Florida International University
FIU Digital Commons

FIU Electronic Theses and Dissertations | University Graduate School

6-24-2021

Understanding Individual Differences within Large-scale Brain Networks across Cognitive Contexts

Katherine L. Bottenhorn
Florida International University, kbott006@fiu.edu

Follow this and additional works at: https://digitalcommons.fiu.edu/etd

Part of the Cognitive Neuroscience Commons, Computational Neuroscience Commons, and the School Psychology Commons

Recommended Citation

Bottenhorn, Katherine L., "Understanding Individual Differences within Large-scale Brain Networks across Cognitive Contexts" (2021). *FIU Electronic Theses and Dissertations*. 4776.
https://digitalcommons.fiu.edu/etd/4776

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

UNDERSTANDING INDIVIDUAL DIFFERENCES WITHIN LARGE-SCALE BRAIN NETWORKS ACROSS COGNITIVE CONTEXTS

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

by

Katherine L. Bottenhorn

2021

To: Dean Michael Heithaus
College of Arts, Sciences, and Education

This dissertation, written by Katherine L. Bottenhorn, and entitled Understanding Individual Differences within Large-Scale Brain Networks across Cognitive Contexts, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Matthew T. Sutherland

Erica D. Musser

Shanna Burke

Timothy Hayes

Angela R. Laird, Major Professor

Date of Defense: June 24, 2021

The dissertation of Katherine L. Bottenhorn is approved.

Dean Michael Heithaus
College of Arts, Sciences, and Education

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2021

© Copyright 2021 by Katherine L. Bottenhorn

All rights reserved.

DEDICATION

To my mother, from whom I inherited my formidable command of the English language. To my father, who urged me not to get a liberal arts degree (so instead I got two). To my family, from whom there was never a doubt that I could succeed and excel in this endeavor. And to a series of wonderful women who mentored me throughout this journey and provided me with all the role models I will ever need. And, finally, to all the girls and women I can inspire and bring up here with me.

ABSTRACT OF THE DISSERTATION

UNDERSTANDING INDIVIDUAL DIFFERENCES WITHIN LARGE-SCALE BRAIN NETWORKS ACROSS COGNITIVE CONTEXTS

by

Katherine L. Bottenhorn

Florida International University, 2021

Miami, Florida

Professor Angela R. Laird, Major Professor

Historically, human neuroimaging has studied brain regions activated during behavior and how they differ between groups of people. This approach has improved our understanding of healthy and disordered brain function, but has two key shortcomings. First, focusing on brain activation restricts how we understand the brain, ignoring vital, behind-the-scenes processing. In the past decade, the focus has shifted to communication between brain regions, or connectivity, revealing networks that exhibit subtle, consistent differences across behaviors and diagnoses. Without activation-focused research's constraints, connectivity-focused neuroimaging research more comprehensively assesses brain function. Second, focusing on group differences ignores substantial within-group heterogeneity and often imposes false dichotomies. Recent findings show that brain network variability within an individual is nearly as great as across a group. Altogether, this illustrates a need for understanding individual variability in brain networks and how it relates to behavior. Therefore, I have developed a pipeline for investigating individual differences in brain connectivity, adapting robust statistical methods to address unique challenges of neuroimaging data analysis. Here, I describe this pipeline and apply it to two datasets. First, I explore between-individual variability in brain connectivity underlying intelligence and academic performance to better understand factors

contributing to student success. Second, I assess the relative contributions of stress, sleep, and hormones to within-individual variability in brain connectivity across the menstrual cycle to illuminate little-studied phenomena affecting the everyday lives of half the population. Finally, I introduce a novel signal processing workflow for cleaning electrophysiological measures of bodily stress and arousal in neuroimaging research.

TABLE OF CONTENTS

CHAPTER PAGE

1. INTRODUCTION 1
1.1 Group vs. individual differences in human neuroimaging research 2
1.2 Challenges to individual differences research in human neuroimaging . . . 4
1.3 Large-scale functional brain networks 5

2. INDIVIDUAL DIFFERENCES IN FUNCTIONAL BRAIN CONNECTIVITY 7
2.1 Statement of need 8
2.2 State of the field 9
2.3 Pipeline 10
2.3.1 Connectivity 10
2.3.2 Networking 11
2.3.3 Data Organization 12
2.3.4 Statistical Inference 12

3. INTELLIGENCE, ACADEMIC PERFORMANCE, AND THE BRAIN . 13
3.1 Introduction 13
3.2 Methods 19
3.2.1 Participants and Study Design 19
3.2.2 Missing Data 20
3.2.3 Behavioral Measures 20
3.2.4 fMRI Tasks 21
3.2.5 fMRI Acquisition and Pre-Processing 22
3.2.6 Inferential Statistics 26
3.2.7 Data and Code Availability 27
3.3 Results 28
3.3.1 WAIS-IV scores increased over a semester of physics learning regardless of sex or classroom 28
3.3.2 Physics task accuracy, but not course grade, is related to WAIS-IV scores differently for male and female students 31
3.4 Discussion 41
3.4.1 Are WAIS gains related to physics education or general college education? 42
3.4.2 Perceptual reasoning improvements underlie physics-related cognition. 43
3.4.3 Brain network organization during physics cognition is related to intelligence, but does not explain its relationship with physics task accuracy. 46
3.4.4 Limitations and future directions 48
3.5 Conclusion 49

4. ENDOCRINE INFLUENCES ON WITHIN-INDIVIDUAL VARIABILITY IN BRAIN AND MOOD . . . 51
4.1 Introduction . . . 51
4.2 Methods . . . 56
4.2.1 Participants and Study Design . . . 56
4.2.2 Hormone Data . . . 57
4.2.3 Behavioral Data . . . 58
4.2.4 Neuroimaging Data Preprocessing . . . 58
4.2.5 Neuroimaging Data Analysis . . . 63
4.2.6 Inferential Statistics . . . 64
4.3 Results . . . 66
4.3.1 Analysis 1: Lifestyle and Endocrine Collinearity . . . 67
4.3.2 Analysis 2: Connectome Similarity . . . 68
4.3.3 Analysis 3: Connectome Variability . . . 70
4.3.4 Analysis 4: Topological Variability . . . 80
4.4 Discussion . . . 80
4.4.1 Menstrual cycle, sleep, and stress . . . 81
4.4.2 Logistical contributions to connectome variability. . . . 81
4.4.3 Endocrine contributions to connectome variability. . . . 82
4.4.4 Lifestyle and endocrine overlap in connectome variability. . . . 85
4.4.5 Limitations . . . 85
4.4.6 Advancing women's health . . . 86

5. CONCURRENT fMRI-PHYSIOLOGICAL STUDY OF PSYCHOPHYSIOLOGICAL VARIABILITY . . . 88
5.1 Introduction . . . 88
5.2 Methods . . . 91
5.2.1 Physiological recordings . . . 91
5.2.2 BOLD EPI Sequences . . . 91
5.2.3 Software tools . . . 93
5.2.4 Denoising electrophysiological recordings . . . 93
5.3 Results . . . 94
5.3.1 Frequencies of MR-related artifacts . . . 94
5.3.2 Denoising electrocardiogram recordings . . . 95
5.3.3 Denoising electrodermal activity recordings . . . 96
5.3.4 Quantitative assessments of denoised data . . . 97
5.3.5 Research Products . . . 103
5.4 Discussion . . . 106
5.4.1 Limitations and Considerations . . . 109
5.4.2 Conclusions . . . 109

6. CONCLUSION . . . 111

VITA 147

LIST OF TABLES

TABLE PAGE

3.1 Average change in WAIS scores pre- to post-instruction 29

3.2 Significant relations between physics task performance and WAIS scores. 34

3.3 Mediation of the relations between changes in full-scale IQ and physics reasoning accuracy by functional connectivity. 39

4.1 Descriptive statistics of participant and lifestyle variables across the sample 66

4.2 Correlations between hormone, lifestyle, and logistical variables 68

4.3 Proportions of variance in functional connectivity explained by endocrine, lifestyle, and logistical variables 71

4.4 Variability in network topology associated with endocrine, lifestyle, and logistical variables 79

LIST OF FIGURES

FIGURE PAGE

2.1 IDConn pipeline illustration. 11

3.1 fMRI tasks performed by students at both pre- and post-instruction time points. 17

3.2 Shen et al. (2013) and Craddock et al. (2012) region-wise brain parcellations. 24

3.3 Pre-instruction, post-instruction, and pre- to post-instruction changes in WAIS-IV scores and subscores. 28

3.4 Change in WAIS scores and subscores pre- to post-instruction. 30

3.5 Associations between post-instruction physics task performance and WAIS scores. 33

3.6 Associations between functional connectivity during physics-related cognition and WAIS scores following a semester of physics instruction. . 38

3.7 Mediation of the association between post-instruction physics reasoning and FSIQ by functional connectivity. 40

4.1 Normalized estradiol and progesterone concentrations. 67

4.2 Resting-state functional connectome similarity across subjects and sessions. 69

4.3 Resting-state brain network similarity across categorical variables. . . . 70

4.4 Resting-state functional connectome variance explained by predictor. . . 72

4.5 Overall resting-state functional connectome variance explained, region-wise. 74

4.6 Resting-state functional connectivity variance explained by predictor, region-wise. 75

4.7 Overall resting-state functional connectome variance explained, system-wise. 76

4.8 Resting-state functional connectivity variance explained by predictor, system-wise. 77

5.1 Single-band, single-echo vs. multiband, multi-echo echo planar imaging pulse sequences. 90

5.2 Electrocardiogram recordings through the denoising process. 95

5.3 Electrodermal activity recordings through the denoising process. 97

5.4 Magnitude squared coherence between raw and cleaned ECG recordings. 98

5.5 Average correlation between multiband, single-echo-concurrent electrocardiogram recordings. 100

5.6 Average correlation between multiband, multi-echo-concurrent electrocardiogram recordings. 101

5.7 Magnitude squared coherence between raw and cleaned EDA recordings. 102

5.8 Average correlation between multiband, single-echo-concurrent electrodermal activity recordings. 104

5.9 Average correlation between multiband, multi-echo-concurrent electrodermal activity recordings. 105

CHAPTER 1

INTRODUCTION

For three decades now, cognitive neuroscience has utilized functional magnetic resonance imaging (fMRI) technology for *in vivo* mapping of human brain activation via the blood-oxygenation level-dependent (BOLD) signal (Ogawa et al., 1990; Kwong et al., 1992; Kwong, 2012). Research using BOLD fMRI has changed our understanding of brain plasticity, furthered our understanding of how memory works, improved our understanding of reward processing, and revealed a persistent, large-scale network organization in humans and nonhuman primates (Rosen and Savoy, 2012). Neuroscientists and neuroimaging researchers have demonstrated that this large-scale network organization shares topological principles not only with brain networks across species, but also with other complex systems from air traffic to social groups. More recently, fMRI research has revealed that while this network organization is largely consistent across individuals, it exhibits temporally stable variations between individuals, in addition to nuanced variability on even a daily basis (Gratton et al., 2018; Seitzman et al., 2019; Pritschet et al., 2020). As interest in personalized, or precision, medicine has grown, so too has an interest in precision neuroimaging (Goldstein-Piekarski et al., 2020; D'Esposito, 2019) and an increased attention to nuanced differences in the brain both within and between people. This interest is supported by recent improvements in MR technology (Lynch et al., 2020), advancements in statistics (Kundu et al., 2017; Gordon et al., 2017), and popularity of dense-sampling approaches (Poldrack et al., 2015; Poldrack, 2017). Understanding individual variability is not only crucial to precision approaches and studies of individual differences, it has implications for neuroimaging research regardless of approach.

