractography is a commonly used technique for inferring the so-called structural connectome (SC) of the brain. An SC is a connectivity matrix (or *graph*) of the anatomical brain network that provides a comprehensive map of the connections in the human brain by mapping long-range white matter fibres between different cortical and subcortical gray matter regions [21][22][23], as shown in the tractograph in *Figure 1.2*. This graph is composed of a set of nodes (brain regions) that are connected by edges (white matter tracts). In the case of structural connectomes derived using DTI-based tractography, the edges are undirected as the diffusion MRI cannot differentiate between afferent and efferent pathways, although other methods can provide directionality to the edges [24]. Additionally, the edges can be binary or weighted, where the edges have different strengths based on a number of factors, such as directionality, number of fibres between brain regions, length of the connection, etc.

A structural connectome is computed by using a tractogram to link the gray matter regions of interest (ROIs) that are defined by a chosen brain parcellation and inferring the white matter connections between them [25]. Thus, the ROIs represent the nodes and the

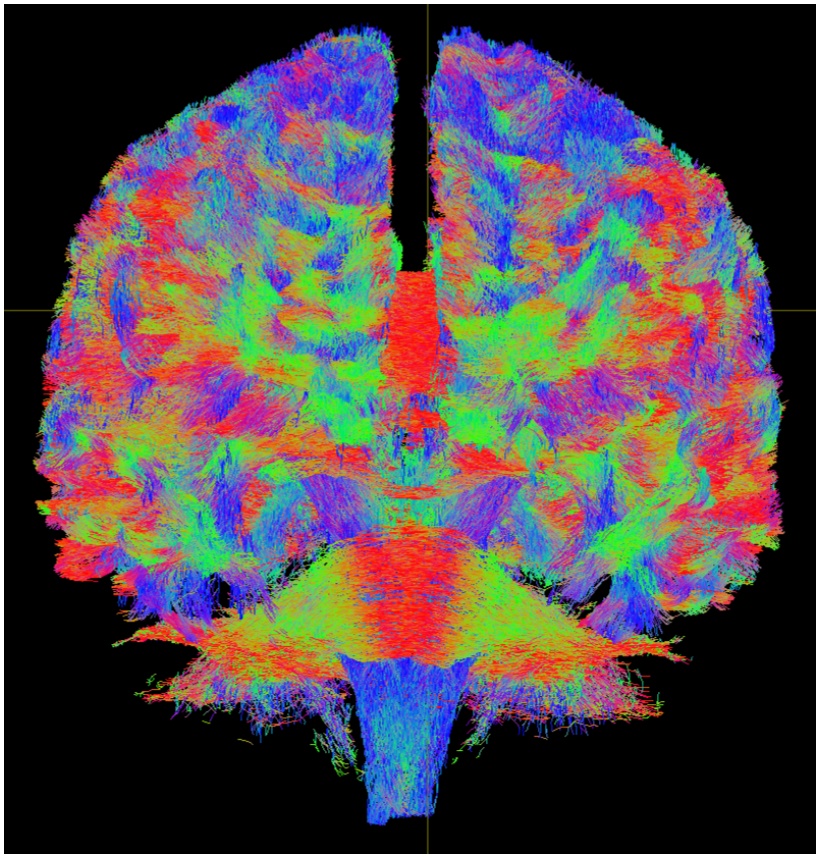


Figure 1.2. Sample tractogram for one HCP subject

inferred white matter connections represent the edges in a structural connectome [26][27]. *Figure 1.3* [23] shows the workflow diagram of the processing pipeline that is commonly employed to derive a structural connectome from raw DWI data. As can be seen, there are several steps that go into obtaining a structural connectome from raw data and thus its quality in terms of reliability heavily depends on the individual decisions taken at each of these steps.

1.2 Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) is an imaging modality that measures the brain activity by detecting changes associated with blood flow in the brain. This imaging technique utilizes the relationship between cerebral blood flow and neuronal activation. When an area of the brain is in use, blood flow to that region increases [28]. fMRI makes use of the blood oxygen level dependent (BOLD) signal to record the changes in the flow

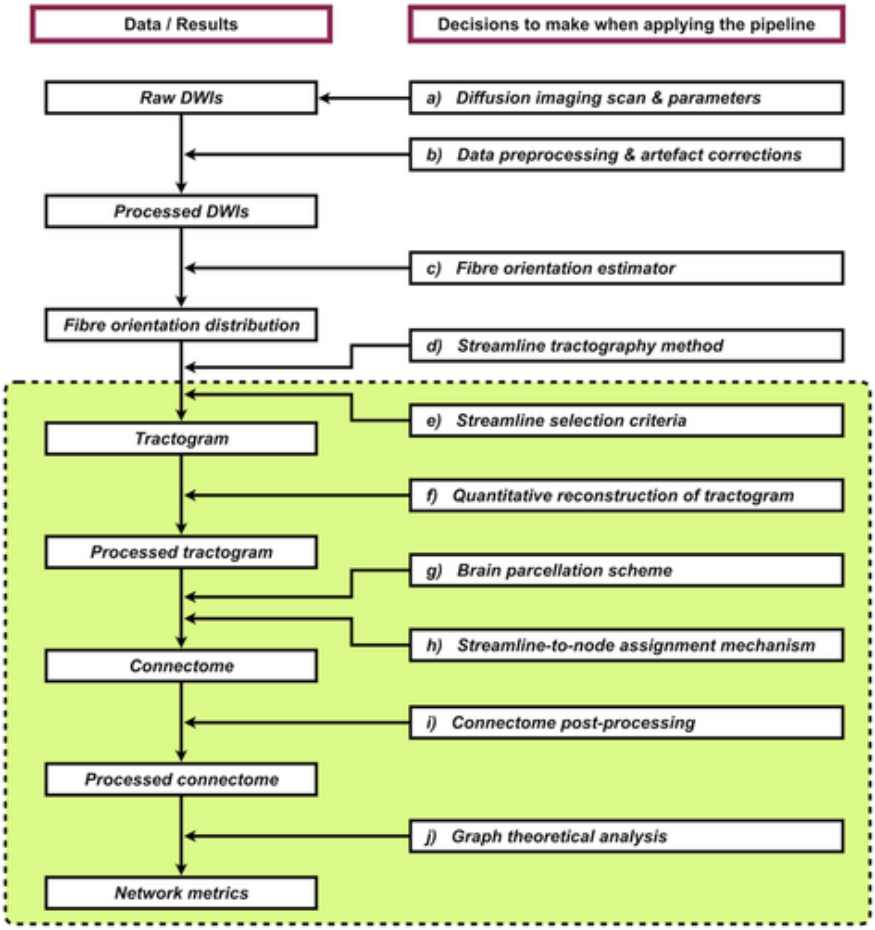


Figure 1.3. Workflow diagram of the processing pipeline commonly employed to derive a structural connectome from raw DWI data. (Figure with permission from Yeh et al., 2020 [23])

of blood to the brain regions that are active across time [29]. As neurons do not contain an internal reserve of energy, their firing causes an increased demand for sugar and oxygen to the particular brain region where this firing is taking place. Blood rushes to these active areas of the brain through a process called *haemodynamic response* in order to release oxygen there. This, in turn, changes the levels of oxyhaemoglobin and deoxyhaemoglobin, both of which have distinct magnetic properties. While oxyhaemoglobin is diamagnetic, deoxyhaemoglobin is paramagnetic which makes it susceptible to changes in the external magnetic field. Thus, the flow of blood to the active areas of the brain causes variation in the magnetic signal, which can then be detected using an MRI scanner [30] with the use of functional MRI acquisition. This process is illustrated in *Figure 1.4* [31].

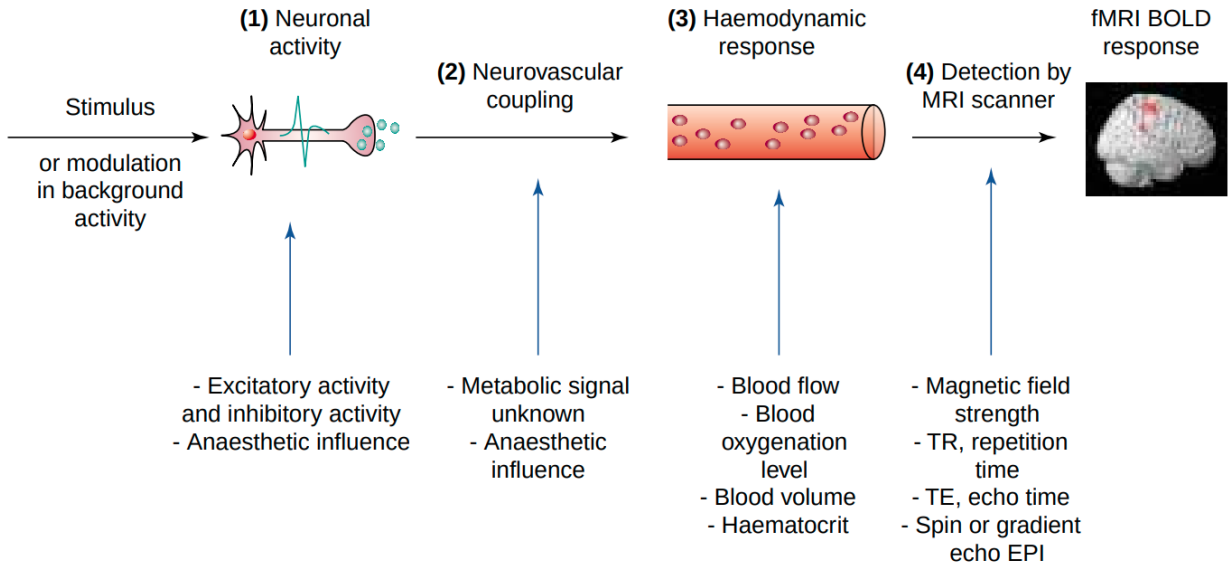


Figure 1.4. The blood oxygenation level-dependent (BOLD) signal has several key determinants. A neuronal response is generated to a stimulus, which in turn triggers a haemodynamic response. This haemodynamic response is detected by an MRI scanner due to the magnetic properties of haemoglobin that is present in the blood. (Figure with permission from Arthurs et al., 2002 [31])

The brain activation maps obtained using fMRI can be graphically represented by color-coding the activation strength across the brain. The main advantage of this technique is that it does not require the subject to undergo any type of invasive procedure, including injection of a dye or being exposed to radiation. fMRI has found extensive uses in research settings and, to a lesser degree, in clinical settings as well. In the field of brain connectomics, it is commonly used in conjunction with DTI to study the relationship between structure and function of the human brain. *Figure 1.5* shows the voxel-wise time series of one subject in the HCP dataset scanned during resting-state fMRI. This time series data can then be processed into parcellated time series and connectomes, as discussed in *Chapter 3*.

1.2.1 Functional Connectivity

Functional connectivity between two brain regions is defined as the temporal relationship between their neuronal activation patterns [32][33][34][35]. Unlike structural connectivity

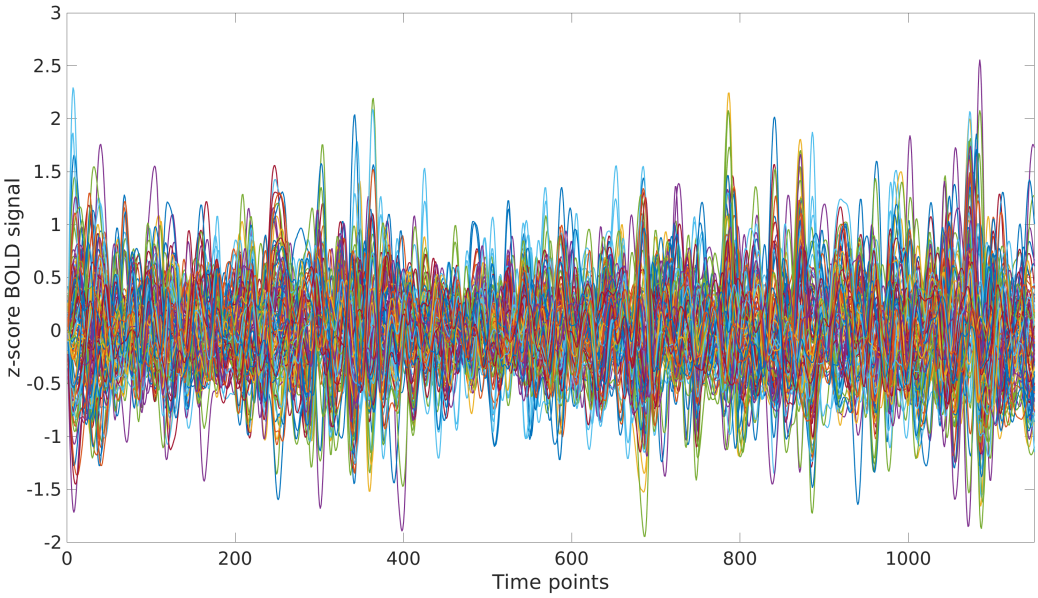


Figure 1.5. Brain region-wise (for Schaefer 100 parcellation) time series of one subject in the HCP dataset scanned during resting-state fMRI

that describes – in an indirect way – the anatomical connections between two brain regions as described in *Section 1.1.1*, functional connectivity examines the interactions or functional coupling between different brain regions in terms of the correlations between fMRI time series. It is worth noting that while the FC of a subject is constantly changing as the subject is performing different tasks (and even during resting state) [36][37][38][39], SC is a much more rigid and slowly evolving topology [40][41][42]. The changes in SC can be observed over the lifespan of a subject but during the few minutes the subject is being scanned the structural architecture of the brain is static. Functional connectomes may be estimated based on electroencephalography (EEG), magnetoencephalography (MEG), local field potentials (LFP), positron emission tomography (PET), or fMRI [43]. Compared to the other imaging modalities, fMRI is a noninvasive technique that provides researchers with an in-vivo representation of brain activation patterns with high spatial resolution. Due to this, in the recent decades, there has been an explosion in the number of neuroimaging studies exploring functional connectivity patterns across the brain by measuring the level of co-activation of resting-state [44] and task-based [45] fMRI time-series. These studies have

uncovered exciting new findings about the functional connectivity patterns of different brain regions along with the overall organization of the brain functional connectome, or FC.

Even though the concept of functional connectivity is fairly straightforward, deriving a functional connectome from raw fMRI data is a long process. And, just as in the case of structural connectivity (see *Section 1.1.1*), there are several processing steps that have to take place in between. Thus, the quality of the resultant functional connectome depends heavily not only on the quality of the raw data, but also on the individual decisions that are being made at every step of data processing. *Chapter 3* will discuss in detail the different steps involved in the processing of the raw fMRI data into individual functional connectomes, as it pertains to the Human Connectome Project (HCP) dataset [46][47][48].

Functional connectivity across the brain also changes over a short time period. This phenomenon can be captured by computing dynamic functional connectivity, or dFC, which is an extension of the traditional FC analysis that typically assumes that functional connectivity profiles are static in time. The most common method used to analyze an fMRI time series to compute dFC is the sliding window analysis [49][50][51][52] shown in *Figure 1.6*. This analysis is performed by computing the FCs using only a set number of timepoints in an fMRI time series, where the number of timepoints is the length of the sliding window. This window is then moved forward a certain number of timepoints depending on the degree of overlap between adjacent windows. The main advantage of this method of conducting dynamic functional connectivity analysis is that it is easy to model, explain, and understand [53]. This method has been used to show different dFC characteristics in diseased and healthy patients [54][55][56], cognitive performance [57][56][58], and between large scale brain states [59][53][60].

A functional connectome is essentially a network representation of the functional connectivity patterns across all brain regions. This network provides the researchers an opportunity to examine the relationships between functional connectivity, information integration, and human behavior using different graph theoretical tools and measures. It also provides for the exploration of how these relationships, and functional connectivity itself, change under different circumstances, such as neurodegeneration, traumatic brain injury, and development, to name a few [32][21][61].

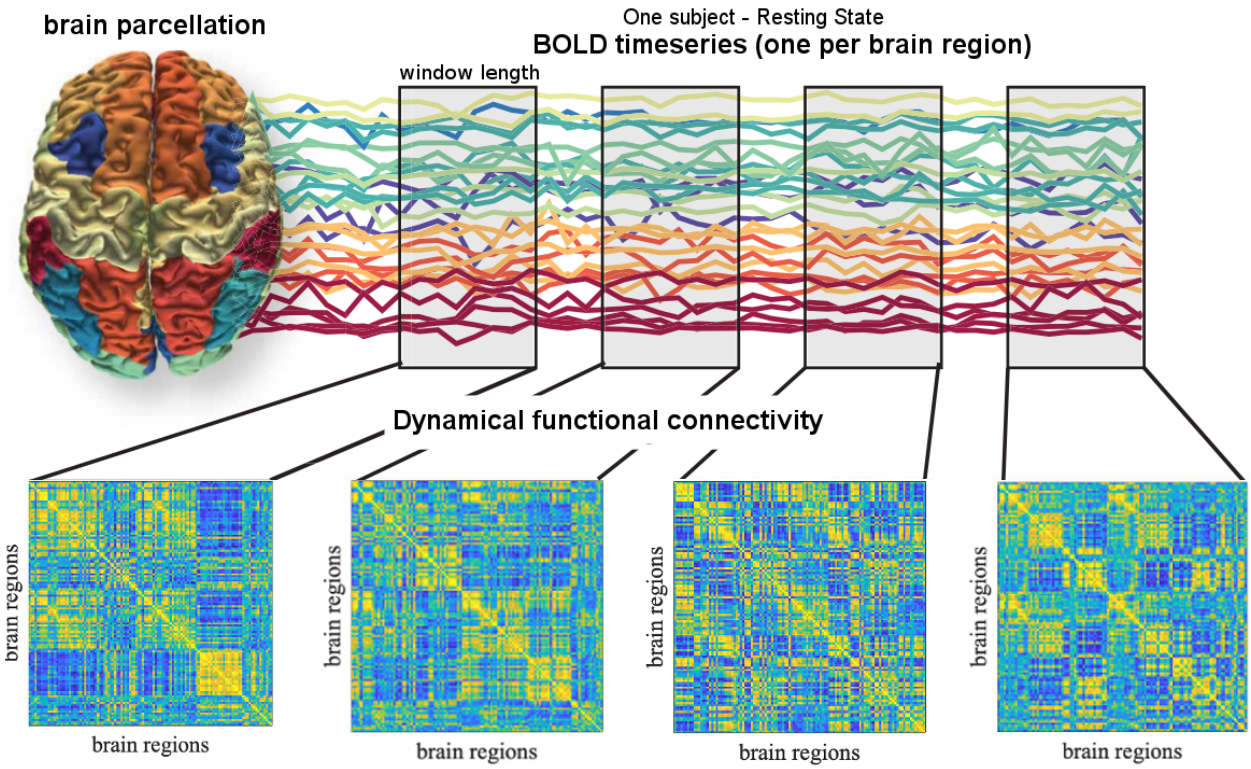


Figure 1.6. Workflow diagram of sliding window analysis to compute dynamic functional connectomes. A window of predefined length is slid along the fMRI time series and FCs are computed using only the timepoints falling within this length. This method captures the dynamic changes happening across time in the functional activity of the brain. The sliding windows can be either overlapping or otherwise.

Every single entry in a functional connectivity matrix, FC_{ij}, corresponds to the functional relationship between two brain regions, i and j. As these are Pearson correlation values, they range between -1 and 1. A high positive value of FC_{ij}, for example a value of 0.8, signifies that the neural activity at the two brain regions i and j is highly in phase. Alternatively, a high negative value, of say -0.8, means that the neural activity at those two brain regions is highly out of phase. In both these cases, it is assumed that there is a high level of communication taking place between the two brain regions i and j. On the other hand, a low value close to zero, either positive or negative, signifies that the two brain regions are out of phase and do no communicate with each other as much. It has also been observed that the longer the shortest path (in terms of structural connectivity) between two brain regions, the lower their functional coupling [62][19][63][22][16][64].

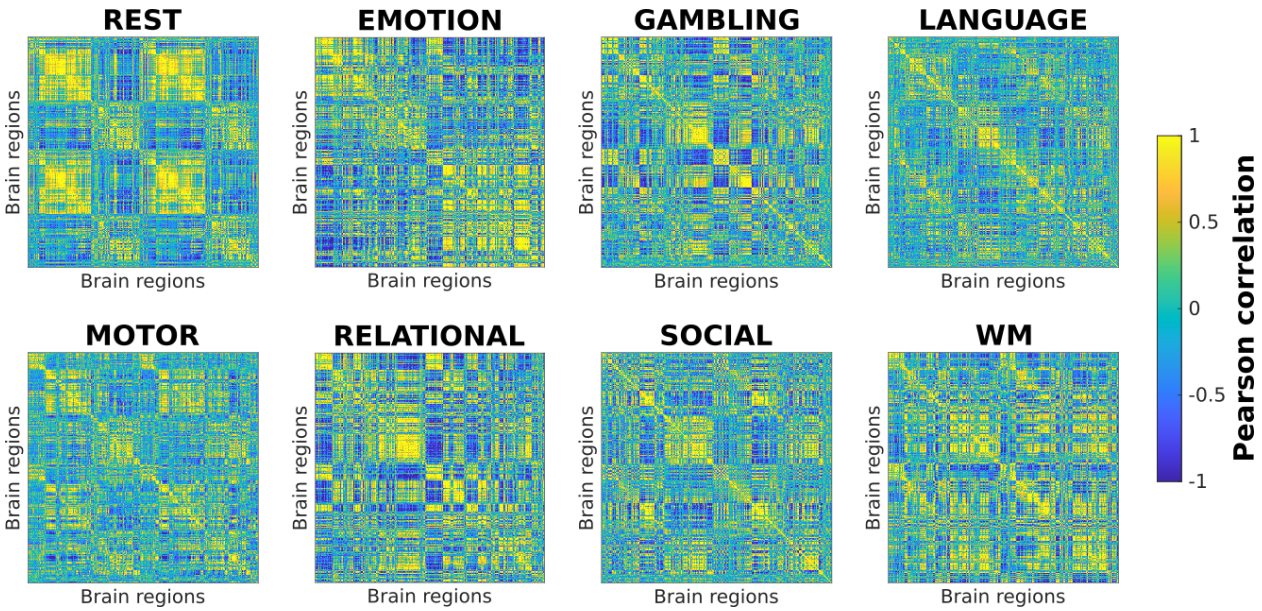


Figure 1.7. Functional connectomes for one subject performing all eight fMRI conditions in the HCP dataset [46][48]

Figure 1.7 shows the functional connectomes of one subject from the Human Connectome Project dataset [46][48] under 8 different fMRI conditions. As can be seen, even though it is the same subject performing these tasks, the FCs look drastically different from each other.

Resting-state functional networks

Across the lifespan of a subject, it has been identified that low-frequency *functional networks* using resting-state fMRI data. These functional networks represent the underlying functional architecture of the brain [65]. The resting-state functional networks have been characterized for aspects of attention [66], memory [67], cognitive control [68][69][70], default mode [71][72], motor [73], and sensory systems [74][75]. Yeo et al., 2011 [76] conducted a comprehensive analysis of functional networks within the human cerebral cortex using resting-state fMRI of 1,000 subjects. They proposed a parcellation of the human brain into seven major resting-state networks: default mode (DMN), dorsal attention, frontoparietal, cinguloopercular (commonly called salience [77] or ventral attention [66]), limbic, visual, and somatomotor networks as shown in *Figure 1.8*. The Yeo networks have become the normative

way of looking at functional communities, not only during rest, but also during tasks. Thus, in this dissertation, we will exclusively use the Yeo networks.

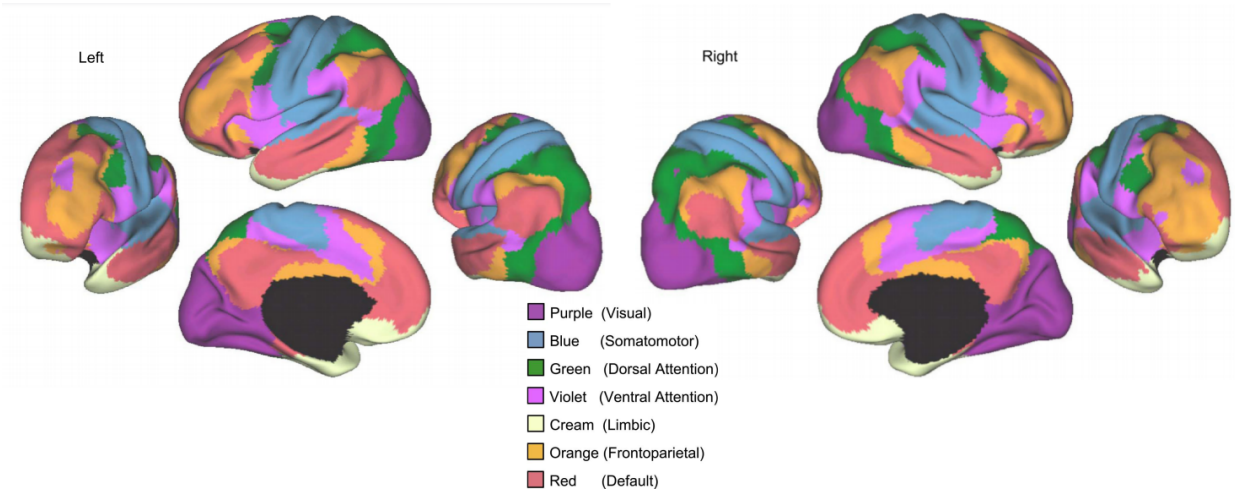


Figure 1.8. A coarse (7-network) parcellation of the human cerebral cortex based on 1,000 subjects. The names and abbreviations for the 7 networks are as follows: visual (VIS, in purple), somatomotor (SM, in blue), dorsal attention (DA, in green), ventral attention (VA, in violet), limbic (LIM, in cream), frontoparietal (FP, in orange), and default mode network (DMN, in red). Additionally, we also add the subcortical brain regions (7 in each hemisphere, 14 total) to the connectomes. (Figure with permission from Yeo et al., 2011 [76])

1.3 Functional connectivity fingerprint gradients

In recent years, functional MRI research has been increasingly finding many applications in clinical settings, including diagnosis of neurological conditions and treatment selection based on the individual diagnosis. Although, one critical caveat with these applications being justified is that the fMRI is capturing the individual differences between subjects in a reliable manner. This has been a bottleneck issue in the research community for a long time which seriously jeopardizes the clinical applications of fMRI [78]. In this section, we will discuss ways to compute and improve upon the concept of functional connectome reliability and fingerprinting. We have also extended the concept of FC fingerprints to capture the similarity between monozygotic and dizygotic twin subjects.

1.3.1 Reliability of repeated measurements

In statistics, *reliability* is defined as the consistency of a measure when an experiment is repeated several times [79]. A measure is said to be highly reliable when the results are highly similar under the same conditions. Thus, a reliability of 100% is achieved when the result is exactly the same every time an experiment is repeated. High reliability is a necessity of research and clinical use of any tool or procedure. Neuroimaging research, specifically, requires that there is a high inter-individual difference so that different individuals are easily identifiable, and low intra-individual differences [80][81][82]. The reliability of repeated measurements in studies can be quantified using intraclass correlation [83][84]. In statistics, the intraclass correlation coefficient (ICC), is a descriptor of how strongly measurements of units in the same group resemble each other. Higher the resemblance between units in the same group, higher the ICC. There are 8 different types of ICC [85] and the two-way random, single score ICC (*ICC-2,1*) is used for computing the reliability of repeated measurements [86][87]. Low reliability in neuroimaging research can lead to low reproducibility [88], erroneously high correlations [89], and a need for huge sample sizes [86][90].

There have been several studies focussed on improving the reliability of resting-state fMRI connectivity measures, specifically by correcting physiological artifacts. Golestani et al., 2017 [91] have showed significant improvements in the reproducibility of resting-state fMRI metrics by the low-frequency physiological correction with end-tidal CO2, while Wang et al., 2017 [92] have demonstrated that test-retest reliability of FCs improves significantly when the impact of sleep is removed using heart rate variability derived from simultaneous electrocardiogram recording. Head motion when the subject is being scanned is yet another source of variability that has been well-studied for its effect on reliability of resting-state fMRI measures by using various preprocessing steps [93][94][95]. It has also been shown that the manner in which these variables are modeled and their order in the preprocessing pipeline also has significant effect on the results [96][97].

1.3.2 Subject fingerprint

Many recent studies have established that functional connectomes have a unique individual *fingerprint* that can be used to identify an individual from a population (a process known as *fingerprinting* or *subject-identification*) [98][99][100][101][102][103][104][105][99][106][107][78]. The concept of brain fingerprints is simple: two functional connectomes of the same subject scanned in different sessions performing the same task should look highly similar. Conversely, two different people performing the same task should look different from each other. As such, subject-level fingerprint is an extension of the idea of reliability of repeated measurements.

There have been several studies that have focussed on quantifying the concept of subject-level fingerprint. Finn et al., 2015 showed that, in FCs, subject fingerprints are reliable and reproducible in high quality datasets [100] (e.g., HCP). They have quantified these subject-level fingerprints through a measure called *identification rate* that relies on the Pearson correlations between test and retest FCs of the subjects. Chiêm at el. [108] have expanded on the concept of identification rate and have introduced a measure called *matching rate*. This measure is equivalent to identification rate, but without replacement.

Amico and Goñi, 2018 showed that this subject-level fingerprint can be improved by using the *differential identifiability framework*, which relies on performing group-level decomposition into principal components followed by an iterative reconstruction adding components in descending order of explained variance until the differential identifiability score reaches an optimal value [98]. Without a high subject-level fingerprint, brain connectomic analyses that are aimed at finding associations between functional connectivity and cognition, behavior, or disease progression are severely compromised [109][110]. Note that the differential identifiability score, I_{diff}, does not maximize the identification rate but rather the group-level contrast of within-subject and between-subject similarity.

Subject fingerprint measurements

Finn et al., 2015 [100] demonstrated that the individual variability in fMRI-based functional connectivity profiles is both robust and reliable, using the the data from HCP. They showed that the FC profiles of individuals act as a *fingerprint* that can be used to accurately

identify subjects from a large set. They showed that this identification was successful across test and retest scanning sessions and even between task and resting-state fMRI by using all the fMRI conditions that are included in the HCP dataset. Thus, they showed that individuals have intrinsic connectivity profiles that can distinguish them from other subjects despite how the brain is engaged during imaging. *Algorithm 1* illustrates the pseudocode for the algorithm developed by Finn et al. to quantify the fingerprint using a metric called *Identification Rate* for test FCs. They assemble two sets of FCs, one for *test* and *retest* each. Next, for each FC in the *test* set (i.e., target FC), they compute the Pearson correlations between the target FC and every FC in the *retest* set. Predicted identity of the target FC is the FC in the *retest* set with the highest correlation value. The prediction is considered successful if the highest correlation is with the retest FC of the same subject. Identification Rate of *test* set with respect to the *retest* set is then defined as the ratio of number of successful predictions with the cohort size. Identification rate can also be computed for the *retest* set with respect to the *test* set, as shown in *Algorithm 2*. Lastly, either both the Identification Rates can be reported separately or an average value can be reported for the entire dataset. Finn et al. have applied this algorithm to the 100 Unrelated Subjects that are a part of the HCP dataset.

Algorithm 1: Computation of Identification Rate (*test* to *retest*)

```
Result: test →retest, Identification Rate
Initialize Success = 0;
Create two sets of FCs, test and retest;
for each FC in test set do
    Pick current target FC from test set;
    for each FC in retest set do
        Compute Pearson correlations between every FC in retest and the target FC;
        Predicted identity is the FC with the maximum Pearson correlation;
        if Prediction successful then
            Success = Success + 1
        end
    end
end
Identification Rate = Success / cohort size
```

Algorithm 2: Computation of Identification Rate (*retest* to *test*)

Result: *retest* →*test*, Identification Rate
Initialize $Success = 0$;
Create two sets of FCs, *test* and *retest*;
for *each FC in retest set* **do**
    Pick current target FC from *test* set;
    for *each FC in test set* **do**
        Compute Pearson correlations between every FC in *test* and the target FC;
        Predicted identity is the FC with the maximum Pearson correlation;
        if *Prediction successful* **then**
            $Success = Success + 1$
        end
    end
end
$Identification\ Rate = \frac{Success}{cohort\ \ size}$

Chiêm et al. [108] have proposed a variant of the identification rate, called the *matching rate*. This measure is equivalent to the concept of identification rate, but without replacement. Matching rate is most applicable in the case when the test FC of a subject is most similar to not only its own retest FC, but to that of other subjects as well. In extreme cases where an individual FC is highly similar to many other FCs in the dataset, the identification rate will be affected as many of the subjects will not be correctly identified. In matching rate, every time a test FC is matched with a retest FC based on the highest Pearson correlation, the matched test-retest pair is removed before the next comparison is made. This way, all FCs are matched only once, regardless of whether they correlate highly with other FCs or not. *Algorithms 3* and *4* illustrate the pseudocode for the algorithm developed by Chiêm et al. to compute the matching rate. As discussed, these algorithms are similar to *Algorithms 1* and *2*, respectively, that compute the identification rates, with the exception of the replacement of the matched FCs.

Amico and Goñi, 2018 [98] have developed another measure of individual fingerprint by improving upon the definition provided by Finn et al. They have proposed a measure called *differential identifiability* that not only looks at the similarity between the test and retest

[1] A_{ii} is Pearson correlation between *test* and *retest* of the same subject and A_{ij} is the Pearson correlation between *test* and *retest* of different subjects i and j

Algorithm 3: Computation of Matching Rate (*test* to *retest*)

```
Result: test →retest, Matching Rate
Initialize Success = 0;
Create two sets of FCs, test and retest;
for each FC in retest set do
    Pick current target FC from test set;
    for each FC in test set do
        Compute Pearson correlations between every FC in test and the target FC;
        Predicted identity is the FC with the maximum Pearson correlation;
        if Prediction successful then
            Success = Success + 1
        end
        Remove the paired test and retest FCs from the datasets;
    end
end
Matching Rate = Success / cohort size
```

connectomes of a subject, but also accounts for the dissimilarity between connectomes of different subjects. Their proposed method, shown in *Algorithm 5*, also works on improving the individual fingerprint using principal component analysis (PCA) on a group level. As can be seen in the pseudocode, the principal components are always added to the reconstruction in the order of descending explained variance of the individual components. On one hand, this guarantees that a large amount of the variance in the original data is preserved upon reconstruction. On the other hand, it limits the potential of finding subsets of principal components that could potentially lead to higher differential identifiability scores. Thus, further exploration into the combinations and permutations of principal components is necessary to see if they have a higher differential identifiability score. This algorithm will be discussed in further detail in *Chapter 3*. In this paper, they have also used the FCs of the 100 Unrelated Subjects part of the HCP dataset to demonstrate that the optimal reconstruction of the individual FCs through connectivity eigenmodes maximizes subject identifiability across all fMRI conditions evaluated. They also show that the differential identifiability of the optimally reconstructed individual FCs increases both at the global and edgewise level.

Algorithm 4: Computation of Matching Rate (*retest* to *test*)

Result: *retest* →*test*, Matching Rate
Initialize $Success = 0$;
Create two sets of FCs, *test* and *retest*;
for *each FC in retest set* **do**
    Pick current target FC from *test* set;
    for *each FC in test set* **do**
        Compute Pearson correlations between every FC in *test* and the target FC;
        Predicted identity is the FC with the maximum Pearson correlation;
        if *Prediction successful* **then**
            $Success = Success + 1$
        end
        Remove the paired *test* and *retest* FCs from the datasets;
    end
end
$Matching\ Rate = \frac{Success}{cohort\ \ size}$

1.3.3 Motivation behind using differential identifiability

The seminal paper by Amico and Goñi, 2018 [98] proposed the differential identifiability framework to uncover fingerprints in same-subject test and retest FCs for all fMRI conditions included in the Human Connectome Project dataset [46]. This algorithm has subsequently been used in several different studies to uncover functional connectivity fingerprints in various other datasets.

In 2019, Bari et al. [111] uncovered FC fingerprints between multiple scanning sites. Thus, in this case, the test and retest fMRI scans were conducted at two different sites and on two different scanners. They conducted this analysis on two independent multi-site datasets. Their results show the existence of individual fingerprints in test-retest visit pairs within and across two sites. Their optimally reconstructed functional connectomes demonstrated a marked improvement in subject fingerprinting within and across the two sites and test-retest visit pairs relative to the original data.

Rajapandian et al., 2020 [104] used the differential identifiability framework to uncover the fingerprints in FC-based network properties. They show that improving across-session fingerprint of FCs also improves fingerprint of derived network measures. They have also demonstrated that, for specific network properties, application of the differential identifi-

Algorithm 5: Computing the differential identifiability profile for subject fingerprint

Result: Differential identifiability vector $I_{diff,vec}$
Initialize $FC_{orig} = [\,]$;
for *subj = 1 to cohort size* **do**
    $test\_vec$ = Fisher transformed and vectorized *test* FC for *subj*;
    $retest\_vec$ = Fisher transformed and vectorized *retest* FC for *subj*;
    $FC_{orig} = [FC_{orig}, test\_vec, retest\_vec]$
end
Decompose FC_{orig} using PCA into as many components as FCs;
Initialize $I_{diff,vec} = [\,]$;
for *comp = 2 to number of components* **do**
    Reconstruct FCs using *1:comp* components;
    Add back the mean of the original FCs;
    Inverse Fisher transform the reconstructed FCs;
    Compute identifiability matrix, A^1;
    $I_{diff} = [avg(A_{ii}) - avg(A_{ij})] \times 100$; where, i $\neq$j, 1 $\leq$i,j $\leq$cohort size;
    $I_{diff,vec} = [I_{diff,vec}, I_{diff}]$
end

ability framework directly on network properties is more effective. They have concluded that applying the framework, on either the FCs or the network properties, increases task sensitivity of the network properties.