1.1 Group vs. individual differences in human neuroimaging research

In functional neuroimaging, a standard approach is to average data across groups of individuals and then either assess this average or compare averages across groups. This approach improves signal-to-noise ratio (SNR) by washing out within-group heterogeneity and allowing researchers to highlight common neural activations underlying a psychological task or understand common connections among brain regions (Friston et al., 1999). Studying the human brain through the lens of group differences has helped to uncover the neural correlates of development (Rubia et al., 2000; Bunge et al., 2002; Konrad et al., 2005), behavior (Lee et al., 2006; Prescott et al., 2010; Graham et al., 2010), diagnoses (Bush et al., 1999; Baron-Cohen et al., 1999; Rauch et al., 2000), and injury (McAllister et al., 2001; Chen et al., 2004), with the added benefit of higher signal-to-noise ratio compared with an individualized approach. The logic underlying group-averaged neuroimaging hinges on the idea that (a) common brain activation across a group sampled from some population reflects functional neuroanatomy "typical" of that population and, thus, (b) differences in common activation between groups reflects differences in functional neuroanatomy between those two populations (Friston et al., 1999). Furthermore, the logic in cognitive neuroscience follows that (c) these differences relate to cognitive differences between the two populations.

This approach has clean benefits when studying the neural correlates of traumatic brain injury and concussion, of discrete diagnoses, and of clearly defined groups. However, in many cases group belonging is decided somewhat arbitrarily, based on statistical properties of the sample or a diagnostic cutoff. For example, the study of intelligence frequently compares groups of individuals with "high" and

"average" or "low" intelligence or intelligence quotient (IQ), based on standard deviations of a psychometric intelligence assessment (Graham et al., 2010, 2017). In yet still more cases, group belonging is decided based on a diagnostic cutoff, as we see in most research concerning autism spectrum disorder (ASD) (Knaus et al., 2008; Bookheimer et al., 2008; Kleinhans et al., 2011; Jones et al., 2010). In cases like these, group division reflects statistical attributes of psychometric measures which are developed largely agnostic of biology. As the brain is a biological system, its structure and function likely do not not reflect such statistically- or psychometrically-based group divisions. On the other hand, group-based analyses have clear benefits for assessing treatment and training effects.

Increasingly, however, cognitive neuroscience is turning its focus toward individual differences, in part brought on by recent research highlighting the high relative contribution of variability in the brain between people (Gratton et al., 2018; Finn et al., 2015; Chamberland et al., 2017; Geerligs et al., 2015; Finn et al., 2017; Vanderwal et al., 2017), and largely made possible by improvements in functional magnetic resonance imaging (fMRI) technology that have yielded increases in SNR (Kirilina et al., 2016; Dubois and Adolphs, 2016). There is growing body of neuroimaging literature highlighting the magnitude of individual variability in the human brain. While interest in individual differences in brain function is not new, even as recently as a decade ago individual differences were considered a nuisance by some researchers (Van Horn et al., 2008; McGonigle, 2012; Zilles and Amunts, 2013). Since then, individual differences research has gained in popularity and trust, leading to findings with implications for all human neuroimaging research. (Gratton et al., 2018) found that the amount of variability in functional brain network connectivity is dominated by *both* group and individual factors, finding these two sources to be of roughly equivalent magnitude. Hawco et al. (2021) took this line of research a

step further, finding instead that "group-based" activation patterns are not borne out on the individual level. Instead, activation patterns in task-based fMRI research demonstrate multiple dimensions of cognitively-relevant variability between participants. Furthermore, functional connectivity demonstrates robust, consistent, "trait-like" variations between individuals (Finn et al., 2015; Seitzman et al., 2019).

1.2 Challenges to individual differences research in human neuroimaging

Although the value of individual differences approaches in human neuroimaging has become more clear, they are not without their drawbacks. The most fundamental of these is validity. In studying the human brain on an individual level, researchers assume that they are measuring the same thing across individuals, whether with questionnaires, cognitive assessments, or neuroimaging methods. Individual differences research in human neuroimaging rests on a basic assumption of functional homology, the same anatomical location in the brain performs the same function across individuals. Findings of functional fingerprinting and trait-like network variants challenge this assumption (Finn et al., 2015; Seitzman et al., 2019), but the subfield of network neuroscience has provided a multitude of solutions for assessing functional connectivity between individuals (Shen et al., 2013; Craddock et al., 2012; Schaefer et al., 2018; Nikolaidis et al., 2020; Salehi et al., 2018; Chong et al., 2017). Another challenge to individual differences research is reliability, from two main sources. The first is the "unmodeled noise" described by Van Horn et al. (2008), sources of which are continually being discovered in both physiological measures that affect the measurement of the BOLD signal (Handwerker et al., 2012; Liu, 2016, 2013; Murphy et al., 2013) and in factors influencing cognition and both

brain function and structure (Taylor et al., 2020; Pritschet et al., 2020; Stewart et al., 2003; Nyberg et al., 2006). The second is statistical power, which is tightly coupled to sample size. In 2018, the median sample size in fMRI studies was 24 participants and since then it appears to be increasing on the order of one participant per year (Szucs and Ioannidis, 2020). This is problematic for individual differences research as small sample sizes artificially inflate effect sizes (Yarkoni, 2009) and with 24 participants, the likelihood that a correlational fMRI study will detect an effect that *is* present in the population is only 10% and the likelihood that effect sizes of significant results are overestimated is 100% (Dubois and Adolphs, 2016). It is only as sample sizes move into the hundreds of participants that statistical results from such studies begin to accurately estimate effects in the population. This issue can be overcome in multiple ways: by moving toward population-level studies with sufficient sample sizes (Alexander et al., 2017; Allen et al., 2014; Casey et al., 2018), by understanding and accounting for sources of previously "unmodeled noise", and by focusing on understanding the brain at an individual level before moving on to population-level neuroscience (Poldrack et al., 2015; Poldrack, 2017).

1.3 Large-scale functional brain networks

In the human brain, there are several anatomically disparate, functional systems, or networks, that are consistently observed across individuals and cognitive contexts (Smith et al., 2009). These networks have been studied at length over the past decade, revealing that they are both behaviorally (Laird et al., 2011) and psychopathologically (Menon, 2011; Menon and Uddin, 2010) relevant. These systems provide a rich opportunity to investigate individual variability in the brain. Although they are predictably coherent across individuals, they do exhibit individual-

specific variability (Gratton et al., 2018; Seitzman et al., 2019). As a network-based study of brain function has become more popular, researchers are applying network science and graph theory to study topological features of brain networks (Bassett and Bullmore, 2006; Bullmore and Sporns, 2012; Rubinov and Sporns, 2010; Sporns et al., 2004; Sporns, 2012). Not only do these approaches provide additional information about brain organization and function beyond what can be gleaned from studies of brain activation or connectivity alone, they exhibit higher retest reliability than individual brain connections (Cao et al., 2014; Welton et al., 2014). Summarizing large-scale characteristics of brain organization with graph theoretic techniques, enables global assessment of inter-individual variability (Tompson et al., 2018; Rubinov and Sporns, 2010; Mears and Pollard, 2016; Li et al., 2009). For example, this work has demonstrated that significant variation in brain network modularity is related to working memory capacity (Stanley et al., 2014; Stevens et al., 2012), network flexibility is related to positive mood (Betzel et al., 2017), and shorter path lengths predict higher IQ scores (Li et al., 2009).

In the following chapters, I present both advances in methodology and research that advances the study of individual variability in the human brain across timescales and domains.

CHAPTER 2

INDIVIDUAL DIFFERENCES IN FUNCTIONAL BRAIN CONNECTIVITY

Historically, human cognitive neuroimaging research has focused on understanding which brain regions are "active" during different mental processes, how these brain regions are associated with behavior, and how these relations may differ across groups of individuals. While this approach has led to advances in our understanding of brain function and functional anatomy, as well as the neurobiology of disease etiology, it has two distinct shortcomings. First, focusing on brain activation substantially limits our ability to understand the brain, as it ignores interactions between brain regions and behind-the scenes brain activity vital to cognition and behavior. In the past decade, the focus has shifted to include functional brain connectivity, which estimates communication between spatially distinct brain regions, and network-based measures of brain organization that are based on this connectivity. Such research has uncovered large-scale brain networks that are coherent across experimental paradigms and behaviors, but which exhibit subtle, yet consistent, differences across development, aging, behavior, and psychiatric and neurological disorders. Without the constraints inherent in activation-focused research, connectivity-focused neuroimaging research more comprehensively assesses neurobiology during experimental paradigms. Second, attempting to view the neurobiology of behavior, development, and disease through the lens of group differences at worst artificially imposes false dichotomies and at best ignores the wealth of information in within-group heterogeneity. Recent research has shown that the relative magnitudes of variability in brain network organization across a group and within each individual are commensurate. Together, these phenomena illustrate a greater need for understanding sources of individual variability in brain network organization and

how it relates to behavior. IDConn is a data analysis workflow for combining methods for assessing individual differences in brain connectivity and organization from functional magnetic resonance imaging (fMRI) data with robust statistical methods for assessing relationships between continuous variables, adjusted to address the unique challenges of fMRI data.

2.1 Statement of need

IDConn is a configurable data analysis pipeline written in Python for assessing individual differences in functional connectivity and derived network measures. It brings together existing tools for neuroimaging data analysis, data science, machine learning, graph theoretic network analysis, statistics, and scientific computing into a unified, streamlined workflow for sophisticated, rigorous computation and extraction of connectivity and graph theoretic measures, for linear regression-based analyses with behavioral, demographic, psychophysiological, or other continuous variables. Furthermore, it conforms with the Brain Imaging Data Structure (BIDS) (Gorgolewski et al., 2016) for optimal redistribution and sharing of pipeline outputs. While there are a host of extant open-source software tools for analyzing fMRI data, many assume a level of code literacy and related skills that are not common in researchers from the psychological sciences and less computational backgrounds, in general. IDConn fills this gap and provides an open, reproducible option for data analysis that incorporates high-quality software tools and up-to-date "best practices" for analyzing fMRI data.

IDConn was designed for applications in human neuroimaging research, providing a flexible, open data analysis stream that takes in preprocessed fMRI data and provides computed graphs, derived graph measures, statistical models and the results thereof, and, optionally, figures presenting these results, all in an organized,

sharing-friendly format for optimal reproducibility and transparency. It has already been used in two scientific publications (Bottenhorn et al., 2021; Gonzalez et al., 2019). Gonzalez et al. (2019) used an early iteration of IDConn to assess how differences in resting state functional connectivity between canonical brain networks are related to generalized anxiety and anxiety around Science, Technology, Engineering, and Math (i.e., STEM anxiety) and how these relationships may differ between male and female university students before and after their first physics class. Similarly, Bottenhorn et al. (2021) used IDConn to assess how task-based functional connectivity and network topology support relationships between intelligence and academic achievement during physics-related cognition. Bringing together robust tools for fMRI data processing, dataset cleaning, and statistical inference, IDConn enables statistically-robust assessments of individual differences in functional brain connectivity and organization by neuroimaging researchers from the Python-naiive to developers.

2.2 State of the field

Currently, there are a wealth of software tools available for analyzing neuroimaging data. Tools such as the FMRIB Software Library (FSL Jenkinson et al., 2012; Smith et al., 2004; Woolrich et al., 2009), Analysis of Functional NeuroImages (AFNI; Cox, 1996), Statistical Parametric Mapping (SPM; Ashburner, 2012; Friston et al., 2004), and Brain Voyager (Goebel, 2012) provide graphical user interfaces to aid researchers in the analysis of fMRI data, along with scripting options to automate and batch-process analyses via command-line interfaces. However, the code bases of these tools have varying degrees of accessibility and some are completely closed. Open source tools for fMRI data analysis exist, too, under the broad umbrella of NiPy

(Millman and Brett, 2007), including Nilearn (Abraham et al., 2014) and Nipype (Gorgolewski et al., 2011). Nipype is an open source data processing framework that provides interfaces for a host of neuroimaging data analysis tools, implementations of a number of analysis algorithms, and a pipeline engine for connecting tools into unified workflows in Python. It allows researchers to integrate tools from existing, and perhaps less "open", software packages in transparent, reproducible, and reusable Python scripts.

2.3 Pipeline

IDConn has a similar goal, though with a narrower scope and more specific applications. Instead of a basis for building data analysis workflows, IDConn brings together existing tools to create a configurable pipeline that is accessible to researchers, regardless of their technical background. It integrates functions from Nilearn (v0.6.2; nilearn.github.io), Nipype (v1.5.1; nipype.readthedocs.io Gorgolewski et al., 2011, 2017; Esteban et al., 2020), scikit-learn (v0.23.1; scikit-learn.org Pedregosa et al., 2011)), SciPy (Virtanen et al., 2020, v1.5.1; scipy.org), Numpy (v1.19.0; numpy.org Harris et al., 2020)), NiBabel (v3.1.1; nipy.org/nibabel/), Pandas (v1.0.5; pandas.pydata.org), Seaborn (v0.10.1; seaborn.pydata.org), statsmodels (v0.11.1; statsmodels.org), NetworkX (v2.5; networkx.github.io), and the Brain Connectivity Toolbox in Python (Rubinov and Sporns, 2010, bctpy v0.5.0) to provide the following workflow (see also Figure 1).

2.3.1 Connectivity

The first step in this pipeline is to create brain connectivity graphs from preprocessed fMRI data (Figure 2.1-1). To do this, IDConn reads in cleaned BOLD fMRI

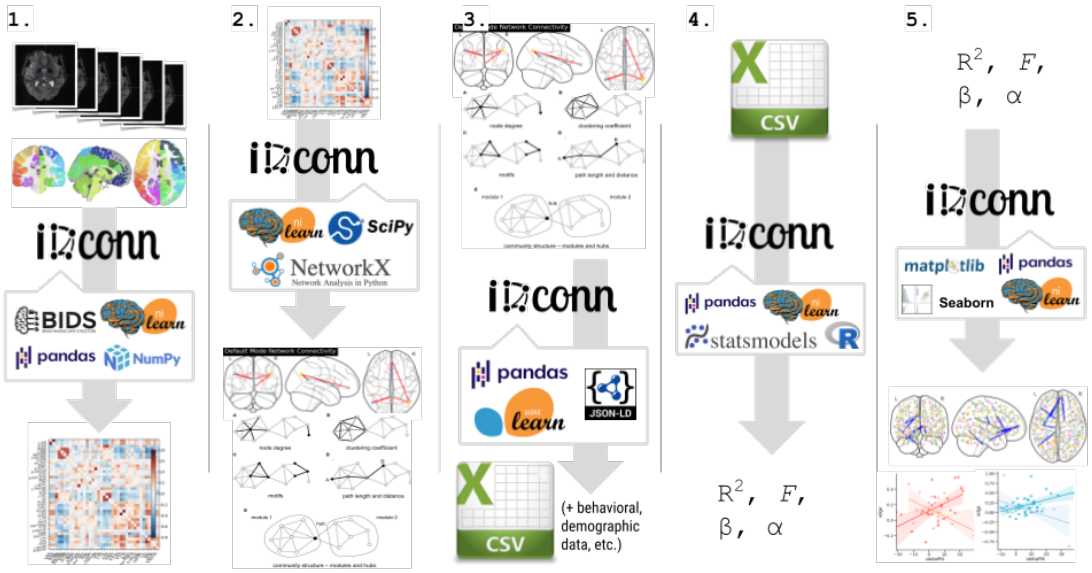

Figure 2.1. IDConn pipeline illustration.

timeseries, masked with a brain atlas or parcellation of the user's choosing, and extracts averaged BOLD signal across voxels within each region or parcel. Connectivity networks are then assembled based on pairwise similarities between averaged BOLD signals per region or parcel. The metric by which pairwise similarity is calculated is up to user specification and can be any of the metrics in scikit-learn's pairwise distance metrics.

2.3.2 Networking

From these connectivity graphs, the pipeline then estimates either (a) connectivity between regions, (b) connectivity between brain networks, or (c) graph theoretic metrics, or topological properties, of these graphs (Figure 2.1-2). To do so, IDConn relies on bctpy and networkX to compute graph theoretic metrics and includes options to standardize such metrics with respect to null distributions based on the empirical networks. Furthermore, IDConn applies known attributes of brain net-

works to estimate thresholds to remove noise and preserve neurally-relevant network characteristics.

2.3.3 Data Organization

Once these measures are computed, they are harmonized with existing demographic, behavioral, and/or biological measures collected per participant (Figure 2.1-3). Furthermore, this module calculates several descriptive statistics and performs an assessment of missing data across measures.

2.3.4 Statistical Inference

With this collated data, the pipeline then performs inferential statistics (Figure 2.1-4), based on a user-input model per the BIDS Statistical Models specification (BEP002). Models can be fit using formula syntax notation popular in the R statistical language, using semopy (Igolkina and Meshcheryakov, 2020) and statsmodels to perform statistical inference, from ordinary least squares and linear mixed effects models to structural mediation models.

CHAPTER 3

INTELLIGENCE, ACADEMIC PERFORMANCE, AND THE BRAIN

3.1 Introduction

Intelligence testing has been the subject of substantial scientific inquiry over the past century, especially with respect to academic performance and overall success (Alloway and Alloway, 2010; Gottfredson, 1998, 2002; Harris, 1940). These lines of inquiry have a pernicious history, as intelligence testing and IQ have been used to deny educational and employment opportunities and as a foundation of eugenics in the United States and abroad in the 20th century, (Frisby and Henry, 2016; Reddy, 2007) though updates to these tests have mitigated bias with respect to racial, ethnic, and gender differences (Reynolds and Suzuki, 2012; Weiss et al., 2010). Both in research and in popular discourse, intelligence is often treated as an inherent, stable, trait-like quality, equated with an intelligence quotient (IQ) (Dzirasa, 2017; Bartels et al., 2002; Canivez and Watkins, 1998; Hertzog and Schaie, 1986; Herrnstein and Murray, 1994). While intellectual abilities are moderately heritable (Plomin and von Stumm, 2018), IQ and other psychometric measures of intellectual ability are also influenced by a number of experiential factors, and the relative influences of genes and environment on IQ change across the lifespan (Reynolds et al., 2005; Finkel and Reynolds, 2014). In general, intelligence is psychometrically assessed via a range of verbal and nonverbal cognitive tests, capturing a general view of ability across domains. One such test is the Wechsler Adult Intelligence Scale (WAIS; Wechsler et al., 2008), which demonstrates moderate stability across adulthood, with increasing stability for shorter intervals and with increasing age (Begovac et al., 2009; Johnstone and Wilhelm, 1996; Moffitt et al., 1993; Pietschnig and Voracek, 2015; Ritchie and Tucker-Drob, 2018; Schneider et al., 2014; Schuerger and Witt,

1989; Tuma and Appelbaum, 1980; Verhaeghen, 2003; Yang et al., 2011; Yu et al., 2018). A history of research and popular discourse presupposes that IQ predicts one's predisposition to academic and life success (Hunter and Hunter, 1984; Deary et al., 2007; Strenze, 2007; Zax and Rees, 2002; Murray, 1998), though in reality, the picture is much more complicated due to a variety of sociocultural factors (von Stumm, 2017; Day et al., 2010; Zuffianò et al., 2013; Bergold and Steinmayr, 2018). Decades of research suggest education and psychometric intelligence are entwined in a bidirectional relationship, as intellectual ability predicts access to and extent of education, through a variety of socioeconomic factors (Neisser et al., 1996; Busato et al., 2000), while years of education predict modest increases in intellectual abilities (Busato et al., 2000; Walker et al., 2009), and some educational interventions likely to improve one's ability to acquire and apply knowledge and skills (Ritchie and Tucker-Drob, 2018; Ding, 2014; Jackson et al., 2008; Roberts et al., 1993; Taasoobshirazi and Sinatra, 2011).

University students who pursue science, technology, engineering, and mathematics (STEM) disciplines are exposed to a rigorous curriculum designed to transform their problem-solving skills (Bicer et al., 2017; Etkina et al., 2006; Gabel and Sherwood, 1983; Stieff and Uttal, 2015), which engages students' perceptual and verbal abilities (Hegarty et al., 2013; Hegarty, 2014; Oswald et al., 2016). The fourth edition of the WAIS (WAIS-IV) provides a full-scale measure of intellectual ability (FSIQ) and four component index scores: Processing Speed, Perceptual Reasoning, Working Memory, and Verbal Comprehension (Weiss et al., 2010). The WAIS is widely used as an extensively validated and researched clinical tool, but less often applied in educational research, though it may be particularly well-suited for exploring associations between education and skill development. Introductory physics presents a prime opportunity for such study, as a gateway course for STEM majors

with a relatively standard curriculum across universities, including instruction on classical Newtonian mechanics and emphasizing the development of quantitative, visuospatial reasoning and problem-solving skills likely captured by WAIS-IV index scores. Unfortunately, female students often perform worse on specific conceptual evaluations in these courses, though not necessarily on overall course grades, due to a host of socioaffective and -cultural factors present in education and physics classrooms, specifically (Day et al., 2016; Kelly, 2016; Lorenzo et al., 2006; Madsen et al., 2013; McCullough, 2013; Nissen and Shemwell, 2016). Female students also constitute a smaller proportion of the student body, compared to their male counterparts (Day et al., 2016; Madsen et al., 2012; McCullough, 2013; Eddy and Brownell, 2016; Gewin, 2011) and ultimately, such disparities can propagate across courses, leading to higher rates of STEM degrees among male students as compared to female students (64% male vs. 36% female in 2015-2016; DoE, 2019). Recently, institutions have sought to improve STEM student success using active learning instructional approaches that yield improved student performance outcomes (Freeman et al., 2014; Hake, 1998) and impact socioemotional aspects of university education, including self-efficacy, science-related anxiety, and identity (Kelly, 2016; Eddy and Brownell, 2016). Together, these effects may mitigate existing sex differences in performance (Lorenzo et al., 2006; Nissen and Shemwell, 2016) though this is contradicted by some findings (Pollock et al., 2007). Altogether, there is a need to better understand sex differences in physics education and potential avenues for mitigating these differences, to ensure all students have the opportunity to succeed.

For as long as we have been trying to understand intelligence, we have been searching for its biological substrates. Recently, neuroscience research has studied the underlying neurobiology of intelligence using neuroimaging techniques such as functional magnetic resonance imaging (Fischer et al., 2014; Heuvel et al., 2009;

Hilger et al., 2017a; Langer et al., 2012; Li et al., 2009, fMRI). Task-based fMRI research has focused on understanding how differences in brain activation during cognition differs relates to intelligence, yielding two theories of the neurobiology of intelligence: the parieto-frontal integration theory (P-FIT; Jung and Haier, 2007) and the neural efficiency hypothesis (NEH; Neubauer and Fink, 2009). The PFIT suggests that interactions between frontal and parietal regions underlie intelligence, while the NEH suggests that higher intelligence is reflected by more "efficient" brain activation during cognitively demanding tasks. An alternative view on "neural efficiency" comes from the application of network science to functional connectivity, to show that more intelligent individuals exhibit greater topological efficiency, which describes ease of information transfer across the brain (Heuvel et al., 2009; Hilger et al., 2017a; Langer et al., 2012; Santarnecchi et al., 2014), rather than activation efficiency. Much of this work has focused on functional brain connectivity during the resting state, i.e., in the absence of a task or externally-directed cognition (Hearne et al., 2016; Dubois et al., 2018; Hilger et al., 2017b), often referred to as "intrinsic" connectivity (Laird et al., 2011; Smith et al., 2009) mirroring the notion of intelligence as an inherent trait. However, recent work shows that individual differences in intelligence are better predicted by task-evoked connectivity (Greene et al., 2018), presenting an opportunity to merge these two lines of research to better understand individual differences in the neurobiology of intellectual abilities.