Lastly, Svaldi et al. [109] showed that the connectome predictive modelling of cognitive deficits in Alzheimer's disease improves when the differential identifiability of the dataset is optimized using the differential identifiability framework. This study combines the frameworks of connectome predictive modelling and differential identifiability in order to show that enhancing the subject-level fingerprint of resting state FCs leads to robust identification of functional networks associated with cognitive outcomes and also improves prediction of cognitive outcomes from the FCs.

In this dissertation, we have extended the differential identifiability framework to uncover the functional connectivity fingerprints between not only test-retest of the same subjects, but also between sets of twin subjects in the HCP dataset. The details of the extension to the algorithm will be discussed in further detail in *Chapter 3*. We have seen in the studies mentioned above that, in case of the same subject test-retest data, PCA decomposes the

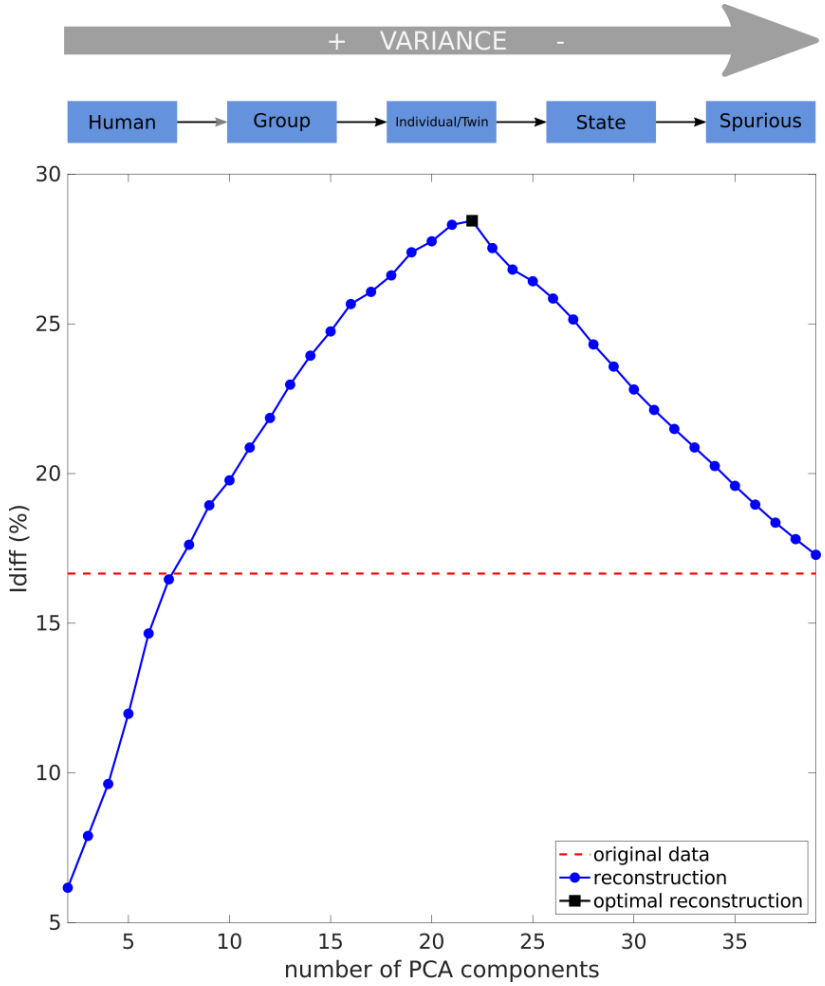


Figure 1.9. The five categories of principal components with their specialization

original data in such a way that the individual principal components (in descending order of their explained variance) can be divided into different categories. Thus, we hypothesize that the principal components presented in descending order of explained variance enable us to distinguish between the different regimes of principal components, as shown in *Figure 1.9*. These regimes can be defined defined as follows:

1. ***Human:*** The first few components represent what is common in all functional connectome by virtue of them being from humans.

2. ***Group:*** The next category of components represents the common characteristics of the cohort under study.

3. ***Individual:*** These principal components specialize in the individual differences between different subjects. No single component specializes in a single subject, but rather these components tell the amount of specific network structures present in the different subjects.

4. ***State:*** These components correspond to the different state of mind that the individual is in at the time of the fMRI scan. These different states bring noise to the data and lead to a lower individual fingerprint.

5. ***Spurious:*** The last few principal components, which also correspond to the smallest explained variance, capture the noise that blurs the individual fingerprint of subjects and the differences between them. The noise can be caused by scanner artifacts, head motion, etc.

We hypothesize that the same trend of the different sets of components corresponding to different information in the dataset will be seen in the case of MZ and DZ twins as well. Except, instead of the components in the middle corresponding to the information about individual subjects will instead correspond to twin subjects, as shown in *Figure 1.9*. We show in *Chapter 3* that that is indeed the case as we see the same trend of an optimal differential identifiability score corresponding to a certain number of principal components as seen in same subject-test retest also be applicable to the MZ and DZ twin cohorts.

It is worth noting that the aim of the differential identifiability framework is not only to assign a fingerprint score to a dataset. It combines that aim with improving the estimations of single-subject, single-session functional connectomes. In that context, the goal behind using this approach is not to find the individual/isolated principal components that may carry high subject or twin identifiability. Instead, our aim is to obtain optimally reconstructed FCs that account for most of the variance explained in the data, while reducing noise at the same time. Thus, we not only provide a fingerprint score at the group level for the dataset – which can also be done by identification rate and matching rate – but also provide back to the user FCs with high signal-to-noise ratio at the session level. This additional achievement is critical so that the larger brain connectomics community can assess optimally reconstructed datasets with high a reliability measured in terms of a fingerprint score.

1.3.4 Twin fingerprint and twin studies

Functional connectivity fingerprints are not unique to test/retest sessions of the same individuals. Subjects sharing genetics and/or environment are expected to have, to some extent, a fingerprint. In particular, similar to subject-level fingerprint in brain functional connectomes, it has been established that a fingerprint also exists in the FCs of monozygotic (MZ) and dizygotic (DZ) twin subjects [112][113][114][115][116][117], albeit to a lower extent than the subject-level fingerprint. *Table 1.1* shows the amount of genetic and environmental similarity between the three datasets – test/retest of the same subject, MZ twins, and DZ twins. As MZ twins are genetically the same, the difference in the two individuals arises from the difference in their environments. For example, one twin might be a hockey player while the other a swimmer. Or, one twin might be a vegan and the other might enjoy their bacon. On the other hand, DZ twins are genetically equivalent to siblings as they share (on average) 50% of their genetic material. But, as they are still twins, their level of genetic similarity is equivalent to that between MZ twins. These different levels of similarities within the datasets can explain why test/retest has the highest level of fingerprint and DZ twins the lowest. *Figure 1.10* [118] shows the contribution of two factors, genetics and environment, to a phenotype. As can be seen, depending on the phenotype being observed, the contribution of these two factors varies. On one side of the spectrum, there are conditions such as Huntington's disease that are caused exclusively by genetic factors, while on the other side are purely environmental conditions such as a traumatic brain injury. The extent genetics and environment dictate a phenotype can be separated using an ACE model that utilizes twin correlations.

Table 1.1. Genetic and environmental similarity between within cohorts

Dataset	**Genetic similarity**	**Environmental similarity**
Test/retest	100%	100%
MZ twins	100%	$< 100\%$
DZ twins	50%	$< 100\%$ ($\simeq$ MZ twins)

Figure 1.11 from Blokland et al., 2013 [118] shows the scatter plots between the heights of MZ and DZ twins. As can be seen, there is a higher correlation between the heights of

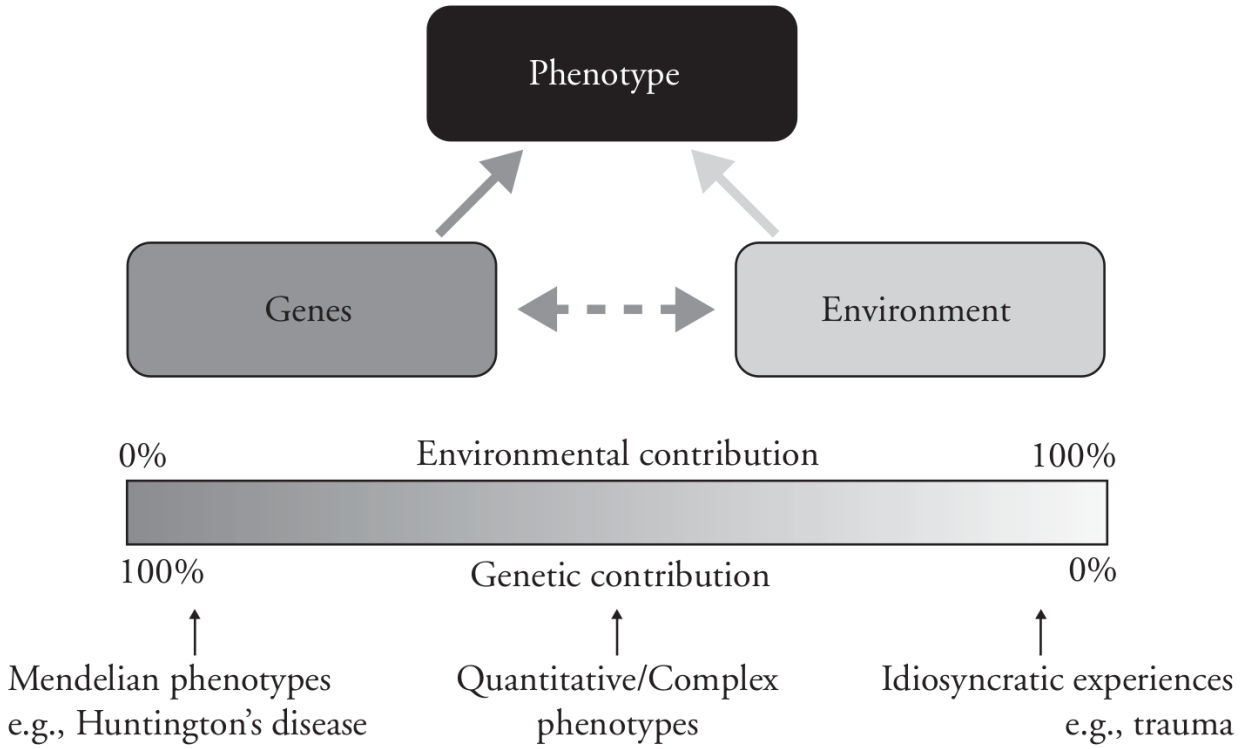


Figure 1.10. Individual differences on traits result from genetic and/or environmental influences, or a combination of both. Mendelian traits, such as Huntington's disease, are (almost) entirely inherited, while traumatic brain injury can be caused by environmental exposures, such as a car accident. Quantitative traits are generally influenced by a combination of genetic and environmental influences. (Figure with permission from Blokland et al., 2013 [118])

MZ twins than that of DZ twins. As explained in *Table 1.1*, this can be explained by the difference in genetic similarity within these two datasets. If we were to produce a similar scatter plot for two measurements, test and retest, of the heights of every individual in a cohort, the correlation would be even higher than MZ twins, and very close to 1. As these are repeat measurements of the same subject, the difference can be chalked up to only measurement error.

Kumar et al. 2018 [114] have presented a framework based on manifold approximation for generating brain fingerprints from multimodal data using T1/T2-weighted MRI, diffusion

[2] A_{ii} is Pearson correlation between FCs of subjects from the same twin pair and A_{ij} is the Pearson correlation between FCs of non-twin subjects i and j

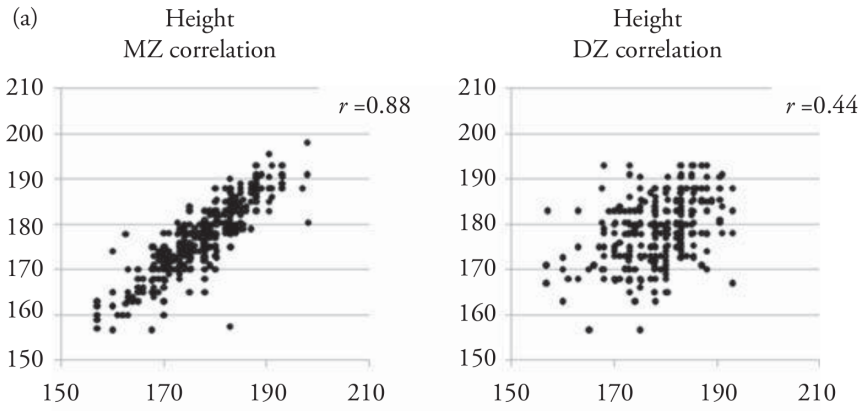


Figure 1.11. Scatter plots showing MZ and DZ twin pair correlations for height in cm (males only) indicating a high heritability for this trait, as the correlation is higher for MZ twins than for DZ twins. Data were provided by the Genetic Epidemiology Laboratory, Queensland Institute of Medical Research. (Figure with permission from Blokland et al., 2013 [118])

MRI, and resting-state fMRI. Their results show a link between amount of fingerprint and genetic proximity as the MZ twins have more prominent fingerprints than DZ or non-twin siblings. Ge et al. 2017 [112] have used a linear mixed effects model to dissociate intra- and inter-subject variation of a phenotype and computed heritability with respect to stable inter-subject variation in fMRI data as the phenotype. Colclough et al. 2018 [116] have investigated the influence of genetics and common environment on functional connectomes of individuals obtained from fMRI and MEG data in HCP. Demeter et al. 2020 [117] have applied support vector machine classifiers on resting state fMRI to predict retest and co-twin pairs from two twin datasets (adult and pediatric) that include repeat scans. Gritsenko et al. 2020 [115] propose a pair-wise twin classification method to identify the zygosity of twin pairs using the resting state fMRI.

The latest release of the HCP-YA [48] dataset includes unrelated subjects, as well as subjects that are related to each other, including MZ and DZ twins. This affords us the opportunity to assess brain connectivity fingerprints not only by comparing test and retest functional connectomes of the same subject (*subject-level fingerprint*), but also by comparing the functional connectomes of MZ and DZ twins (*twin fingerprint*) across different fMRI conditions. We have used an extended version of the algorithm proposed by Amico and Goñi,

Algorithm 6: Computing the differential identifiability profile for twin fingerprint

Result: Differential identifiability vector $I_{diff,vec}$
Initialize $FC_{orig} = [\,]$;
for *i = 1 to number of twin pairs* **do**
  $twin_1\_vec$ = Fisher transformed and vectorized FC for $twin_1$ from pair i;
  $twin_2\_vec$ = Fisher transformed and vectorized FC for $twin_2$ from pair i;
  $FC_{orig} = [FC_{orig}, twin_1\_vec, twin_2\_vec]$
end
Decompose FC_{orig} using PCA into as many components as FCs;
Initialize $I_{diff,vec} = [\,]$;
for *comp = 2 to number of components* **do**
  Reconstruct FCs using *1:comp* components;
  Add back the mean of the original FCs;
  Inverse Fisher transform the reconstructed FCs;
  Compute identifiability matrix, A^2;
  $I_{diff} = [avg(A_{ii}) - avg(A_{ij})] \times 100$; where, i $\neq$j, 1 $\leq$i,j $\leq$cohort size;
  $I_{diff,vec} = [I_{diff,vec}, I_{diff}]$
end

2018 [98], as shown in *Algorithm 6*. As can be seen, we can compute the *twin fingerprint* by simply replacing the test/retest FCs in *Algorithm 5* with one connectome each from the twins in a pair.The rest of the algorithm is identical to the one used for computing subject-level fingerprint. This has been discussed in further detail in *Chapter 3*.

1.4 Dissertation outline

This dissertation document is divided into three main parts. The first part, covered in *Chapter 2*, focuses on studying the communication taking place in the structural network of human brain [62]. We have made use of a collaborative spreading algorithm inspired by the natural behaviour of an ant colony, called the *ant-colony algorithm*. We also report the associations of the results of this algorithm with measured functional connectivity in the human brain under different conditions. This provides us valuable information about whether the behaviour of ants is a good aggregate of the communication in the brain.

Chapter 3 [119] is dedicated to the processing pipeline we have developed to process the fMRI data included in the HCP dataset [46][48] into functional connectomes using the local-global Schaefer parcellations that have multiple granularities [120]. We have processed the

fMRI data for all the subjects and have made it available to the scientific community. Along with the data processing, we have also analyzed the resulting functional connectomes for their subject and twin fingerprint. We have extended the *differential identifiability framework* by Amico and Goñi [98] for this analysis. As the Schaefer parcellations have varying granularities, we have been able to study the effect of granularities on the fingerprints.

In the last part of this dissertation, *Chapter 4*, we have used the data processed in the previous chapter to disentangle the effects of genetics and environment on functional connectome. As the HCP dataset has extensive family structures, including MZ and DZ twins, we have used the classical twin model, also known as the ACE model [118][121], for this purpose with a proposed extension to this model. This extension makes use of the repeat measurements (test and retest) included in the HCP dataset to also compute the measurement error. We can then use this information to further disentangle the effects of uncommon environment from pure error that cannot be explained.

2. MODELLING COMMUNICATION PROCESSES IN THE HUMAN CONNECTOME THROUGH COOPERATIVE LEARNING

2.1 Introduction

Shortest paths in any system that can be described as a network, such as the human brain or roads in a state, are considered important as routing through them could minimize communication delays [122]. The concept of a shortest path is defined as the sequence of edges between a node pair in the network corresponding to the shortest distance between them. In the case of a road network, shortest paths may be used by a traveler who has knowledge of the network topology. On the other hand, in a biological network, such as the human brain, it is not assumed that the signal has knowledge of the network topology and thus does not necessarily take the shortest path. Even so, shortest paths are highly important in a biological network as they inform us of the most efficient routing.

Additionally, shortest paths and walks also have practical importance in artificial systems that can be represented as networks. For example, this measure has proven useful in identification of keywords in textual applications [123], distinguishing literary styles in books published over four centuries [124], and probing the topology of language networks [125] [123] [126], among other applications. The measure also has applications in pattern recognition and machine vision [127].

In many natural systems, the concept of shortest paths can be relaxed to a natural selection of one or more communication paths (potentially, an ensemble of paths) chosen through a collaborative effort, as in the case of foraging behaviour of ants. Deneubourg et al. [128] conducted experiments on Argentine ants (*I. humilis*) to study their pheromone-driven foraging behaviour. In order to study how the indirect collaboration, also called *stigmergy*, evolves over time from random exploration for food by an ant colony, the authors set up a two-bridge environment (from a network perspective, two nodes connected through two different edges, as in the case of multigraphs) that the ant colony explored. Both the edges (here representing possible paths) were of the exact same length and the passage of ants over the edges was observed over time. It was observed that at the beginning the

individual ants randomly chose one of the two possible paths in search of food. However, as the pheromones dropped by the ants on their way started accumulating and affecting the environment over time, the ants eventually converged to using only one of the edges. As the experiment was run repeatedly, either of the paths emerged as the one on which the ants randomly converged. This is explained by the slight fluctuation in the number of ants taking each of the paths, increasing the pheromone concentration on that path. In the second part of the experiment, the ratio of lengths of the two bridges was 2:1. As a result, the ants always converged on the shortest path every time the experiment was run.

The foraging behaviour of many species of ants is driven by the indirect communication between them through chemicals, called *pheromones*, as their visual faculties are not well developed [129]. When a colony of ants starts searching for food, individual ants do so in a completely random and uncooperative manner [129]. As soon as an ant finds a food source, it takes some of it and carries it back to the nest and then starts again towards the food source. Thus, the ants keep moving back and forth between the nest and the food source until the food is completely depleted. During this journey, the ants leave a trail of pheromones on their path that other ants can smell. As the concentration of pheromones increases, ants gradually change their exploration behavior from unbiased random walks to an exploration biased by the concentration of pheromones. This is a critical characteristic of the ant-colony that allows the individual agents to initially explore many possible paths while subsequently converging to a potentially optimal path or ensemble of paths. This behavior of ant colonies seems to emerge in open two-dimensional spaces [128] as well as in constrained, network-like structures [130][128].

Ant colony optimization (ACO) algorithm takes its motivation from this indirectly collaborative behaviour of ants that allows them to find the shortest paths [131]. The ACO is typically used to find solutions to NP-hard problems that can be modeled as shortest path problems in a graph, e.g., travelling salesman problem (TSP), scheduling problems, assignment problems, amongst others.

It is well known that neuronal structure in the brain forms a complex structural network [132]. This network is a static representation of white matter connections between brain regions. As such, it is a very slowly evolving topology. This structural network, called

structural connectivity (SC), dictates how different parts of the brain communicate with each other, which is known as functional connectivity (FC). FC between two brain regions is the correlation between their fMRI time-series data. Thus, it is a fast evolving topology. Structural topology of the Human Connectome has been extensively assessed through measures that provide *static* views of the underlying connectivity of the brain network, such as shortest path between regions, search information (SI), modularity, and degree distribution, amongst others. As these measures are static, they do not provide much explanation of how the SC and FC, which is a dynamic topology, might be related. Also, only a fraction of the edges in a network form the shortest paths. Thus, by assuming that communication in the brain takes place through shortest paths, one is essentially ignoring large parts of the network (discussed further in *Section 2.3.1*).

In order to overcome the problems discussed above, De Vico Fallani et al. [133] and Avena-Koenigsberger et al. [134] have proposed that the communication between brain regions does not take place through shortest paths. Instead, Avena-Koenigsberger et al. have suggested that communication between brain regions takes place through an ensemble of k-shortest paths, while De Vico Fallani et al. have suggested investigating all possible paths between a pair of brain regions consisting of a certain number of edges. Thus, in both of these methods the paths that are investigated are pre-defined by the user.

Communication efficiency as defined by Estrada and Hatano [135] takes into account all potential paths between a pair of source-target nodes instead of using only the shortest paths. The authors propose a function based on this measure as a way to quantify how much information can flow from one node to another in a network. Furthermore, Chavez et al. [136] have studied the accessibility of different cortical areas to measure how well an area can be reached by the rest of the network. The method proposed in the current paper is a relaxation of the measure proposed by Estrada and Hatano, as we consider only the paths that are traversed by the ants. It can also be compared to the accessibility measure proposed by Chavez et al., as it measures the hiddenness of brain regions from each other.

In this paper, we propose a method to model signal propagation and communicability between brain regions through the use of an ant colony-inspired algorithm. We test this novel framework on the functional and structural data provided by the Human Connectome

Project [137][138]. When exploring the network topology of the human structural connectome, the ants trace the ensemble of paths between each source-target pair of brain regions by traversing them probabilistically. By tuning two main communicability parameters related to the ant colony behavior (i.e., *pheromone perception* and *edge perception*), we investigate four different communication scenarios on SC: independent random walk, weakly coupled random walk, collaborative spreading preferentially along weak structural connections (*side roads*), and collaborative spreading preferentially along strong structural connections (*main roads*). For each scenario, we define two network measures extracted from the path ensembles traveled by the ants, namely *effective path length* (EPL) and *arrival rate* (AR). We show how these two node pair-wise measurements are good predictors of task-based FCs and partly of resting-state FCs, for different optimal choices of *pheromone perception* and *edge perception*. The predictive power of AR and EPL is even more noteworthy when considering communication scenarios within different functional subnetworks. We conclude by discussing the potentials of this new model for describing communication in large-scale brain networks and new directions for the investigation of communicability and signal propagation regimes in the human connectome, a new exciting avenue in brain connectomics.

2.2 Methods

2.2.1 Human Connectome Project Data Processing

The functional and structural dataset used in this work is from the Human Connectome Project (HCP, http://www.humanconnectome.org/), Release Q3. Below is the full description of the acquisition protocol and processing steps. We employed the Freesurfer parcellation into 164 brain regions [139][140].

HCP: Structural Data

We used DWI runs from the 100 unrelated subjects of the HCP 900 subjects data release [137], [138]. The diffusion acquisition protocol is covered in detail elsewhere [47][141][142]. Below we mention the main characteristics. Very high-resolution acquisitions (1.25 mm isotropic) were obtained by using a Stejskal–Tanner (monopolar) [143] diffusion-encoding

scheme. Sampling in q-space was performed by including 3 shells at b = 1000, 2000 and 3000 s/mm2. For each shell corresponding to 90 diffusion gradient directions and 5 b = 0's acquired twice were obtained, with the phase encoding direction reversed for each pair (i.e., LR and RL pairs).

The HCP DWI data were processed following the MRtrix3 [144] guidelines (http://mrtrix.readthedocs.io/en/latest/tutorials/hcp_connectome.html), as done in recent paper [145]. In summary, we first generated a tissue-segmented image appropriate for anatomically constrained tractography (ACT [146], MRtrix command `5ttgen`); we then estimated the multi-shell multi-tissue response function [147](MRtrix command `dwi2response msmt_5tt`) and performed the multi-shell, multi-tissue constrained spherical deconvolution [148] (MRtrix `dwi2fod msmt_csd`); afterwards, we generated the initial tractogram (MRtrix command `tckgen`, 10 million streamlines, maximum tract length = 250, FA cutoff = 0.06) and applied the successor of Spherical-deconvolution Informed Filtering of Tractograms (SIFT2, [149]) methodology (MRtrix command `tcksift2`). Both SIFT [149] and SIFT2 [150] methods provide more biologically meaningful estimates of structural connection density. SIFT2 allows for a more logically direct and computationally efficient solution to the streamlines connectivity quantification problem by determining an appropriate cross-sectional area multiplier for each streamline rather than removing streamlines altogether, biologically accurate measures of fibre connectivity are obtained whilst making use of the complete streamlines reconstruction [149]. Finally, we mapped the SIFT2 outputted streamlines onto the 164 chosen brain regions [139], [140] to produce a structural connectome (MRtrix command `tck2connectome`). Finally, a log10 transformation [63] was applied on the structural connectomes to better account for differences at different magnitudes. In consequence, SC values ranged between 0 and 5 on this dataset.

HCP: Functional Data

The description below is a summary of the detailed fMRI processing presented in Amico and Goñi, 2018 [98]. We used fMRI runs from the 100 unrelated subjects of the HCP 900 subjects data release [137], [138]. The fMRI resting-state runs (HCP filenames: rfMRI_REST1 and rfMRI_REST2) were acquired in separate sessions on two different days, with two

different acquisitions (left to right or LR and right to left or RL) per day [137], [138], [47]. The seven fMRI tasks were the following: gambling (tfMRI_GAMBLING), relational (tfMRI_RELATIONAL), social (tfMRI_SOCIAL), working memory (tfMRI_WM), motor (tfMRI_MOTOR), language (tfMRI_LANGUAGE, including both a story-listening and arithmetic task), and emotion (tfMRI_EMOTION). The working memory, gambling, and motor task were acquired on the first day, and the other tasks were acquired on the second day [47], [151]. The HCP scanning protocol was approved by the local Institutional Review Board at Washington University in St. Louis. For all sessions, data from both the left-right (LR) and right-left (RL) phase-encoding runs were averaged to calculate connectivity matrices. Full details on the HCP dataset have been published previously [47], [151], [152].

The HCP functional preprocessing pipeline [47], [152] was used for the employed dataset. This pipeline included artifact removal, motion correction and registration to standard space. Full details on the pipeline can be found in [47], [152],[153].

For the resting-state fMRI data, we also added the following steps: global gray matter signal was regressed out of the voxel time courses [154]; a bandpass first-order Butterworth filter in forward and reverse directions [0.001 Hz, 0.08 Hz] [154] was applied (Matlab functions butter and filtfilt); the voxel time courses were z-scored and then averaged per brain region, excluding outlier time points outside of 3 standard deviation from the mean, using the workbench software [155] (workbench command - `cifti-parcellate`). For task fMRI data, we applied the same above mentioned steps, with a less restrictive range for the bandpass filter [0.001 Hz, 0.25 Hz].

Pearson correlation coefficients between pairs of nodal time courses were calculated (MATLAB command `corr`), resulting in a symmetric connectivity matrix for each fMRI session of each subject. Functional connectivity matrices from the left-right (LR) and right-left (RL) phase-encoding runs were averaged to improve signal-to-noise ratio. The functional connectomes were kept in its signed weighted form, hence neither thresholded nor binarized.

Finally, group average matrices were obtained from the resulting individual structural and functional connectivity (rest and 7 tasks) matrices. These were then grouped (rows and columns) according to the 7 cortical functional networks (FNs) as proposed by Yeo et al. [76] based on resting state.

2.2.2 Ant-colony Inspired Algorithm

Ant-colony algorithm is an optimisation algorithm used to solve NP-hard problems [131], although in this paper the algorithm has been modified so that the goal is not optimisation anymore. Instead of finding the shortest path between brain regions, which can be done by better and much faster algorithms [156][157], the modified ant colony algorithm is aimed at exploring the brain network and at assessing different communication regimes between source and target nodes at different cognitive states. At the beginning, a source node and a target node are fixed, with the colony of ants located at the source and their goal is to find the target node. The network is unmarked by any pheromones in the beginning and so the ants start exploring the network in a random fashion. The probability of an ant taking a certain edge in the network is calculated as a transition probability and the ant randomly chooses between the neighbors of a node. The probability of an ant taking the edge ij at time-step t is calculated as follows:

$$P_{ij}^{(t)} = \frac{(\tau_{ij}^{\alpha})(\eta_{ij}^{\beta})}{\sum_{j}(\tau_{ij}^{\alpha})(\eta_{ij}^{\beta})} \tag{2.1}$$

Where,

τ_{ij} = amount of existing pheromone on edge ij at time-step t

α = pheromone perception

η_{ij} = fiber density of edge ij, i.e., the underlying brain structural connectivity matrix

β = edge perception

Note that exponents α and β characterize the individual perception of the ants. In particular, α represents the sensitivity to follow edges previously used by other ants, whereas β represents how influential is the strength of the edge. Both exponents maybe sublinear, linear (equal to 1), and supralinear, and hence reflect the perception of the ants with respect to the underlying topology (SC network) and to the collaborative architecture through pheromones. Hence we will refer to α as *pheromone (ph) perception* and to β as *edge perception*.

In *Equation 2.1*, we are essentially calculating the transition probability of an ant going from node i to node j when accounting for two dimensions or layers of information (i.e.

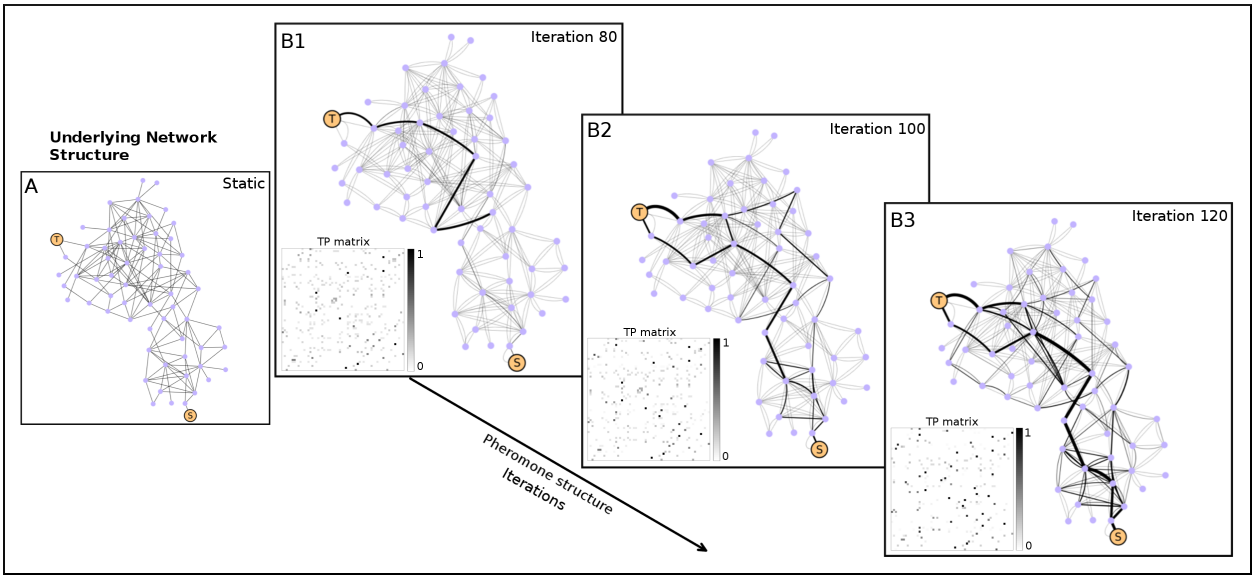


Figure 2.1. Ants' collective collaborative behaviour exhibited on the undirected Dolphin social network with 62 nodes. As ants start finding the target node, they deposit pheromones on some of the edges. Thus, the pheromone structure changes at every iteration. Shown here is the pheromone structure at three different iterations, along with the changing transition probability (TP) matrix.

Table 2.1. Effect of the different configurations of *pheromone* and *edge perception* on the behaviour of the ant colony

Pheromone Perception	**Edge Perception**	**Ant Colony Behaviour**	**Communication Regime**
$\alpha > 1$	$\beta > 1$	Highly collaborative and communication through most prominent edges only	Enforcing the use of *main roads*
$\alpha < 1$	$\beta > 1$	No collaborative and communication through most prominent edges only	Weakly-coupled random walkers with preference for *main roads*
$\alpha > 1$	$\beta < 1$	Highly collaborative and extensive use of network	Enforcing the use of *side roads*
$\alpha < 1$	$\beta < 1$	Weak collaboration and extensive use of network	Weakly-coupled random walkers with almost no preference for roads
$\alpha = 0$	$\beta = 0$	Not collaborative and extensive use of network	Independent random walkers with no preference for roads

structural topology and pheromones). Thus, the denominator is a normalizing factor in order to keep the row-wise sum in the transition probability (TP) matrix equal to 1. As can be seen in *Equation 2.1*, the calculation of the transition probability also involves the amount of existing pheromone on the edge *ij*, which changes at every iteration. Hence, the TP matrix is dynamic across time. Thus, we cannot resort to analytical approaches, such as mean first passage time (MFPT) [158], in order to study this changing topology.

While running the simulations, the exponents in *Equation 2.1*, *pheromone perception* and *edge perception*, are the variables that can be controlled within a certain range. Each separate combination of these variables characterizes a different collaborative spreading regime or configuration (expressed in terms of the two parameters) that can be associated with different task and resting state FCs. As η is the underlying structural connectivity matrix in terms of fiber density, all the values in this matrix are lower than 1. Thus, when the value of *edge perception* is below 1, the combined effect is to give more importance to all the values in the η matrix. On the other hand, when the value of *edge perception* is over 1, the combined effect is to shrink the values in η even further, but the smaller values shrink more than the higher ones. For example, the effect of *edge perception* $= 2$ on $\eta_{ij} = 0.05$ and 0.5 is to make

them 0.0025 and 0.25 respectively. Thus, the effect is similar to almost completely wiping out the weak connections in the structural connectivity matrix. The interaction between the pheromone matrix, τ, and its exponent *pheromone perception* is different as τ has entries that are above and below 1. The interactions between the *ph* and *edge* perception is summarized in *Table 2.1*. As the underlying structural connectivity matrix, η, remains static throughout the simulation and the pheromone matrix changes at every iteration, it can be thought of as a *double-layered* network structure, where the interaction between the two layers of the network is regulated by *Equation 2.1*. Even though the pheromone matrix changes at every step of the simulation, the way it changes depends on the SC, as the ants only take the existent edges. Thus, only the edges existent in the SC appear in the pheromone matrix.

The global behavior of the ants is a result of the individual dynamics. Thus, the simulation is set up in such a way that the ants are always in one of the three different states as follows:

- **Explorer ant:** The ants all start at the source node acting as random walkers (*explorer ants*) and do not deposit any pheromone on the edges. Hence, for each explorer ant, the step at time t will be determined by the transition probability matrix at that time (see *Equation 2.1*). The ants remain in this state until they find the target node.

- **Ant at target:** When an explorer ant reaches the target, it becomes an *ant at target*. The ants are in this state only for one time step. At this step, the label of the ant changes from *explorer ant* to *homebound ant*.

- **Homebound ant:** The ants are in this state when they are coming back to the source node after visiting the target. In order to do so, they trace the same path back to the source that they took to get to the target. On their way, they deposit pheromones on the path as a signal to the other *explorer ants*. This effectively increases the probability of other ants taking the same edges to reach the target in future. The amount of pheromone deposited by ants on their return journey is inversely proportional to the length of the path taken. This pheromone update mechanism rewards the shorter paths over longer ones. When the *homebound ant* makes its way back to the source node, it again starts a new journey as an *explorer ant*.

As the algorithm is designed so that all the ants move one edge at every step of the simulation, it can provide us with the number of ants that have made it to the target at least once at every step.

In cases where no pheromone is deposited by the ants on the edges of the network, for any value of the *edge perception* (β), the system behaves as a set of random walkers independently exploring the underlying SC. Such SC is in our case a representation of a complex network. This is not necessarily the case, and more simple models such as two dimensional lattices (representing a landscape) can be used as well. This would be equivalent to a coarse-grain scenario of the exploration of a two-dimensional open space, as widely assessed in foraging theory [128].

In order to demonstrate the mechanism by which the ant colony algorithm cooperatively learns the topology over time, we executed it on a small toy network. *Figure 2.1* shows the ant colony's collective behaviour exhibited on the undirected dolphin social network with 62 nodes [159] for one source-target pair. As can be seen, the pheromone structure evolves with iterations as more and more ants reach the target and multiple paths emerge between the node pair. Note that the running time of an ant colony algorithm with n ants is $O(n^2 m.log(n)/\rho)$, for networks with n nodes and m edges, where ρ is the evaporation rate [160].

2.2.3 Ant Colony Simulations

One run of the simulation consists of running the ant colony algorithm for every source-target pair in a 164-region parcellation for the structural connectivity (SC) of the group average of 100 unrelated HCP subjects. The algorithm runs in discrete time steps, i.e., the ants move one edge at a time. Before starting a simulation run, the following parameters can be controlled, along with the range of values that have been explored for each of them:

- α = Pheromone perception = [0.01,0.05,0.1,0.5:0.5:4]
- β = Edge perception = [0.1,0.5:0.5:4]
- Amount of pheromone deposited by each ant = $1/Length_{path}$

- Number of ants in the colony = 200
- Number of simulation steps = 1000
- Number of simulation runs per configuration = 10

The simulation does not necessarily run for 1000 steps for every source-target pair as it stops when at least 95% of the ants have made it to the target. This termination condition is added in order to optimize the running time of a full simulation run.