Here, we build on the knowledge that the WAIS measures cognitive abilities that are subject to influence by educational interventions, and leverage the WAIS to study individual differences across student performance in an introductory physics course, and potential roles of sex and pedagogy therein. Then, we build on prior neuroimaging research suggesting task-evoked brain organization can explain individual differences in intellectual abilities, to search for a biological substrate for associations

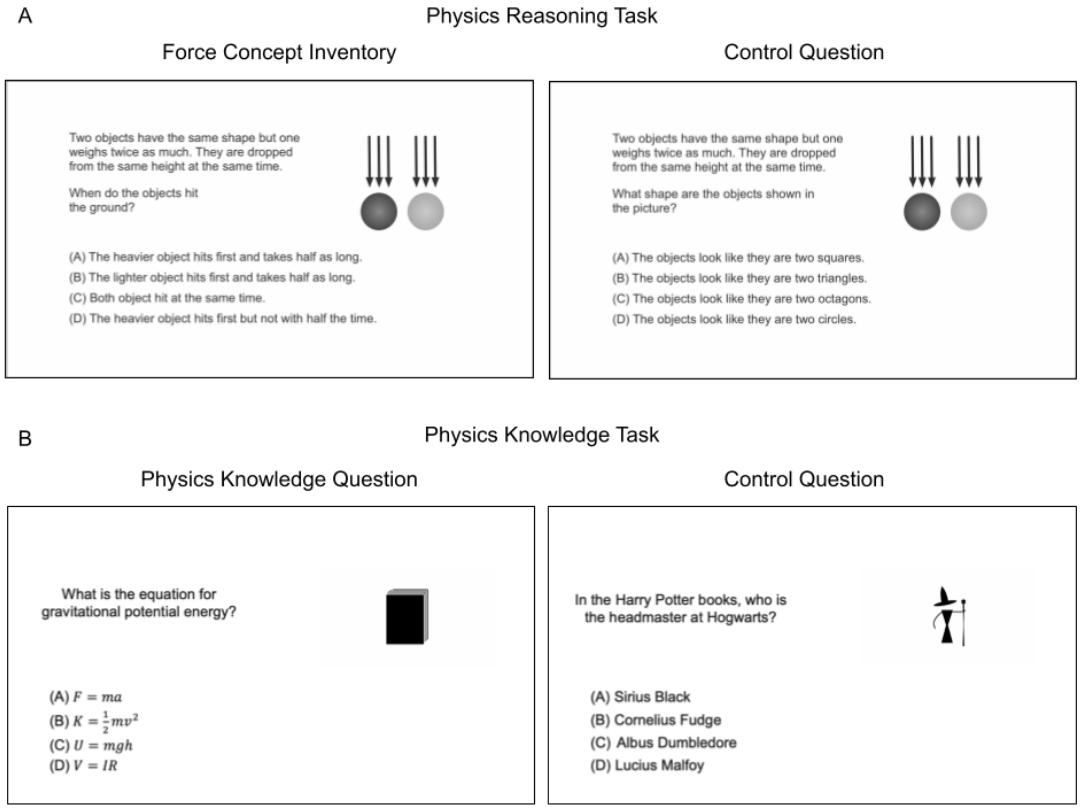

Figure 3.1. fMRI tasks performed by students at both pre- and post-instruction time points.

between ability and student performance. To do so, we collected data from undergraduate students enrolled in either a lecture-based or active learning section of an introductory physics course. Both pre- and post-instruction, students completed the WAIS-IV alongside a robust fMRI protocol, including two physics-related tasks (Figure 3.1) with different demands on cognition. The first task engaged students' reasoning skills and conceptions about forces at work in the natural world (Figure 3.1A), based on the Force Concept Inventory (FCI; Hestenes et al., 1992, see Bartley et al. (2019) for detailed task results). In this physics reasoning task, students viewed questions about forces on and the movement of objects, along with answer choices that included the correct (e.g., Newtonian) explanation, and choices that reflect common but incorrect (e.g., non-Newtonian) conceptions about forces and motion. The second task required students to recall concepts and equations taught in an introductory physics course (Figure 3.1B). In this physics knowledge task, students engage semantic memory to recognize equations or definitions of physics concepts learned in the course from a list of possible answer choices presented. Here, we used these data to, first, assess changes in WAIS-IV scores (both FSIQ and index scores) over the course of the semester, then applied a series of linear regressions to assess associations between post-instruction and pre- to post-instruction changes in WAIS-IV scores and post-instruction student performance (i.e., task accuracy and final course grade). Finally, we assessed associations between WAIS-IV scores, and changes therein, and post-instruction functional brain organization during the two tasks, and the degree to which these associations provide a common neural substrate supporting the role of cognitive abilities in student performance. We hypothesized that, while FSIQ itself is stable, different WAIS-IV index scores are differentially associated with performance on physics-related assessments with different cognitive demands. Further, we hypothesized these differences would be reflected in brain

organization during these domain-specific tasks (e.g., physics reasoning and physics knowledge), providing a neurobiological explanation for ability-performance relationships across physics-related cognition. Finally, including sex and pedagogy in these assessments may provide an insight into sex differences in physics education and the potential role of active learning in their mitigation, though we did not expect different brain-IQ relations across active learning and lecture classrooms.

3.2 Methods

3.2.1 Participants and Study Design

One hundred and thirty healthy right-handed undergraduate students (mean age = 20.03 ± 2.25 years, range = 18-25 years; 61 females) who completed a semester of introductory calculus-based physics at Florida International University (FIU), a Hispanic Serving Institution, took part in this study. Participants were not currently using psychoactive medications and reported that they had not been diagnosed with any cognitive impairments or neurological or psychiatric conditions. The physics course emphasized problem solving skill development and covered topics in classical Newtonian mechanics, including motion along straight lines and in two and three dimensions, Newton's laws of motion, work and energy, momentum and collisions, and rotational dynamics. Students were either enrolled in a lecture class or an active learning, "Modeling Instruction", class, which bases course content in conceptual scientific models and instructs students to appropriate scientific models for their own use. Students completed behavioral assessments and MRI scans in separate appointments at two time points: at the beginning ("pre-instruction") and conclusion ("post-instruction") of the 15-week semester. Pre-instruction data collection

sessions were acquired no later than the fourth week of classes and post-instruction sessions were completed no more than two weeks after the final exam. Written informed consent was obtained in accordance with FIU's Institutional Review Board approval.

3.2.2 Missing Data

A missing value analysis indicated that, of the variables of interest in this study, missingness ranged from 2% to 17%. Data were more often missing from MRI data than behavioral or demographic data and more often missing from post-instruction data than from pre-instruction. Assessment of the relations between missingness on each variable and values of each other variable of interest revealed that missingness in the data is unrelated to values of the other variables. Behavioral and brain network efficiency data were imputed using iterated Bayesian ridge regression implemented in scikit-learn (v. 0.23.1; scikit-learn.org/). Due to its high dimensionality, missingness in edgewise functional connectivity data was addressed using distance-weighted K-Nearest Neighbors approach (K = 100, where p = 71,824) implemented in scikit-learn, which is robust to missingness up to 20% (Troyanskaya et al., 2001).

3.2.3 Behavioral Measures

During pre- and post-instruction behavior sessions, participants were administered the fourth edition of the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler et al., 2008), a standardized intelligence test for adults, in addition to other assessments not used here. The WAIS-IV provides scores in four domains, in addition to an overall score of intellectual functioning. The Verbal Comprehension index measures application of verbal skills in problem solving. The Perceptual Reasoning index

measures the ability to detect the underlying conceptual relationship among visual objects and use reasoning to identify and apply rules. The Working Memory index measures short-term memory with auditory and visual stimuli. The Processing Speed index measures speed of mental operations and visual-motor coordination. Lastly, the Full-Scale intelligence quotient (IQ) represents a global estimate of intellectual or cognitive ability. All instruments were administered by researchers for the purpose of this research and not professionally or clinically (i.e., for diagnostic or instructional purposes).

3.2.4 fMRI Tasks

In the scanner, participants performed two different physics tasks, each probing different aspects of physics learning and problem solving (Figure 3.1).

Participants completed three runs of a physics reasoning task, which uses questions from the Force Concept Inventory (FCI; Hestenes et al., 1992) to assess domain-specific problem solving. This task includes two conditions, FCI and control, presented in a block design with self-paced trials. FCI and control questions were presented in three screens (Figure 3.1A), between which participants advanced by the press of a button. The first screen presented a written description of a physical scenario and corresponding figure; the second, a question relating to the scenario; and the third, four answer choices from which the participants were instructed choose the correct answer while mentally justifying their choice.

Participants additionally completed two runs of a physics knowledge task, which probed physics-related memory retrieval and included physics, general, and control conditions. Participants were asked a series of multiple-choice questions and instructed to respond by indicating their choice with the press of a button (Figure

3.1B). In the physics condition, participants were asked to recall definitions and formulas taught in the physics course (e.g., "What does the 'SI' in SI units stand for?" or "What is the value of the acceleration due to gravity?"). In the general condition, participants were asked to recall general trivia (e.g., "Which of these is not an automobile brand?" or, "Who is the President of the United States?"). The low-level control condition asked participants to press the button corresponding to a letter or symbol. Conditions were organized into blocks and each run included three blocks per condition.

3.2.5 fMRI Acquisition and Pre-Processing

Neuroimaging data were acquired on a GE 3T Healthcare Discovery 750W MRI scanner at the University of Miami. Functional MRI (fMRI) data were acquired with an interleaved gradient-echo, echo planar imaging (EPI) sequence (TR/TE = 2000/30ms, flip angle = 75°, field of view [FOV] = 220x220mm, matrix size = 64x64, voxel dimensions = 3.4×3.4×3.4mm, 42 axial oblique slices). A T1-weighted series was also acquired using a 3D fast spoiled gradient recall brain volume (FSPGR BRAVO) sequence with 186 contiguous sagittal slices (TI = 650ms, bandwidth = 25.0kHz, flip angle = 12°, FOV = 256x256mm, and slice thickness = 1.0mm). A 2-mm isotropic MNI152 template image was nonlinearly oriented to each participant's structural T1-weighted image using FMRIB's Software Library's (FSL; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki; Jenkinson et al., 2012) nonlinear registration tool (FNIRT; Andersson et al., 2007). Then, each participant's T1-weighted image was coregistered to the middle volume of each functional run, using FSL's linear registration tool (FLIRT; Jenkinson et al., 2002). These two transformations were concatenated and used to align regionwise parcellations to each subject's functional

images. Tissue-type masks for white matter, gray matter, and cerebrospinal fluid (CSF) were created from each subject's T1-weighted images using FSL's automated segmentation tool (FAST; Zhang et al., 2001).

Task-based fMRI preprocessing began with FSL's MCFLIRT with spline interpolation, per run per functional task, to align all volumes of each subject's fMRI time series with that middle volume. To further correct for in-scanner motion effects, functional volumes unduly affected by motion were identified using fsl_motion_outliers, with a framewise displacement threshold of 0.9mm for functional scans (Siegel et al., 2014). Data were standardized, detrended, and high-pass filtered, according to the period of each task. The physics knowledge task was high-pass filtered at 0.018Hz and the physics reasoning task was thresholded according to each participant's individual timing.

Regionwise Parcellation and Brain Connectivity Analyses.