At every step of the simulation, the number of ants that have made it to the target at least once is saved, along with the paths that each individual ant took to reach there. Due to the effect of pheromones, convergence of ants for each source and target is typically observed on multiple paths, although a path is discarded at the end of the simulation if it is used less than 10 times. The remaining paths are saved in a path ensemble, along with the number of ants that have taken each path to reach the target. The data regarding the different paths taken by the ants is saved in order to study the backbone and centrality of the brain structural connectivity in terms of effective path length and arrival rate (measurements explained in Section 2.4) of the target from the source and edge centrality that drives the spreading of ants in the network.

One important factor that should be remembered here is that the pheromone structure is not updated at every iteration in the entire network until all the ants have finished their move. Thus, an ant does not see another ants' freshly deposited pheromones until everyone has finished taking one step. This is to ensure that only the pheromones existing before the start of the iteration affect an ant's decision to take a certain edge.

2.2.4 Network Analysis

The data generated through the simulations contains very valuable information about the system, the evolving communicability, and the effects of collaborative spreading. Such information cannot be summarized by means of static measurements on the shortest-path, or even on a fixed set of paths. Hence, we developed two network-based measurements that account for the collective behavior of the ant colony as a single entity. One of the most

important outputs of the ant colony simulations is the use of different paths by different number of ants to reach the target. The different paths taken by the ants between each source-target pair are kept track of, along with the number of times those paths are used. Thus, for each source-target pair, we may isolate the subnetwork within the underlying SC matrix that only includes edges belonging to the ensemble of paths used by the ants under each configuration.

One such network measure that we have defined is *effective path length* (EPL), which is defined as the sum of the length of each path multiplied by the number of ants taking that path as a fraction of the total number of ants that have ever made it to the target. This can be represented mathematically as:

$$EPL_{\text{ij}} = \frac{\sum_{p=1}^{n_p} (L_p \times Traffic_p)}{\sum_{p=1}^{n_p} Traffic_p} \tag{2.2}$$

Where,

ij = source-target pair for which effective path length is being calculated

n_p = number of different paths taken by the ants

L_p = length of path p based on the fiber density

$Traffic_p$ = number of times the ants took path p to reach the target

Note that *Equation 2.2* is normalized by dividing by the sum of all the ants that have reached the target by taking the saved paths. Thus, the contribution of paths taken less than 10 times by the ants to the EPL would be negligibly small. Hence, ignoring those paths does not significantly affect the calculation of EPL. A high EPL reflects that communication through the path ensemble involves longer paths, whereas a low EPL suggests the involvement of shorter paths.

Another network measurement that we have defined is *arrival rate* (AR). For every source-target pair, this is the ratio of the number of arrivals to the maximum number of arrivals that could have taken place (defined as an ant exclusively taking the shortest path back and forth between source and target). Mathematically, this can be shown as:

$$AR_{\text{ij}} = log_{10}\left(\frac{2 \times Arrivals_{\text{ij}} \times SPL_{\text{ij}}}{numAnts \times (iter_{arrival} + SPL_{\text{ij}})}\right) \tag{2.3}$$

Where,

$Arrivals_{\text{ij}}$ = number of times ants have successfully reached target j from source i by using any path in the path ensemble

$numAnts$ = number of ants used in the simulation = 200

SPL_{ij} = number of edges in the shortest path length between i and j

$iter_{arrival}$ = iteration number when at least 95% of the ants reach the target

AR represents the log_{10} transformation of the percentage of arrivals (hence, ranging between 0 and 1) to the target with respect to the maximum number of arrivals that could possibly happen as bounded by communication solely happening through shortest paths. A high value of AR represents that the communication between a node pair is efficient, whereas a low value indicates that it is inefficient with few arrivals to the target.

2.2.5 Null Models Based on Structural Connectivity

We tested two different null models based on randomizations of SC [161]. In order to do so, we evaluated the FC predictive power of the ant colony-derived measurements on two different randomized topologies. Note that the randomization procedure used for both null models have density, degree-distribution, and degree-sequence as topological invariants.

- **Whole-brain Randomization:**

  The entire SC network was randomized using the *xswap* method [161] [162]. This iterative procedure was performed until full randomization was achieved. This was evaluated by measuring dissimilarity between the original and the increasingly randomized network until a plateau was reached. Dissimilarity reflects the percentage of entries that are different between a network and a reference network. Note that upper boundaries of dissimilarity are dependent on the density of the network, and hence a perfect dissimilarity (value of 1) may only be reached in networks where density is 50%. *Figure 2.2.A* and *B* show dissimilarity as a function of number of xswaps and the randomized network, respectively. *Figure 2.2.A* shows that dissimilarity reaches a plateau

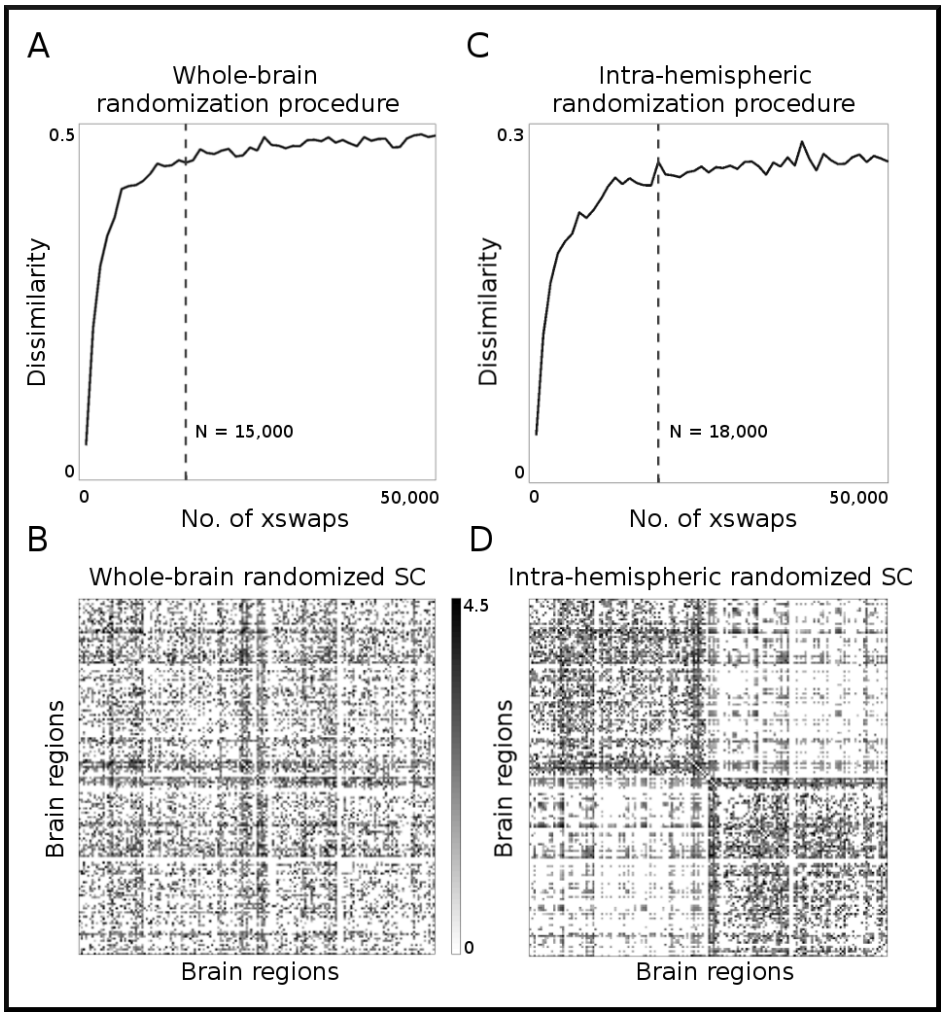


Figure 2.2. Iterative randomization procedure on SC network. Dissimilarity as a function of the number of xswaps for whole-brain (A) and intra-hemispheric randomizations (C). Adjacency matrices for final randomized networks are shown for whole-brain (B) and intra-hemispheric (D).

around 0.5 after 15,000 xswaps. Hence, we chose this configuration for the analysis of the null model based on whole-brain randomization. This whole-brain randomized network will henceforth be referred to as SC^{rand}_{whole}.

- **Intra-hemispheric Randomization:**

  The second null model used introduced another invariant by always preserving the inter-hemispheric connections, i.e., neither deleting any existing nor adding any new edges when performing xswaps. Analogously to the procedure performed in the first null model, it was found that dissimilarity reached a plateau around 0.3 after 18,000 xswaps. *Figure 2.2.C* and *D* are the dissimilarity as a function of the number of xswaps and the randomized network, respectively. This intra-hemispheric randomized network will henceforth be referred to as SC^{rand}_{intra}.

We run the simulations as explained in *Section 2.2.3* on the two null models explained above. The same network analysis as in *Section 2.2.4* was conducted out on the data generated by these simulations. The *Results* section discusses the results of these null model-based simulations in detail.

2.3 Results

As described earlier, 100 unrelated subjects in the HCP [137][138] dataset were used to construct the group average structural and functional connectomes. The ant colony simulations were run on the group average structural connectome. The data that is saved from these simulations consists of the different paths taken by the ants between each node pair, along with the number of ants taking each of the paths. In order to characterize different aspects of communication for different α-β configurations, we calculated EPL and AR for each source-target pair (see *Section 2.2.4* for details). These two measures were then associated, individually and together (in a multilinear regression), with the group average task-based and resting-state functional connectivity patterns, estimated as per the protocol described in *Section 2.2.1*. This section reports the results obtained from the path ensembles and the associations with the FCs.

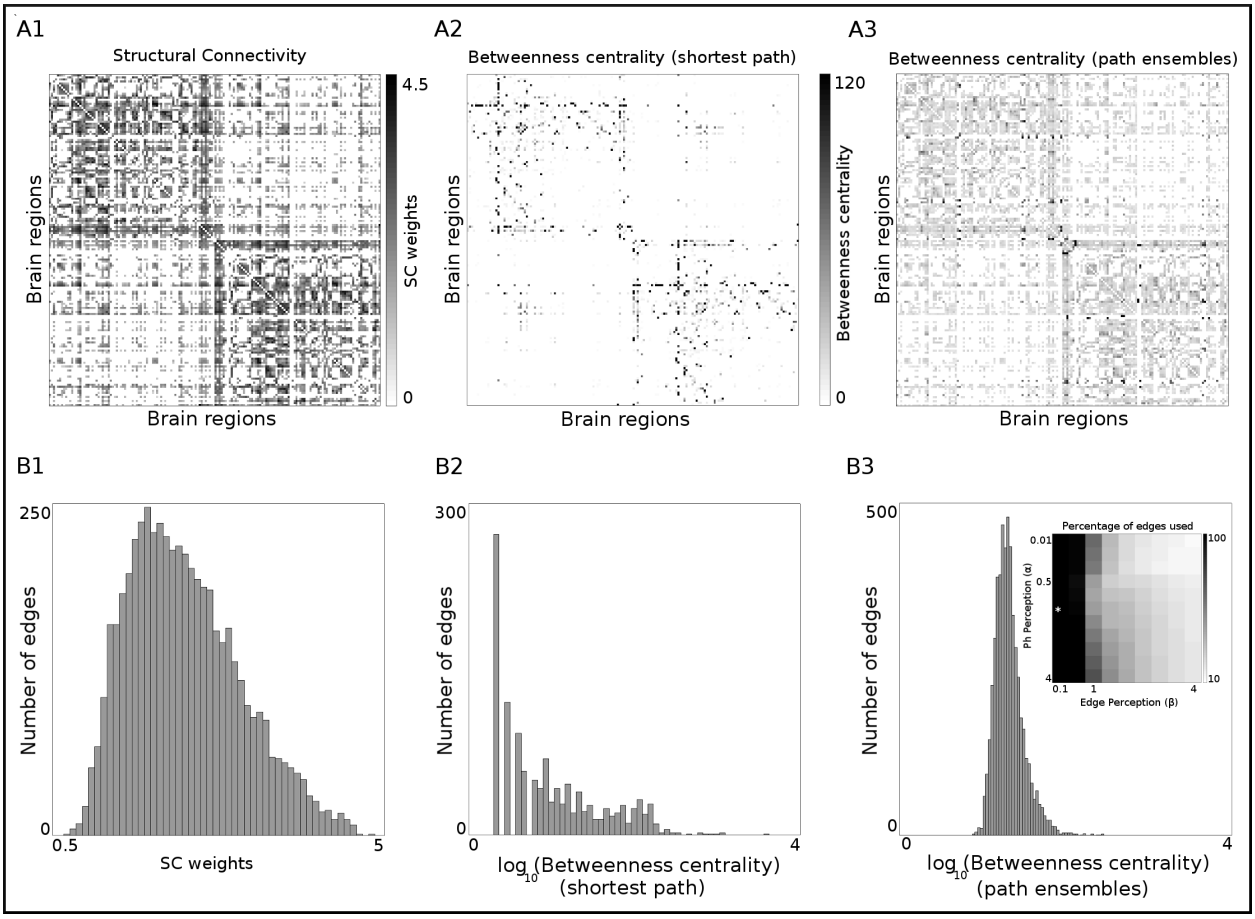


Figure 2.3. *A1* is the group average weighted structural connectivity (SC). *A2* is the edgewise shortest-path betweenness centrality on SC. Note that 23% of the edges participate in at least one shortest path. *A3* is the equivalent of the shortest-path betweenness centrality in the path ensembles obtained through the ant colony algorithm (results correspond to the configuration $\alpha = 1.5$ and $\beta = 0.1$). 100% of the edges participate in at least one path ensemble. The *B1, B2*, and *B3* are the corresponding histograms for each measurement. The plot inset inside *B3* shows the percentage of edges used for each α-β configuration. The * indicates the configuration to which *A3* and *B3 belong.*

2.3.1 Evaluation of Path Ensembles and Betweenness Centrality

As discussed in *Section 2.2*, running the ant colony inspired algorithm allows us to identify the ensemble of paths most widely used by the ants for each source-target pair in the brain structural network. *Figure 2.3.A1* shows the group-average weighted structural connectome used in this study whereas *Figure 2.3.B1* shows the distribution of the non-zero weights. *Figure 2.3.A2* shows the edgewise betweenness centrality as measured on the shortest paths

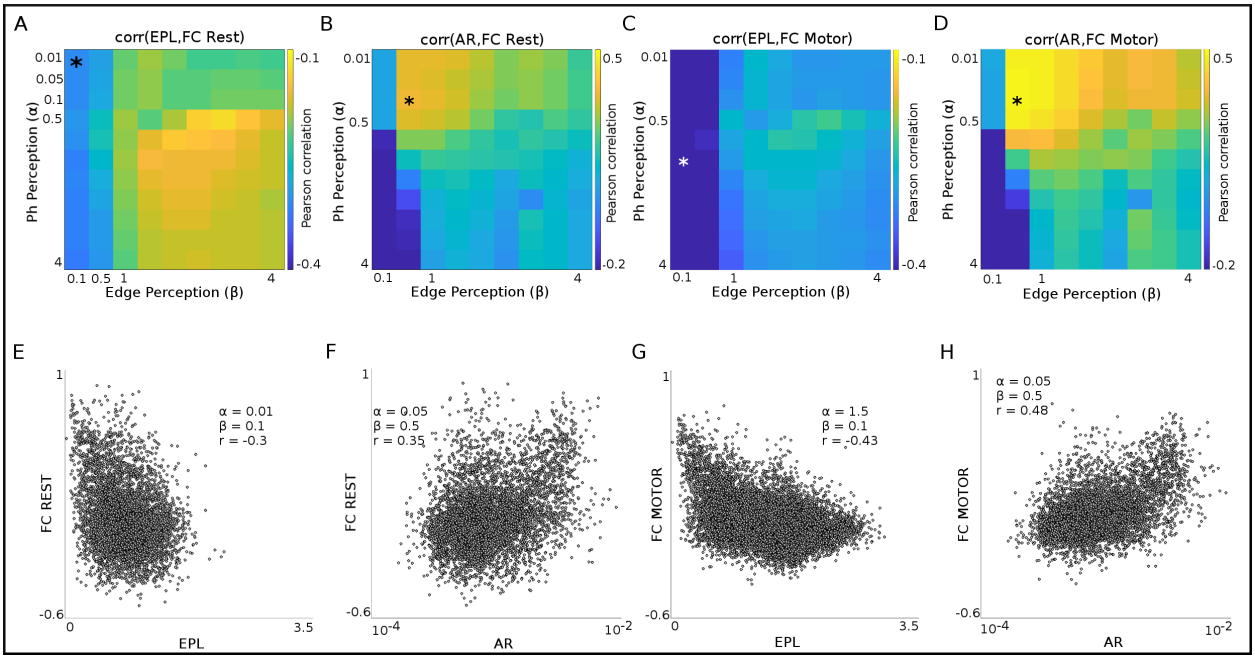


Figure 2.4. For every configuration of the ant colony, Effective Path Length and Arrival Rate are calculated for every source-target pair. The Pearson correlations between these measures and different task-based and resting state functional connectivities are calculated. *A* and *B* show the correlations of resting state FC with EPL and AR, while *C* and *D* show these correlations with Motor FC. The * in each of these matrices shows the configuration for which the correlation is highest. *E, F, G*, and *H* are the scatter plots between the FCs and EPL and AR for the configuration with the highest correlation.

and *Figure 2.3.B2* shows the distribution of such centrality. Different configurations of α-β yield potentially different path ensembles. As an example of it, we show in *Figure 2.3.A3* the pairwise path ensemble centrality (as obtained from the simulations) for $\alpha = 2$, $\beta = 0.1$, as well as its distribution (*Figure 2.3.B3*). *Figures 2.3.A2* and *A3* show that only 23% of the edges participate in shortest paths, whereas 100% of the edges participate in the path ensembles for this particular collaborative regime. The inset in *Figure 2.3.B3* shows the percentage of edges used for every α-β configuration.

2.3.2 Associations Between Functional Connectivity and Path Ensemble-Derived Measures

The two network measures defined in *Section 2.2.3*, effective path length (EPL) and arrival rate (AR), were calculated for every source-target pair in every configuration of

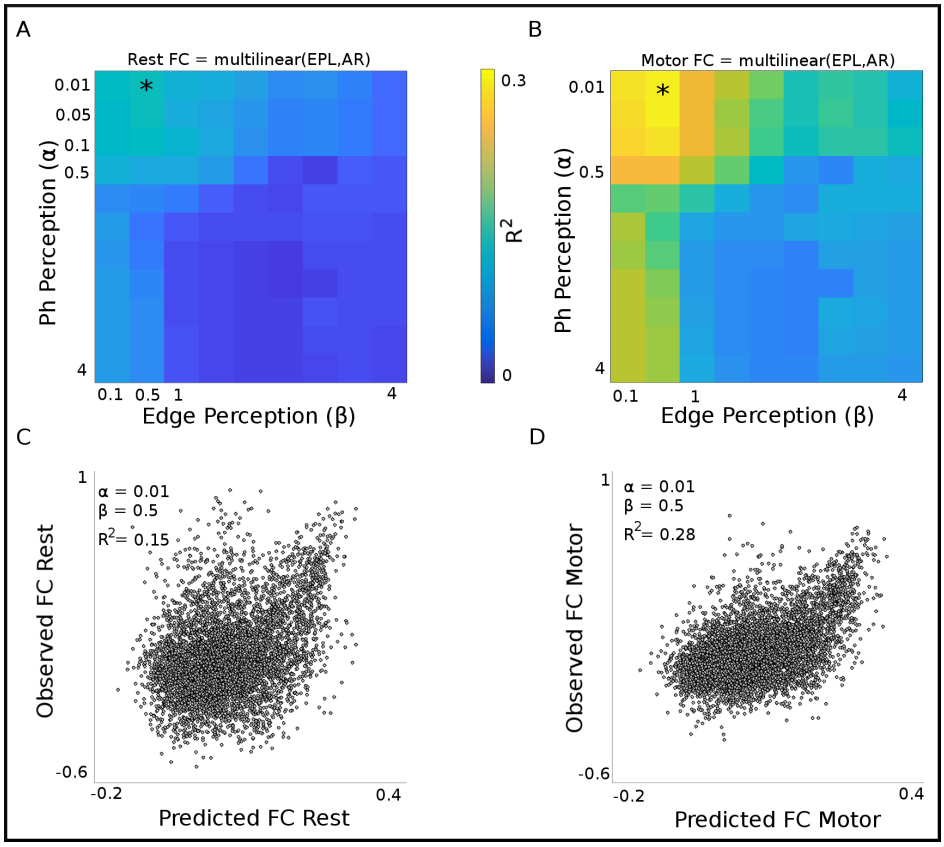


Figure 2.5. For every configuration of the ant colony, multilinear regression is carried out with EPL and AR as the predictor variables and the different task-based and resting state FCs as the predicted variable. *A* and *B* show the R^2 values of the regression for the different configurations, while the * highlights the configuration for which the R^2 is highest. *C* and *D* are the scatter plots between the predicted and observed FCs for resting state and Motor task respectively.

Table 2.2. Pearson correlation coefficients between path length measurements on SC and task-based FCs. Path length measurements were obtained from SC, and from two null models derived from SC, namely SC^{rand}_{intra}, and SC^{rand}_{whole}. EPL is the effective path length, calculated from the collaborative behaviour of the ant colony, whereas SPL is the shortest path length on SC (as a baseline for EPL)

	Path Length Measurements			
Task	**SC (EPL)**	**SC^{rand}_{intra} (EPL)**	**SC^{rand}_{whole} (EPL)**	**SC (SPL)**
REST	-0.31	-0.07	0.03	-0.14
LANGUAGE	-0.41	-0.16	0.02	-0.27
EMOTION	-0.40	-0.15	0.01	-0.28
GAMBLING	-0.39	-0.15	0.01	-0.27
MOTOR	-0.45	-0.20	-0.02	-0.33
RELATIONAL	-0.37	-0.15	0.01	-0.27
SOCIAL	-0.37	-0.13	0.01	-0.24
WM	-0.42	-0.17	0.01	-0.31

ph and *edge perception*. Then, EPL and AR for each configuration were correlated with each task-based and resting state functional connectivity (FCs). *Figure 2.4.A-D* show the correlation values of EPL and AR with resting state and motor-task FC for all configurations of *ph* (α) and *edge perception* (β). *Figure 2.4.E-H* illustrate the scatter plots of EPL and AR with the resting state and motor-task FCs corresponding to the configurations with highest correlations in *Figure 2.4.A-D*. Note that for resting state FC, the highest correlation with EPL is achieved at ($\alpha = 0.01$, $\beta = 0.1$) and with AR at ($\alpha = 0.05$, $\beta = 1$). Analogously, for motor-task FC, the highest correlations are attained at ($\alpha = 1.5$, $\beta = 0.1$) and ($\alpha = 0.5$, $\beta = 0.5$) with EPL and AR respectively. In order to test the variation in EPL and AR values across different simulation runs, we have calculated the node pair-wise coefficient of variation values of these measures as shown in *Figure A.1*.

Tables 2.2 shows the Pearson correlation coefficients between the path length measurements and task-based FCs. The path length measurements were calculated on SC and two null models, SC^{rand}_{intra} (intra-hemispheric randomization) and SC^{rand}_{whole} (whole-brain randomization, see *Section 2.2.5* for details). The table also reports the associations between FCs and shortest path length (SPL) as a baseline for EPL. Note that, as expected, EPL calculated

Table 2.3. Pearson correlation coefficients between arrival measurements on SC and task-based FCs. Arrival measurements were obtained from SC and from two null models derived from SC, namely SC^{rand}_{intra}, and SC^{rand}_{whole}. AR is the arrival rate calculated from the collaborative behaviour of the ant colony, whereas MF is the maximum feasible flow between node pairs (baseline for AR).

	Arrival Measurements			
Task	**SC (AR)**	**SC^{rand}_{intra} (AR)**	**SC^{rand}_{whole} (AR)**	**SC (MF)**
REST	0.35	0.13	-0.03	-0.0043
LANGUAGE	0.43	0.17	-0.03	0.0123
EMOTION	0.43	0.16	-0.01	0.0109
GAMBLING	0.42	0.16	0.02	0.0054
MOTOR	0.48	0.20	-0.02	0.0290
RELATIONAL	0.42	0.16	0.01	-8×10^{-5}
SOCIAL	0.39	0.15	-0.01	-0.0011
WM	0.45	0.18	-0.02	0.0078

Table 2.4. Multi-linear models using path-length and arrival measurements as predictors of FC for different tasks. Values indicate explained variance. SC and two subsequent null models, SC^{rand}_{intra}, and SC^{rand}_{whole} are evaluated. SPL and MF are evaluated on SC as baseline models for EPL and AR respectively.

	Combined Predictors			
Task	**SC (EPL,AR)**	**SC^{rand}_{intra} (EPL,AR)**	**SC^{rand}_{whole} (EPL,AR)**	**SC (SPL,MF)**
REST	0.14	0.02	0.0001	0.03
LANGUAGE	0.22	0.04	0.0001	0.10
EMOTION	0.21	0.04	0.0004	0.11
GAMBLING	0.19	0.03	0.0004	0.11
MOTOR	0.28	0.05	0.0010	0.14
RELATIONAL	0.18	0.03	0.0003	0.11
SOCIAL	0.17	0.03	0.0005	0.08
WM	0.23	0.04	0.0007	0.13

on SC is negatively correlated with all task and resting-state FCs. This indicates that the longer the EPL for a node pair, the less functionally coupled it is. Also, for all tasks and resting-state FCs, the EPL computed on SC outperforms EPL on the two null models as well as SPL on SC.

Analogously, *Tables 2.3* shows the Pearson correlation coefficients between the arrival measurements and task-based FCs. The arrival measurements were calculated on SC and two null models, SC^{rand}_{intra} (intra-hemispheric randomization) and SC^{rand}_{whole} (whole-brain randomization, see *Section 2.5* for details). The table also details the associations between FCs and max flow (MF) [163] as a baseline for AR. Note that, as expected, AR calculated on SC is positively correlated with all task FCs. This indicates that the higher the AR for a node pair, the higher the functional coupling between them. Also, for all tasks and resting-state FCs, the AR computed on SC outperforms AR on the two null models as well as MF on SC.

In order to evaluate the joint predictive capacity of EPL and AR, we conducted multilinear regression analysis for every α-β configuration. *Figure 2.5.A* and *B* show the explained variance (R^2) for all configurations for the resting-state and motor-task FCs respectively. *Figure 2.5.C* and *D* are the scatter plots between the predicted and observed FCs for resting-state and motor-task respectively for the optimal configurations. *Table 2.4* summarizes the R^2 values for all FCs when the underlying topology is SC, SC^{rand}_{intra}, and SC^{rand}_{whole}. The predictive capacity of the EPL and AR when using the SC null models is negligible. Also observe that EPL and AR calculated on SC outperform the baseline predictors, SPL and MF.

2.3.3 Associations within Functional Networks

Lastly, we further tested the joint predictive capabilities of EPL and AR within the 7 brain functional networks (FN) as defined by Yeo et al. [76]. This was achieved by carrying out multilinear regression analysis across all α-β configurations for node pairs within the 7 FNs (i.e., visual, somatomotor, dorsal attention, ventral attention, limbic, frontoparietal, and default-mode network). For each FN-FC combination, the highest explained variance (R^2) and its corresponding α-β configuration was saved. *Figure 2.6.A* shows maximum R^2 values reached for each FC-FN combination, while *Figure 2.6.B* shows the corresponding α and β values. We further used these α and β values in order to summarize the configuration

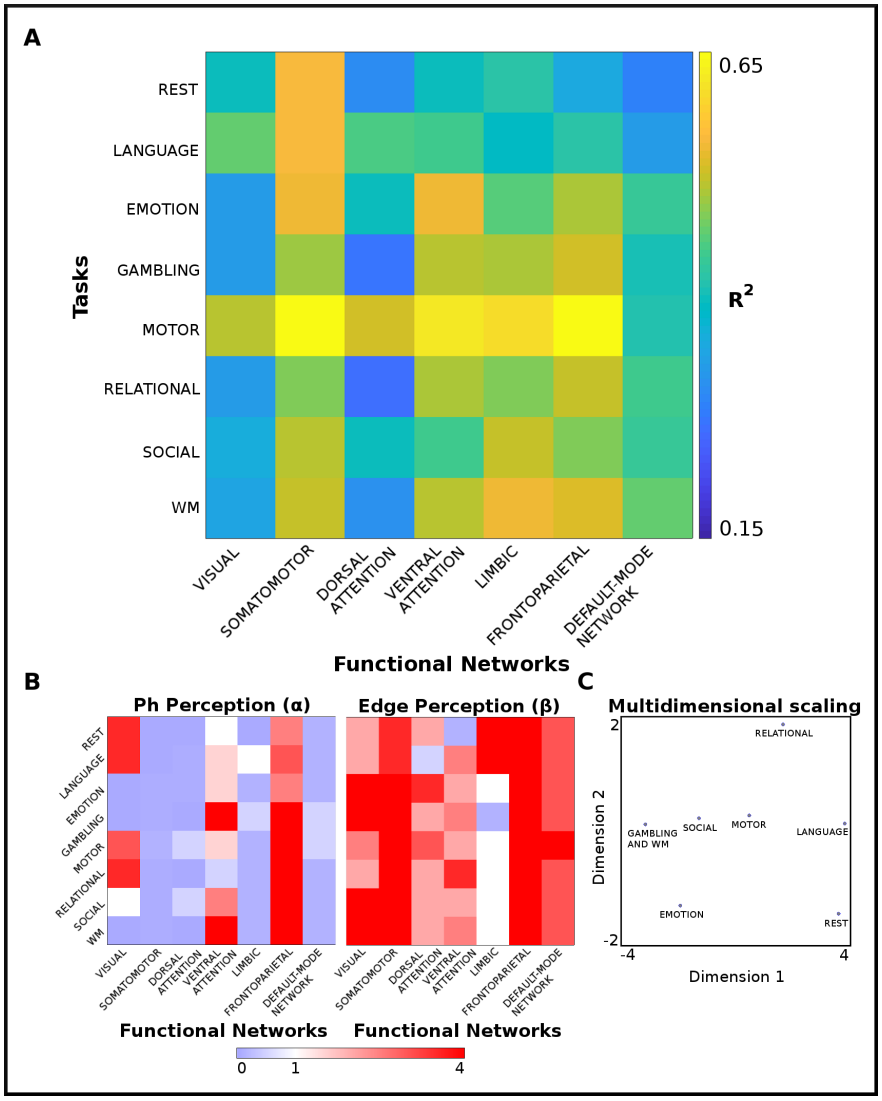


Figure 2.6. *A.* Optimal configurations of ant-colony parameters, *ph* and *edge perception*. Each entry denotes the optimal configuration (highest R^2) for a task and a functional network (FN). Note that for each FN, only the node pairs involving brain regions of that FN are considered. *B.* shows the values of *pheromone* and *edge perception* associated with the optimal R^2 for each task-FN combination. *C.* shows the dimensionality reduction of the task states to two dimensions. It can be seen that, except for *Gambling* and *WM*, all the other task FCs occupy unique places in the reduced dimension space.

distances amongst all tasks into a two-dimensional space. This was done by using the pairwise euclidean distances between the vectors (corresponding rows of the matrices in *Figure 2.6.B*) as the input for multidimensional scaling (MATLAB command `mdscale`). *Figure 2.6.C* shows the result of this dimensionality reduction. It can be seen that, except for *Gambling* and *WM*, all the other FCs occupy unique spaces inside this reduced two-dimensional space.

Observe that Motor-task has the highest R^2 within frontoparietal FN, followed closely by somatomotor FN. Even though these R^2 values are very close, the big difference between them is the α-β configuration for which they are observed. Note that both α and β are above 1 for frontoparietal region, while for somatomotor the configuration is α below 1 and β above 1. Referring to *Table 2.1*, we can see that a highly collaborative regime and communication through the most prominent edges is essential for motor task within frontoparietal FN. On the other hand, within somatomotor FN, the ant colony regime is not collaborative but the communication takes place through the prominent edges in the network. This suggests that, for a single task, the communication within brain regions imitates different regimes of the ant colony algorithm within different FNs. Similar phenomena can be observed for other FC-FN combinations as well.

2.4 Discussion

There have been several studies in the recent past focused on a better understanding of the communication mechanisms of the human brain [122]. The present paper delves into this topic by proposing a framework inspired from the collaborative foraging behaviour of a colony of ants in order to simulate communication as a collaborative spreading phenomenon on top of an underlaying complex network. This framework allowed characterization of source-target communications, not as a single estimate through a single (shortest) path but through path ensembles whose identification is sensitive to the pheromone-based activity of the ants.

We found important differences when looking at the betweenness centrality of the edges based on the path ensembles as compared to the shortest path-based centrality. As shown in *Figures 2.3.A2* and *A3*, path ensembles used by the ant colony involve many more of the structural edges than the small portion (23%) that participates in any shortest path. Indeed,

configurations with $\beta < 1$ involved all structural edges for any α value evaluated. *Figures 2.3.B1*, *B2*, and *B3* show the histograms of the values for each of the three corresponding plots. Also notice that the distribution of the path ensemble-based betweenness centrality displays a log-normal behaviour.

The plot inset inside *Figures 2.3.A2* shows the percentage of edges used for each of the collaborative regimes that have been explored. It can be seen that this percentage is significantly higher for low values of β, as it is the exponent that determines the edge weights. Higher values of β penalize the weaker edges and reward the stronger ones, thus incentivizing the ants to opt for paths with higher weights. When there is no collaboration, i.e. $\alpha = 0$ and $\beta = 1$, the ants act as pure (independent) random walkers. As such, they make their routing decisions based only on the edge weights and do no converge on any single path. This is evident in the fact that, for $\alpha = 0$ and $\beta = 1$, all path ensembles were empty at the end of the simulations, as only the paths used at least 10 times by the ants are considered.

From *Tables 2.2*, *2.3*, and *2.4* we can observe that the associations between the different task-based and resting state FCs and the two measures that we have defined, EPL and AR, are systematically higher when the algorithm is run on SC as compared to the two different null models and the baseline models in terms of SPL and MF. *Figure A.2* in *Appendix A* shows the different R^2 values for the multilinear regression models evaluated for every level of collaboration, i.e. α-β configuration, for the task FCs not represented in *Figure 2.5*. The panel also provides the scatter plot of predicted versus observed FCs for the optimal configurations. It can be seen from *Figures 2.5* and *A.1* that the associations are consistently higher for the task FCs as compared to the Rest FC. This might suggest that the ants' collective foraging behaviour is a better representation of the communication processes taking place in the brain when subjects are engaged in a task, as opposed to when the subjects are at rest.

Additionally, it was found that the associations are higher within certain functional networks for specific FCs (see *Section 2.3.3* for details). This result in particular might be an indication that when a person is performing a task, a specific set of brain regions is more active than the other parts of the brain and that this activity is simulated well by the ant colony algorithm. Further, the communication regimes used by the ants in terms of α and

β are different for different FC-FN combinations. This suggests that the communication dynamic within different FNs varies when a person is performing different tasks.

In the next step, we performed a multidimensional scaling (MATLAB command `mdscale`) on the results obtained in *Section 2.3.3*. The inputs to the multidimensional scaling algorithm were the optimal α and β values per task FC, i.e. the rows of the two matrices shown in *Figure 2.6.B*. The result of this scaling, shown in *Figure 2.6.C*, show that all the functional states occupy unique positions on the reduced two-dimensional space, with the exception of *Gambling* and *WM*. This suggests that the α and β values per functional state per functional network can be successfully used as predictors of the functional state.

An added value of the method proposed in this paper is that it allows for simple parametrization of any system between two layers, its structural and functional sides, using just two parameters, namely *pheromone perception* and *edge perception*. Another point to note is that this framework obtains, in a data-driven fashion based on simulations, the path ensembles representing the most important communication pathways between each source and each target, as opposed to fixing the number of paths [134] as a constant value for all path ensembles or fixing the number of steps [133]. In consequence, the presented framework allows for communication between different sources and targets to have different number of paths involved depending on the topology and the dynamics of the ant-colony as defined by *pheromone perception* and *edge perception*. The impact of these two exponents on the behavior of the ant colony and on the communication regime is summarized in *Table 2.1*. The presented framework allows for very different communication regimes occurring under the same topology, from independent random walkers that perceive the network as binary to highly collaborative walkers biased towards using either the *main roads* or the *side roads*.

The ant colony-inspired algorithm presented here as a dynamical model on top of a network topology is a framework that may be used as a testbed for evaluating different communication scenarios. Previously, Mišić et al. [164] have used a cascade spreading model to study the network properties of the human brain that facilitates spreading of signal through any possible path. In order to do so, they activated two or more *seed nodes* in the SC and studied how the perturbation spreads through the network in a collaborative and also in a competitive manner. The primary difference between our approach to modeling

communication in the brain as opposed to Mišić et al.'s approach is the use of pheromones as a means of indirect communication between the ants. Another difference in the two approaches is that while Mišić et al. do not allow a node that has already received a signal once to be perturbed by the same or another signal again, the ant colony simulations allow for multiple uses of the same paths. Indeed, this is how the ants strengthen certain paths more than the others in their search for the target node.

This study has several limitations, most of them related to the computational power and time required to run the ant colony simulations on the brain network (structural connectivity). This has restricted the simulations in different manners, including the number of simulation runs per configuration, size of the ant colony population, values of pheromone amount explored, and the brain parcellation used. It has also prevented the use of individual connectomes, hence preventing us from assessing inter-subject differences that should be explored in future work. All these factors have ultimately hampered the exploration of the solution space in a more thorough manner. In order to study the stability of results, future work shall be focused on exploring the effect of parameters such as number of ants and values of pheromone amount over a more constricted range of α-β exponents. Additionally, a larger number of simulations per configuration should be performed to assess the stability of the solutions (i.e. the path ensembles obtained). Even with all the limitations described above, it can be seen from the coefficient of variation values and their distributions for EPL and AR (see *Figure A.2* for an example) that the results did not vary extensively from simulation to simulation for the same α-β configuration.