Each participant's fMRI data were parcellated according to two functionally-derived, whole-brain parcellations with similar numbers of regions. Here, we used a 268-region parcellation computed via multigraph k-way clustering, without spatial constraints, henceforth referred to as the Shen parcellation (Shen et al., 2013). To ensure that our results were not artifacts of node definition, we additionally performed all analyses with an atlas generated from resting-state fMRI data by performing normalized-cut spectral clustering on voxelwise functional connectivity data to define homogeneous, spatially-constrained clusters (i.e., regions), henceforth referred to as the Craddock parcellation (Craddock et al., 2012), which includes a range of atlas sizes from which we chose a 268-region solution, to match the granularity of the Shen parcellation. Connectivity graphs were computed using two different brain parcellations (Figure 3.2.5). Following preprocessing, data analysis continued

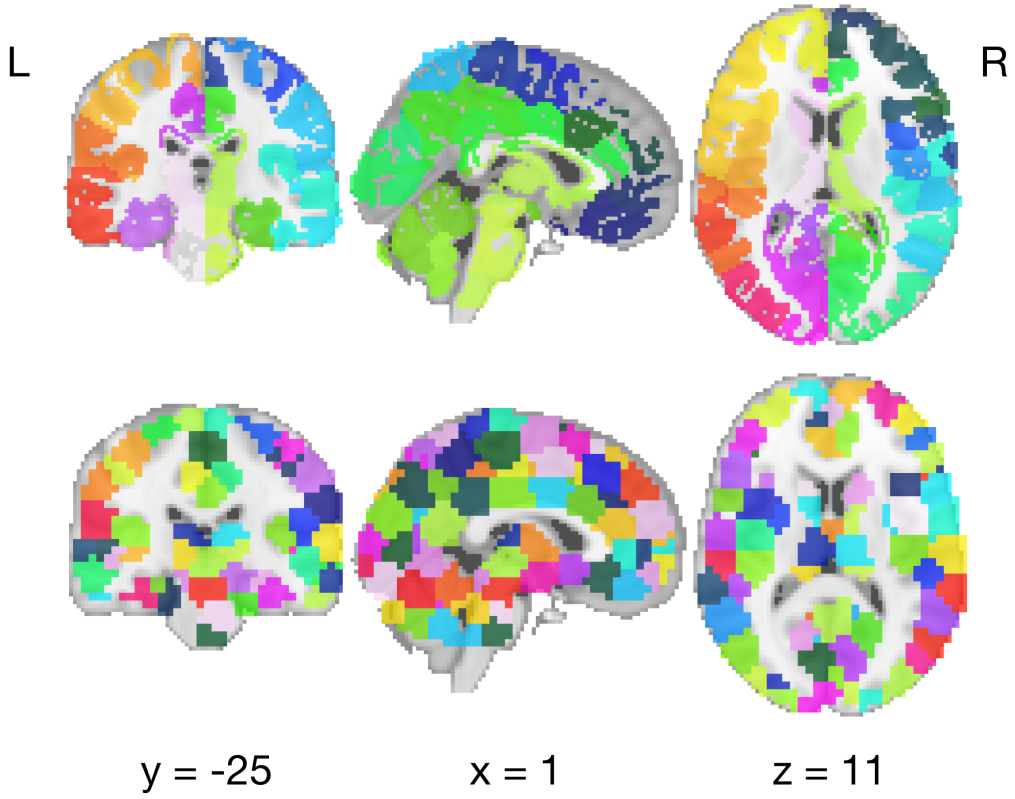

Figure 3.2. Shen et al. (2013) and Craddock et al. (2012) region-wise brain parcellations.

in two parallel streams, one with the Shen parcellation (top row) and one with the Craddock parcellation (bottom row), to ensure that any results were not artifacts of parcellation schema. Presented results include brain-IQ relations observed for both parcellations, but values displayed are derived from the Craddock parcellation.

For each region, a single time series was computed as an average of the fMRI time series from all voxels within the region, after further regressing out six motion parameters (from MCFLIRT) and censoring high-motion volumes (framewise displacement ¿0.9mm), as well as the immediately preceding volume and two following volumes, following recommendations from Power et al. (2014). Functional tasks' regionwise time series were standardized (i.e., z-scored), divided by condition per task per run and spliced together across runs, creating separate time series per condition per task for each participant. Adjacency matrices were constructed with each parcellation per participant, per functional task, per session (pre- and post-instruction) using Nilearn (v. 0.3.1, http://nilearn.github.io/index.html; Abraham et al., 2014; Pedregosa et al., 2011), a Python (v 2.7.13) module, built on scikit-learn, for the statistical analysis of neuroimaging data, by computing the pairwise Pearson's correlations between each pair of regions, resulting in a 268x268 region-wise correlation matrix for each subject per condition per task per session (pre- and post-instruction). Graph theoretic, topological measures were calculated across a range of density-based thresholds. The lowest thresholds at which each network (a) approximated a scale-free degree distribution and (b) ceased to be node-connected were calculated for each adjacency matrix. From these values, a lower- and upper-bound for network thresholding were estimated, following recommendations from Lynall et al. (2010) and Ginestet and Simmons (2011), such that networks would remain node-connected, meaning there are no brain regions completely separate from the rest of the brain, and spurious connections would be removed while main-

taining the scale-free degree distribution expected of the brain per prior research (Gong et al., 2009; Hayasaka and Laurienti, 2010; He et al., 2007).

All topological measures were calculated using bctpy, a Python toolbox intended to replicate the functionality of the Brain Connectivity Toolbox, a MATLAB toolbox for graph theoretic analysis of functional and structural brain connectivity (brain-connectivity-toolbox.net; Rubinov and Sporns, 2010). From each correlation matrix described above, we calculated global efficiency, characteristic path length, and modularity for across the range of proportional thresholds as calculated above ($\kappa = [0.21, 0.31]$ at steps of 0.01), then calculated the area under the curve (AUC) of each measure (Bassett et al., 2006, 2012). These AUCs were used in all following statistical tests assessing the relationship between brain network organization and IQ and will henceforth be referred to per the topological measure from which they were calculated. All topology-related results reported here are significant per topology values calculated from graphs generated from both parcellations.

3.2.6 Inferential Statistics

Statistical inference was performed using the lavaan R package, and the Python modules SciPy (v. 1.2.1; scipy.org; Harris et al., 2020; McKinney, 2010; Millman and Brett, 2007; Oliphant, 2007) statsmodels (v. 0.9.0; statsmodels.org), and nilearn (v. 0.6.2; nilearn.github.io).

Paired t-tests were used to assess changes in WAIS-IV scores pre- to post-instruction. Two-sample t-tests were used to assess differences in the changes in WAIS-IV scores (post- minus pre-instruction) between male and female students, as well as between students enrolled in the active learning and lecture classes.

Ordinary least squares (OLS) regressions implemented in R were used to regress measures of academic and task performance on WAIS-IV scores, sex, class, age, and years in university, per Equation 3.3.2. Significance of individual models was assessed by comparing models' p-values to a significance threshold adjusted for multiple comparisons via Šidák correction (Šidák, 1967).

A similar procedure, using mass-univariate OLS regressions with permutation testing as implemented by the Python package Nilearn (Anderson and Robinson, 2001; Freedman and Lane, 1983; Winkler et al., 2014), was used to regress topological measures and functional connectivity (thresholded at $\kappa = 0.31$) on WAIS-IV scores, sex, class, age, years in university, head size, average framewise displacement (calculated by fsl_motion_outliers, per run, per task), per Equation 3.3.2. To correct for multiple comparisons across these regressions, we used the Šidák correction as mentioned above. All reported results were significant in both parcellations, to minimize the effects of brain parcellation on our interpretations.

Mediation models to assess whether brain connectivity explained the relationship between WAIS-IV scores and task performance were run using the R package lavaan (Rosseel, 2012).

3.2.7 Data and Code Availability

A GitHub repository was created at github.com/62442katieb/physics-learning-iq to archive the code and source files for this study, including data preprocessing and analysis scripts and behavioral data. Significant neuroimaging results are available at neurovault.org/collections/9385/.

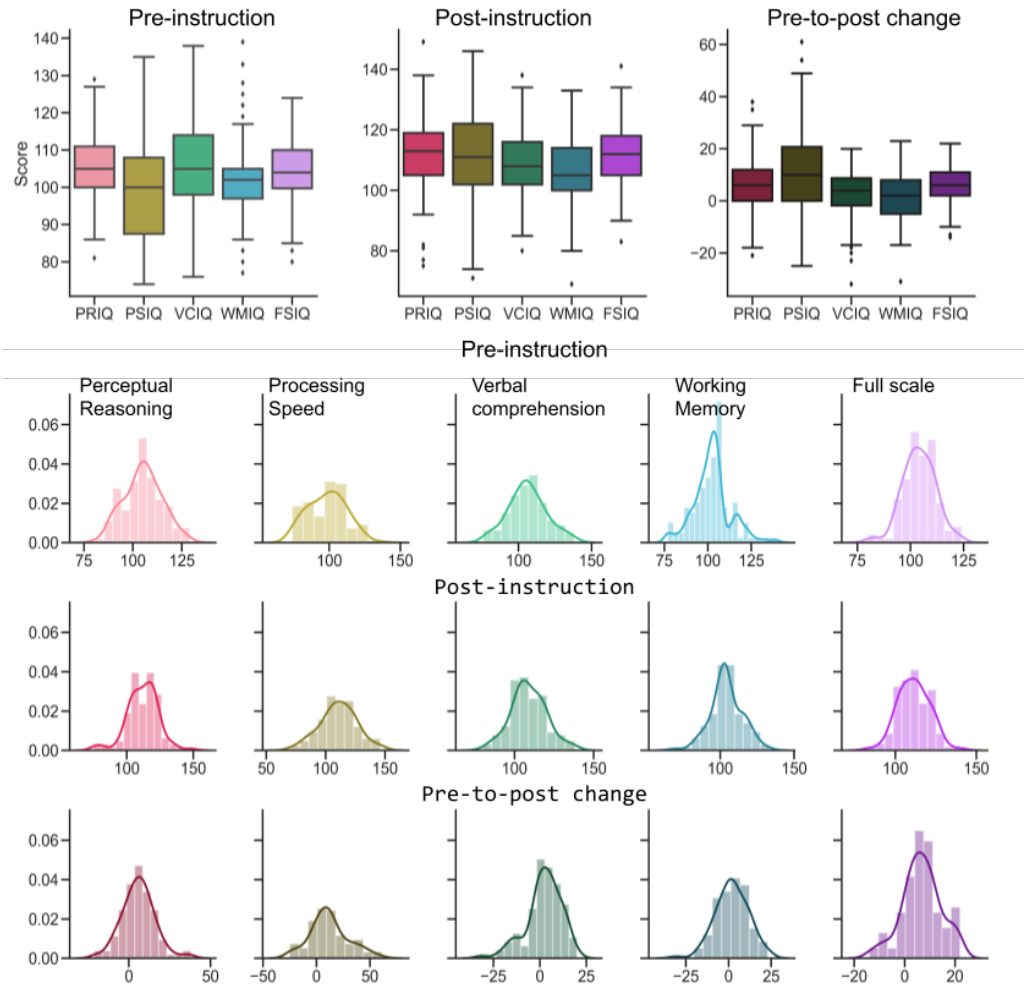

Figure 3.3. Pre-instruction, post-instruction, and pre- to post-instruction changes in WAIS-IV scores and subscores.

3.3 Results

3.3.1 WAIS-IV scores increased over a semester of physics learning regardless of sex or classroom

First, to understand characteristics of WAIS-IV scores (i.e., FSIQ and index scores) in this sample, we administered the WAIS-IV before and after a semester of physics instruction from 110 students who completed undergraduate introductory physics in

Table 3.1

Average change in WAIS scores pre- to post-instruction

WAIS Score	Pre-instruction	Post-instruction	Pre-to-post change	Estevis et al., 2012	Change - Estevis t-test (p)
Full score WAIS-IV Score	103.9 ± 7.6	110.8 ± 9.4	**7.0 ± 7.3**	**6.7 ± 5.2**	-0.12 (0.90)
Perceptual Reasoning Index	105.0 ± 9.6	111.5 ± 11.1	**6.0 ± 9.5**	**3.6 ± 5.6**	2.39 (0.02)
Processing Speed Index	98.3 ± 13.3	110.6 ± 14.5	**12.6 ± 17.5**	**10.5 ± 9.5**	0.71 (0.48)
Verbal Comprehension Index	106.0 ± 13.0	108.5 ± 10.6	**2.5 ± 9.0**	**4.2 ± 7.6**	-1.28 (0.20)
Working Memory Index	102.3 ± 10.8	103.2 ± 10.5	2.0 ± 9.1	3.2 ± 4.8	-2.36 (0.02)

Note. Bolded changes indicate significan Wilcoxon Signed Rank tests at $p_{FWE-corr} < 0.05$ ($\alpha_{adjusted} = 0.014$), after controlling for familywise error using the Šidák correction (Šidák, 1967). The *Change - Estevis t-test (p)* column (far right) provides the results of a Student's t-test for independent samples, comparing the changes in WAIS scores seen in this sample with those reported by Estevis et al. (2012) to determine whether the changes in our sample are comparable with those reported elsewhere. Bold values in this column indicate significant differences between the two samples at $p_{FWE-corr} < 0.05$, after controlling for familywise error using the Šidák correction.