Future work shall be focused on exploring the inter-subject differences in connectomes through the algorithm proposed in this paper. As this paper used data from healthy subjects, another avenue is to study the behaviour of the ant colony on the connectomes of patients of neurodegenerative disorders, such as Alzheimer's or multiple sclerosis. As this framework is source-target oriented, it could also be linked with experiments where the concept of a source is very well-defined, such as Transcranial Magnetic Stimulation (TMS) [165][166].

The framework presented here combines a complex network topology tested by an ant-colony algorithm that, by means of two perception exponents, namely *ph* and *edge perception*, allows to simulate different communication regimes and to capture the most important path

ensembles involved on the communication of each pair of source and target nodes. This framework has shown evidence of being able to establish associations between SC and FC when subjects are in different cognitive states as they are performing different tasks. This methodology allows for compression of the communicability happening to a reduced two-parameter space. We have presented important foundations on how these parameters mimic different communication regimes that might better explain different functional states.

3. FUNCTIONAL CONNECTOME FINGERPRINT GRADIENTS IN YOUNG ADULTS

3.1 Introduction

With the advent of improved neuroimaging acquisition techniques, there has been a surge in the availability of high quality neuroimaging data in recent years. Data repositories such as the different Human Connectome Project (HCP) [138] datasets (HCP Young Adult [138], HCP Aging [167], HCP Development [168], etc.), 1000 Functional Connectomes Project (http://fcon_1000.projects.nitrc.org/), UK Biobank [169][170], and the Alzheimer's Disease Neuroimaging Initiative (ADNI) [171] among others are openly available to the scientific community. These data repositories, although highly valuable, do not always provide the users with ready-to-use processed subject-level whole-brain functional connectomes (FCs). Instead, they provide raw data or data that has been only minimally processed [47][172]. Hence, it is typically up to the researcher to estimate single session whole-brain functional connectomes from fMRI and T1 data. This step is critical [173][94][174][154][175][176] for subsequent brain connectivity and network neuroscience analyses [63][19][64]. This can be a difficult task due to the knowledge required as well as the amount of computational power necessary to process large datasets such as HCP.

These open-source datasets are usually shared with the community with either no or minimal artefact and/or noise removal. This is an efficient and suitable strategy for neuroimaging data sharing. One of the main reasons is that MRI data processing is constantly evolving, with registration, processing and denoising methods constantly evolving [173][94][174][154][175][176] as well as new brain atlases [120][48][177] being provided to the community. Providing raw or minimally processed datasets allows for up-to-date processing techniques to be applied to the dataset.

Amongst the many different choices one has to make while processing raw neuroimaging data to obtain subject-level, single-session, whole-brain functional connectomes, the choice of the parcellation is very important [120][48][177]. The subsequent analysis of the connectomes depends on the level of granularity of a parcellation. The importance of the parcellation granularity has been shown, for instance, when evaluating brain fingerprints [100][178]. Schaefer

et al., 2018 [120] recently published a scheme of parcellations that gives the user the ability to assess up to 10 different levels of granularity (atlases include 100 to 1,000 brain regions, in steps of 100). Another added advantage of this parcellation scheme is that all ten levels of granularity are further divisions of the resting state functional networks proposed by Yeo et al., 2011 [76].

There is no standard procedure to decide on the brain parcellation to estimate functional connectomes. This is also true for the artefact and noise removal steps in the fMRI data processing. To that end, the amount of subject fingerprint present in the resultant functional connectomes is a useful proxy, as a whole, of the measure of quality of the experimental design, acquisition parameters, and the ultimate estimation of the functional connectomes. Many recent studies have established that functional connectomes have an individual fingerprint that can be used to identify an individual from a population (a process known as *fingerprinting* or *subject-identification*) [98][99][100][101][102][103][104][105][99][106][107]. Subject-level fingerprints in the FCs have been found to be reliable and reproducible in high quality datasets [100] (e.g., HCP). Moreover, this fingerprint can be improved by using the differential identifiability framework (If), which relies on performing group-level decomposition into principal components followed by an iterative reconstruction adding components in descending order of explained variance until the differential identifiability score reaches an optimal value [98]. Without a high subject-level fingerprint, brain connectomic analyses that are aimed at finding associations between functional connectivity and cognition, behavior, or disease progression are severely compromised [109][110].

Fingerprints are not unique to test/retest sessions of the same individuals. Subjects sharing genetics and/or environment are expected to have, to some extent, a fingerprint. In particular, similar to subject-level fingerprint in brain functional connectomes, it has been established that a fingerprint also exists in the FCs of monozygotic (MZ) and dizygotic (DZ) twin subjects [112][113][114][115][116][117], albeit to a lower extent than the subject-level fingerprint. Kumar et al. 2018 [114] have presented a framework based on manifold approximation for generating brain fingerprints from multimodal data using T1/T2-weighted MRI, diffusion MRI, and resting-state fMRI. Their results show a link between amount of fingerprint and genetic proximity as the MZ twins have more prominent fingerprints than DZ

or non-twin siblings. Ge et al. 2017 [112] have used a linear mixed effects model to dissociate intra- and inter-subject variation of a phenotype and computed heritability with respect to stable inter-subject variation in fMRI data as the phenotype. Colclough et al. 2018 [116] have investigated the influence of genetics and common environment on functional connectomes of individuals obtained from fMRI and MEG data in HCP. Demeter et al. 2020 [117] have applied support vector machine classifiers on resting state fMRI to predict retest and co-twin pairs from two twin datasets (adult and pediatric) that include repeat scans. Gritsenko et al. 2020 [115] propose a pair-wise twin classification method to identify the zygosity of twin pairs using the resting state fMRI. The latest release of the HCP-YA [48] dataset includes unrelated subjects, as well as subjects that are related to each other, including MZ and DZ twins. This affords us the opportunity to assess brain connectivity fingerprints not only by comparing test and retest functional connectomes of the same subject (*subject-level fingerprint*), but also by comparing the functional connectomes of MZ and DZ twins (*twin fingerprint*) across different fMRI conditions.

The aim of this study is to provide state-of-the-art processed whole-brain, single-session FCs to the scientific community for conducting research in brain connectomics [64][63][179]. We provide FCs corresponding to all 10 levels of granularity (100 to 1,000) of the Schaefer parcellations. In terms of artefact/noise removal processing steps, we provide FCs at different level of processing/denoising (e.g. with and without global signal regression [180][181][182][183][184][185]). In addition, we assess the amount of subject-level fingerprint and twin-fingerprint (MZ and DZ) in each fMRI condition (resting-state and 7 tasks), at different levels of granularity of the Schaefer parcellations. We estimate and uncover these fingerprints using an extended version of the differential identifiability framework (If) [98][111].

3.2 Methods

3.2.1 The HCP-YA dataset

The functional MRI data processed as a part of this study is available in the Human Connectome Project-Young Adult (HCP-YA) repository [138]. The HCP-YA data consists

of behavioural and 3T MRI data from 1206 healthy young adult subjects collected between August 2012 and October 2015. 3T MR structural scans are available for 1113 subjects, out of which 889 subjects have fully complete data for all four 3T MRI modalities: structural (T1w and T2w) data, resting state fMRI, task fMRI, and high angular resolution diffusion MRI data. The HCP-YA dataset also has extensive family structures, including siblings and twin pairs (monozygotic and dizygotic). All the subjects are within the age range of 22-37 years at the time of scanning. *Table 3.1* summarizes the number of unrelated subjects and monozygotic and dizygotic twin pairs for resting state and all 7 tasks included in HCP-YA. The term *fMRI Condition* would be used to indicate both resting state and tasks that are included in the dataset.

Table 3.1. Summary of the number of unrelated subjects, MZ and DZ twin pairs corresponding to each of the fMRI conditions in the HCP-YA dataset

fMRI Condition	Unrelated subjects	MZ twin pairs	DZ twin pairs
REST1	435	131	76
REST2	435	131	76
EMOTION	416	124	70
GAMBLING	438	135	74
LANGUAGE	417	129	72
MOTOR	438	134	76
RELATIONAL	414	125	69
SOCIAL	417	128	71
WORKING MEMORY	436	133	77

HCP-YA fMRI conditions

We have used the fMRI data from the HCP-YA 1200 subjects release [137][138]. The fMRI resting-state data (HCP-YA filenames: rfMRI_REST1 and rfMRI_REST2) were acquired in separate sessions on two different days, with two different phase acquisitions (left to right or LR and right to left or RL) per day [137][138][47]. The seven fMRI tasks are the following: gambling (tfMRI_GAMBLING), relational (tfMRI_RELATIONAL), social (tfMRI_SOCIAL), working memory (tfMRI_working memory), motor (tfMRI_MOTOR), language (tfMRI_LANGUAGE, including both a story-listening and arithmetic task), and emotion (tfMRI_EMOTION). Two runs (LR and RL) were acquired for each task. Working

memory, gambling, and motor task were acquired on the first day, and the other tasks on the second day [47][151]. *Table 3.2* summarizes the run time and number of frames per condition:

Table 3.2. Summary of the number of runs, run time (in minutes and seconds), and number of frames per run for resting state and 7 tasks included in the HCP-YA dataset

fMRI Condition	#Runs	Run time (min:sec)	#Frames
REST1	2	14:33	1,200
REST2	2	14:33	1,200
EMOTION	2	2:16	176
GAMBLING	2	3:12	253
MOTOR	2	3:34	284
LANGUAGE	2	3:57	316
RELATIONAL	2	2:56	232
SOCIAL	2	3:27	274
WORKING MEMORY	2	5:01	405

The following is a brief description of each fMRI condition. This description relies on the documentation provided in the HCP S1200 Release. Minimum edits were done in order to preserve the description of the original experimental design of the fMRI protocols and acquisitions. More extensive information may be found in the HCP S1200 Release Reference Manual [47][46][48].

- **REST:** Resting state fMRI (rs-fMRI) data was acquired in four runs of approximately 15 minutes each, two runs in each session. The subjects were instructed to keep their eyes open with relaxed fixation on a projected bright cross-hair on a dark background presented in a darkened room. Within each session, oblique axial acquisitions alternated between phase encoding in a right-to-left (RL) direction in one run and phase encoding in a left-to-right (LR) direction in the other run. 1200 frames were obtained per run at 720 ms TR.

- **EMOTION:** This task was adapted from the one developed by Hariri et al. [186]. Subjects are shown blocks of trials that either ask them to decide which of two faces on the bottom of the screen match the face at the top, or which of two shapes at the bottom match the shape at the top of the screen. The faces have either an angry or

fearful expression. Trials are presented in blocks of 6 trials of the same task (face or shape), with the stimulus presented for 2000 ms and a 1000 ms ITI. Each block is preceded by a 3000 ms task cue ("shape" or "face"). Each of the two runs includes three face blocks and three shape blocks, with 8 seconds of fixation at the end of each run. In total, the emotion processing task fMRI had a run duration of 2:16 minutes per run, with 176 frames per run.

- **GAMBLING:** This task has been adapted from the one developed by Delgado and Fiez [187]. Subjects are asked to play a card guessing game wherein they are asked to guess the number on a mystery card in order to win or lose money. Subjects are told that potential card numbers are between 1 and 9 and to indicate whether they think the mystery card number is more or less than 5 by a button press. Feedback to the subject is the actual number on the card and either a green arrow up with "$1" for reward or red arrow down with "-$0.5" for loss. If the mystery number is equal to 5, the trial is considered neutral and a grey double headed arrow is shown. The subjects have 1500 ms to respond with button press, followed by 1000 ms of feedback. If the subject responds before the 1500 ms is over, a fixation cross is displayed. The task is presented in blocks of 8 trials that are either mostly reward or mostly loss. In each of the two runs, there are 2 mostly reward and 2 mostly loss blocks, interleaved with 4 fixation blocks (15 seconds each). In total, the gambling task fMRI had a run duration of 3:12 minutes per run, with 253 frames per run.

- **LANGUAGE:** This task was developed by Binder el al. [188] and uses the E-prime scripts provided by them. The task consists of two runs that each interleave four blocks of a story task and four blocks of a math task. The lengths of the blocks vary (average of approximately 30 seconds), but the task was designed so that the math task blocks match the length of the story task blocks, with some additional math trials at the end of the task to complete the 3.8 minute run as needed. The story blocks present participants with brief auditory stories (5-9 sentences) adapted from Aesop's fables, followed by a 2-alternative forced choice question that asks participants about the topic of the story. The math task also presents trials aurally and requires subjects

to complete addition and subtraction problems. Participants push a button to select either the first or the second answer from the options presented. The math task is adaptive to try to maintain a similar level of difficulty across participants. In total, the language task fMRI had a run duration of 3:57 minutes per run, with 316 frames per run.

- **MOTOR:** This task was adapted from the one developed by Yeo et al., 2011 [76]. In this task, subjects are shown visual cues asking them to either tap their left or right fingers, or squeeze their left or right toes, or move their tongue to map different motor areas in the brain. There are total 10 movements and each movement type lasted 12 seconds, preceded by a 3 second cue. In each of the two runs, there are 13 blocks, with two of tongue movements, four of hand movements (2 right and 2 left), and four of foot movements (2 right and 2 left). In addition, there are three 15-second fixation blocks per run. In total, the motor task fMRI had a run duration of 3:34 minutes per run, with 284 frames per run.

- **RELATIONAL:** This task has been adapted from the work done by Smith et al. [189]. The stimuli are six different shapes filled with one out of six different textures. Subjects are presented with 2 pairs of objects, one at the top of the screen and the other at the bottom. They have to first decide whether shape or texture differs across the top pair and then they have to decide whether the bottom pair also has the same difference. In the control matching condition, participants are shown two objects at the top of the screen and one at the bottom, and a word in the middle of the screen (either "shape" or "texture"). They are told to decide whether the bottom object matches either of the top two objects on that dimension (e.g., if the word is "shape", is the bottom object the same shape as either of the top two objects. For both conditions, the subject responds with a button press. For the relational condition, the stimuli are presented for 3500 ms, with a 500 ms inter-task interval, with four trials per block. In the matching condition, stimuli are presented for 2800 ms, with a 400 ms inter-task interval, and there are 5 trials per block. Each type of block (relational or matching) lasts a total of 18 seconds. In each of the two runs of this task, there are three

relational blocks, three matching blocks, and three 16-second fixation blocks. In total, the relational processing task fMRI had a run duration of 2:56 minutes per run, with 232 frames per run.

- **SOCIAL:** Subjects were shown 20 second video clips of objects (squares, circles, triangles) that either interacted in some way, or moved randomly on the screen. These videos were developed by either Castelli et al. [190][191] or Martin et al. [192][193]. After each video clip, subjects were asked to judge whether the objects had a mental interaction (an interaction that appears as if the shapes are taking into account each other's feelings and thoughts), Not Sure, or No interaction (i.e., there is no obvious interaction between the shapes and the movement appears random). Each of the two task runs has 5 video blocks (2 Mental and 3 Random in one run, 3 Mental and 2 Random in the other run) and 5 fixation blocks (15 seconds each). In total, the social cognition task fMRI had a run duration of 3:27 minutes per run, with 274 frames per run.

- **WORKING MEMORY:** The working memory task has been adapted from the one developed by Drobyshevsky et al., 2006 [194] and Caceres et al., 2009 [195]. Category specific representation task [196][197][198][199] and working memory (working memory) task [196][199][200][197] were combined into a single task paradigm. Subjects were presented with blocks of trials consisting of pictures of places, tools, faces, and non-mutilated body parts. Within each run, the four different stimulus types were presented in separate blocks. Also, within each run, half of the blocks use a 2-back working memory task and the other half use a 0-back working memory task (as a working memory comparison). A 2.5 second cue indicates the task type (and target for 0-back) at the start of the block. Each of the two runs contains 8 task blocks (10 trials of 2.5 seconds each, for 25 seconds) and 4 fixation blocks (15 seconds). On each trial, the stimulus is presented for 2 seconds, followed by a 500 ms inter-task interval (ITI). In total, the working memory task fMRI had a run duration of 5:01 minutes per run, with 405 frames per run.

3.2.2 HCP-YA preprocessing: FC pipeline

The HCP-YA minimal processing pipeline overview

Our starting point to process the HCP-YA data is the denominated *minimally processed* dataset, as provided by HCP [47]. The pipeline includes artifact removal, motion correction, and registration to standard space. The main steps of this pipeline are spatial (*minimal*) preprocessing, in standard volumetric and combined volume and surface spaces. By taking care of the necessary spatial preprocessing once in a standardized fashion, rather than expecting each user to repeat this processing, the minimal preprocessing pipeline avoids duplicate effort and ensures a minimum standard of data quality. The main steps of this minimal processing functional pipeline [152][47][152] are described in this section.

In total, there are six minimal preprocessing pipelines included in the HCP, three structural (*PreFreeSurfer*, *FreeSurfer*, and *PostFreeSurfer*), two functional (*fMRIVolume* and *fMRISurface*), and a *Diffusion Preprocessing* (not covered in this work) pipeline. Following is a brief description of the structural pipelines:

1. ***PreFreeSurfer:*** This produces an undistorted "native" structural volume space for each subjects, aligns the T1w and T2w images, performs a bias field correction, and registers the subject's native structural volume space to MNI space.

2. ***FreeSurfer:*** This pipeline is based on FreeSurfer version 5.2. It segments the volume into predefined structures, reconstructs white and pial cortical surfaces, and performs FreeSurfer's standard folding-based surface registration to their surface atlas (`fsaverage`).

3. ***PostFreeSurfer:*** This pipeline produces all of the NIFTI volume and GIFTI surface files necessary for viewing the data in Connectome Workbench, applies the surface registration to the Conte69 surface template [201], downsamples registered surfaces for connectivity analysis, and creates the final brain mask and myelin maps.

There are two volume spaces and three surface spaces in the HCP-YA data. The volume spaces are the subject's undistorted native volume space and the standard MNI space, which

is useful for comparisons across subjects and studies. The surface spaces are the native surface mesh for each individual (136k vertices, most accurate for volume to surface mapping), the high resolution Conte69 registered standard mesh (164k vertices, appropriate for cross-subject analysis of high resolution data like myelin maps), and the low resolution Conte69 registered standard mesh (32k vertices, appropriate for cross-subject analysis of low resolution data like fMRI or diffusion). The 91,282 standard grayordinate (CIFTI) space is made up of a standard subcortical segmentation in 2 mm MNI space and the 32k Conte69 mesh of both hemispheres. The functional and diffusion pipelines can be run after completing the structural pipelines described above. Following is a brief description of the two functional pipelines:

1. ***fMRIVolume:*** This pipeline removes the spatial distortions, carries out motion correction by realigning volumes, reduces the bias field, normalizes the 4-dimensional image to a global mean, and masks the data with the final brain mask. There is no overt volume smoothing in this pipeline as the output of this pipeline is in the volume space and can be used for volume-based fMRI analysis.

2. ***fMRISurface:*** In this pipeline, the volume-based time series is brought into the CIFTI grayordinate standard space. The voxels within the cortical gray matter ribbon are mapped onto the native cortical surface. This transforms the voxels according to the surface registration onto the 32k Conte69 mesh and maps the set of subcortical gray matter voxels from each subcortical region in each subject to a standard set of voxels in each atlas parcel. This gives a standard set of grayordinates in every subject with 2 mm average surface vertex and subcortical volume spacing. This data is then smoothed with surface and parcel constrained smoothing of 2 mm FWHM (full width at half maximum) to regularize the mapping. This pipeline outputs a CIFTI dense time-series (denominated `{TASK}_{ACQ}_Atlas_MSMAll.dtseries.nii`, where `{TASK}` refers to the fMRI condition and `{ACQ}` is the acquisition, either LR or RL) that can be used for surface-based fMRI analysis.

For the resting-state data, in addition to the minimal processing pipeline described above, a 24-parameter motion regression and ICA-FIX [189][202][203] have also been ap-

plied in the data provided by HCP-YA. The 24 parameters included in the motion regression step are the 6 rigid-body parameter time series, their backwards-looking temporal derivatives, and all squared 12 resulting regressors. The motion regressesion and ICA-FIX step applied on resting state fMRI data has produced time series denominated `rfMRI_REST_{ACQ}_Atlas_hp2000_clean.dtseries.nii`.

Additional processing steps

We perform the following additional steps on the fMRI data (denominated `{TASK}_{ACQ}_Atlas_MSMAll_ hp2000_clean.dtseries.nii` for resting state and `{TASK}_{ACQ}_Atlas_MSMAll.dtseries.nii` for task-based fMRI in HCP-YA data):

- ***Nuisance regression:*** This step is carried out for the task fMRI data only, where we regress out the 24-parameter motion regressors (6 rigid-body parameter time series, their backward-looking derivatives, and all squared resulting regressors), average time-series from the cerebro-spinal fluid (CSF), and the average time-series from the white matter. We have also provided processed data where this step is not performed.
- ***Global signal regression (GSR):*** This step involves the removal of the global (or average) signal from the time series of each voxel using linear regression [180]. We have produced two sets of connectomes, one where GSR has been performed and one where it has not been, as there is a lack of consensus in the scientific community regarding whether it should be performed or not [181][182][183][184][185][204]. The results shown in the main text are based on connectomes where GSR has been performed as part of the preprocessing. This step was performed for all fMRI conditions (if specified).
- ***Bandpass filtering:*** Lastly, we bandpass filter the time-series data using the following parameters:
  - Minimum frequency (f_{min}) = 0.009 Hz
  - Maximum frequency (f_{max}) = 0.08 Hz for resting state and 0.25 Hz for task-based fMRI

– Repetition time (TR) = 0.72 s

– Butterworth filter order = 4

This step was always performed on all the fMRI conditions.

Brain atlases

The brain atlases used in this paper have been developed by Schaefer et al., 2018 [120]. The Schaefer parcellations are further divisions of the resting state functional networks described by Yeo et al., 2011 [76] and have different levels of granularity (100 to 1,000 brain regions, in steps of 100). We have also added the 14 subcortical regions as provided in the HCP-YA dataset (denominated `Atlas_ROIs.2.nii.gz`) to each of these atlases, thus making them 114, 214,..., 1,014 brain regions. All the results described in this paper correspond to the Schaefer parcellations.

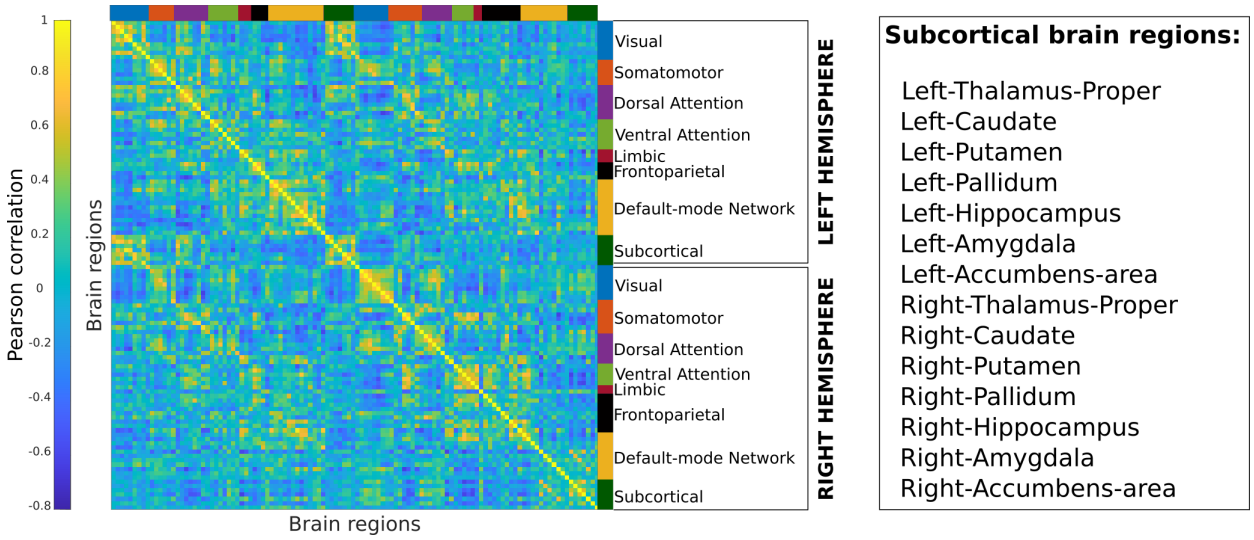


Figure 3.1. Example of a single-session, single-subject, whole-brain functional connectome (FC) using the Schaefer100 cortical atlas together with 14 subcortical regions. Functional couplings between brain regions are estimated through Pearson's correlation coefficients between their corresponding BOLD time-series. Rows and columns of the FC are ordered by hemisphere (Left and Right), and further divided into resting-state functional networks denoted by different colors.

Estimation of functional connectomes

The next step is to extract the functional time series corresponding to each brain region of the parcellation, z-score them, and ultimately estimate the functional connectomes. This step is conducted in Connectome Workbench (freely available at https://www.humanconnectome.org/software/connectome-workbench) with the commands `cifti-reduce`, `cifti-math`, and `cifti-parcellate`. The command `cifti-reduce` is used to compute the mean and standard deviation of time series data, while `cifti-math` computes the z-scored time series data using the mean and standard deviation previously computed. Lastly, `cifti-parcellate` parcellates the voxel-wise time series data into different brain regions as per the Schaefer parcellation, by averaging all the voxel-level time series belonging to each brain region. These parcellated time series have been made available and can be used to perform brain-region level activity and connectivity analysis.

Finally, whole-brain functional connectomes are generated by computing the Pearson's correlation coefficients between the time series of every pair of brain regions in the parcellated time series data computed in the earlier step. These connectomes are square and symmetric matrices that have been made available and can be used to perform functional connectome analyses. *Figure 3.1* shows a sample connectome parcellated into the Schaefer100 atlas, with 114 brain regions (100 cortical + 14 subcortical). The figure names all the subcortical regions.

3.2.3 The differential identifiability framework (If)

Identifiability matrix

In order to quantify the subject-level fingerprint from a cohort of functional connectomes, Amico and Goñi [98] proposed an object called the *identifiability matrix*. This is a non-symmetric correlation matrix that compares the all-to-all test-retest functional connectomes from a cohort of unrelated subjects. Thus, every entry in this matrix is the Pearson's correlation between test and retest functional connectomes in their vector form. Typically, the x-axis of the identifiability matrix represents the test (or first run, or day 1) and y-axis the retest (or second run, or day 2). Importantly, the ordering of the subjects is kept the

same in the rows and columns of this matrix (i.e., test and retest sessions), and hence the main diagonal contains the correlation values between the test and retest connectomes of the same subject. The higher the values in the main diagonal compared to the off-diagonal elements, the better the subject-level fingerprint of the dataset.

We have further expanded on this intuitive interpretation of the identifiability matrix by using the connectomes from twin pairs instead of test-retest of the same subject. In order to do so, we are extending the concept of identifiability and fingerprints beyond test/retest of the same subjects. To do so, we have computed the identifiability matrix for two different cohorts of twins, monozygotic (MZ) twins and dizygotic (DZ) twins. In this case, rows and columns of the identifiability matrix represent each of the two twins respectively, with the main diagonal values being the Pearson correlations between the FCs of the twin pair and the off-diagonal being the correlations between the FCs of unrelated subjects from the twin-cohort. *Figure 3.2* shows the differential identifiability matrices for sample cohorts of 20 test-retest, MZ twin pairs, and DZ twin pairs.

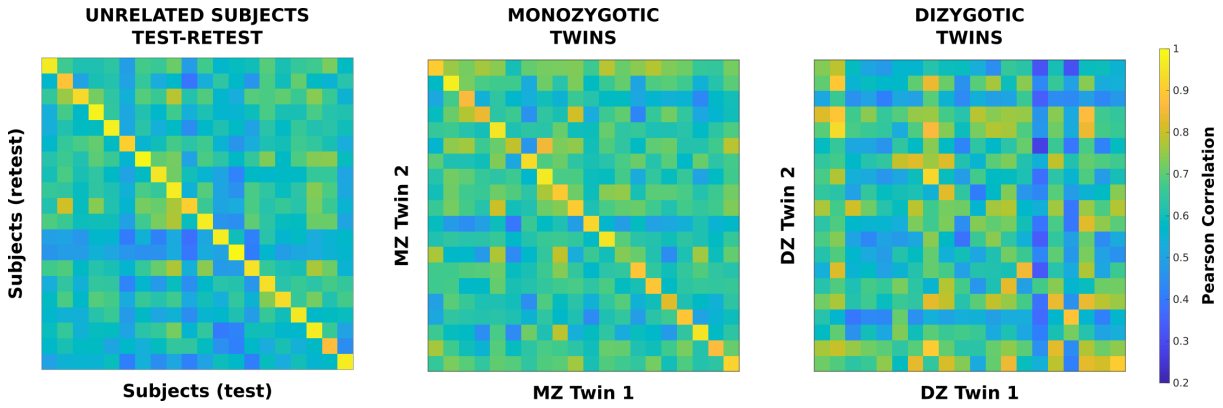


Figure 3.2. Differential identifiability matrices for sample cohorts of 20 Unrelated subject test-retest, Monozygotic twin pairs, and Dizygotic twin pairs

Differential identifiability score

Using the identifiability matrix, Amico and Goñi [98] have proposed a measure called *differential identifiability* (I_{diff}) to quantify the subject-level fingerprint. I_{diff} quantifies the contrast between the self-similarity (main diagonal) and the similarity between different subjects (off diagonal). I_{diff} can be computed as:

$$I_{diff}^{subject} = (I_{self} - I_{others}) \times 100 \tag{3.1}$$

where,

I_{self} = self similarity, mean of the main diagonal values in the identifiability matrix

I_{others} = similarity between different subjects, mean of the off-diagonal elements in the identifiability matrix

As discussed above, the differential identifiability score can also be calculated for a twin cohort by pairing FCs of twin subjects instead of test-retest. In this case, the main diagonal elements of the identifiability matrix will be the correlations between the FCs of twin subjects (MZ and DZ) and the off-diagonal elements will be the correlations between unrelated subjects. In this case, the *twins* differential identifiability can be expressed as:

$$I_{diff}^{twin} = (I_{twin} - I_{others}) \times 100 \tag{3.2}$$

This can be repeated separately for monozygotic (or identical) twins and dizygotic (or fraternal) twins. Monozygotic (MZ) twins are genetically 100% identical whereas dizygotic (DZ) twins have, on average, 50% genetic material in common [205][118].

PCA-based differential identifiability framework

In order to assess and compare the different Schaefer parcellations with each other in terms of their fingerprints, we have adapted the identifiability framework put forth by Amico and Goñi, 2018 [98]. They used group-level principal component analysis (PCA) to decompose functional connectomes into orthogonal *principal components* and then subsequently reconstructed with fewer and fewer principal components in order to find the reconstruction level where the differential identifiability score was maximum. Further developments based on the differential identifiability framework have been recently used to improve FC fingerprints across different scanning sites [111] as well as in network-derived measurements [104]. PCA is a statistical procedure that transforms a set of observations of possibly correlated variables into a set of linearly uncorrelated variables, i.e., *principal components*. PCA as a

tool is widely used in the exploratory analysis of the underlying structure of data in pattern recognition [206][207] and denoising [208][209], among other areas.

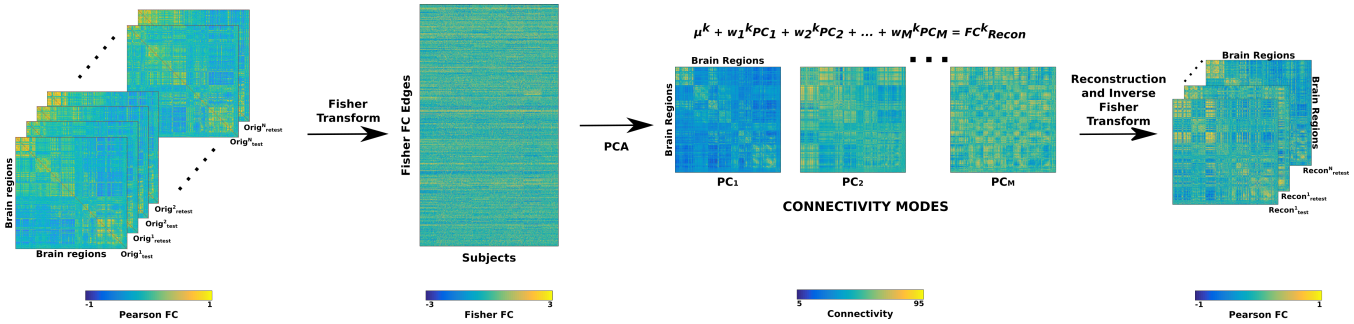


Figure 3.3. Workflow scheme of the group-level principal component analysis (PCA) reconstruction procedure of individual functional connectomes. The upper triangular values (as the matrices are symmetrical) of the test and retest FCs are vectorized, z-transformed using Fisher transform (MATLAB function `atanh`), and stacked into a matrix. This matrix is then decomposed using PCA to get as many components as connectomes in the cohort. The next step is to incrementally add principal components to the reconstruction, undo the Fisher transform (MATLAB function `tanh`) to get reconstructed functional connectomes, and compute the differential identifiability at each step.

We noticed that during the partial reconstructions of the functional connectomes using subsets of principal components, the FCs were not pure correlation matrices as some of the values in the FCs fell outside the [-1 1] range. To avoid this numerical issue, we have adapted the identifiability framework as proposed by Amico and Goñi for this study by using Fisher transform [210] as shown in *Figure 3.3*. As can be seen in the figure, we vectorize upper triangular (as the matrices are symmetrical) values in each FC (two FCs per subject for test-retest and one FC per subject for twins) before assembling them into a matrix where the columns are separate FCs and rows are the vectorized connectivity patterns. Before this vectorization, we z-transform the FCs by employing Fisher transform (MATLAB function `atanh`). Fisher transform has also been employed previously in several studies focusing on functional connectivity in the human brain [211][212][213][214]. The assembled matrix of Fisher transformed FCs is then decomposed using PCA into as many principal components as input FCs. In the next step, we reconstruct the FCs by incrementally adding one principal component at a time (in descending order of explained variance) and employing the inverse Fisher transform (MATLAB function `tanh`) in order to get back the Pearson correlation-

based FCs. These FCs at each step of the reconstruction are then used to compute the identifiability matrix and, by extension, the differential identifiability score (I_{diff}). Through this procedure, we obtain a curve of I_{diff} values for the whole range of principal components used in the reconstruction. It should also be noted that when all the principal components are used to reconstruct the FCs, we obtain the original input FCs, and thus the resulting I_{diff} score corresponds to that of the original FCs.

3.2.4 Assessment of brain fingerprints

Using the identifiability framework, we have assessed the brain fingerprints for three different cohorts – test-retest of a group of unrelated subjects (for the rest of the paper, we refer to this as *Unrelated subjects*), MZ twin pairs, and DZ twins pairs – at all the different levels of granularity afforded to us by the Schaefer parcellations, and for all fMRI conditions. We have set up different experimental designs to evaluate the fingerprints at the whole brain level and resting state functional network level. We also examine the effect of scanning length and repetition time (TR) on the subject-level and twin fingerprints for resting state fMRI. This section contains the description of the different experimental designs.

For consistency, we have only included those subjects in each of these three datasets in the PCA identifiability framework (unrelated subjects, MZ twins, and DZ twins) for whom all the task test/retest (both runs) functional connectomes are available. Thus, there are 428 unrelated subjects, 116 MZ twins pairs, and 63 DZ twin pairs included in the PCA reconstruction.

Whole brain differential identifiability

For a given parcellation granularity and a given fMRI condition, whole-brain FCs are used to compute the differential identifiability profiles.

Comparison of individual and twin fingerprint

To facilitate meaningful comparison between the differential identifiability profiles between the Unrelated subjects, MZ twins, and DZ twins, we have run the identifiability

framework on a subset of Unrelated subjects and MZ twin data so that the number of FCs in these cohorts is equal to the number of FCs in the DZ twin dataset (as DZ is the smallest dataset). This analysis was performed only for the Schaefer400 parcellation for all fMRI conditions. We have conducted the analysis for 100 bootstrap runs of 80% connectome pairs for each of the three cohorts. We hypothesize an ordinal presence of fingerprints that is highest for test-retest, lower for MZ twins, and lowest for DZ twins.

For each cohort separately, we test a null model where the rows of the identifiability matrix are shuffled before computing the identifiability score. We perform the bootstrap runs for the null models as well. This is to test whether the identifiability values we obtain from the identifiability framework applied to the three cohorts are a matter of chance.

Functional network-specific differential identifiability

In order to quantify the amount of fingerprint specific to a functional network, we assess the differential identifiability profiles by considering only the brain regions inside a specific functional network. For the network definitions, we use the 7 resting-state networks (RSNs) provided by Yeo et al., 2011 [76]. We should highlight that the PCA decomposition for this experiment is identical to when we explore whole-brain differential identifiability as described above, but for the differential identifiability calculation we only include the brain regions that belong to a specific RSN.

Effect of scanning length on differential identifiability

In order to test the effect of scanning length (in term of number of frames) on differential identifiability, we compute the differential identifiability profiles by gradually increasing the sequential number of fMRI volumes used in constructing the FCs. For this analysis, we have used resting-state scans as they have the longest scan duration (approx. 15 minutes) and the highest number of frames (1200). This allows us to study the effect of scanning length on differential identifiability for a wide range (50 to 1200, in steps of 50).

3.3 Results

3.3.1 The HCP-YA Functional Connectomes Data Release

The results of the processing performed in *Section 3.2* have been made publicly available at https://rdl-share.ucsd.edu/message/0Y3GKJM7a2CR2FgMSbk4st (functional connectomes) and https://rdl-share.ucsd.edu/message/Lqi0Oj0fALIrh4lv0Z5L06 (parcellated time-series). Here, we outline the specific data products that have been produced and how to access them. The data release includes FCs and time-series parcellated according to the Schaefer atlases [120] with different levels of granularity. For ease of downloading, we have created separate compressed files for each of the Schaefer parcellations, GSR/non-GSR status, and connectome/time series data. Additionally, for the task-based fMRI, we also have FCs with and without 26 regressors that correspond to motion and average signals from white matter and CSF (see details in *Section 3.2.2*). Each compressed file includes the FCs or time-series (depending on the selection) of all the fMRI conditions (resting state and 7 tasks) included in the HCP dataset that we have parcellated into the selected granularity of Schaefer parcellation. *Figure 3.4* shows an example of the data structure for connectome and time series data, respectively, for data processed with GSR.

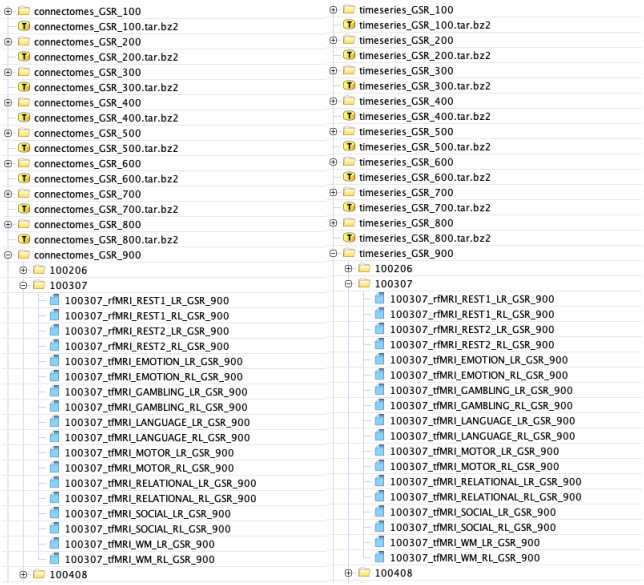


Figure 3.4. Sample data structure for functional connectomes and parcellated time series

3.3.2 Whole brain differential identifiability

Unrelated subjects test-retest

We first ran and assessed the differential identifiability framework on Unrelated subjects. This assessment is an extension with respect to Amico and Goñi [98] where a small cohort was evaluated (100 unrelated subjects, as opposed to 428) by using a single parcellation scheme [48]. *Figure 3.5* shows the I_{self}, I_{others}, and I_{diff} profiles for all the Schaefer parcellations for the 7 tasks and resting state included in the HCP-YA dataset.

Please note that in all the plots (I_{self}, I_{others}, and I_{diff}), the values for the maximum number of principal components correspond to the full reconstruction of the FCs with all the variance retained, i.e., original FCs. The number of principal components for which I_{diff} is highest is considered the optimal point of reconstruction. As can be seen, the I_{self} value decreases with increasing granularity for original FCs. However, as we reconstruct with fewer principal components, the I_{self} values pass through a point of inflection (where I_{self} values are approximately equal for all granularities) and leads to the optimal point of reconstruction where the reverse is true; i.e. I_{self} increases with increasing granularity. I_{others} on the other hand, is consistently lower for higher granularity across the whole range of principal components. Finally, I_{diff} values for original FCs are either approximately equal (e.g.n in MOTION) or increase negligibly with increasing granularity. However, at the optimal point of reconstruction, the I_{diff} score always increases with increasing granularity and the difference between the I_{diff} curves is much more pronounced.

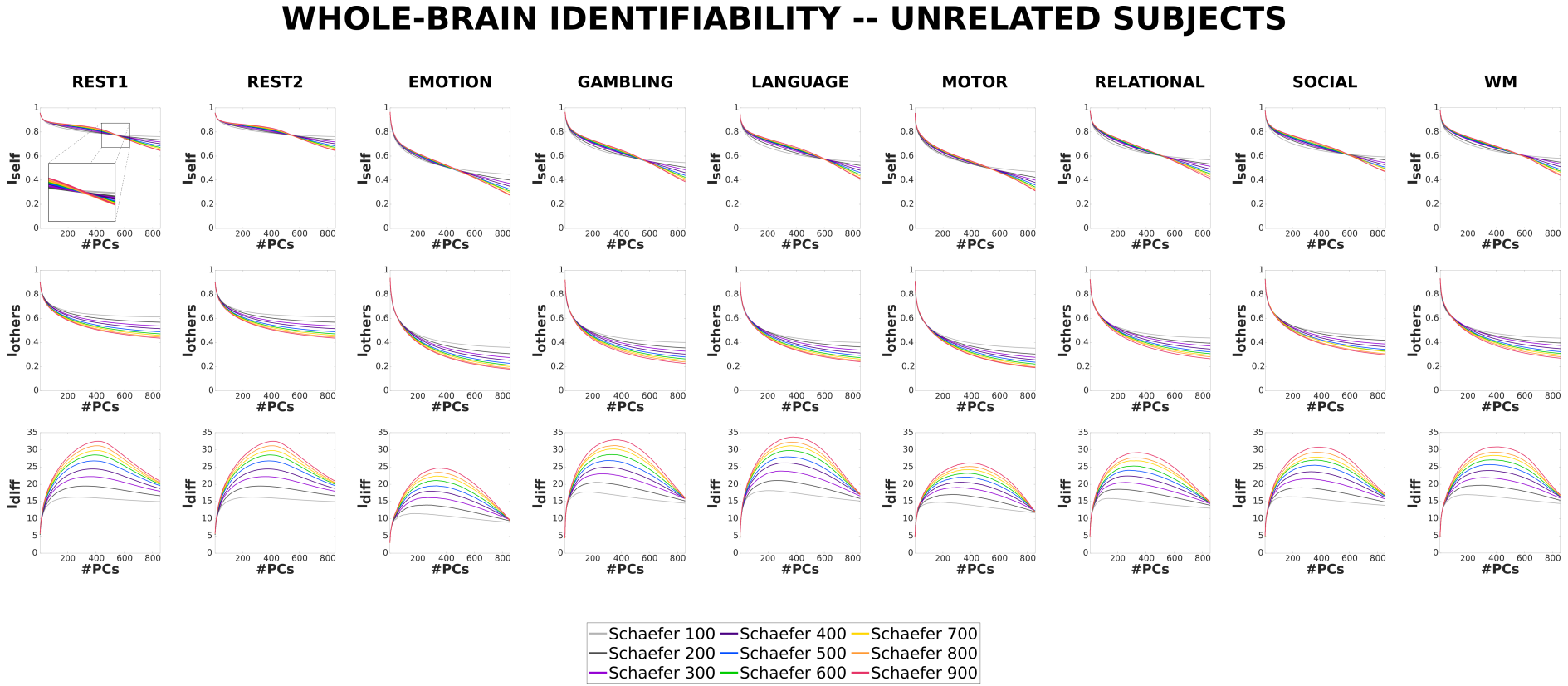


Figure 3.5. I_{self}, I_{others}, and I_{diff} curves for Schaefer 100 to 900 parcellations for all the fMRI conditions in HCP, for unrelated subjects. The higher the granularity of the Schaefer parcellation, the higher the test-retest identifiability regardless of the fMRI condition.

Monozygotic twins

Monozygotic (or identical) twins share 100% of their genetic material [118][205]. The differences between the MZ twins thus arise from their having different environments.

Both I_{twins} and I_{others} for MZ twins decrease with increasing granularity for the original FCs. I_{diff} scores are approximately equal across all granularities for original FCs. However, similarly as in Unrelated subjects, I_{diff} score increases with increasing granularity and the difference between the I_{diff} profiles for MZ twins is prominent at the optimal reconstruction.

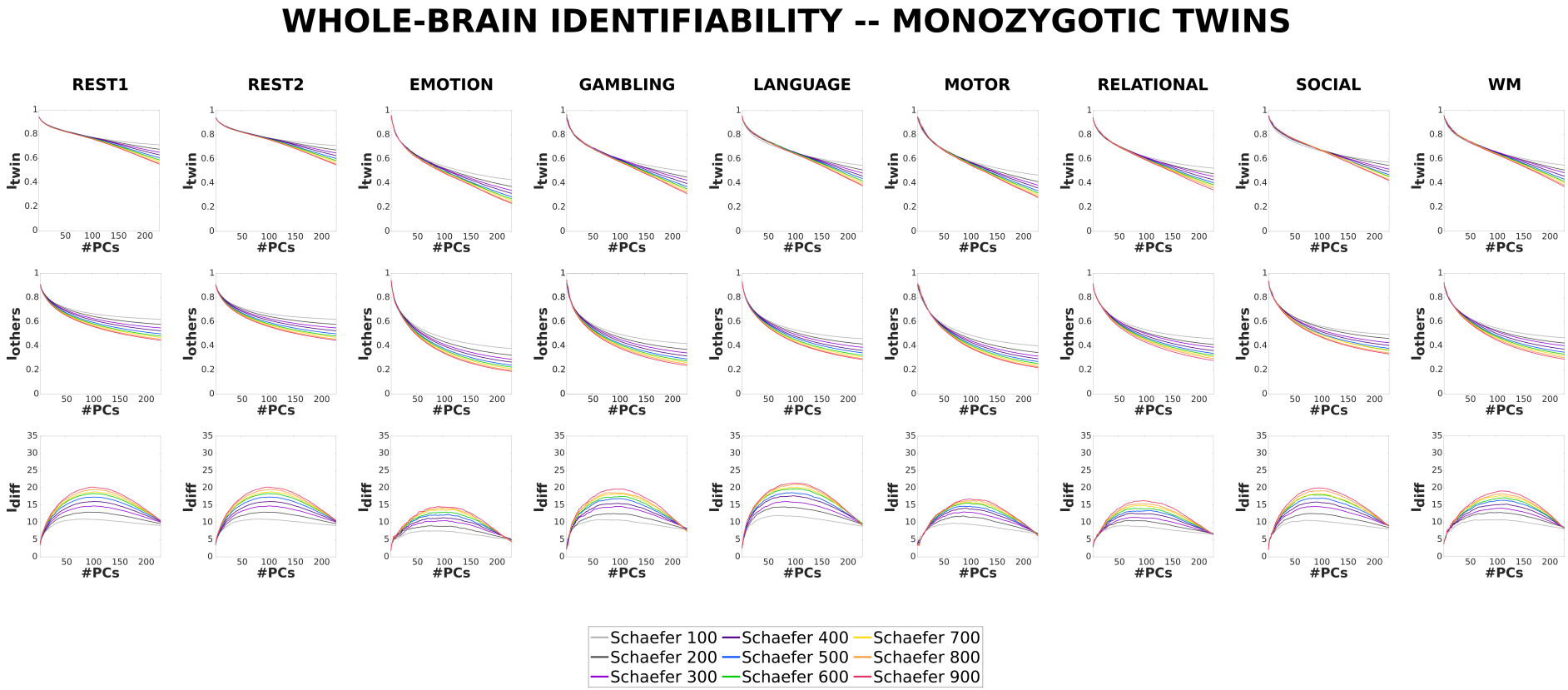


Figure 3.6. I_{self}, I_{others}, and I_{diff} curves for Schaefer 100 to 900 parcellations for all the fMRI conditions in HCP, for monozygotic (MZ) twin subjects. The higher the granularity of the Schaefer parcellation, the higher the MZ twin identifiability regardless of the fMRI condition, although the differential identifiability of MZ twins is lower than that of test-retest of the same subject.

Dizygotic twins

Dizygotic (or fraternal) twins are, on average, share 50% of their genetic material [118][205]. Thus, genetically speaking, they are siblings, but often their environment is more similar than that of non-twin siblings as they are born at the same time [215].

Similar to the results of MZ twins, both I_{twins} and I_{others} for DZ twins also decrease with increasing granularity for the original FCs. I_{diff} scores are approximately equal across all granularities for original FCs and, similar to Unrelated subjects and MZ twins, I_{diff} score increases with increasing granularity at the optimal point of reconstruction. However, the difference between the I_{diff} profiles for DZ twins is not as prominent as that for Unrelated subjects or MZ twins, but is still noticeable.

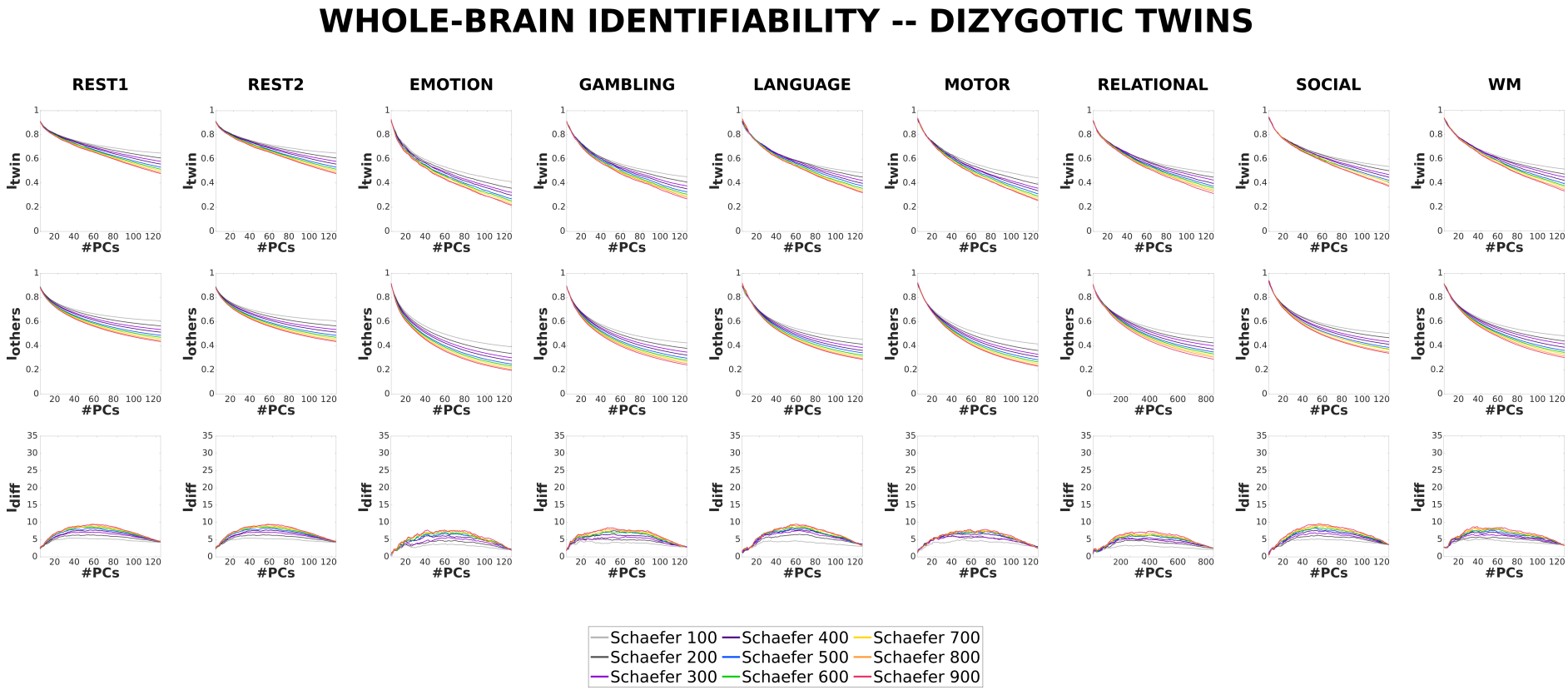


Figure 3.7. I_{self}, I_{others}, and I_{diff} curves for Schaefer 100 to 900 parcellations for all the fMRI conditions in HCP, for dizygotic (DZ) twin subjects. The higher the granularity of the Schaefer parcellation, the higher the DZ twin identifiability regardless of the fMRI condition. The differential identifiability of DZ twins is lower than that of both MZ twins and test-retest of the same subject.

Comparison of individual and twin fingerprint

In order to facilitate a meaningful comparison across the three cohorts (Unrelated subjects, MZ twins, and DZ twins), we chose a random subset of Unrelated subjects and a separate random subset of MZ twins so that their numbers match the sample size of DZ twins (DZ twins cohort has the smallest sample size out of the three cohorts). In *Figure 3.8*, we have plotted the I_{diff} profiles for Unrelated subjects, MZ twins, and DZ twins cohorts for all tasks and resting state for the Schaefer400 parcellation.

As can be seen in *Figure 3.8*, the I_{diff} scores are the highest for Unrelated subjects across all the fMRI conditions for the entire range of principal components. These are then followed by the I_{diff} scores for MZ twins and DZ twins, respectively. The three curves at the bottom of the figures are results for the null models. As can be seen, the I_{diff} scores for these null models are approximately zero for the entire range of the principal components (two of the curves are mostly hidden behind the third).

Functional network-specific differential identifiability

In order to assess the level of fingerprint in specific functional networks of the brain, we have computed the I_{diff} profiles of each of the 7 resting-state networks, as proposed by Yeo et al. [76]. *Figure 3.9* shows the differential identifiability profiles for the 7 RSNs using resting-state FCs for all Schaefer parcellations. Please note that the decomposition/reconstruction based on PCA is carried out on whole-brain FCs and not on isolated functional networks. In other words, results presented in this section belong to the same decomposition/reconstruction procedure as the ones shown in *Figure 3.5*.

When assessing I_{diff} in an isolated fashion on each RSN, it can be observed that granularity of the parcellations increases I_{diff} at any level of reconstruction. Notice that some RSNs present higher levels of fingerprints than others for all granularities. For any RSN (with the only exception of VISUAL), at the optimal level of reconstruction the I_{diff} scores for any given Schaefer parcellation are higher than that achieved when computing the I_{diff} scores using whole-brain FCs (see *Figure 3.5*).

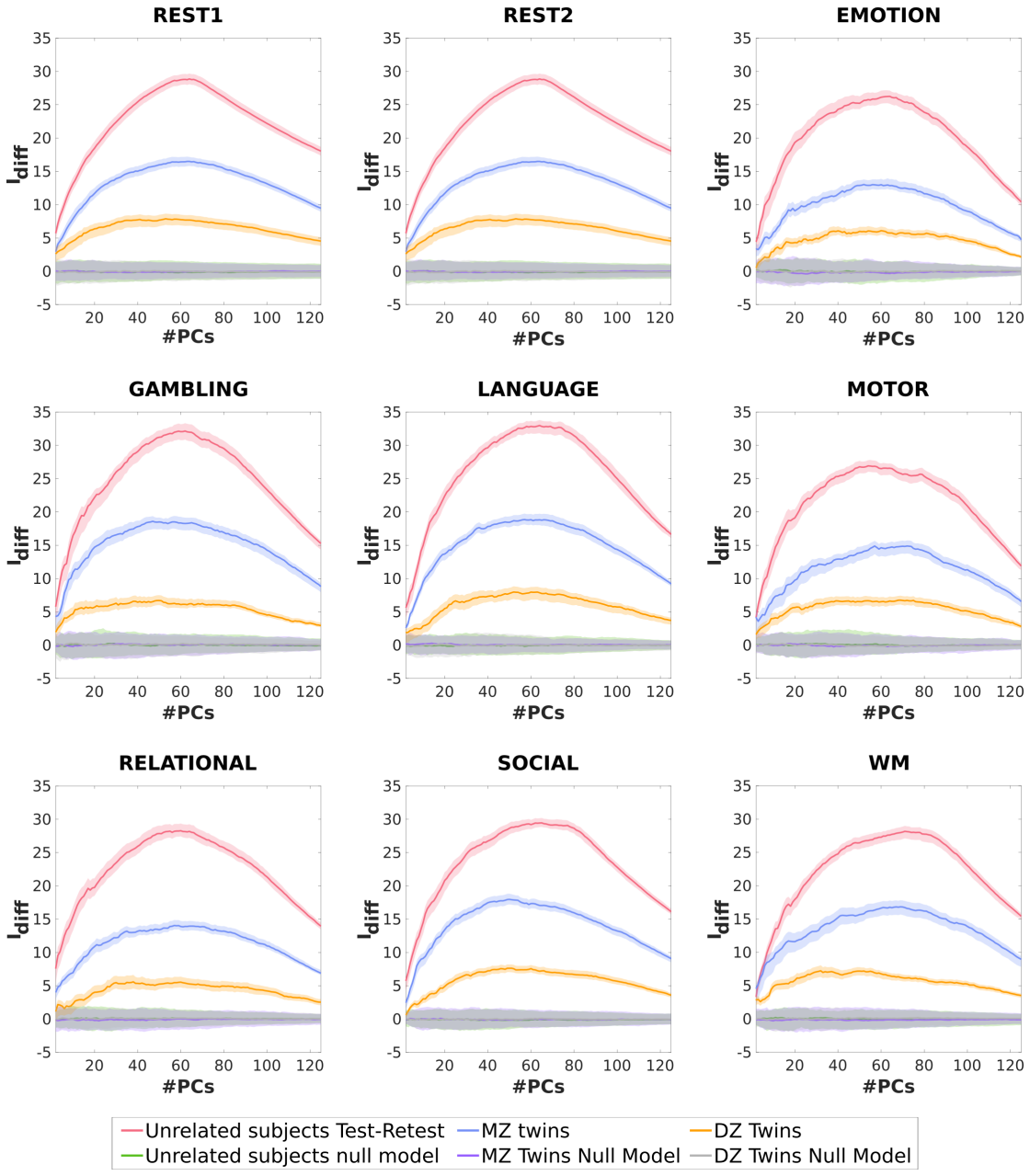


Figure 3.8. I_{diff} profiles for the three cohorts – Unrelated subject test-retest (red), Monozygotic twins (blue), and Dizygotic twins (orange) – for all fMRI conditions using Schaefer400 parcellation. The cohort sizes have been matched in order to facilitate comparisons between them. The figure also includes results for the null models based on the three cohorts. Shaded areas represent the variability (5-95 percentile) of I_{diff} scores across the 100 samples without replacement.

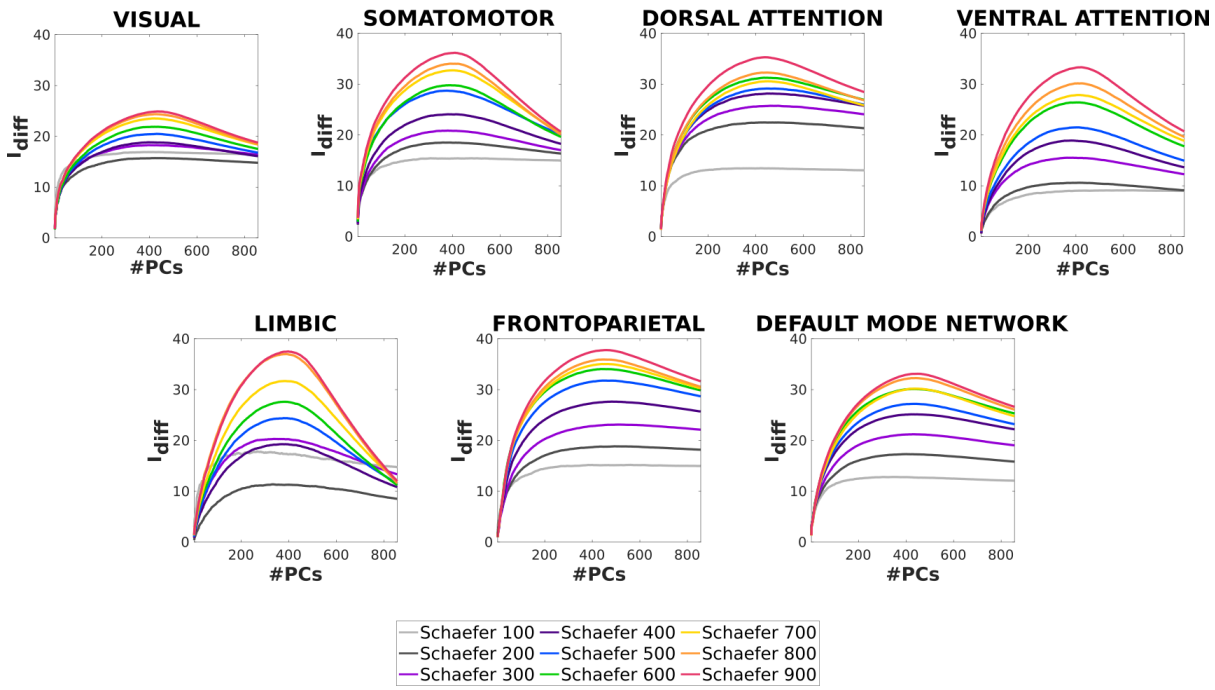


Figure 3.9. Functional network-specific I_{diff} curves for Schaefer 100 to 900 parcellations for resting state connectomes. The higher the granularity, the higher the differential identifiability in most cases. This does not hold true when the number of brain regions included in a functional network is too few. For example, there are less than 10 brain regions included in the limbic functional network for the Schaefer 100 parcellation, which causes the I_{diff} curve to be very unstable.

Effect of scanning length on differential identifiability

For the next analysis, we assessed the effect of different lengths of acquisition time on the differential identifiability of functional connectomes. In order to do so, we subsampled different lengths of time series from the resting state fMRI acquisition and applied the identifiability framework on the resulting connectomes. We repeated this experiment for all the Schaefer parcellations. *Figure 3.10* shows the original and optimal differential identifiability for different number of time points and for all Schaefer parcellations, along with the difference between them. Also observe that the difference between original and optimal I_{diff} increases with the granularity of the brain atlas. From the plot showing the difference between the original and optimal I_{diff} scores, we can also note that, for every parcellation and scanning length combination, the differential identifiability is always higher. The different levels of granularity are more distinguishable from each other in terms of their I_{diff} scores for shorter scanning lengths (>150 timepoints) at the optimal reconstruction as compared to the original FCs (>300 timepoints).

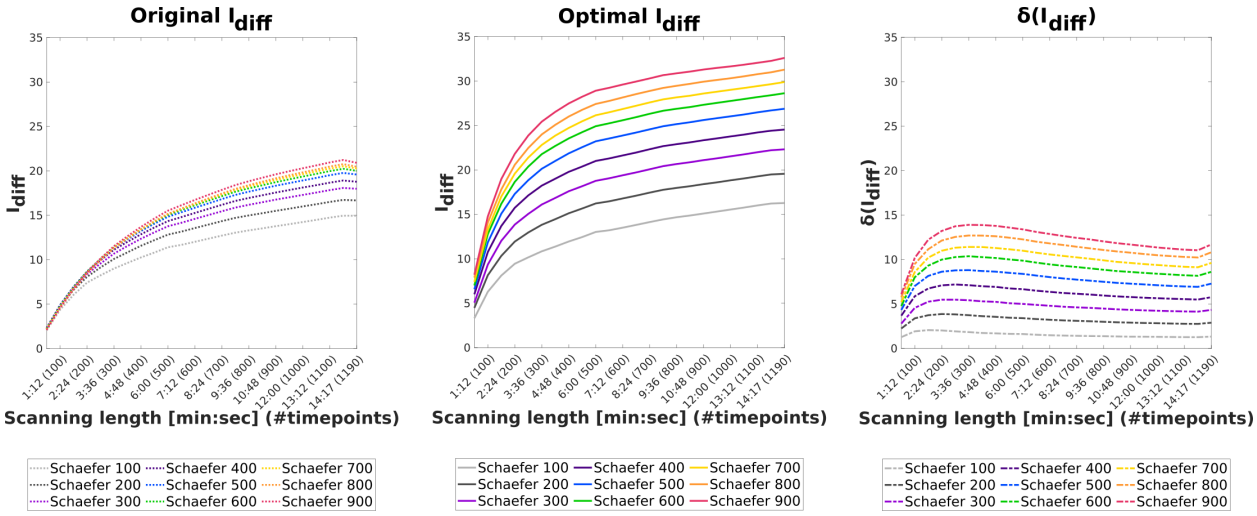


Figure 3.10. Original and optimal I_{diff} values for resting state in all Schaefer parcellations for different scanning lengths, along with the difference between the two. For every Schaefer parcellation, we mimic a shorter scanning length by sampling from the entire rs-fMRI scan (50:50:1190 timepoints), construct functional connectomes from these shortened scanning lengths, and run the PCA identifiability framework in order to study their stability. The x-axes of the plots show the scanning length, both in terms of minutes and seconds and the number of timepoints.

3.4 Discussion

In this paper, we have discussed the processing pipeline we have developed to extract brain region-level time-series and the subsequent functional connectomes for each session and all the fMRI conditions of all the subjects in the Human Connectome Project – Young Adult (HCP-YA) dataset. We have made these time-series and functional connectomes datasets available, parcellated according to the Schaefer atlases [120] that afford us different levels of granularity (100 to 900 brain regions, in steps of 100), combined with subcortical regions (7 regions in each hemisphere; 14 in total). We have also provided a quantification of the individual and twin (monozygotic and dizygotic) fingerprint present in the FCs (of each fMRI condition separately) using an extension of the identifiability framework proposed by Amico and Goñi, 2018 [98]. Briefly, results show the presence of fingerprints at three different levels of genetic and environmental similarity (as depicted by Unrelated subjects greater than MZ twins, greater than DZ twins; *Figure 3.8*).These results are present for all fMRI conditions evaluated and with different sensitivity to parcellation granularity. We also found that the identifiability framework not only uncovers individual and twin-fingerprints in FCs, but, importantly, also enables us to benefit from the higher levels of fingerprints present in FCs corresponding to higher levels of granularity (*Figures 3.5, 3.6*, and *3.7*). Subsequently, we discovered that different levels of fingerprint are present for the various resting-state networks, with the same pattern of higher levels of granularity enabling us to uncover higher fingerprints (*Figure 3.9*). Finally, we found that the amount of individual fingerprint in resting-state FCs increases with increasing scanning length, but saturating after ∼13 minutes of scanning (*Figure 3.10*).

We have assessed the individual-level fingerprint between the test and retest FCs of the cohort of Unrelated subjects using the identifiability framework (see *Figure 3.5*). Consistent with previous investigations [98][100][111][99][216][217][218][219][220][221][222][223][105], we have found that FCs have a recurrent and reproducible individual fingerprint across all the fMRI conditions. This is all the more important as in this study, the sample size of upwards of 400 Unrelated subjects is considerably bigger than the previous studies (usually 100 unrelated subjects from the HCP-YA dataset). As can be seen in *Figure 3.5*, without implementing the

identifiability framework and hence assessing original FCs, there is little to no difference in I_{diff} scores between the different parcellations for each of the fMRI conditions. This implies that the granularity of the parcellation is inconsequential in terms of individual fingerprint, prompting one to use the smallest parcellation as it would lead to lower computational load. However, the difference between the fingerprints of the different parcellations only becomes apparent when the identifiability framework is applied (see *Figure 3.5*). In particular, it can be observed that the higher the granularity of a parcellation, the higher the optimum test-retest fingerprint achieved for the cohort of Unrelated subjects. In other words, the potential to uncover fingerprints by fine-grained parcellations of the cortex is only unleashed when using the identifiability framework. On a related note, higher granularity was associated with a larger number of principal components leading to the highest I_{diff} scores. This might be an indication of higher granularity parcellations containing more information about the individual fingerprint.

A subset of the HCP-YA dataset is made up of monozygotic (MZ) and dizygotic (DZ) twin pairs (see *Table 3.1*), so the next step was to utilize this and quantify the twin-fingerprint in the dataset which has not been done before. In this analysis, we adapt the identifiability framework by using, for each fMRI condition, one single-session functional connectome from each of the twins in a pair (MZ and DZ, separately) in lieu of test and retest of the same subject. We found the presence of a twin-fingerprint (both for MZ and DZ twins) in all fMRI conditions. It is noteworthy that the twin-fingerprint was much higher than expected by chance, at the same time being lower than individual fingerprint (based on test/retest Unrelated subjects; see *Figure 3.8*). Similar to the cohort of Unrelated subjects, the identifiability framework not only contributed in uncovering higher twin-fingerprints in all fMRI conditions, but also enabled us to utilize the higher granularity of the parcellations to achieve higher I_{diff} scores. In particular, the I_{diff} profiles are similar across the three cohorts, with the cohort of Unrelated subjects achieving the highest peaks, followed by MZ and DZ twins, respectively (see *Figures 3.5, 3.6, 3.7*, and *3.8*). This ordinal structure of the fingerprint across the three cohorts (namely Unrelated subjects, MZ twins, and DZ twins) can be explained in terms of the genetic and environmental similarity between the pairs of connectomes across two sessions. In particular, for Unrelated subjects, the genetic informa-

tion is 100% equal and the environment is also highly shared across sessions as the scans belong to the same subject. On the other hand, the MZ twins, even though they share 100% of their genetic information, they environment is shared to a much lower degree as the MZ twins are two separate individuals. Lastly, DZ twins share (on average) 50% of their genetic information and the environment shared is similar to that of MZ twins [224]. Please note that, in *Figure 3.8*, we selected a subset of the Unrelated subjects and MZ twins cohorts in order to match number to DZ twins (69 pairs) in order to facilitate a meaningful comparison.

The presence of a substantial twin-fingerprint (MZ and DZ) is a compelling argument for utilizing only a cohort of unrelated subjects when conducting studies that rely on brain fingerprinting and differences between individuals. This is because including twin pairs or siblings in such studies can confound the results, as evidenced by the findings of this paper. An alternative strategy could be to keep both twins from a pair either in the training or the validation dataset. If one of the twins is used for training and the other for validation, it might lead to a false increase in the prediction accuracy of the model under consideration [225].

In their seminal work, Finn et al., 2015 [100] observed that some of the functional networks of the brain contained a higher fingerprint than the whole-brain fingerprint at rest and between fMRI tasks (used as test/retest). In order to replicate and extend on this result, we quantified the amount of individual fingerprint for the 7 resting-state networks (or RSNs, as proposed by Yeo et al., 2011 [76]) using the (single-session) resting-state fMRI condition across all available levels of granularity. Consistent with previous findings, *Figure 3.9* shows that some of the RSNs achieve a higher (e.g., somatomotor and frontoparietal) fingerprint than others (e.g., visual and limbic) using original FCs. Similar to the whole-brain scenarios, the identifiability framework uncovers the individual fingerprint in all the RSNs and enables us extract the higher fingerprint present in the higher granularity of the parcellations. Another effect of using the identifiability framework is that the amount of fingerprint present in each RSN becomes much more uniform, with all RSNs (except visual) reaching around $I_{diff} \approx 35$ for Schaefer900.

Lastly, *Figure 3.10* shows the effect of scanning length on the identifiability of resting-state FCs. We have chosen to run this experiment only on resting-state data as it is the

longest fMRI acquisition in the HCP-YA dataset. As can be seen from the profiles of original I_{diff} scores for all the Schaefer parcellations, the difference between the identifiability of different parcellations is negligible up to 300 timepoints. The same is not true for the optimal I_{diff}, as the profiles start diverging from each other as early as 150 timepoints. The optimal I_{diff} achieved is also higher than the original I_{diff} at every scanning length, but the difference is more pronounced for higher granularity parcellations. Overall, we observe that a longer scanning duration leads to a higher fingerprint, but saturating at around ∼13 minutes for resting-state fMRI. In addition, identifiability framework not only allows us to uncover higher fingerprint for the same scanning duration, but also enables us to utilize the higher granularity parcellation to achieve higher fingerprint.

Amongst the limitations of this work is the fact that a dataset such as HCP-YA inherently does not have a large cohort of twin subjects or different age groups. This has been a limitation in terms of not being able to ask specific research questions to study the fingerprint between twins or across the lifespan. Also, specific to HCP-YA, the task fMRI acquisition lengths are heterogeneous and not as long as resting state acquisition (see *Table 3.2*). The availability of longer task-based fMRI sequences would allow researchers to analyze the effect of scanning length and TR on the stability of the connectomes, similar to the analysis done on resting state fMRI data carried out in *Sections 3.3.2*. Lastly, we have not studied the effect of different processing pipelines on identifiability measures [204][226].

The signal-to-noise ratio (SNR) of an fMRI time series is the ratio of the amount of BOLD signal to the amount of background noise. In higher the granularity of a brain parcellation, the smaller the size of the brain regions and the lesser the number of voxels per brain region. This leads to a decrease in the SNR as the estimation of the average time series representing the brain region is based on a smaller sample. In other words, the time series of the brain region is a mean of the time series of a smaller set of voxels. The estimation is also more susceptible to head motion as the number of voxels per brain region is small. Thus, we hypothesize that, as the parcellation granularity increases there would be an upper limit on the FC fingerprint that can be achieved. In our study, up to the Schaefer 900 parcellation the time series SNR does not go along the SNR of the FC fingerprint. Thus, we do not

know what this upper limit on the FC fingerprint is and it is an open question that can be explored in further detail in the future.

Additionally, similar efforts could be pursued to also process the diffusion weighted imaging data included in the Human Connectome Project Young Adult dataset and make it available for public use of the corresponding subject-level structural connectomes in the future. Researchers could also provide processed versions other state-of-the-art brain connectivity datasets such as Alzheimer's Disease Neuroimaging Initiative (ADNI) [171], Adolescent Brain Cognitive Development (ABCD), HCP–Lifespan, HCP–Aging, among others together with their corresponding fingerprint analyses.