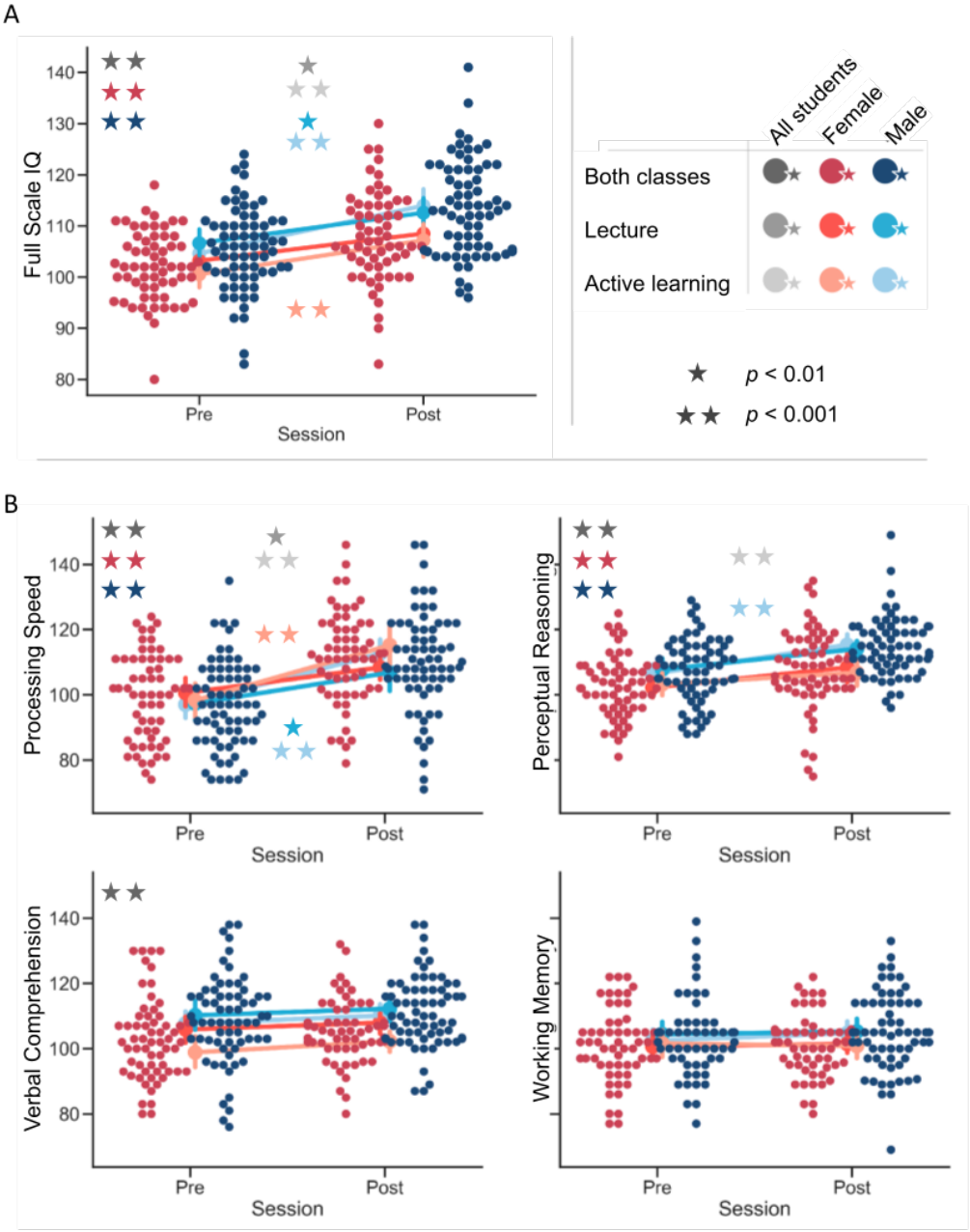

Figure 3.4. Change in WAIS scores and subscores pre- to post-instruction.

lecture-based (25 female, 28 male) or active-learning (24 female, 33 male) classrooms. Pre- to post-instruction changes in WAIS-IV scores were assessed via Wilcoxon signed-rank tests, due to the presence of outliers (Figure 3.3.1). Full-scale WAIS-IV scores increased an average of 7.03 points ($W+ = 822$, $p < 0.001$; Table 3.1; Figure 3.4A). This change was driven by significant increases in three of the four index scores of the WAIS (Table 3.1; Figure 3.4B), greatest in Processing Speed (PSI) and Perceptual Reasoning (PRI). Importantly, however, there were no significant differences in the change in WAIS-IV scores with respect to sex and only PSI changes varied with respect to classroom (i.e., active learning, lecture), evidenced by a significant time by class interaction. The changes in WAIS-IV scores were commensurate with previously reported retest gains among college students across a similar time period.81

3.3.2 Physics task accuracy, but not course grade, is related to WAIS-IV scores differently for male and female students

To assess associations between course performance and intellectual ability, we separately regressed each WAIS-IV score (both post-instruction and pre- to post-instruction changes) on post-instruction course grade and physics-related task accuracies (denoted "performance" below), controlling for students' sex, classroom environment, and other demographics, for a total of 30 separate regressions (con-

trolling for familywise error rate with the Šidák correction).

$$performance = \beta_0 + \beta_1 IQ + \beta_2 IQ \times Sex + \beta_3 IQ \times Class + \beta_4 IQ \times Sex \times Class$$
$$+ \beta_5 Sex \times Class + \beta_6 Sex + \beta_7 Class + \beta_8 Age + \beta_9 Years\ in\ Univ.$$

Here, task accuracy refers to the proportion of correct answers given while participants performed a physics task in the MRI scanner. In addition to the full models (Equation 3.3.2), we considered nested models without the WAIS interaction terms, none of which explained significantly more variance in task accuracy that the full interaction models, as detailed in the following paragraphs.

Post-instruction accuracy on the physics reasoning task was significantly related to post-instruction PRI scores (post PRI; $F(9,\ 120) = 5.122$, $p < 0.001$; Figure 3.5A, C) and FSIQ scores (post FSIQ; $F(9,\ 120) = 5.770$, $p < 0.001$; Figure 3.5B), in addition to the pre- to post-instruction changes in both PRI (ΔPRI; $F(9,\ 120) = 5.034$, p < 0.001; Figure 3.5D, F) and FSIQ scores (ΔFSIQ; $F(9,\ 120) = 4.498$, p < 0.001; Figure 3.5E, G). After controlling for demographics and interactions, post PRI, post FSIQ, ΔPRI, and ΔFSIQ all significantly predicted physics reasoning task accuracy, implying that they were not wholly dependent on students' sex and classroom environment. However, relations between each post PRI, ΔPRI, and ΔFSIQ and performance were moderated by students' sex, such that female students exhibited a more positive relationship between WAIS-IV scores and performance than male students (Figure 3.5B, C, and D; Table 3.3.2, Physics Reasoning Accuracy). Conversely, male students demonstrated overall higher accuracy on the task, in line with previous research (Docktor and Heller, 2008). Of these, the regression of task accuracy on ΔPRI that included class- and sex-interaction terms (i.e., per Equation

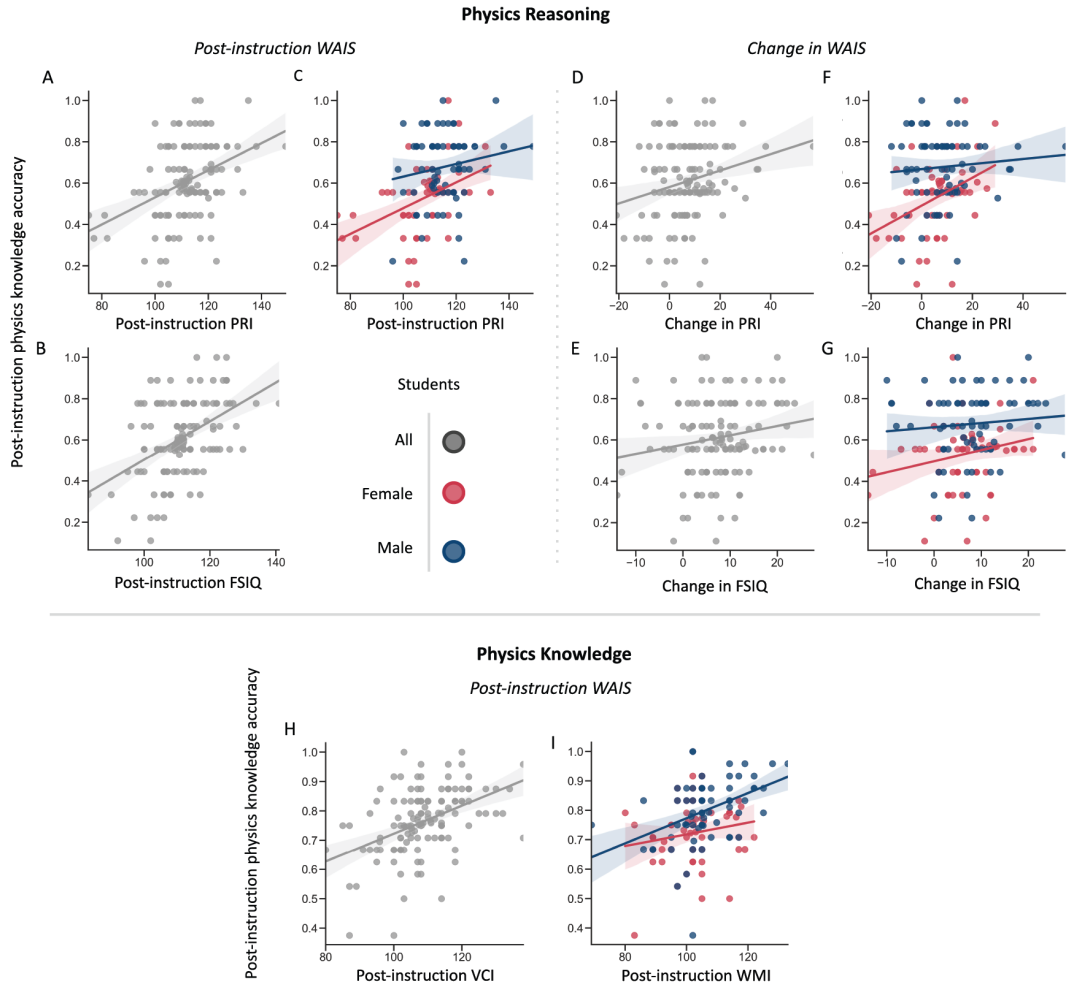

Figure 3.5. Associations between post-instruction physics task performance and WAIS scores.

Table 3.2

Significant relations between physics task performance and WAIS scores.

	Physics Reasoning				Physics Knowledge	
	Post PRI	PostFSIQ	ΔPRI*	ΔFSIQ	Post VCI	Post WMI
WAIS	**0.005**	**0.007**	**0.006**	**0.005**	**0.003**	0.001
WAIS × Sex	**-0.009**	-0.003	***-0.010***	***-0.010***	0.004	**0.005**
WAIS × Class	0.002	0.003	-0.001	-0.001	-0.002	0.002
WAIS × Sex × Class	0.006	-0.001	0.008	0.008	-0.002	-0.004
Sex × Class	-0.764	0.038	-0.109	-0.109	0.270	0.498
Sex (M)	**1.190**	0.458	***0.232***	***0.225***	-0.434	**-0.507**
Class (A)	-0.212	-0.319	0.004	-0.028	0.142	-0.235
Age	**-0.022**	-0.014	**-0.026**	**-0.029**	-0.001	-0.004
Year in Univ.	0.011	0.001	0.020	0.016	-0.014	-0.013

Note. Regression coefficients are shown for each variable on which physics reasoning and physics knowledge were regressed, for OLS regressions of the form shown in Equation 3.3.2 that were significant at $p_{FWE-corr} < 0.05$ and in which a WAIS term or interaction parameter significantly related to performance. Bold text indicates parameters significant at $p < 0.05$; bold and italicized text, parameters significant at $p < 0.01$. *Indicates that the full interaction model (i.e., of the form shown in Equation 1) explained significantly more variance than a smaller model without interactions between each WAIS scores, class (A for active learning), and sex (M for male).