4. ESTIMATION OF EFFECTS OF HERITABILITY AND ENVIRONMENT IN FUNCTIONAL CONNECTOMES

4.1 Introduction

Heritability is defined as the proportion of phenotypic variance in a trait that is attributable to the effects of genetic variation. The environment that an individual experiences in their lifetime also typically has an effect on this phenotypic variance [118]. In order to distinguish between the effects of genetics and environment on a phenotype, the classic twin model, also called ACE model, is frequently employed [227][118]. The ACDE model compares the similarity between monozygotic (MZ) or *identical* twins and dizygotic (DZ) or *fraternal* twins. MZ twins are formed when a fertilized egg splits into two *in utero*, while DZ twins are formed when two separate eggs are fertilized by two different sperms at the same time. Thus, MZ twins are always the same gender while DZ twins can be either the same gender or opposite genders. This is shown graphically in *Figure 4.1* [118]. As can be expected, MZ twins are genetically identical as they are formed due to a duplication of an embryo, while DZ twins share on average 50% of their genetic material [228][121]. Thus, DZ twins are genetically similar to siblings, with the exception that they are born at the same time and thus share a higher amount of environment between themselves than pure siblings. The differences between MZ twins are supposed to be a result of their individual environments while that between DZ twins is a combination of genetics and individual environments. This is summarized in *Table 4.1* below, along with the information for the genetic and environmental similarity for the test/retest of the same subject as reference:

Table 4.1. Genetic and environmental similarity between within cohorts

Dataset	**Genetic Similarity**	**Environmental Similarity**
Test/retest	100%	100%
MZ twins	100%	$< 100\%$
DZ twins	50%	$< 100\%$ ($\simeq$ MZ twins)

In this chapter, we will implement the ACDE model in order to determine the effect of genetics and environment on the human functional connectome. We also study the effect of

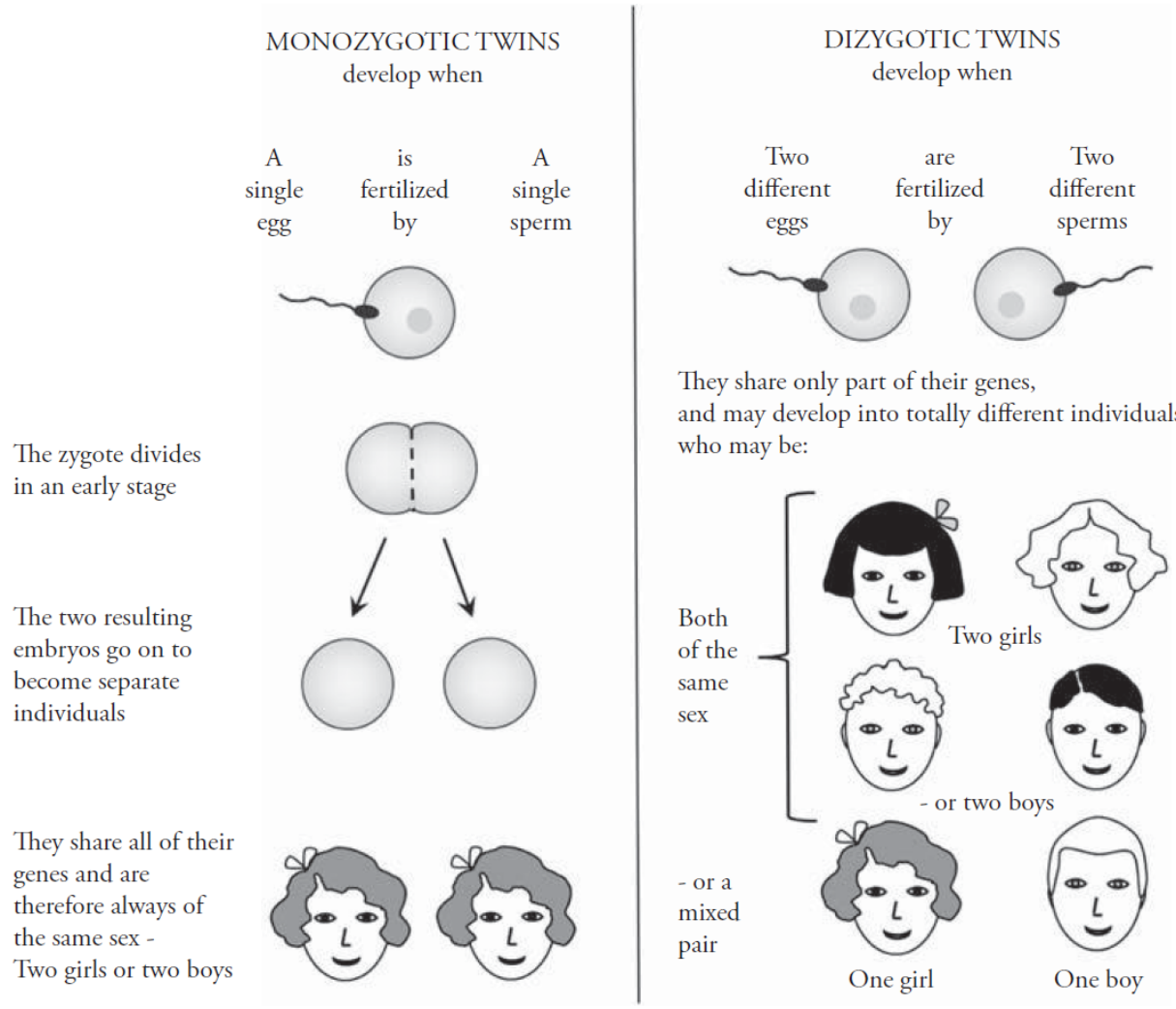


Figure 4.1. Development of monozygotic versus dizygotic twins (Figure with permission from Blokland et al., 2013 [118])

increasing the fingerprints between the two different cohorts, MZ and DZ twins, on the ACDE model. For this purpose, we will make use of the optimally reconstructed FCs obtained in *Chapter 3* for this purpose.

4.2 Methods

4.2.1 The Classic ACDE Model

The ACDE model [121][227][118], also known as the Classical Twin Model, draws its explanatory power from the differences in genetic and environmental similarity between

MZ and DZ twin pairs. This model uses simultaneous equations to partition the variance observed in a phenotype into that caused as a result of additive genetic (A, or a^2), dominant genetics (D or d^2), common environmental (C or c^2), and unshared environmental (E or e^2) influences. Additive and dominant genetic influences are the cumulative effects of genes presenting in an additive or dominant manner. Additive genetic effects take place when two or more genes *add up* to contribute to a phenotype. An example of this type of effect is skin color, which depends on the amount of melanin in a person has in their skin [229]. There are three genes that affect the amount of melanin and they each have an equal impact. Thus, none of the three genes is dominant or regressive. On the other hand, dominant genetic effects take place when a person receives two versions of each gene, known as alleles, from each of the parents. If the alleles of a gene are different, one allele is expressed while the effect of the other allele are masked. The allele that is expressed is called the dominant gene and the other allele is called the recessive gene. Examples of this type of genetic effect are hair and eye color, where red/blonde hair and blue/green eyes are regressive traits. Thus, if one allele is that of red hair or blue eyes and the other is for brown hair or brown eyes, the person will have brown hair and brown eyes [230][231]. Common environmental influences refer to the experiences shared by twins in a pair, including the *in utero* environment, and the social and cultural environment they might have experienced, i.e., same socioeconomic status, parents, diet, etc. Unshared environmental factors include all aspects of the physical and social environment experienced differently by individuals in a family. Examples of this include illness, physical and psychological trauma, peers, teachers, etc. This component also includes measurement error and gene–environment interactions, which are not accounted for in this model [232][233]. Structural equation modelling is commonly used to estimate the effect of these four factors on a phenotype [234][232].

Figure 4.2 shows the structural equation diagram for the ACDE model with all four factors – additive genetics, dominant genetics, common environment, and unshared environment – are included. This is known as the univariate ACDE model. Here, p_1 and p_2 are phenotypes of the same trait for the two twins in a pair. Additionally, *a, c, d,* and *e* are the additive genetics, common environment, dominant genetics, and unshared environment, respectively. It is worth noting that the only known variables in this model are the cor-

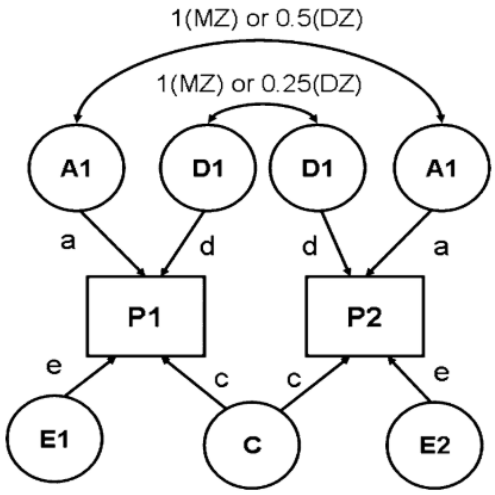


Figure 4.2. Path diagram depicting the classical ACDE model. p_1 and P_2 are the phenotypes in twin-1 and twin-2, respectively, while MZ stands for monozygotic twins and DZ for dizygotic twins. A is additive genetic influences, D is dominant genetic influences, C is common environment influences, and E is unique environmental influences. *a, c, d*, and *e* show the additive genetic, common environment, dominant genetics, and unique environment path coefficient, respectively. Correlations between additive genetics are set at 1 for MZ twin pairs and 0.5 for DZ twins pairs, as MZ twins share 100% of the genetic material and DZ twins on average 50%. Correlations between common environment are 1 for both MZ and DZ twins, as both types of twins share 100% of their familial environment. Uncommon environment is uncorrelated in both MZ and DZ twins as they are unique for each individual. (Figure with permission from Ozaki et al., 2011 [121])

relations between MZ twin pairs (r_{MZ}) and DZ twin pairs (r_{DZ}). The ACDE model also assumes that the variation in a phenotype arises from the linear sum of the four sources of phenotypic variation as follows:

$$a^2 + c^2 + d^2 + \mathrm{e}^2 = 1 \tag{4.1}$$

As MZ twins share 100% of their genetic material, additive and dominant, and on average 100% of their shared environment as well, the covariance between the MZ twin pair can be written as:

$$r_{MZ} = a^2 + c^2 + d^2 + \mathrm{e}^2 \tag{4.2}$$

On the other hand, DZ twins share 50% of their genetic material on average, and 25% of the time they also share the same alleles from both the parents, which are required to share the dominant genetic effects. DZ twins are also assumed to share the same amount of environment as MZ twins, which is approximately 100%. Thus, the covariance between DZ twins can be written as:

$$r_{DZ} = 0.5a^2 + c^2 + 0.25d^2 + \mathrm{e}^2 \tag{4.3}$$

As can be observed from *Equations 4.1, 4.2*, and *4.3*, the information contained within the classical twin model is insufficient to estimate the contribution of all four factors to the variation in a phenotype. As a result of this, twin studies tend to estimate either the effect of common environment (c^2) or that of dominant genetics (d^2), as these measures are negatively confounded [118][121]. This is due to the fact that dominant genetic effects tend to typically make MZ twins more similar, in turn making the DZ correlation lower than the MZ correlation. On the other hand, common environment has the opposite effect, where it makes DZ twins more similar than MZ twins. Thus, either c^2 or d^2 are computed depending on whether the DZ correlation is greater than or less than half the MZ correlation. Thus, the ACDE model can be split into two separate models, ACE model and ADE model, as follows:

ACE Model

$$\begin{aligned} r_{MZ} &= a^2 + c^2 + \mathrm{e}^2 \\ r_{DZ} &= 0.5a^2 + c^2 + \mathrm{e}^2 \end{aligned} \tag{4.4}$$

ADE Model

$$\begin{aligned} r_{MZ} &= a^2 + d^2 + \mathrm{e}^2 \\ r_{DZ} &= 0.5a^2 + 0.25d^2 + \mathrm{e}^2 \end{aligned} \tag{4.5}$$

Where,

r_{MZ} = correlation between MZ twins

r_{DZ} = correlation between DZ twins

a^2 = additive genetics

c^2 = common environment

d^2 = dominant genetics

e^2 = uncommon/unique environment and error

Equations 4.4 belong to the ACE model which is computed when $r_{MZ} \leq r_{DZ}$, whereas *Equations 4.5* belong to the ADE model which is computed when $r_{MZ} > r_{DZ}$. We can obtain the formulae for estimating the effect of the a^2, c^2, d^2, and e^2 factors by solving these two sets of equations. These formulae are listed in *Table 4.2* below:

Table 4.2. Formulae for ACE and ADE models

	ACE Model	**ADE Model**
$\mathbf{a^2}$	$2(r_{MZ} - r_{DZ})$	$4r_{DZ} - r_{MZ}$
$\mathbf{c^2}$	$2r_{DZ} - r_{MZ}$	N/A
$\mathbf{d^2}$	N/A	$2r_{MZ} - 4r_{DZ}$
$\mathbf{e^2}$	$1 - r_{MZ}$	$1 - r_{MZ}$

The classical twin model has several underlying assumptions [118]. Following is a brief explanation of these assumptions:

1. **Generalizability:** The model assumes that the results can be generalized to the general, non-twin population. The experience of being a twin, including the sharing of limited space and resources during gestation, and the differences in the birth process, may cause twins to be different from non-twins. Twins differ from non-twins, especially in traits related to prenatal growth, although most studies generally do not find differences in personality and social traits [235]. If this assumption is violated, additional twin-specific effects have to be incorporated in the model.

2. **Random mating:** The assumption that DZ twins share on average 50% of their genetic material does not hold in the case of *assortative mating*. Assortative mating is said to happen when individuals with similar phenotypes mate with one another more frequently than would be expected under a random mating pattern. As these phenotypes are be at least partially caused by similar gene variants, their children are likely to share more than 50% of their genetic information.

3. **Degree of genetic similarity between MZ twins:** MZ twins are assumed to be genetically identical, i.e., sharing 100% of their genetic material. Although, a study by Bruder et al., 2008 [236] of 19 MZ twin pairs detected subtle differences in copy number variations of the DNA that could take place when a set of coding neucleotide bases in the DNA are missing or have extra copies.

4. **Equal environments:** The classic twin model divides the environment into that which is shared and that which is unshared between a twin pair. Shared environment refers to the in-utero environment and the effects of growing up in the same household. This assumes that the effect of environmental influences on a phenotype is the same on MZ and DZ twins. But it has been found that MZ twins have a higher level of shared environment growing up than DZ twins [237][238][239]. However, it is not clear whether having greater environmental similarity translates into greater phenotypic similarity.

5. **Genotype-environment interaction:** The classical twin model ignores the possible effects of genotype-environment interaction, which takes place when environments have differential effects on different genotypes. For example, Parkinson's disease is a result of a combination of multiple gene variants, environmental exposures, and lifestyle choices [240].

6. **Genotype-environment correlation:** Gene–environment correlation is said to happen when subjects expose themselves to different environments depending on their genotype, or when individuals' genotypes affect their social interactions or influence the responses they elicit from other individuals [241][242]. The classic twin model assumes this effect to be non-existent.

4.2.2 Extended ACE Model

As discussed in *Section 4.2.1*, the classical ACDE model combines the effect of unique environment and measurement error in the term e^2 [118][121]. Here, we propose an extension to the classical model that will disentangle the effect of unique environment from simple

measurement error. As the HCP dataset contains repeat measurements for every subject, we can estimate the measurement error as:

$$\epsilon = 1 - r_{TR} \tag{4.6}$$

Where,

ϵ = measurement error, and

r_{TR} = correlation between the two measurements of the same subject (test-retest)

Thus, we can extend the classical ACDE model by subtracting this new term, ϵ, from the uncommon environment, e^2, as follows:

$$\begin{aligned} u^2_{ext} &= e^2 - \epsilon \\ &= 1 - r_{MZ} - (1 - r_{TR}) \\ u^2_{ext} &= r_{TR} - r_{MZ} \end{aligned} \tag{4.7}$$

Note that *Equation 4.7* applies to both ACE and ADE models, as the formula for computing the effect of uncommon environment is the same for both the case.

4.2.3 Computation of ACE model for original and optimally reconstructed connectomes

In this chapter, we will use the extended ACE model to estimate the effect of genetics and environmental factors on the brain functional connectivity. We will compute this measure for every edge in two sets of connectomes: original and optimally reconstructed using the PCA-based differential identifiability algorithm as shown in *Chapter 3*. For this purpose, we use the results generated in *Section 3.3.2* where we compare the differential identifiability profiles of the test/retest, MZ twins, and DZ twins cohorts. We first obtain the the number of PCA components, say m, required for optimum differential identifiability achieved for the test/retest cohort. Then, we reconstruct the FCs of all three cohorts using m number of components from their respective PCA decompositions. This way, we impose the number of components on the reconstructed MZ and DZ twin FCs based on test/retest FCs. Then, we compute the edgewise correlations between test/retest, MZ twins, and DZ twins separately.

This gives us one matrix each for r_{TR}, r_{MZ}, and r_{DZ}, respectively which we then use to compute the extended ACE model for the entire functional connectome.

4.3 Results

As mentioned in *Section 4.2*, we have conducted edgewise ACE model [121][118] computations for all fMRI conditions present in the HCP dataset, parcellated according to the mutligranular Schaefer atlas scheme [120]. This section will provide the results of these computations for resting state and motor task-based FCs, while the results for the remaining tasks have been provided in the Appendix.

The top block of *Figure 4.3* shows the edgewise additive genetics, a^2, for each of the Schaefer parcellations. The first row corresponds to the a^2 for the original data while the second row shows a^2 for the reconstructed data using the optimum number of principal components for the test/retest cohort. As can be seen, the overall value of the a^2 increases for all the Schaefer parcellations after PCA-based reconstruction. The other three blocks in *Figure 4.3* are the corresponding common environment (c^2), uncommon environment (u^2), and measurement error (ϵ) matrices, respectively. The same pattern as a^2, that of increased values after reconstruction, can be observed in the c^2 and u^2 matrices, although the gain achieved in c^2 values is very low. On the other hand, the values in the ϵ matrices reduce noticeably after optimal reconstruction.

Figure 4.4 shows the corresponding extended ACE model results for the original and reconstructed motor task-based FCs and the violin plots for the heritability values, respectively. As can be seen from *Figure 4.4*, the pattern of increased a^2, c^2, and u^2 is similar to the ones observed in resting-state FCs, although not as pronounced. Additionally, notice that, for both rest and motor FCs, the effect of common environment on the FC is low, with or without the PCA-based reconstruction. Another observation that can be made is the high measurement error in motor task, which does not reduce as much as in rest FCs after reconstruction. These results hold true for the other task-based fMRI conditions as well, shown in the *Appendix B*.

4.4 Discussion

In this chapter, we estimated the effects of genetics and environment through the use of classical twin ACE model on every edge of the functional connectome. We use the extensive monozygotic and dizygotic twin data included in the Human Connectome Project dataset for this purpose. We have also proposed an extension on the classical ACE model to disentangle the effects of measurement error and unique environments with the use of the test-retest data available in the Human Connectome Project.

Going forward, we plan to also employ the APACE model [243][116] to estimate the effects of genetics and environment on functional connectivity in the human brain. APACE algorithm speeds up the computation time by inferring the ACE model from the squared differences of paired individuals. This algorithm also gives us the ability to include nuisance variables such as age, sex, age^2, (age$\times$sex), etc.

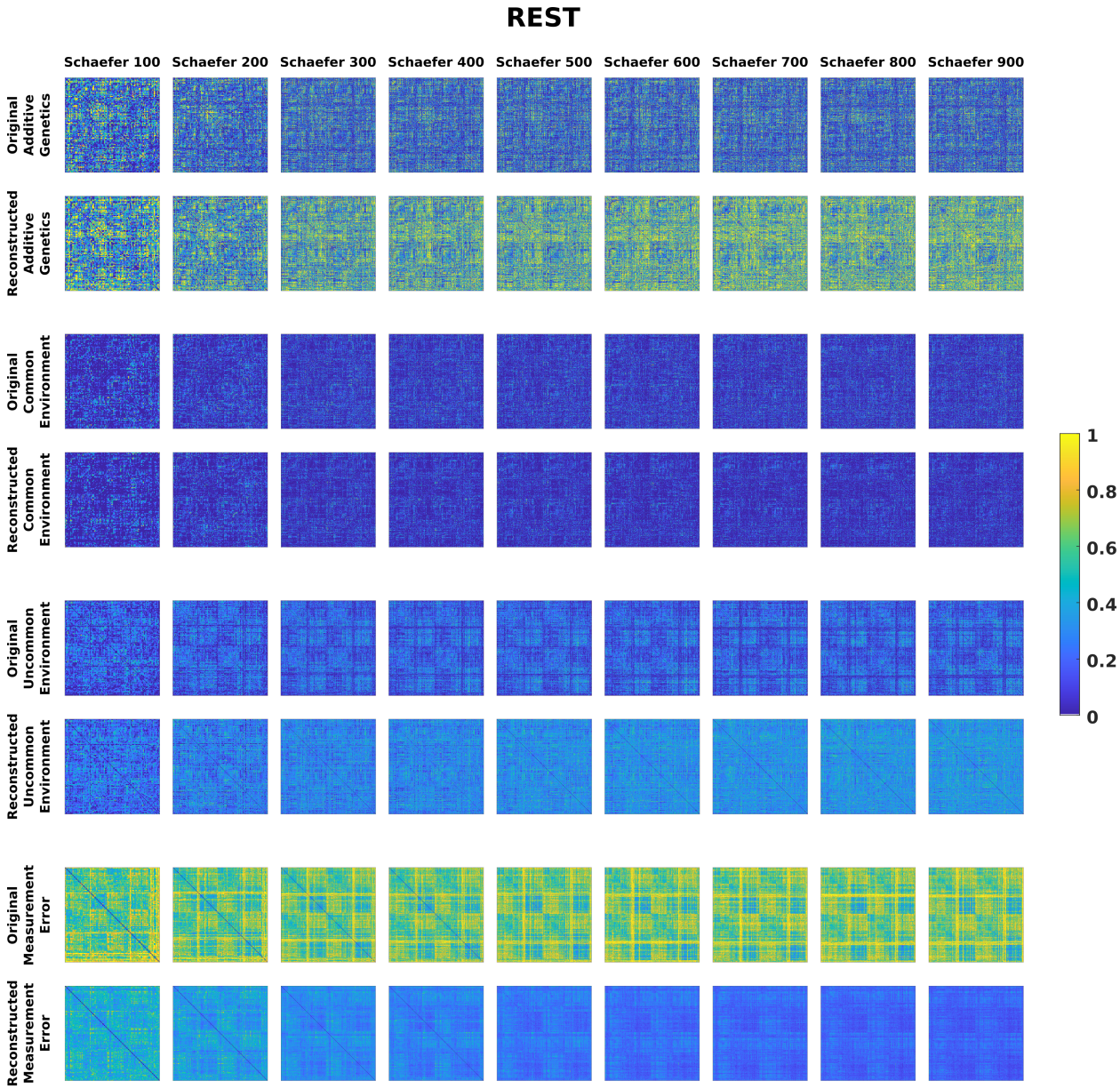


Figure 4.3. Edgewise extended ACE model results for the Schaefer parcellations with increasing granularity for resting-state FCs

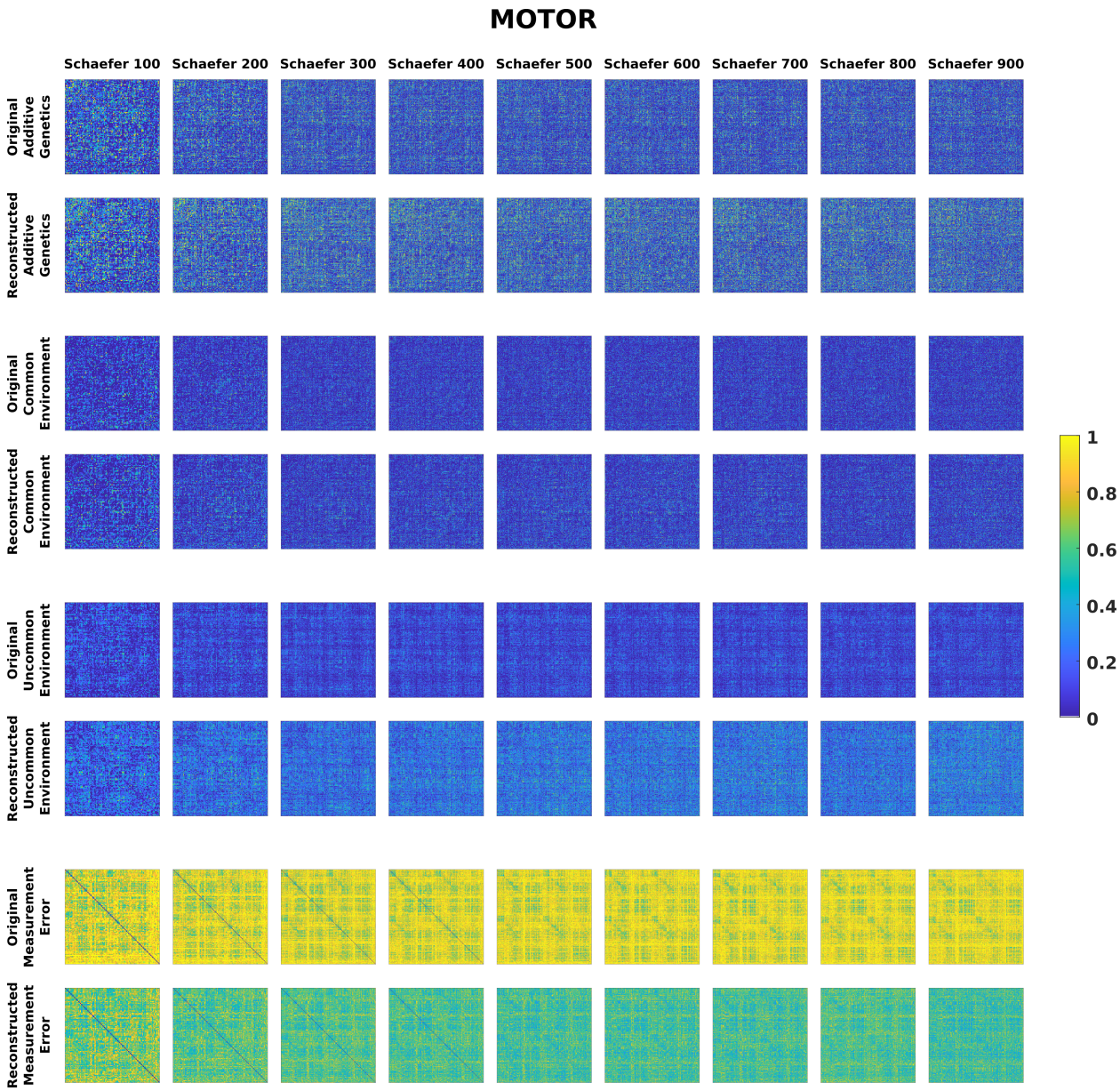


Figure 4.4. Edgewise extended ACE model results for the Schaefer parcellations with increasing granularity for motor task-based FCs

5. SUMMARY AND FUTURE WORK

Chapter 2 [62] of this dissertation demonstrated that the ant colony algorithm is an effective tool to emulate the communication processes taking place in the human brain. The framework presented in this paper combines a complex network topology tested by an ant-colony algorithm that, by means of two perception exponents, namely *ph* and *edge perception*, allows to simulate different communication regimes and to capture the most important path ensembles involved on the communication of each pair of source and target nodes in a brain network. With the use of this cooperative learning algorithm, which models the communication in a network as a nest of ants foraging for food and communicating with each other indirectly through pheromones, we have shown evidence of being able to establish associations between SC and FC when subjects are in different cognitive states as they are performing different tasks. Future work for this project will focus on running the algorithm on higher resolution parcellations and on individual connectomes instead of a group-average connectome. The framework can also be applied to other systems that can be modeled as networks, such as social interactions, to study the communication processes in them.

Chapter 3 [119] focussed on the processing pipeline that we have developed to extract brain region-level time-series and the subsequent functional connectomes for each session and all the fMRI conditions of all the subjects in the Human Connectome Project dataset. We also assessed the subject-level and twin fingerprint between the FCs of the cohort of Unrelated subjects, MZ twins, and DZ twins using the identifiability framework (see *Figures 3.5*, *3.6*, and *3.7*). The future work for this project will focus on processing and analyzing the structural data included in the Human Connectome Project.

Lastly, in *Chapter 4*, we estimated the effects of genetics and environment through the use of classical twin ACE model on every edge of the functional connectome. We use the extensive monozygotic and dizygotic twin data included in the Human Connectome Project dataset for this purpose. We have also proposed an extension on the classical ACE model to disentangle the effects of measurement error and unique environments with the use of the test-retest data available in the Human Connectome Project. Going forward, we plan to also employ the APACE model [243][116] to estimate the effects of genetics and environment on

functional connectivity in the human brain. APACE algorithm speeds up the computation time by inferring the ACE model from the squared differences of paired individuals. This algorithm also gives us the ability to include nuisance variables such as age, sex, age^2, (age$\times$sex), etc.

REFERENCES

[1] R. Bammer, "Basic principles of diffusion-weighted imaging," *European journal of radiology*, vol. 45, no. 3, pp. 169–184, 2003.

[2] J. S. Phillips, A. S. Greenberg, J. A. Pyles, S. K. Pathak, M. Behrmann, W. Schneider, and M. J. Tarr, "Co-analysis of brain structure and function using fmri and diffusion-weighted imaging," *Journal of visualized experiments: JoVE*, no. 69, 2012.

[3] Y. Assaf, H. Johansen-Berg, and M. Thiebaut de Schotten, "The role of diffusion mri in neuroscience," *NMR in Biomedicine*, vol. 32, no. 4, e3762, 2019.

[4] M. Moseley, J. Kucharczyk, J. Mintorovitch, Y. Cohen, J. Kurhanewicz, N. Derugin, H. Asgari, and D. Norman, "Diffusion-weighted mr imaging of acute stroke: Correlation with t2-weighted and magnetic susceptibility-enhanced mr imaging in cats.," *American Journal of Neuroradiology*, vol. 11, no. 3, pp. 423–429, 1990.

[5] M. G. Lansberg, A. M. Norbash, M. P. Marks, D. C. Tong, M. E. Moseley, and G. W. Albers, "Advantages of adding diffusion-weighted magnetic resonance imaging to conventional magnetic resonance imaging for evaluating acute stroke," *Archives of neurology*, vol. 57, no. 9, pp. 1311–1316, 2000.

[6] M. Nakahara, K. Ericson, and B. Bellander, "Diffusion-weighted mr and apparent diffusion coefficient in the evaluation of severe brain injury," *Acta Radiologica*, vol. 42, no. 4, pp. 365–369, 2001.

[7] M. Filippi, M. Cercignani, M. Inglese, M. Horsfield, and G. Comi, "Diffusion tensor magnetic resonance imaging in multiple sclerosis," *Neurology*, vol. 56, no. 3, pp. 304–311, 2001.

[8] R. Bammer and F. Fazekas, "Diffusion imaging in multiple sclerosis," *Neuroimaging Clinics*, vol. 12, no. 1, pp. 71–106, 2002.

[9] T. Klingberg, M. Hedehus, E. Temple, T. Salz, J. D. Gabrieli, M. E. Moseley, and R. A. Poldrack, "Microstructure of temporo-parietal white matter as a basis for reading ability: Evidence from diffusion tensor magnetic resonance imaging," *Neuron*, vol. 25, no. 2, pp. 493–500, 2000.

[10] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi, "In vivo fiber tractography using dt-mri data," *Magnetic resonance in medicine*, vol. 44, no. 4, pp. 625–632, 2000.

[11] M. Hulkower, D. Poliak, S. Rosenbaum, M. Zimmerman, and M. L. Lipton, "A decade of dti in traumatic brain injury: 10 years and 100 articles later," *American Journal of Neuroradiology*, vol. 34, no. 11, pp. 2064–2074, 2013.

[12] K. Oishi, M. M. Mielke, M. Albert, C. G. Lyketsos, and S. Mori, "Dti analyses and clinical applications in alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 26, no. s3, pp. 287–296, 2011.

[13] K. O. Lim and J. Helpern, "Neuropsychiatric applications of dti–a review," *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo*, vol. 15, no. 7-8, pp. 587–593, 2002.

[14] M. Breakspear and V. K. Jirsa, "Neuronal dynamics and brain connectivity," in *Handbook of brain connectivity*, Springer, 2007, pp. 3–64.

[15] M. D'Amelio and P. M. Rossini, "Brain excitability and connectivity of neuronal assemblies in alzheimer's disease: From animal models to human findings," *Progress in neurobiology*, vol. 99, no. 1, pp. 42–60, 2012.

[16] O. Sporns, "Brain connectivity," *Scholarpedia*, vol. 2, no. 10, p. 4695, 2007.

[17] S. W. Oh, J. A. Harris, L. Ng, B. Winslow, N. Cain, S. Mihalas, Q. Wang, C. Lau, L. Kuan, A. M. Henry, *et al.*, "A mesoscale connectome of the mouse brain," *Nature*, vol. 508, no. 7495, pp. 207–214, 2014.

[18] H. Zeng, "Mesoscale connectomics," *Current opinion in neurobiology*, vol. 50, pp. 154–162, 2018.

[19] O. Sporns, *Networks of the Brain*. MIT press, 2010.

[20] P. Hagmann, "From diffusion mri to brain connectomics," EPFL, Tech. Rep., 2005.

[21] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature reviews. Neuroscience*, vol. 10, no. 3, p. 186, 2009.

[22] O. Sporns, G. Tononi, and R. Kötter, "The human connectome: A structural description of the human brain," *PLoS Comput Biol*, vol. 1, no. 4, e42, 2005.

[23] C.-H. Yeh, D. K. Jones, X. Liang, M. Descoteaux, and A. Connelly, "Mapping structural connectivity using diffusion mri: Challenges and opportunities," *Journal of Magnetic Resonance Imaging*, 2020.

[24] R. Bakker, R. Cannon, F. Howell, and R. Kötter, "The new cocomac database: A community project," *Front. Neuroinform*, 2008.

[25] P. Hagmann, L. Cammoun, X. Gigandet, S. Gerhard, P. E. Grant, V. Wedeen, R. Meuli, J.-P. Thiran, C. J. Honey, and O. Sporns, "Mr connectomics: Principles and challenges," *Journal of neuroscience methods*, vol. 194, no. 1, pp. 34–45, 2010.

[26] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns, "Mapping the structural core of human cerebral cortex," *PLoS Biol*, vol. 6, no. 7, e159, 2008.

[27] P. Hagmann, M. Kurant, X. Gigandet, P. Thiran, V. J. Wedeen, R. Meuli, and J.-P. Thiran, "Mapping human whole-brain structural networks with diffusion mri," *PloS one*, vol. 2, no. 7, e597, 2007.

[28] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fmri signal," *nature*, vol. 412, no. 6843, pp. 150–157, 2001.

[29] S. Ogawa, T.-M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *proceedings of the National Academy of Sciences*, vol. 87, no. 24, pp. 9868–9872, 1990.

[30] I.-h. Chou, "Read my mind," *Nature Physics*, vol. 4, no. 1, S17–S17, 2008.

[31] O. J. Arthurs and S. Boniface, "How well do we understand the neural origins of the fmri bold signal?" *TRENDS in Neurosciences*, vol. 25, no. 1, pp. 27–31, 2002.

[32] M. P. Van Den Heuvel and H. E. H. Pol, "Exploring the brain network: A review on resting-state fmri functional connectivity," *European neuropsychopharmacology*, vol. 20, no. 8, pp. 519–534, 2010.

[33] A. Aertsen, G. Gerstein, M. Habib, and G. Palm, "Dynamics of neuronal firing correlation: Modulation of" effective connectivity"," *Journal of neurophysiology*, vol. 61, no. 5, pp. 900–917, 1989.

[34] K. Friston, C. Frith, P. Liddle, and R. Frackowiak, "Functional connectivity: The principal-component analysis of large (pet) data sets," *Journal of Cerebral Blood Flow & Metabolism*, vol. 13, no. 1, pp. 5–14, 1993.

[35] L. Lee, L. M. Harrison, and A. Mechelli, "A report of the functional connectivity workshop, dusseldorf 2002," *Neuroimage*, vol. 19, no. 2, pp. 457–465, 2003.

[36] L. Q. Uddin, K. S. Supekar, S. Ryali, and V. Menon, "Dynamic reconfiguration of structural and functional connectivity across core neurocognitive brain networks with development," *Journal of Neuroscience*, vol. 31, no. 50, pp. 18 578–18 589, 2011.

[37] I. Stitt, K. J. Hollensteiner, E. Galindo-Leon, F. Pieper, E. Fiedler, T. Stieglitz, G. Engler, G. Nolte, and A. K. Engel, "Dynamic reconfiguration of cortical functional connectivity across brain states," *Scientific reports*, vol. 7, no. 1, pp. 1–14, 2017.

[38] A. Brovelli, J.-M. Badier, F. Bonini, F. Bartolomei, O. Coulon, and G. Auzias, "Dynamic reconfiguration of visuomotor-related functional connectivity networks," *Journal of Neuroscience*, vol. 37, no. 4, pp. 839–853, 2017.

[39] D. Duong-Tran, E. Amico, B. Corominas-Murtra, K. Abbas, M. Dzemidzic, D. Kareken, M. Ventresca, and J. Goñi, "A morphospace framework to assess configural breadth based on brain functional networks," *arXiv preprint arXiv:1901.10962*, 2019.

[40] R. F. Betzel, L. Byrge, Y. He, J. Goñi, X.-N. Zuo, and O. Sporns, "Changes in structural and functional connectivity among resting-state networks across the human lifespan," *Neuroimage*, vol. 102, pp. 345–357, 2014.

[41] A. M. Fjell, M. H. Sneve, H. Grydeland, A. B. Storsve, I. K. Amlien, A. Yendiki, and K. B. Walhovd, "Relationship between structural and functional connectivity change across the adult lifespan: A longitudinal investigation," *Human brain mapping*, vol. 38, no. 1, pp. 561–573, 2017.

[42] C. E. Han, L. R. Peraza, J.-P. Taylor, and M. Kaiser, "Predicting age across human lifespan based on structural connectivity from diffusion tensor imaging," in *2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*, IEEE, 2014, pp. 137–140.

[43] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

[44] M. J. Lowe, M. Dzemidzic, J. T. Lurito, V. P. Mathews, and M. D. Phillips, "Correlations in low-frequency bold fluctuations reflect cortico-cortical connections," *Neuroimage*, vol. 12, no. 5, pp. 582–587, 2000.

[45] R. Sala-Llonch, D. Bartrés-Faz, and C. Junqué, "Reorganization of brain networks in aging: A review of functional connectivity studies," *Frontiers in psychology*, vol. 6, p. 663, 2015.

[46] M. Glasser, T. Coalson, E. Robinson, C. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Anderson, C. Beckmann, M. Jenkinson, *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, 2015.

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

[48] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, pp. 171–178, 2016.

[49] U. Sakoglu, A. Michael, and V. Calhoun, "Classification of schizophrenia patients vs healthy controls with dynamic functional network connectivity," *Neuroimage*, vol. 47, no. 1, S39–41, 2009.

[50] U. Sakoglu and V. Calhoun, "Dynamic windowing reveals task-modulation of functional connectivity in schizophrenia patients vs healthy controls," in *Proc. Intl. Soc. Mag. Reson. Med*, vol. 17, 2009, p. 3675.

[51] U. Sakoglu and V. Calhoun, "Temporal dynamics of functional network connectivity at rest: A comparison of schizophrenia patients and healthy controls," *NeuroImage*, no. 47, S169, 2009.

[52] Ü. Sakoğlu, G. D. Pearlson, K. A. Kiehl, Y. M. Wang, A. M. Michael, and V. D. Calhoun, "A method for evaluating dynamic functional network connectivity and task-modulation: Application to schizophrenia," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 23, no. 5-6, pp. 351–366, 2010.

[53] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo, *et al.*, "Dynamic functional connectivity: Promise, issues, and interpretations," *Neuroimage*, vol. 80, pp. 360–378, 2013.

[54] J. Schumacher, L. R. Peraza, M. Firbank, A. J. Thomas, M. Kaiser, P. Gallagher, J. T. O'Brien, A. M. Blamire, and J.-P. Taylor, "Dynamic functional connectivity changes in dementia with lewy bodies and alzheimer's disease," *NeuroImage: Clinical*, vol. 22, p. 101 812, 2019.

[55] E. Fiorenzato, A. P. Strafella, J. Kim, R. Schifano, L. Weis, A. Antonini, and R. Biundo, "Dynamic functional connectivity changes associated with dementia in parkinson's disease," *Brain*, vol. 142, no. 9, pp. 2860–2872, 2019.

[56] M. Díez-Cirarda, A. P. Strafella, J. Kim, J. Peña, N. Ojeda, A. Cabrera-Zubizarreta, and N. Ibarretxe-Bilbao, "Dynamic functional connectivity in parkinson's disease patients with mild cognitive impairment and normal cognition," *NeuroImage: Clinical*, vol. 17, pp. 847–855, 2018.

[57] J. R. Cohen, "The behavioral and cognitive relevance of time-varying, dynamic changes in functional connectivity," *NeuroImage*, vol. 180, pp. 515–525, 2018.

[58] L. Douw, D. G. Wakeman, N. Tanaka, H. Liu, and S. M. Stufflebeam, "State-dependent variability of dynamic functional connectivity between frontoparietal and default networks relates to cognitive flexibility," *Neuroscience*, vol. 339, pp. 12–21, 2016.

[59] E. Damaraju, E. A. Allen, A. Belger, J. M. Ford, S. McEwen, D. Mathalon, B. Mueller, G. Pearlson, S. Potkin, A. Preda, *et al.*, "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia," *NeuroImage: Clinical*, vol. 5, pp. 298–308, 2014.

[60] S. Shakil, C.-H. Lee, and S. D. Keilholz, "Evaluation of sliding window correlation performance for characterizing dynamic functional connectivity and brain states," *Neuroimage*, vol. 133, pp. 111–128, 2016.

[61] M. Greicius, "Resting-state functional connectivity in neuropsychiatric disorders," *Current opinion in neurology*, vol. 21, no. 4, pp. 424–430, 2008.

[62] U. Tipnis, E. Amico, M. Ventresca, and J. Goni, "Modeling communication processes in the human connectome through cooperative learning," *IEEE Transactions on Network Science and Engineering*, 2018.

[63] A. Fornito, A. Zalesky, and E. Bullmore, *Fundamentals of Brain Network Analysis*. Academic Press, 2016.

[64] O. Sporns, *Discovering the human connectome*. MIT press, 2012.

[65] H. K. Hausman, A. O'Shea, J. N. Kraft, E. M. Boutzoukas, N. D. Evangelista, E. J. Van Etten, P. K. Bharadwaj, S. G. Smith, E. Porges, G. A. Hishaw, *et al.*, "The role of resting-state network functional connectivity in cognitive aging," *Frontiers in Aging Neuroscience*, vol. 12, p. 177, 2020.

[66] M. D. Fox, M. Corbetta, A. Z. Snyder, J. L. Vincent, and M. E. Raichle, "Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems," *Proceedings of the National Academy of Sciences*, vol. 103, no. 26, pp. 10 046–10 051, 2006.

[67] J. L. Vincent, A. Z. Snyder, M. D. Fox, B. J. Shannon, J. R. Andrews, M. E. Raichle, and R. L. Buckner, "Coherent spontaneous activity identifies a hippocampal-parietal memory network," *Journal of neurophysiology*, vol. 96, no. 6, pp. 3517–3531, 2006.

[68] N. U. Dosenbach, D. A. Fair, F. M. Miezin, A. L. Cohen, K. K. Wenger, R. A. Dosenbach, M. D. Fox, A. Z. Snyder, J. L. Vincent, M. E. Raichle, *et al.*, "Distinct brain networks for adaptive and stable task control in humans," *Proceedings of the National Academy of Sciences*, vol. 104, no. 26, pp. 11 073–11 078, 2007.

[69] J. L. Vincent, I. Kahn, A. Z. Snyder, M. E. Raichle, and R. L. Buckner, "Evidence for a frontoparietal control system revealed by intrinsic functional connectivity," *Journal of neurophysiology*, vol. 100, no. 6, pp. 3328–3342, 2008.

[70] M. W. Cole, S. Pathak, and W. Schneider, "Identifying the brain's most globally connected regions," *Neuroimage*, vol. 49, no. 4, pp. 3132–3148, 2010.

[71] M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman, "A default mode of brain function," *Proceedings of the National Academy of Sciences*, vol. 98, no. 2, pp. 676–682, 2001.

[72] R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter, "The brain's default network: Anatomy, function, and relevance to disease.," 2008.

[73] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar mri," *Magnetic resonance in medicine*, vol. 34, no. 4, pp. 537–541, 1995.

[74] M. De Luca, S. Smith, N. De Stefano, A. Federico, and P. M. Matthews, "Blood oxygenation level dependent contrast resting state networks are relevant to functional activity in the neocortical sensorimotor system," *Experimental brain research*, vol. 167, no. 4, pp. 587–594, 2005.

[75] J. S. Damoiseaux, S. Rombouts, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, and C. F. Beckmann, "Consistent resting-state networks across healthy subjects," *Proceedings of the national academy of sciences*, vol. 103, no. 37, pp. 13 848–13 853, 2006.

[76] B. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, *et al.*, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, vol. 106, no. 3, pp. 1125–1165, 2011.

[77] W. W. Seeley, V. Menon, A. F. Schatzberg, J. Keller, G. H. Glover, H. Kenna, A. L. Reiss, and M. D. Greicius, "Dissociable intrinsic connectivity networks for salience processing and executive control," *Journal of Neuroscience*, vol. 27, no. 9, pp. 2349–2356, 2007.

[78] M. P. Milham, J. Vogelstein, and T. Xu, "Removing the reliability bottleneck in functional magnetic resonance imaging research to achieve clinical utility," *JAMA psychiatry*,

[79] W. M. Trochim and J. P. Donnelly, "Research methods knowledge base," 2001.

[80] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical methods for rates and proportions*. john wiley & sons, 2013.

[81] X. Zuo and X. Xing, "Neuroscience and biobehavioral reviews test-retest reliabilities of resting-state fmri measurements in human brain functional connectomics: A systems neuroscience perspective," *Neuroscience and Biobehavioral Reviews*, vol. 45, pp. 100–118, 2014.

[82] X.-N. Zuo, B. B. Biswal, and R. A. Poldrack, "Reliability and reproducibility in functional connectomics," *Frontiers in neuroscience*, vol. 13, p. 117, 2019.

[83] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: Uses in assessing rater reliability.," *Psychological bulletin*, vol. 86, no. 2, p. 420, 1979.

[84] B.-E. INEQUALITY and I. BIRNBAUM-RAYMOND-ZUCKERMAN, "Encyclopedia of statistical sciences (9 vols.), samuel kotz and norman l. johnson (eds.-in-chief) and campbell," *Encyclopedia of Statistical Sciences*, p. 830, 1989.

[85] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of chiropractic medicine*, vol. 15, no. 2, pp. 155–163, 2016.

[86] C. Hedge, G. Powell, and P. Sumner, "The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences," *Behavior Research Methods*, vol. 50, no. 3, pp. 1166–1186, 2018.

[87] D. L. Streiner, G. R. Norman, and J. Cairney, *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA, 2015.

[88] R. A. Poldrack, C. I. Baker, J. Durnez, K. J. Gorgolewski, P. M. Matthews, M. R. Munafò, T. E. Nichols, J.-B. Poline, E. Vul, and T. Yarkoni, "Scanning the horizon: Towards transparent and reproducible neuroimaging research," *Nature reviews neuroscience*, vol. 18, no. 2, p. 115, 2017.

[89] E. Vul, C. Harris, P. Winkielman, and H. Pashler, "Puzzlingly high correlations in fmri studies of emotion, personality, and social cognition," *Perspectives on psychological science*, vol. 4, no. 3, pp. 274–290, 2009.

[90] D. Streiner, G. R. Norman, and J. Cairney, "Health measurement scales: A practical guide to their development and use," *Aust NZJ Public Health*, 2016.

[91] A. M. Golestani, J. B. Kwinta, Y. B. Khatamian, and J. J. Chen, "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*, vol. 11, p. 546, 2017.

[92] J. Wang, J. Han, V. T. Nguyen, L. Guo, and C. C. Guo, "Improving the test-retest reliability of resting state fmri by removing the impact of sleep," *Frontiers in neuroscience*, vol. 11, p. 249, 2017.

[93] C.-G. Yan, B. Cheung, C. Kelly, S. Colcombe, R. C. Craddock, A. Di Martino, Q. Li, X.-N. Zuo, F. X. Castellanos, and M. P. Milham, "A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics," *Neuroimage*, vol. 76, pp. 183–201, 2013.

[94] L. Parkes, B. Fulcher, M. Yücel, and A. Fornito, "An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional mri," *Neuroimage*, vol. 171, pp. 415–436, 2018.

[95] R. Ciric, D. H. Wolf, J. D. Power, D. R. Roalf, G. L. Baum, K. Ruparel, R. T. Shinohara, M. A. Elliott, S. B. Eickhoff, C. Davatzikos, *et al.*, "Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity," *Neuroimage*, vol. 154, pp. 174–187, 2017.

[96] M. A. Lindquist, S. Geuter, T. D. Wager, and B. S. Caffo, "Modular preprocessing pipelines can reintroduce artifacts into fmri data," *Human brain mapping*, vol. 40, no. 8, pp. 2358–2376, 2019.

[97] J. E. Chen, H. Jahanian, and G. H. Glover, "Nuisance regression of high-frequency functional magnetic resonance imaging data: Denoising can be noisy," *Brain connectivity*, vol. 7, no. 1, pp. 13–24, 2017.

[98] E. Amico and J. Goñi, "The quest for identifiability in human functional connectomes," *Scientific reports*, vol. 8, no. 1, p. 8254, 2018.

[99] K. Abbas, E. Amico, D. O. Svaldi, U. Tipnis, D. A. Duong-Tran, M. Liu, M. Rajapandian, J. Harezlak, B. M. Ances, and J. Goñi, "Geff: Graph embedding for functional fingerprinting," *NeuroImage*, p. 117 181, 2020.

[100] E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable, "Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity," *Nature neuroscience*, vol. 18, no. 11, pp. 1664–1671, 2015.

[101] T. Kaufmann, D. Alnæs, N. T. Doan, C. L. Brandt, O. A. Andreassen, and L. T. Westlye, "Delayed stabilization and individualization in connectome development are related to psychiatric disorders," *Nature neuroscience*, vol. 20, no. 4, pp. 513–515, 2017.

[102] O. Miranda-Dominguez, B. D. Mills, S. D. Carpenter, K. A. Grant, C. D. Kroenke, J. T. Nigg, and D. A. Fair, "Connectotyping: Model based fingerprinting of the functional connectome," *PloS one*, vol. 9, no. 11, e111048, 2014.

[103] S. Noble, M. N. Spann, F. Tokoglu, X. Shen, R. T. Constable, and D. Scheinost, "Influences on the test–retest reliability of functional connectivity mri and its relationship with behavioral utility," *Cerebral Cortex*, vol. 27, no. 11, pp. 5415–5429, 2017.

[104] M. Rajapandian, E. Amico, K. Abbas, M. Ventresca, and J. Goñi, "Uncovering differential identifiability in network properties of human brain functional connectomes," *Network Neuroscience*, vol. 4, no. 3, pp. 698–713, 2020.

[105] R. B. Mars, R. E. Passingham, and S. Jbabdi, "Connectivity fingerprints: From areal descriptions to abstract spaces," *Trends in cognitive sciences*, vol. 22, no. 11, pp. 1026–1037, 2018.

[106] D. Hu, F. Wang, H. Zhang, Z. Wu, L. Wang, W. Lin, G. Li, D. Shen, U. B. C. P. Consortium, *et al.*, "Disentangled intensive triplet autoencoder for infant functional connectome fingerprinting," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2020, pp. 72–82.

[107] G. H. Ngo, M. Khosla, K. Jamison, A. Kuceyeski, and M. R. Sabuncu, "From connectomic to task-evoked fingerprints: Individualized prediction of task contrasts from resting-state functional connectivity," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2020, pp. 62–71.

[108] B. Chiêm, K. Abbas, E. Amico, D. A. Duong-Tran, F. Crevecoeur, and J. Goñi, "Improving functional connectome fingerprinting with degree-normalization," *arXiv preprint arXiv:2011.10079*, 2020.

[109] D. O. Svaldi, J. Goñi, K. Abbas, E. Amico, D. G. Clark, C. Muralidharan, M. Dzemidzic, J. D. West, S. L. Risacher, A. J. Saykin, *et al.*, "Optimizing differential identifiability improves connectome predictive modeling of cognitive deficits in alzheimer\'s disease," *arXiv preprint arXiv:1908.06197*, 2019.

[110] C. Sripada, A. Taxali, M. Angstadt, and S. Rutherford, "Boost in test-retest reliability in resting state fmri with predictive modeling," *BioRxiv*, p. 796 714, 2020.

[111] S. Bari, E. Amico, N. Vike, T. M. Talavage, and J. Goñi, "Uncovering multi-site identifiability based on resting-state functional connectomes," *NeuroImage*, vol. 202, p. 115 967, 2019.

[112] T. Ge, A. J. Holmes, R. L. Buckner, J. W. Smoller, and M. R. Sabuncu, "Heritability analysis with repeat measurements and its application to resting-state functional

connectivity," *Proceedings of the National Academy of Sciences*, vol. 114, no. 21, pp. 5521–5526, 2017.

[113] L. De Gennaro, C. Marzano, F. Fratello, F. Moroni, M. C. Pellicciari, F. Ferlazzo, S. Costa, A. Couyoumdjian, G. Curcio, E. Sforza, *et al.*, "The electroencephalographic fingerprint of sleep is genetically determined: A twin study," *Annals of neurology*, vol. 64, no. 4, pp. 455–460, 2008.

[114] K. Kumar, M. Toews, L. Chauvin, O. Colliot, and C. Desrosiers, "Multi-modal brain fingerprinting: A manifold approximation based framework," *NeuroImage*, vol. 183, pp. 212–226, 2018.

[115] A. Gritsenko, M. Lindquist, and M. K. Chung, "Twin classification in resting-state brain connectivity," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2020, pp. 1391–1394.

[116] G. L. Colclough, S. M. Smith, T. E. Nichols, A. M. Winkler, S. N. Sotiropoulos, M. F. Glasser, D. C. Van Essen, and M. W. Woolrich, "The heritability of multi-modal connectivity in human brain activity," *Elife*, vol. 6, e20178, 2017.

[117] D. V. Demeter, L. E. Engelhardt, R. Mallett, E. M. Gordon, T. Nugiel, K. P. Harden, E. M. Tucker-Drob, J. A. Lewis-Peacock, and J. A. Church, "Functional connectivity fingerprints at rest are similar across youths and adults and vary with genetic similarity," *Iscience*, vol. 23, no. 1, p. 100 801, 2020.

[118] G. A. Blokland, M. A. Mosing, K. H. Verweij, and S. E. Medland, "Twin studies and behavior genetics," *The Oxford handbook of quantitative methods in psychology*, vol. 2, pp. 198–218, 2013.

[119] U. Tipnis, K. Abbas, E. Tran, E. Amico, L. Shen, A. D. Kaplan, and J. Goñi, "Processed functional connectomes for the hcp young adult: Data release and assessment on brain fingerprints," *arXiv preprint arXiv:2011.05212*, 2020.

[120] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. Yeo, "Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri," *Cerebral Cortex*, vol. 28, no. 9, pp. 3095–3114, 2018.

[121] K. Ozaki, H. Toyoda, N. Iwama, S. Kubo, and J. Ando, "Using non-normal sem to resolve the acde model in the classical twin design," *Behavior genetics*, vol. 41, no. 2, pp. 329–339, 2011.

[122] A. Avena-Koenigsberger, B. Misic, and O. Sporns, "Communication dynamics in complex brain networks," *Nature Reviews Neuroscience*, vol. 19, no. 1, p. 17, 2018.

[123] D. R. Amancio, "Comparing the topological properties of real and artificially generated scientific manuscripts," *Scientometrics*, vol. 105, no. 3, pp. 1763–1779, 2015.

[124] D. R. Amancio, O. N. Oliveira Jr, and L. da Fontoura Costa, "Identification of literary movements using complex networks to represent texts," *New Journal of Physics*, vol. 14, no. 4, p. 043 029, 2012.

[125] A. E. Motter, A. P. De Moura, Y.-C. Lai, and P. Dasgupta, "Topology of the conceptual network of language," *Physical Review E*, vol. 65, no. 6, p. 065 102, 2002.

[126] D. R. Amancio, "Probing the topological properties of complex networks modeling short written texts," *PloS one*, vol. 10, no. 2, e0118394, 2015.

[127] H. Bunke and K. Shearer, "A graph distance metric based on the maximal common subgraph," *Pattern recognition letters*, vol. 19, no. 3-4, pp. 255–259, 1998.

[128] J.-L. Deneubourg, S. Aron, S. Goss, and J. M. Pasteels, "The self-organizing exploratory pattern of the argentine ant," *Journal of Insect Behavior*, vol. 3, no. 2, pp. 159–168, 1990.

[129] S. Goss, S. Aron, J.-L. Deneubourg, and J. M. Pasteels, "Self-organized shortcuts in the argentine ant," *Naturwissenschaften*, vol. 76, no. 12, pp. 579–581, 1989.

[130] M. Dorigo, E. Bonabeau, and G. Theraulaz, "Ant algorithms and stigmergy," *Future Generation Computer Systems*, vol. 16, no. 8, pp. 851–871, 2000.

[131] A. Colorni, M. Dorigo, V. Maniezzo, *et al.*, "Distributed optimization by ant colonies," in *Proceedings of the First European Conference on Artificial Life*, Paris, France, vol. 142, 1991, pp. 134–142.

[132] S. R. y Cajal, *Histology of the Nervous System of Man and Vertebrates*. Oxford University Press, USA, 1995, vol. 1.

[133] F. D. V. Fallani, F. A. Rodrigues, L. da Fontoura Costa, L. Astolfi, F. Cincotti, D. Mattia, S. Salinari, and F. Babiloni, "Multiple pathways analysis of brain functional networks from eeg signals: An application to real data," *Brain Topography*, vol. 23, no. 4, pp. 344–354, 2011.

[134] A. Avena-Koenigsberger, B. Mišić, R. X. Hawkins, A. Griffa, P. Hagmann, J. Goñi, and O. Sporns, "Path ensembles and a tradeoff between communication efficiency and resilience in the human connectome," *Brain Structure and Function*, vol. 222, no. 1, pp. 603–618, 2017.

[135] E. Estrada and N. Hatano, "Communicability in complex networks," *Physical Review E*, vol. 77, no. 3, p. 036 111, 2008.

[136] M. Chavez, F. D. V. Fallani, M. Valencia, J. Artieda, D. Mattia, V. Latora, and F. Babiloni, "Node accessibility in cortical networks during motor tasks," *Neuroinformatics*, vol. 11, no. 3, pp. 355–366, 2013.

[137] D. C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S. W. Curtiss, *et al.*, "The human connectome project: A data acquisition perspective," *Neuroimage*, vol. 62, no. 4, pp. 2222–2231, 2012.

[138] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium, *et al.*, "The wu-minn human connectome project: An overview," *Neuroimage*, vol. 80, pp. 62–79, 2013.

[139] B. Fischl, A. Van Der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, *et al.*, "Automatically parcellating the human cerebral cortex," *Cerebral cortex*, vol. 14, no. 1, pp. 11–22, 2004.

[140] C. Destrieux, B. Fischl, A. Dale, and E. Halgren, "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature," *Neuroimage*, vol. 53, no. 1, pp. 1–15, 2010.

[141] S. N. Sotiropoulos, S. Jbabdi, J. Xu, J. L. Andersson, S. Moeller, E. J. Auerbach, M. F. Glasser, M. Hernandez, G. Sapiro, M. Jenkinson, *et al.*, "Advances in diffusion mri acquisition and processing in the human connectome project," *Neuroimage*, vol. 80, pp. 125–143, 2013.

[142] K. Uğurbil, J. Xu, E. J. Auerbach, S. Moeller, A. T. Vu, J. M. Duarte-Carvajalino, C. Lenglet, X. Wu, S. Schmitter, P. F. Van de Moortele, *et al.*, "Pushing spatial and temporal resolution for functional and diffusion mri in the human connectome project," *Neuroimage*, vol. 80, pp. 80–104, 2013.

[143] E. O. Stejskal and J. E. Tanner, "Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient," *The Journal of Chemical Physics*, vol. 42, no. 1, pp. 288–292, 1965.

[144] J. Tournier, F. Calamante, A. Connelly, *et al.*, "Mrtrix: Diffusion tractography in crossing fiber regions," *International Journal of Imaging Systems and Technology*, vol. 22, no. 1, pp. 53–66, 2012.

[145] E. Amico and J. Goñi, "Mapping hybrid functional-structural connectivity traits in the human connectome," *Network Neuroscience*, pp. 1–17, 2018.

[146] R. E. Smith, J.-D. Tournier, F. Calamante, and A. Connelly, "Anatomically-constrained tractography: Improved diffusion mri streamlines tractography through effective use of anatomical information," *Neuroimage*, vol. 62, no. 3, pp. 1924–1938, 2012.

[147] D. Christiaens, M. Reisert, T. Dhollander, S. Sunaert, P. Suetens, and F. Maes, "Global tractography of multi-shell diffusion-weighted imaging data using a multi-tissue model," *Neuroimage*, vol. 123, pp. 89–101, 2015.

[148] B. Jeurissen, J.-D. Tournier, T. Dhollander, A. Connelly, and J. Sijbers, "Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data," *NeuroImage*, vol. 103, pp. 411–426, 2014.

[149] R. E. Smith, J. D. Tournier, F. Calamante, and A. Connelly, "Sift2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography," *NeuroImage*, vol. 119, pp. 338–351, 2015.

[150] R. E. Smith, J.-D. Tournier, F. Calamante, and A. Connelly, "Sift: Spherical-deconvolution informed filtering of tractograms," *Neuroimage*, vol. 67, pp. 298–312, 2013.

[151] D. M. Barch, G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt, *et al.*, "Function in the human connectome: Task-fmri and individual differences in behavior," *Neuroimage*, vol. 80, pp. 169–189, 2013.

[152] S. M. Smith, C. F. Beckmann, J. Andersson, E. J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D. A. Feinberg, L. Griffanti, M. P. Harms, *et al.*, "Resting-state fmri in the human connectome project," *Neuroimage*, vol. 80, pp. 144–168, 2013.

[153] E. Amico, D. Marinazzo, C. Di Perri, L. Heine, J. Annen, C. Martial, M. Dzemidzic, M. Kirsch, V. Bonhomme, S. Laureys, *et al.*, "Mapping the functional connectome traits of levels of consciousness," *NeuroImage*, vol. 148, pp. 201–211, 2017.

[154] J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Methods to detect, characterize, and remove motion artifact in resting state fmri," *Neuroimage*, vol. 84, pp. 320–341, 2014.

[155] D. S. Marcus, J. Harwell, T. Olsen, M. Hodge, M. F. Glasser, F. Prior, M. Jenkinson, T. Laumann, S. W. Curtiss, and D. C. Van Essen, "Informatics and data mining tools and strategies for the human connectome project," *Frontiers in Neuroinformatics*, vol. 5, 2011.

[156] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[157] R. W. Floyd, "Algorithm 97: Shortest path," *Communications of the ACM*, vol. 5, no. 6, p. 345, 1962.

[158] C. M. Grinstead and J. L. Snell, *Introduction to Probability*. American Mathematical Soc., 2012.

[159] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, "The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations," *Behavioral Ecology and Sociobiology*, vol. 54, no. 4, pp. 396–405, 2003.

[160] N. Attiratanasunthron and J. Fakcharoenphol, "A running time analysis of an ant colony optimization algorithm for shortest paths in directed acyclic graphs," *Information Processing Letters*, vol. 105, no. 3, pp. 88–92, 2008.

[161] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science*, vol. 296, no. 5569, pp. 910–913, 2002.

[162] S. Hanhijärvi, G. C. Garriga, and K. Puolamäki, "Randomization techniques for graphs," in *Proceedings of the 2009 SIAM International Conference on Data Mining*, SIAM, 2009, pp. 780–791.

[163] T. Harris and F. Ross, "Fundamentals of a method for evaluating rail net capacities," RAND CORP SANTA MONICA CA, Tech. Rep., 1955.

[164] B. Mišić, R. F. Betzel, A. Nematzadeh, J. Goñi, A. Griffa, P. Hagmann, A. Flammini, Y.-Y. Ahn, and O. Sporns, "Cooperative and competitive spreading dynamics on the human connectome," *Neuron*, vol. 86, no. 6, pp. 1518–1529, 2015.

[165] A. Pascual-Leone, C. Freitas, L. Oberman, J. C. Horvath, M. Halko, M. Eldaief, S. Bashir, M. Vernet, M. Shafi, B. Westover, *et al.*, "Characterizing brain cortical plasticity and network dynamics across the age-span in health and disease with tms-eeg and tms-fmri," *Brain Topography*, vol. 24, no. 3-4, p. 302, 2011.

[166] E. Amico, O. Bodart, M. Rosanova, O. Gosseries, L. Heine, P. Van Mierlo, C. Martial, M. Massimini, D. Marinazzo, and S. Laureys, "Tracking dynamic interactions between structural and functional connectivity: A tms/eeg-dmri study," *Brain Connectivity*, vol. 7, no. 2, pp. 84–97, 2017.

[167] S. Y. Bookheimer, D. H. Salat, M. Terpstra, B. M. Ances, D. M. Barch, R. L. Buckner, G. C. Burgess, S. W. Curtiss, M. Diaz-Santos, J. S. Elam, *et al.*, "The lifespan human connectome project in aging: An overview," *NeuroImage*, vol. 185, pp. 335–348, 2019.

[168] L. H. Somerville, S. Y. Bookheimer, R. L. Buckner, G. C. Burgess, S. W. Curtiss, M. Dapretto, J. S. Elam, M. S. Gaffrey, M. P. Harms, C. Hodge, *et al.*, "The lifespan human connectome project in development: A large-scale study of brain connectivity development in 5–21 year olds," *Neuroimage*, vol. 183, pp. 456–468, 2018.

[169] N. E. Allen, C. Sudlow, T. Peakman, R. Collins, *et al.*, *Uk biobank data: Come and get it*, 2014.

[170] K. L. Miller, F. Alfaro-Almagro, N. K. Bangerter, D. L. Thomas, E. Yacoub, J. Xu, A. J. Bartsch, S. Jbabdi, S. N. Sotiropoulos, J. L. Andersson, *et al.*, "Multimodal population brain imaging in the uk biobank prospective epidemiological study," *Nature neuroscience*, vol. 19, no. 11, pp. 1523–1536, 2016.

[171] R. C. Petersen, P. Aisen, L. A. Beckett, M. Donohue, A. Gamst, D. J. Harvey, C. Jack, W. Jagust, L. Shaw, A. Toga, *et al.*, "Alzheimer's disease neuroimaging initiative (adni): Clinical characterization," *Neurology*, vol. 74, no. 3, pp. 201–209, 2010.

[172] A. Makropoulos, E. C. Robinson, A. Schuh, R. Wright, S. Fitzgibbon, J. Bozek, S. J. Counsell, J. Steinweg, K. Vecchiato, J. Passerat-Palmbach, *et al.*, "The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction," *Neuroimage*, vol. 173, pp. 88–112, 2018.

[173] J. D. Power, C. J. Lynch, B. Adeyemo, and S. E. Petersen, "A critical, event-related appraisal of denoising in resting-state fmri studies," *Cerebral Cortex*, 2020.

[174] J. D. Power, M. Plitt, S. J. Gotts, P. Kundu, V. Voon, P. A. Bandettini, and A. Martin, "Ridding fmri data of motion-related influences: Removal of signals with distinct spatial and physical bases in multiecho data," *Proceedings of the National Academy of Sciences*, vol. 115, no. 9, E2105–E2114, 2018.

[175] J. D. Power, B. L. Schlaggar, and S. E. Petersen, "Recent progress and outstanding issues in motion correction in resting state fmri," *Neuroimage*, vol. 105, pp. 536–551, 2015.

[176] G. C. Burgess, S. Kandala, D. Nolan, T. O. Laumann, J. D. Power, B. Adeyemo, M. P. Harms, S. E. Petersen, and D. M. Barch, "Evaluation of denoising strategies to address motion-correlated artifacts in resting-state functional magnetic resonance imaging data from the human connectome project," *Brain connectivity*, vol. 6, no. 9, pp. 669–680, 2016.

[177] M. Salehi, A. S. Greene, A. Karbasi, X. Shen, D. Scheinost, and R. T. Constable, "There is no single functional atlas even for a single individual: Functional parcel definitions change with task," *NeuroImage*, vol. 208, p. 116 366, 2020.

[178] K. Abbas, M. Liu, M. Venkatesh, E. Amico, J. Harezlak, A. D. Kaplan, M. Ventresca, L. Pessoa, and J. Goñi, "Regularization of functional connectomes and its impact on geodesic distance and fingerprinting," *arXiv preprint arXiv:2003.05393*, 2020.

[179] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[180] K. Murphy and M. D. Fox, "Towards a consensus regarding global signal regression for resting state functional connectivity mri," *Neuroimage*, vol. 154, pp. 169–173, 2017.

[181] T. T. Liu, A. Nalci, and M. Falahpour, "The global signal in fmri: Nuisance or information?" *NeuroImage*, vol. 150, pp. 213–229, 2017.

[182] S. Hayasaka, "Functional connectivity networks with and without global signal correction," *Frontiers in human neuroscience*, vol. 7, p. 880, 2013.

[183] H. Xu, J. Su, J. Qin, M. Li, L.-L. Zeng, D. Hu, and H. Shen, "Impact of global signal regression on characterizing dynamic functional connectivity and brain states," *Neuroimage*, vol. 173, pp. 127–145, 2018.

[184] S. J. Gotts, Z. S. Saad, H. J. Jo, G. L. Wallace, R. W. Cox, and A. Martin, "The perils of global signal regression for group comparisons: A case study of autism spectrum disorders," *Frontiers in human neuroscience*, vol. 7, p. 356, 2013.

[185] Z. S. Saad, S. J. Gotts, K. Murphy, G. Chen, H. J. Jo, A. Martin, and R. W. Cox, "Trouble at rest: How correlation patterns and group differences become distorted after global signal regression," *Brain connectivity*, vol. 2, no. 1, pp. 25–32, 2012.

[186] A. R. Hariri, S. M. Brown, D. E. Williamson, J. D. Flory, H. De Wit, and S. B. Manuck, "Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity," *Journal of Neuroscience*, vol. 26, no. 51, pp. 13 213–13 217, 2006.

[187] M. R. Delgado, L. E. Nystrom, C. Fissell, D. Noll, and J. A. Fiez, "Tracking the hemodynamic responses to reward and punishment in the striatum," *Journal of neurophysiology*, vol. 84, no. 6, pp. 3072–3077, 2000.

[188] J. R. Binder, W. L. Gross, J. B. Allendorfer, L. Bonilha, J. Chapin, J. C. Edwards, T. J. Grabowski, J. T. Langfitt, D. W. Loring, M. J. Lowe, *et al.*, "Mapping anterior temporal lobe language areas with fmri: A multicenter normative study," *Neuroimage*, vol. 54, no. 2, pp. 1465–1475, 2011.

[189] R. Smith, K. Keramatian, and K. Christoff, "Localizing the rostrolateral prefrontal cortex at the individual level," *Neuroimage*, vol. 36, no. 4, pp. 1387–1396, 2007.

[190] F. Castelli, F. Happé, U. Frith, and C. Frith, "Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns," *Neuroimage*, vol. 12, no. 3, pp. 314–325, 2000.

[191] F. Castelli, C. Frith, F. Happé, and U. Frith, "Autism, asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes," *Brain*, vol. 125, no. 8, pp. 1839–1849, 2002.

[192] T. Wheatley, S. C. Milleville, and A. Martin, "Understanding animate agents: Distinct roles for the social network and mirror system," *Psychological science*, vol. 18, no. 6, pp. 469–474, 2007.

[193] S. J. White, D. Coniston, R. Rogers, and U. Frith, "Developing the frith-happé animations: A quick and objective test of theory of mind for adults with autism," *Autism Research*, vol. 4, no. 2, pp. 149–154, 2011.

[194] A. Drobyshevsky, S. B. Baumann, and W. Schneider, "A rapid fmri task battery for mapping of visual, motor, cognitive, and emotional function," *Neuroimage*, vol. 31, no. 2, pp. 732–744, 2006.

[195] A. Caceres, D. L. Hall, F. O. Zelaya, S. C. Williams, and M. A. Mehta, "Measuring fmri reliability with the intra-class correlation coefficient," *Neuroimage*, vol. 45, no. 3, pp. 758–768, 2009.

[196] P. E. Downing, Y. Jiang, M. Shuman, and N. Kanwisher, "A cortical area selective for visual processing of the human body," *Science*, vol. 293, no. 5539, pp. 2470–2473, 2001.

[197] M. V. Peelen and P. E. Downing, "Within-subject reproducibility of category-specific visual activation with functional mri," *Human brain mapping*, vol. 25, no. 4, pp. 402–408, 2005.

[198] J. C. Taylor, A. J. Wiggett, and P. E. Downing, "Functional mri analysis of body and body part representations in the extrastriate and fusiform body areas," *Journal of neurophysiology*, vol. 98, no. 3, pp. 1626–1633, 2007.

[199] C. J. Fox, G. Iaria, and J. J. Barton, "Defining the face processing network: Optimization of the functional localizer in fmri," *Human brain mapping*, vol. 30, no. 5, pp. 1637–1651, 2009.

[200] C.-C. Kung, J. J. Peissig, and M. J. Tarr, "Is region-of-interest overlap comparison a reliable measure of category specificity?" *Journal of Cognitive Neuroscience*, vol. 19, no. 12, pp. 2019–2034, 2007.

[201] D. C. Van Essen, M. F. Glasser, D. L. Dierker, J. Harwell, and T. Coalson, "Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases," *Cerebral cortex*, vol. 22, no. 10, pp. 2241–2262, 2012.

[202] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith, "Automatic denoising of functional mri data: Combining independent component analysis and hierarchical fusion of classifiers," *Neuroimage*, vol. 90, pp. 449–468, 2014.

[203] L. Griffanti, G. Salimi-Khorshidi, C. F. Beckmann, E. J. Auerbach, G. Douaud, C. E. Sexton, E. Zsoldos, K. P. Ebmeier, N. Filippini, C. E. Mackay, *et al.*, "Ica-based artefact removal and accelerated fmri acquisition for improved resting state network imaging," *Neuroimage*, vol. 95, pp. 232–247, 2014.

[204] K. M. Aquino, B. D. Fulcher, L. Parkes, K. Sabaroedin, and A. Fornito, "Identifying and removing widespread signal deflections from fmri data: Rethinking the global signal regression problem," *NeuroImage*, vol. 212, p. 116 614, 2020.

[205] C. A. Prescott and K. S. Kendler, "Twin study design," *Alcohol Health and Research World*, vol. 19, no. 3, p. 200, 1995.

[206] X. Deng and X. Tian, "Nonlinear process fault pattern recognition using statistics kernel pca similarity factor," *Neurocomputing*, vol. 121, pp. 298–308, 2013.

[207] P.-C. Hsieh and P.-C. Tung, "A novel hybrid approach based on sub-pattern technique and whitened pca for face recognition," *Pattern Recognition*, vol. 42, no. 5, pp. 978–984, 2009.

[208] J. V. Manjón, P. Coupé, L. Concha, A. Buades, D. L. Collins, and M. Robles, "Diffusion weighted image denoising using overcomplete local pca," *PloS one*, vol. 8, no. 9, e73021, 2013.

[209] A. De Cheveigné and J. Z. Simon, "Denoising based on time-shift pca," *Journal of neuroscience methods*, vol. 165, no. 2, pp. 297–305, 2007.

[210] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507–521, 1915.

[211] M. Negishi, R. Martuzzi, E. J. Novotny, D. D. Spencer, and R. T. Constable, "Functional mri connectivity as a predictor of the surgical outcome of epilepsy," *Epilepsia*, vol. 52, no. 9, pp. 1733–1740, 2011.