3.3.2) explained significantly more variance in physics reasoning task accuracy than did the corresponding model without interaction terms. Thus, the relation between task accuracy and ΔPRI is better understood in the context of students' sex and classroom environments.

Post-instruction accuracy on the physics knowledge task was associated with post-instruction Verbal Comprehension Index (post VCI; $F(9,\ 120) = 6.474$, $p < 0.001$; Figure 3.5H) and Working Memory Index (post WMI; $F(9,\ 120) = 6.008$, $p < 0.001$; Figure 3.5I). Physics knowledge accuracy was related to post VCI after controlling for potential moderations of this relation by sex and class type (Figure 3.5H). The converse was true of relations between accuracy and post WMI. Post WMI displayed sex-dependent relations with task accuracy, such that greater accuracy in male students' was associated with greater increases in WAIS scores than those of female students (Figure 3.5I). Notably, this moderation was in the opposite direction of those between WAIS scores and physics reasoning accuracy noted above. Of these, regression of task accuracy on WAIS-IV scores with class- and sex-interaction terms (i.e., of the form displayed in Equation 3.3.2) did not explain significantly more of the variance in physics knowledge task accuracy than did the models without interaction terms. This indicates that, unlike in the case of physics reasoning, these sex and class interactions do not significantly add to our understanding of these relations.

We found no relations between students' final course grade and any WAIS-IV score or change therein.

These data indicate that WAIS-IV scores were clearly, but differentially associated with performance on physics-related assessments, suggesting a distinction between skills related to physics conceptual reasoning and content knowledge recall. Significant associations between physics reasoning accuracy and each ΔPRI and

ΔFSIQ suggest that the development of perceptual reasoning ability and general intellectual ability underscore performance in physics reasoning. Similarly, relations between post-instruction VCI and WMI scores and physics knowledge accuracy suggest that working memory and verbal comprehension at post-instruction support successful physics knowledge retrieval, not necessarily the development of those skills (i.e., pre- to post-instruction). Nonetheless, post PRI and full-scale WAIS scores remain relevant for students' performance on physics conceptual reasoning and problem solving tasks, in addition to the development of such skills.

Functional brain network efficiency and connectivity differentially support component intelligence across contexts

To further investigate associations between intellectual ability and task performance, we assessed brain organization during physics-related cognition using regressions of the same form as Equation 3.3.2 above. Specifically, we combined theories of the neurobiology of intelligence (i.e., P-FIT and NEH) with methods for studying individual differences in the brain organization (i.e., connectomics and network science) to search for a common neural substrate underlying the relations between WAIS-IV scores and accurate physics cognition. Measures of functional connectivity and network efficiency (denoted "topology" below) were regressed on WAIS-IV scores, while students' sex, classroom environment, demographics, and head movement (i.e., framewise displacement, "fd").

$$\begin{aligned} topology = {} & \beta_0 + \beta_1 IQ + \beta_2 IQ \times Sex + \beta_3 IQ \times Class + \beta_4 IQ \times Sex \times Class \\ & + \beta_5 Sex \times Class + \beta_6 Sex + \beta_7 Class + \beta_8 Age + \beta_9 Years\ in\ Univ. + \beta_{10} fd \end{aligned}$$

Topological measures were calculated from functional connectivity graphs computed from fMRI data collected while participants performed the physics reasoning and physics knowledge tasks, using two brain parcellations to ensure that results are not parcellation-induced artifacts. In these graphs, individual brain regions comprise nodes and the pairwise correlation of their BOLD signals comprise edge weights, representing functional connectivity. We regressed, separately, (a) global efficiency calculated during each task, (b) local efficiency of each brain region during each task, and (c) connectivity between each pair of brain regions, during each task, on only the WAIS-IV scores significantly related to performance on said task. Significance thresholds of $\alpha < 0.05$ were adjusted to control the familywise error rate using the Šidák procedure, adjusted to account for dependence of correlated measures (Li and Ji, 2005; Šidák, 1967).

These analyses found that WAIS-IV scores were not significantly associated with global efficiency or with local efficiency across the brain during either task. Across both tasks, only head movement was associated with brain network efficiency.

Of the WAIS-IV scores associated with physics reasoning task accuracy, only post FSIQ was additionally associated with task connectivity, of the right anterior insula (Figure 3.6A, 3.6C), depending on students' sex and classroom environment (Figure 3.6B, 3.6D). For female students greater FSIQ was associated with increased connectivity for those enrolled in lecture-based classes, but decreased connectivity for those enrolled in active learning classes. Meanwhile, for male students, the direction of associations between post-instruction FSIQ and connectivity did not differ due to classroom environment, though it did across parcellations (Figure 3.6B, 3.6D). Furthermore, the two parcellations both indicate significant associations of FSIQ and functional connectivity during the physics reasoning task; they did not converge on a particular network or region, providing no specific neuroanatomical locus. Across

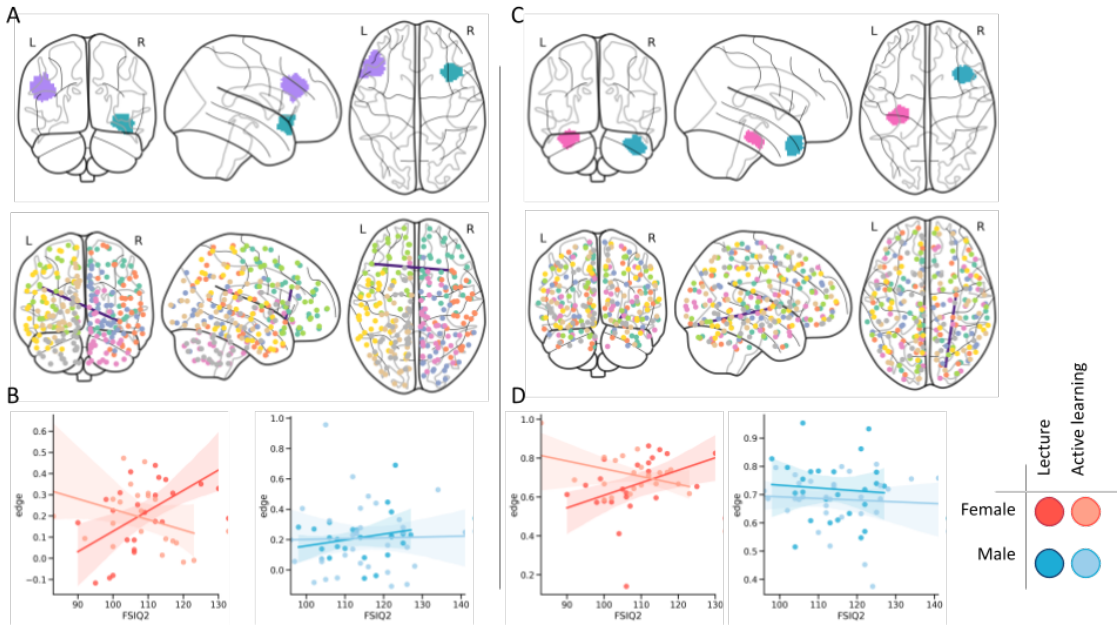

Figure 3.6. Associations between functional connectivity during physics-related cognition and WAIS scores following a semester of physics instruction.

both parcellations, there was no consistent association between functional connectivity during the physics knowledge task and either post WMI or VCI scores, only with students' head movement during this task.

Functional brain networks are not a common neural substrate supporting the role of intelligence in physics-related cognition.

We sought to explore possible common neural substrates for WAIS-IV and physics reasoning task accuracy. To this end, we assembled mediation models to assess whether brain connectivity significantly associated with WAIS-IV scores (Table 3.3.2, Figure 3.6) explains shared variance between WAIS-IV scores and task accuracy (Table 3.3.2, Figure 3.5), accounting for the interactions and covariates in Equations 3.3.2 and 3.3.2. In Figure 3.3.2, lighter red paths indicate significant predictors of post-instruction physics reasoning connectivity, while lighter blue paths indicate significant predictors of post-instruction physics reasoning accuracy. Sin-

Table 3.3

Mediation of the relations between changes in full-scale IQ and physics reasoning accuracy by functional connectivity.

	Connectivity		Accuracy	
	Parameter Estimate	p-value	Parameter Estimate	p-value
FSIQ	-0.009	0.695	0.054	0.123
FSIQ × Sex	**0.062**	0.025	0.011	0.805
FSIQ × Class	0.004	0.894	0.003	0.938
FSIQ × Sex × Class	**-0.071**	0.043	0.023	0.691
Sex × Class	0.046	0.199	0.053	0.355
Sex	-0.037	0.26	***-0.135***	0.001
Class	-0.033	0.154	-0.056	0.191
Age	-0.001	0.931	-0.013	0.168
Year in Univ.	0.002	0.467	0	0.967
Head motion	0.021	0.098	–	–
Connectivity	–	–	0.020	0.891

Note. Bold indicates significant parameters at $\alpha <$, bold and italicized, at $\alpha < 0.01$. Values shown are for the connectivity significant in the Craddock parcellation.

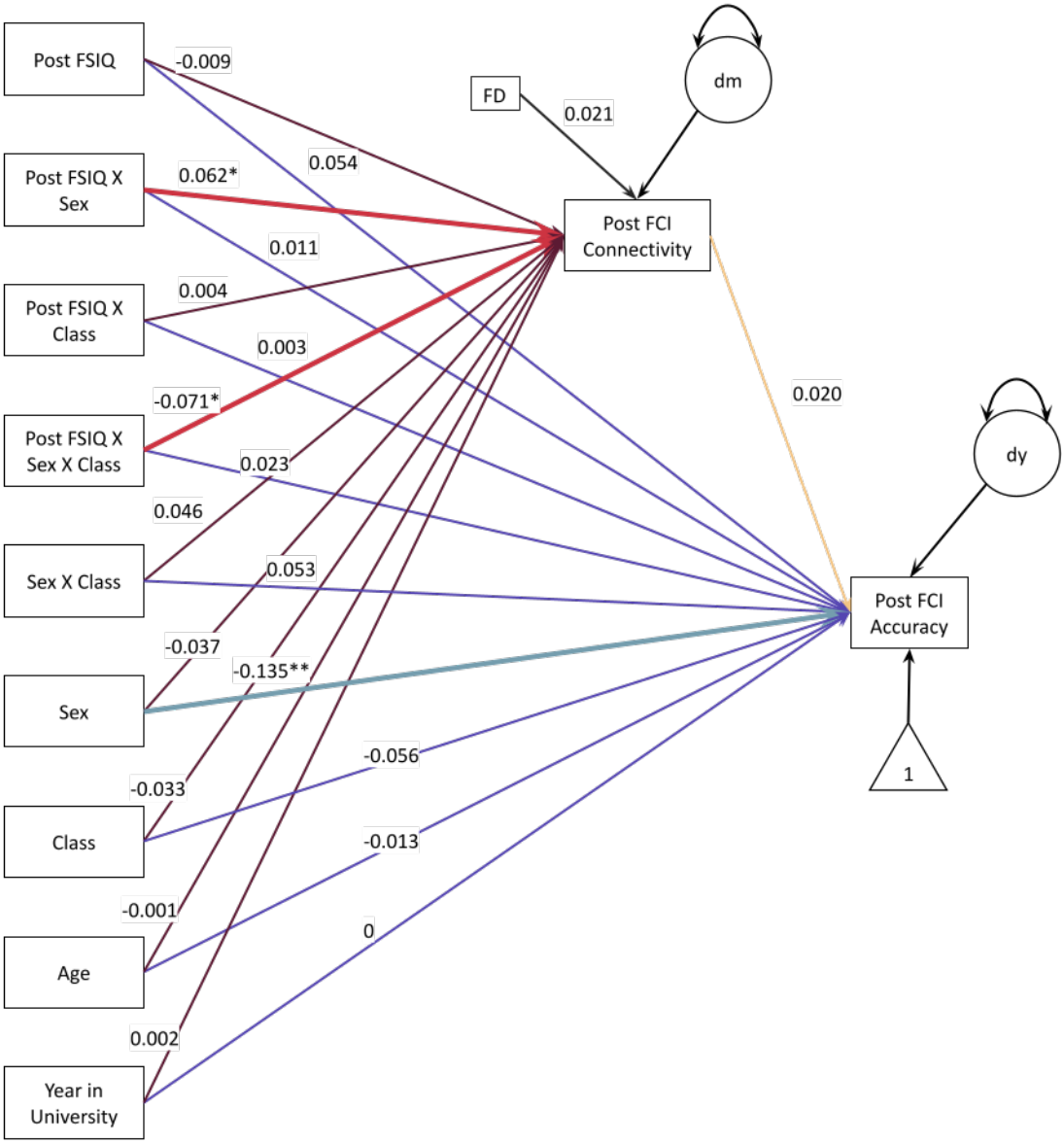

Figure 3.7. Mediation of the association between post-instruction physics reasoning and FSIQ by functional connectivity.