[212] M. Hampson, N. Driesen, J. K. Roth, J. C. Gore, and R. T. Constable, "Functional connectivity between task-positive and task-negative brain areas and its relation to working memory performance," *Magnetic resonance imaging*, vol. 28, no. 8, pp. 1051–1057, 2010.

[213] D. Tomasi and N. D. Volkow, "Resting functional connectivity of language networks: Characterization and reproducibility," *Molecular psychiatry*, vol. 17, no. 8, pp. 841–854, 2012.

[214] M. D. Fox, H. Liu, and A. Pascual-Leone, "Identification of reproducible individualized targets for treatment of depression with tms based on intrinsic connectivity," *Neuroimage*, vol. 66, pp. 151–160, 2013.

[215] K. M. Mark, A. Pike, R. M. Latham, and B. R. Oliver, "Using twins to better understand sibling relationships," *Behavior genetics*, vol. 47, no. 2, pp. 202–214, 2017.

[216] V. Pallarés, A. Insabato, A. Sanjuán, S. Kühn, D. Mantini, G. Deco, and M. Gilson, "Extracting orthogonal subject-and condition-specific signatures from fmri data using whole-brain effective connectivity," *Neuroimage*, vol. 178, pp. 238–254, 2018.

[217] J. Liu, X. Liao, M. Xia, and Y. He, "Chronnectome fingerprinting: Identifying individuals and predicting higher cognitive functions using dynamic brain connectivity patterns," *Human brain mapping*, vol. 39, no. 2, pp. 902–915, 2018.

[218] L. Byrge and D. P. Kennedy, "High-accuracy individual identification using a "thin slice" of the functional connectome," *Network Neuroscience*, vol. 3, no. 2, pp. 363–383, 2019.

[219] S. Mueller, D. Wang, M. D. Fox, B. T. Yeo, J. Sepulcre, M. R. Sabuncu, R. Shafee, J. Lu, and H. Liu, "Individual variability in functional connectivity architecture of the human brain," *Neuron*, vol. 77, no. 3, pp. 586–595, 2013.

[220] J. Faskowitz, F. Z. Esfahlani, Y. Jo, O. Sporns, and R. F. Betzel, "Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture," Nature Publishing Group, Tech. Rep., 2020.

[221] M. Venkatesh, J. Jaja, and L. Pessoa, "Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification," *NeuroImage*, vol. 207, p. 116 398, 2020.

[222] T. D. Satterthwaite, C. H. Xia, and D. S. Bassett, "Personalized neuroscience: Common and individual-specific features in functional brain networks," *Neuron*, vol. 98, no. 2, pp. 243–245, 2018.

[223] C. Gratton, T. O. Laumann, A. N. Nielsen, D. J. Greene, E. M. Gordon, A. W. Gilmore, S. M. Nelson, R. S. Coalson, A. Z. Snyder, B. L. Schlaggar, *et al.*, "Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation," *Neuron*, vol. 98, no. 2, pp. 439–452, 2018.

[224] G. Koeppen-Schomerus, F. M. Spinath, and R. Plomin, "Twins and non-twin siblings: Different estimates of shared environmental influence in early childhood," *Twin Research and Human Genetics*, vol. 6, no. 2, pp. 97–105, 2003.

[225] C. Seguin, Y. Tian, and A. Zalesky, "Network communication models improve the behavioral and functional predictive utility of the human structural connectome," *bioRxiv*, 2020.

[226] E. Výtvarová, J. Fousek, M. Bartoň, R. Mareček, M. Gajdoš, M. Lamoš, M. Nováková, T. Slavíček, I. Peterlik, and M. Mikl, "The impact of diverse preprocessing pipelines on brain functional connectivity," in *2017 25th European Signal Processing Conference (EUSIPCO)*, IEEE, 2017, pp. 2644–2648.

[227] K. J. Holzinger, "The relative effect of nature and nurture influences on twin differences.," *Journal of Educational Psychology*, vol. 20, no. 4, p. 241, 1929.

[228] D. Carmelli, G. E. Swan, D. Robinette, and R. Fabsitz, "Genetic influence on smoking—a study of male twins," *New England Journal of Medicine*, vol. 327, no. 12, pp. 829–833, 1992.

[229] R. Rieger, A. Michaelis, and M. M. Green, *Glossary of genetics and cytogenetics: classical and molecular*. Springer Science & Business Media, 2012.

[230] A. J. Griffiths, W. M. Gelbart, R. C. Lewontin, and J. H. Miller, *Modern genetic analysis: integrating genes and genomes*. Macmillan, 2002, vol. 1.

[231] D. Bourguet, "The evolution of dominance," *Heredity*, vol. 83, no. 1, pp. 1–4, 1999.

[232] L. J. Eaves, K. Last, N. G. Martin, and J. L. Jinks, "A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences," *British Journal of Mathematical and Statistical Psychology*, vol. 30, no. 1, pp. 1–42, 1977.

[233] J. L. Jinks and D. W. Fulker, "Comparison of the biometrical genetical, mava, and classical approaches to the analysis of the human behavior.," *Psychological bulletin*, vol. 73, no. 5, p. 311, 1970.

[234] K. A. Bollen, "Structural equations with latent variables wiley," *New York*, 1989.

[235] D. M. Evans, N. Gillespie, and N. Martin, "Biometrical genetics," *Biological psychology*, vol. 61, no. 1-2, pp. 33–51, 2002.

[236] C. E. Bruder, A. Piotrowski, A. A. Gijsbers, R. Andersson, S. Erickson, T. D. de Ståhl, U. Menzel, J. Sandgren, D. von Tell, A. Poplawski, *et al.*, "Phenotypically concordant and discordant monozygotic twins display different dna copy-number-variation profiles," *The American Journal of Human Genetics*, vol. 82, no. 3, pp. 763–771, 2008.

[237] J. C. Loehlin and R. C. Nichols, *Heredity, environment, and personality: A study of 850 sets of twins*. University of Texas Press, 2012.

[238] D. J. Cohen, E. Dibble, J. M. Grawe, and W. Pollin, "Separating identical from fraternal twins," *Archives of General Psychiatry*, vol. 29, no. 4, pp. 465–469, 1973.

[239] K. S. Kendler, A. C. Heath, N. G. Martin, and L. J. Eaves, "Symptoms of anxiety and symptoms of depression: Same genes, different environments?" *Archives of general psychiatry*, vol. 44, no. 5, pp. 451–457, 1987.

[240] A. Priyadarshi, S. A. Khuder, E. A. Schaub, and S. Shrivastava, "A meta-analysis of parkinson's disease and exposure to pesticides.," *Neurotoxicology*, vol. 21, no. 4, pp. 435–440, 2000.

[241] D. Falconer and T. Mackay, "Introduction to quantitative genetics. longman, harlow, england.," *Introduction to quantitative genetics. 4th ed. Longman, Harlow, England.*, 1996.

[242] R. Plomin, J. C. DeFries, and J. C. Loehlin, "Genotype-environment interaction and correlation in the analysis of human behavior.," *Psychological bulletin*, vol. 84, no. 2, p. 309, 1977.

[243] X. Chen, T. Nichols, *et al.*, "Apace: Accelerated permutation inference for the ace model," *OHBM 2014*, 2014.

A. APPENDIX

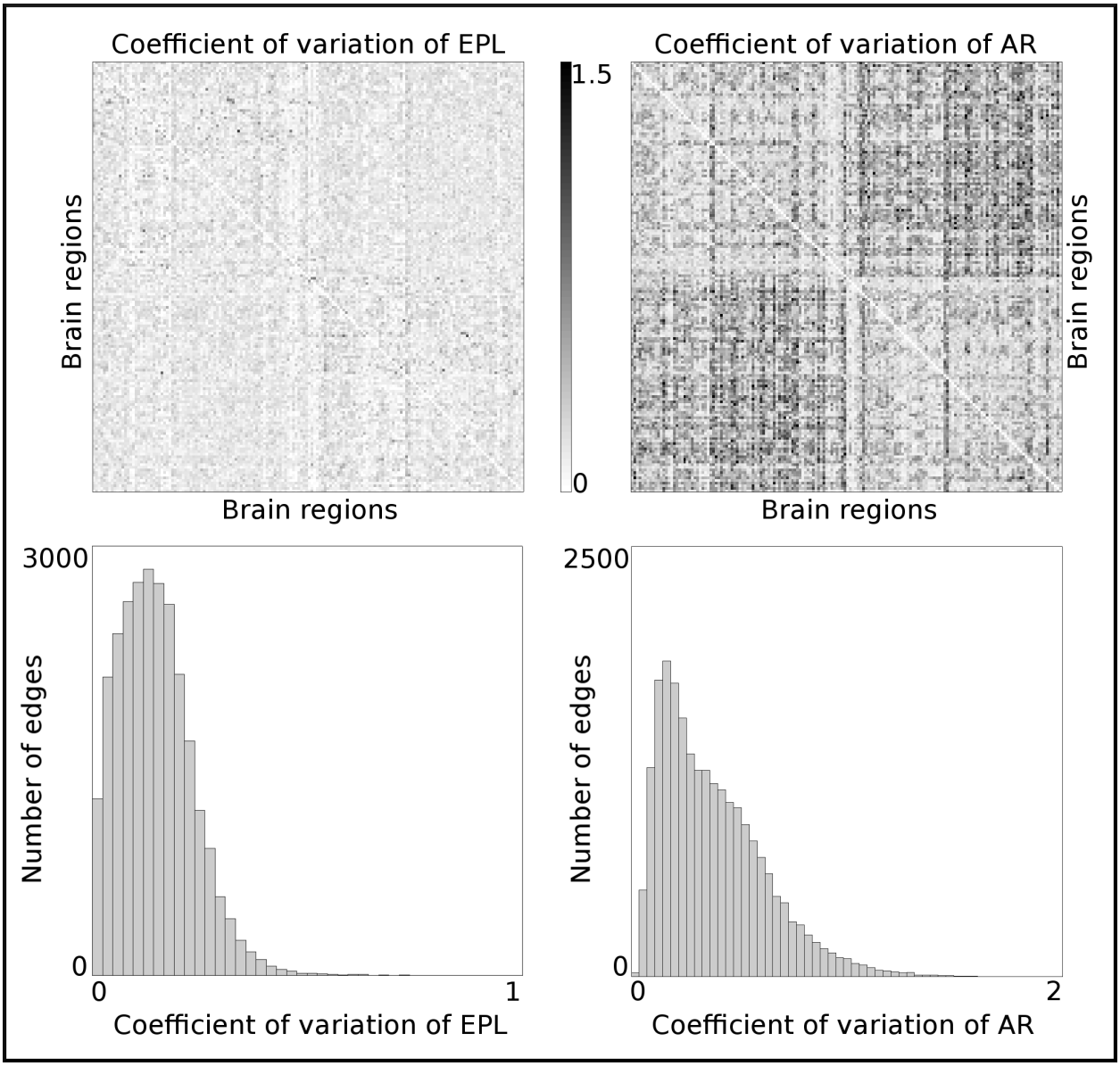


Figure A.1. This panel shows the node pairwise coefficient of variation for EPL and AR (for $\alpha = 1.5$ and $\beta = 0.1$) and the distributions of these values based on the 10 simulation runs.

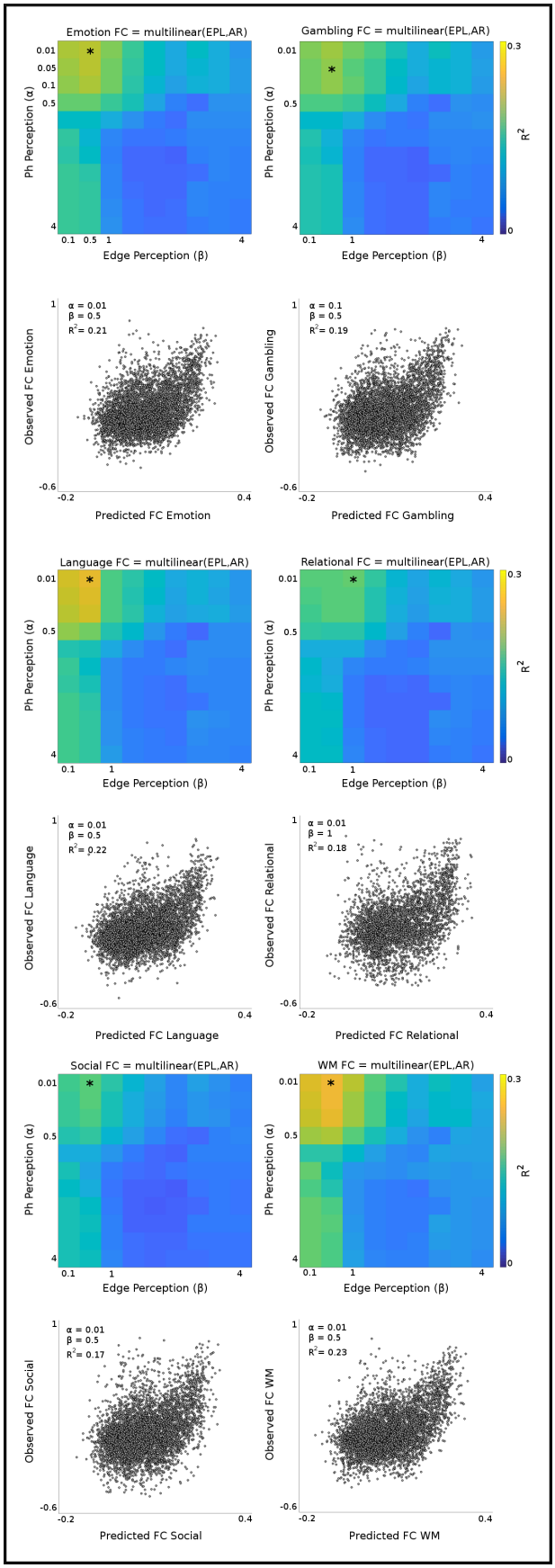


Figure A.2Multi-linear regression analyses with EPL and AR as predictor variables and task-based FCs (Emotion, Gambling, Language, Relational, Social, and Working Memory) as dependent variables (one for each model). Heatmaps denote R^2 obtained for different configurations of Ph- and Edge-Perception (α and β respectively). Subsequent scatter plots (predicted vs observed FC values) corresponding to the highest R^2 for each dependent variable (denoted by a * in the heatmap) are included as well.

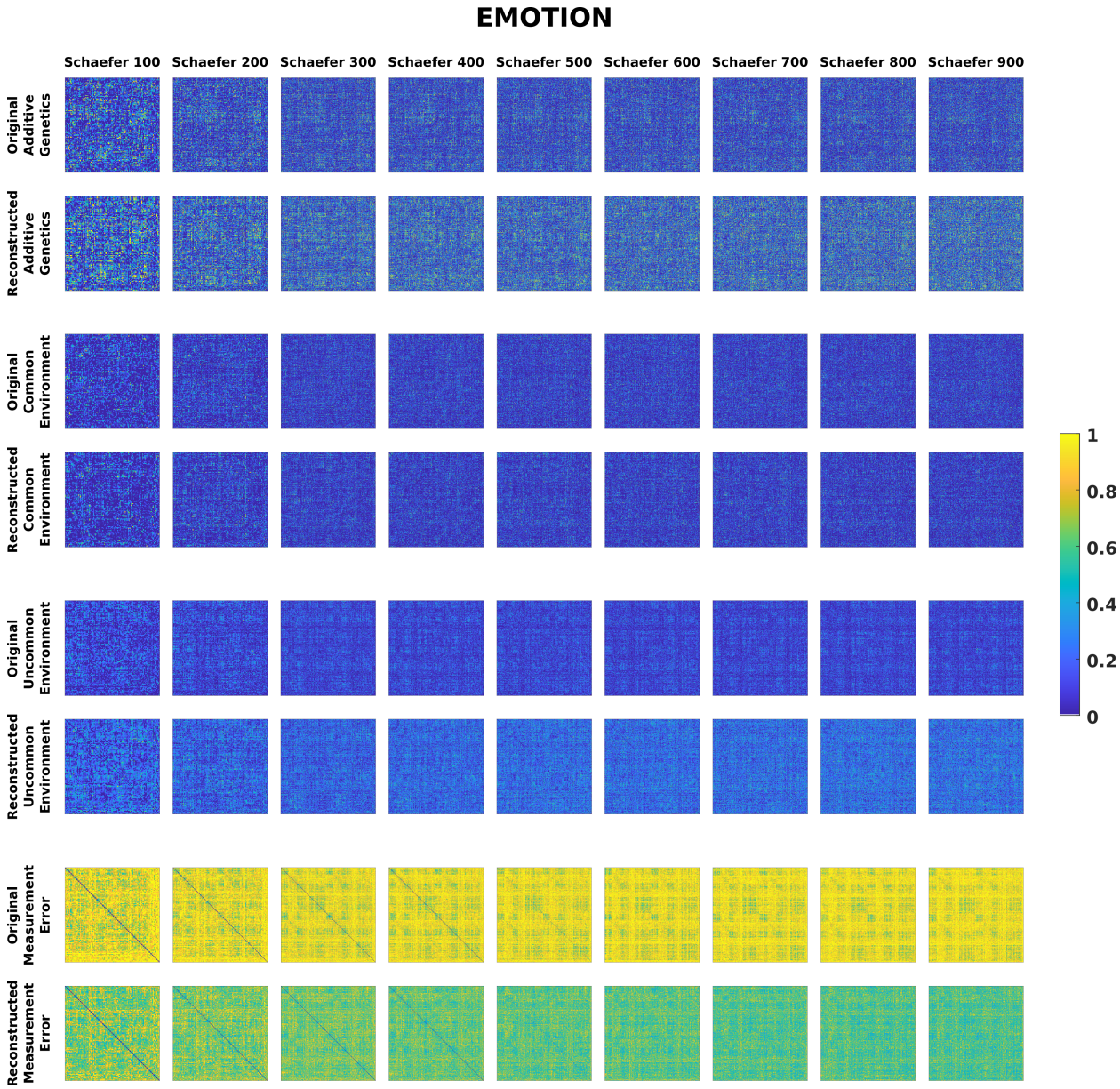


Figure B.1. Edgewise extended ACE model results for the Schaefer parcellations with increasing granularity for emotion task-based FC

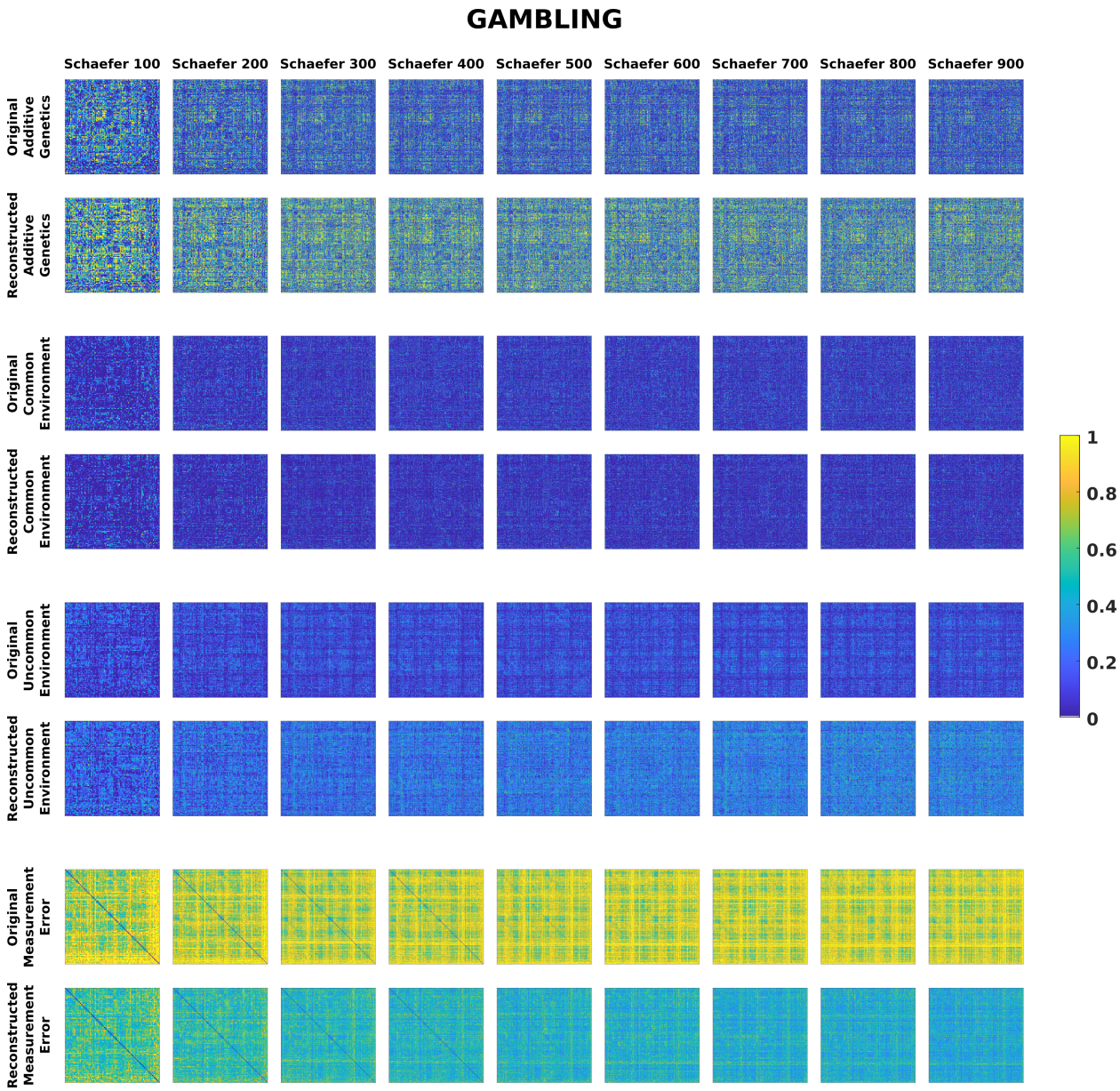


Figure B.2. Edgewise extended ACE model results for the Schaefer parcellations with increasing granularity for gambling task-based FC

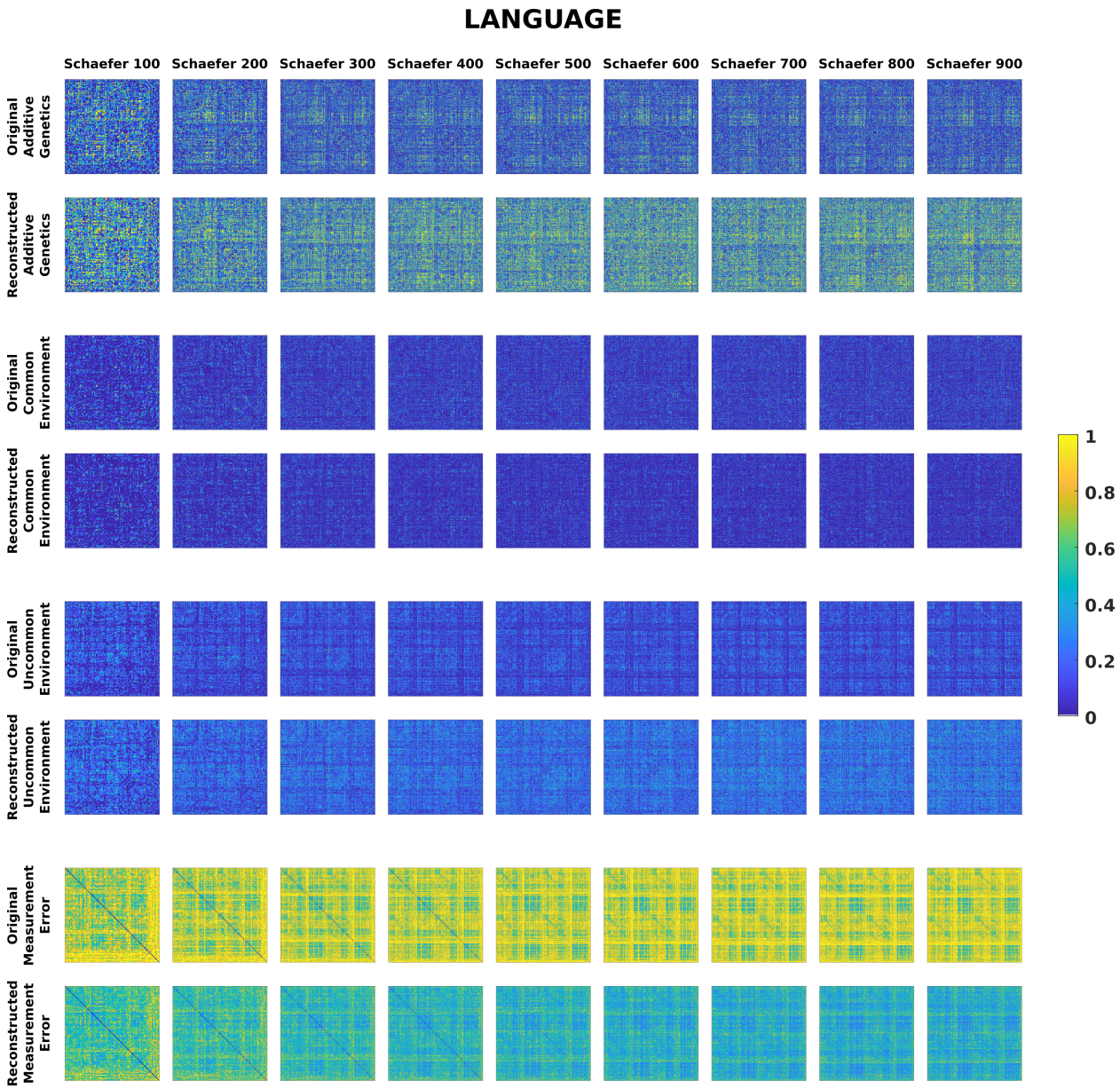


Figure B.3. Edgewise extended ACE model results for the Schaefer parcellations with increasing granularity for language task-based FC

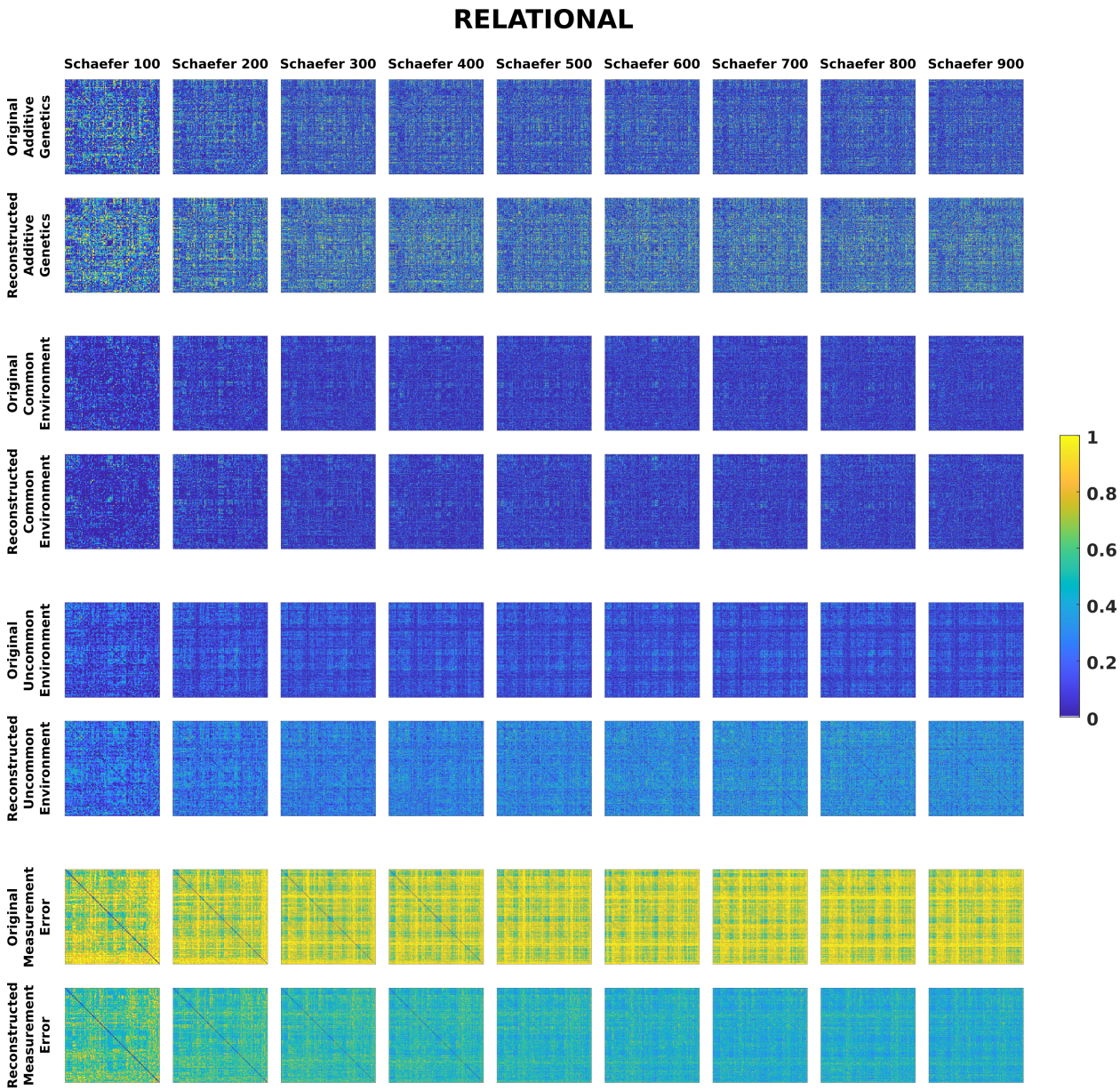


Figure B.4. Edgewise extended ACE model results for the Schaefer parcellations with increasing granularity for relational task-based FC

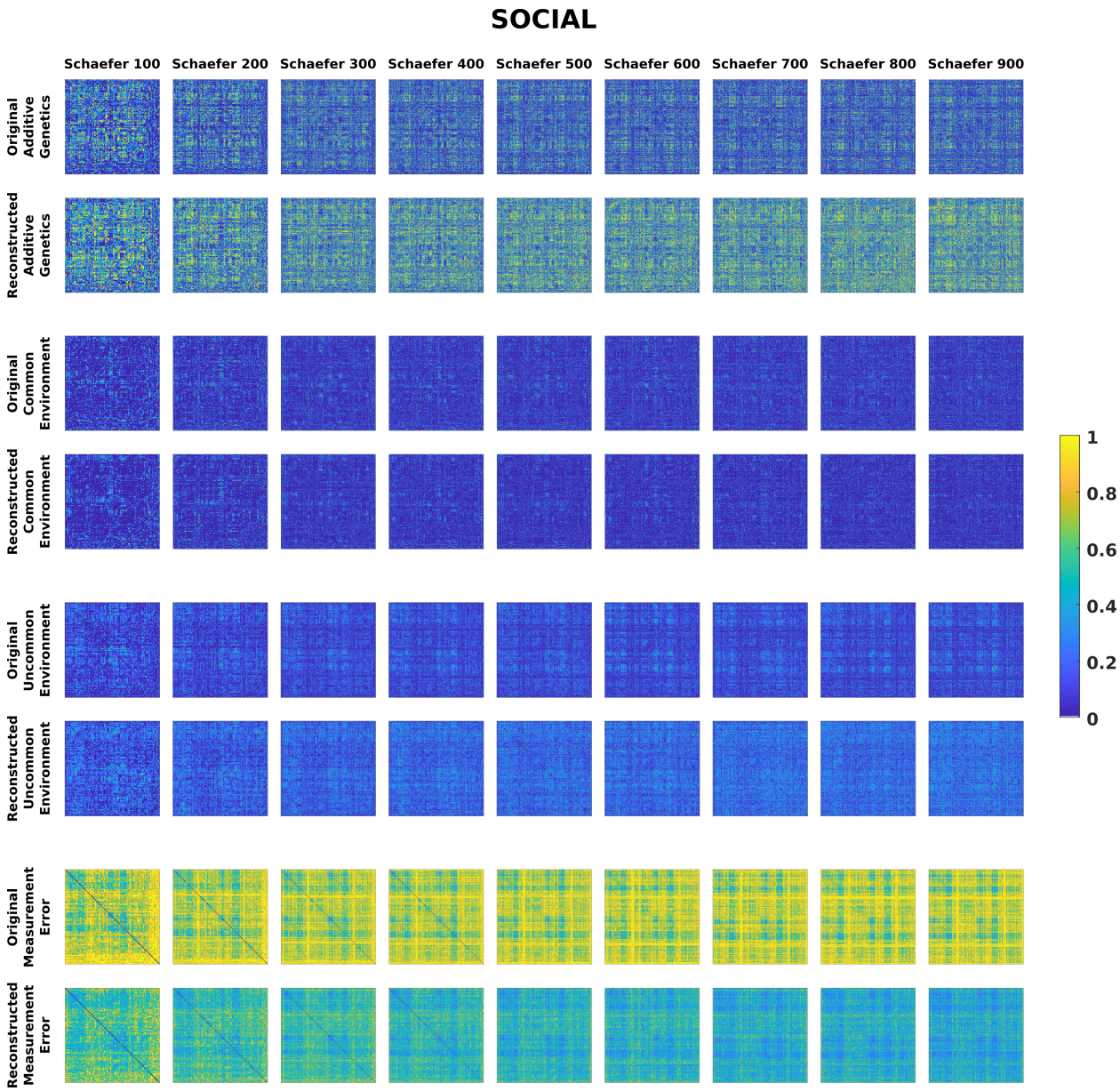


Figure B.5. Edgewise extended ACE model results for the Schaefer parcellations with increasing granularity for social task-based FC

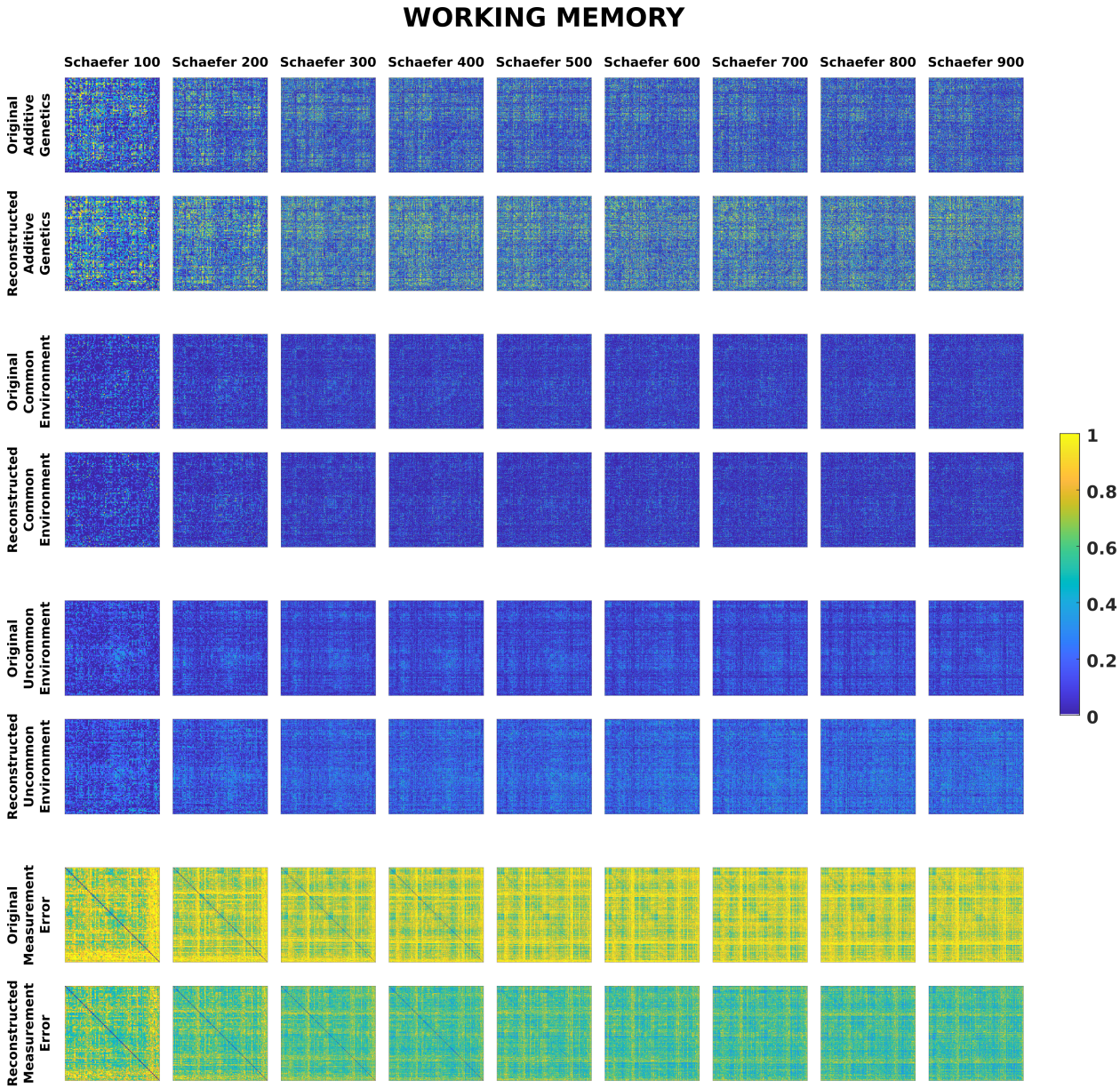


Figure B.6. Edgewise extended ACE model results for the Schaefer parcellations with increasing granularity for working memory task-based FC

VITA

Uttara Tipnis is a PhD candidate at the school of Industrial Engineering, Purdue University, West Lafayette, IN. She received her Bachelors degree in Mechanical Engineering from Pune University, India in 2013 and Masters degree in Industrial Engineering and Management from Oklahoma State University, Stillwater, OK in 2016. Her research interests include network neuroscience, data science, image and signal processing, neuroimaging, machine learning, and information theory.