gle asterisks indicate significance at $p < 0.05$; double, at $p < 0.01$. The model was significant at $p_{FWE-corr} < 0.01$. Covariance between exogenous variables is not shown in this model, but was assessed. These models (Figure 3.3.2) indicated that the functional connectivity was unrelated to students' accuracy on the task (yellow path), and while FSIQ scores continued to explain a significant proportion of variability in connectivity (red paths), they no longer significantly accounted for variability in physics reasoning task accuracy (blue paths). Across parcellations, anterior insula connectivity (Figure 3.6) was significantly associated with FSIQ, but not task accuracy, and provided no significant mediation of the IQ-accuracy relationship. While WAIS scores and task accuracy seem to capture related behavioral phenomena, FSIQ-accuracy relations were weakened by the inclusion of functional connectivity in the model, and connectivity did not significantly mediate the FSIQ-accuracy association (Table 3.3.2).

3.4 Discussion

We present evidence of complex relations between intellectual ability and physics learning, behaviorally and neurally. Among these are significant increases in cognitive ability (i.e., WAIS-IV scores) over a semester of physics instruction, corresponding with previously reported increases in college students tested twice over a similar three-month period. These changes did not differ based on students' sex or based on course pedagogy and classroom environment. Gains in PRI were among the largest across index scores, and positively related to physics reasoning accuracy, suggesting the WAIS-IV components driving FSIQ increases represent physics-related skill development. Our data indicate that WAIS-IV measures of cognitive ability related to task performance were also related to brain connectivity, but not efficiency, during

the task. Although we found no evidence of common neural underpinnings for the performance-ability relationship, we did uncover a moderation of brain-ability associations by students' sex and physics classroom environment. Therefore, the neurobiology supporting skill acquisition is context-, and perhaps pedagogy-dependent, instead of intrinsic, underscoring the importance of experience and environment in associations between ability and performance.

3.4.1 Are WAIS gains related to physics education or general college education?

The study of intelligence lacks concrete definitions,84,85 but compensates with extensive psychometrics (Neisser et al., 1996; Weinberg, 1989; Howe, 1997; Hunt, 1997; Richardson, 2002; Schönemann, 1983; STERNBERG, 1984). Here, we observed gains in intellectual ability, per the WAIS-IV, that outpaced retest gains reported in the measure's standardization sample, (Wechsler et al., 2008; Holdnack et al., 2013; Cullum and Larrabee, 2010). These gains were seen across WAIS-IV index scores, greatest in PSI and PRI though minimal in WMI, but did not significantly differ from retest effects over a similar time period in an independent sample of college students (Estevis et al., 2012). WAIS-IV score increases were not related to overall class performance (i.e., course grade), but instead to physics reasoning ability at course completion (i.e., post-instruction FCI accuracy). This suggests they capture skills related to students' grasp of Newtonian mechanics. While these relations are moderated by students' class type, we cannot definitively link WAIS-IV score increases to physics instruction, and pedagogical differences therein, without a control group of participants who were not exposed to a semester of physics instruction. Perceptual reasoning seems directly related to skills developed by the

specific demands of the course, though it and other components of intelligence are likely developed and engaged broadly across university coursework. To clarify how university instruction may impact different components of intelligence across disciplines, future studies should extend assessments across a range of curricula in the sciences, arts, and humanities and compare retest effects across a semester of education with those seen in the broader population. This is especially relevant given (a) the present evidence of education-related gains in WAIS-IV scores that appear to capture domain-specific skills and, importantly, (b) the widespread belief in the significance of intelligence and IQ for student and life success.

3.4.2 Perceptual reasoning improvements underlie physics-related cognition.

Following a semester of physics instruction, physics knowledge task accuracy was related to post-instruction verbal comprehension and working memory abilities, while physics reasoning task accuracy was related to pre- to post-instruction changes in perceptual reasoning and full-scale WAIS-IV scores (i.e., intellectual or cognitive ability), in addition to their values post-instruction. These differences between the tasks may reflect their unique cognitive demands and the manner in which the associated skills are acquired and exercised.

While physics knowledge reflects memorization of formulae and definitions of physics concepts, physics reasoning reflects the development of accurate conceptual understanding of Newtonian mechanics and the macro-scale forces at work in the physical world (e.g., gravity, friction). Students are unlikely to know definitions and formulae learned in class before enrolling in a physics class, but working memory and reading comprehension skills captured by WMI and VCI are not domain-specific

but domain-general abilities exercised across curricula (Beilock and DeCaro, 2007; Piacente, 2012; Helwig et al., 1999; Simmons et al., 2012; Chen and Whitehead, 2009; Staver and Jacks, 1988; Cassels and Johnstone, 1984; Daneman and Hannon, 2001; Andreassen and Bråten, 2010). Therefore, it is the associated post-instruction scores that capture the skills utilized in accurately recalling physics knowledge, i.e., recalling definitions and formulae learned in class, that were not necessarily developed during the class.

Conversely, physics reasoning draws on the conceptions of physical phenomena that students bring into their first physics class, including pre-existing "common-sense beliefs" acquired over a lifetime of interacting with the physical world that are at odds with scientific explanations (Halloun and Hestenes, 1985). Decades of physics education research suggests that these pre-existing conceptions are difficult for students to overcome, even with formal instruction (Eaton et al., 2019; Halloun and Hestenes, 1985; Neidorf et al., 2020). As this instruction relies heavily on visual representations of the movement of macroscale objects to teach Newtonian mechanics, encouraging students to rely on visuospatial skills and mental imagery, likely developing their perceptual reasoning skills throughout the course (Hegarty, 2014; Buffler et al., 2008; Kozhevnikov et al., 2002; Kozhevnikov and Thornton, 2006; Madsen et al., 2012; Pallrand and Seeber, 1984). In this respect, our results indicate that students with more accurate conceptions of Newtonian mechanics following physics instruction were those demonstrating larger increases in perceptual reasoning skills and greater absolute perceptual reasoning ability, post-instruction. It may be that either (a) students who acquire more perceptual reasoning skills in the course are better able to align their conceptions of how the world works with Newtonian explanations or (b) students who are able to overcome previous conceptions about the mechanisms of the physical world gain better perceptual reasoning

skills than their peers who have more difficulty doing so. In the former case, these findings might illustrate another approach for physics instructors to help students develop more accurate conceptions of Newtonian mechanics: focusing on students' perceptual reasoning skills.

Insofar as changes in WAIS-IV index scores reflect skill acquisition, it follows that physics reasoning accuracy, reflecting conceptualizations of Newtonian mechanics honed throughout physics instruction, is associated with changes in perceptual reasoning and full-scale intelligence, though absolute levels of these skills remain relevant. Conversely, physics knowledge accuracy, reflecting correct recall of definitions and formulae learned throughout physics instruction, is associated with absolute levels of verbal comprehension and working memory skills, but not the changes therein. This may highlight an opportunity for correcting students' pre-existing conceptions of physical phenomena, by focusing on students' perceptual reasoning skills throughout their physics instruction. However, the sex- and classroom-differences suggest that this is not a one-size-fits-all opportunity, and that students may benefit differently from such learning interventions. For female students, who historically perform poorer on the Force Concept Inventory (our physics reasoning task) due to a host of sociocognitive factors, increases in PRI might provide a means of "leveling the playing field" compared to their male counterparts.

3.4.3 Brain network organization during physics cognition is related to intelligence, but does not explain its relationship with physics task accuracy.

Contrary to previous research (Hearne et al., 2016; Dubois et al., 2018; Hilger et al., 2017b), our data did not indicate associations between students' WAIS-IV scores and either global or local brain network efficiency and did not support the neural efficiency hypothesis, which suggests more intelligent individuals benefit from "more efficient" brains. The concept of neural "efficiency" is, at best, unclear, with varying definitions across the years, and at worst, empty and misleading.110 Here, we use the network science definition of efficiency, defined mathematically in terms of connections between nodes of a graph (i.e., connectivity between brain regions) (Rubinov and Sporns, 2010). These findings add domain-specific insight to what has previously been a study of intrinsic abilities and neurobiological processes, failing to find support in a sample demonstrating intelligence gains related to an extrinsic manipulation (i.e., physics instruction).

Representing a small proportion of all possible connections in the brain, sparse connectivity during physics reasoning was related to post-instruction full-scale WAIS-IV scores, the full complement of measured verbal, perceptual, working memory, and processing speed skills. Different roles of pedagogy and classroom environment on brain network connectivity, cognitive abilities, and relations between the two with respect to sex point out a potentially significant sex difference in classroom experience. In a heavily male-dominated field like physics, it is a reasonable assumption that male and female students would have differential classroom experiences (Gewin, 2011; Charles and Bradley, 2009; Baird, 2018; Richman et al., 2011; Barone, 2011). For example, it is a commonly-held stereotype that men are good at math and that

women are not (Nosek et al., 2002; Spencer et al., 1999), subjecting women and female students to stereotype threat in physics classrooms, where beliefs negatively affect classroom experience and performance (Keller and Dauenheimer, 2003; Keller, 2007). The data presented here indicate no sex differences in overall course performance, but persistent sex differences across classroom environments in associations between cognitive abilities and not only performance on physics assessments, but brain connectivity during those assessments. Together, this literature and our findings indicate meaningful neurobiological consequences of classroom experience on the basis of sex. However, our data show no sex- or classroom-related difference in changes in intelligence or cognitive abilities, though the brain-WAIS relationships point to a multitude of neurobiological representations of intelligence. Neural phenomena related to intelligence differ based not only on cognitive context,but on sociological and pedagogical contexts, as well.

Although intelligence measures associated with physics task accuracy also explained certain functional connectivity during these tasks, we found no evidence of a common neural substrate for intelligence and accuracy. On the whole, task-based connectivity related to intelligence was unrelated to participants' accuracy on said task, and no connectivity associated with intelligence accounted for any significant portion of the relations between intelligence and accuracy. This may indicate a lack of a common neural substrate for these two phenomena, suggesting that, while they are measuring a similar skill or capacity, they are not measuring the same skill or capacity. On the other hand, the neural substrate of a common skill measured by PRI or full-scale WAIS scores and physics reasoning accuracy may merely lie beyond a linear relation with brain network connectivity or topology. In any case, understanding the neural instantiation of a common reasoning skill to perceptual reasoning and physics reasoning requires further study.

Finally, rather than a global property of brain network organization, as indicated in prior research (Heuvel et al., 2009; Langer et al., 2012; Zhao et al., 2021), these data indicate that sparse, coordinated interactions of disparate brain regions underlie intelligence, in this domain-specific context. That connections across the brain during physics-related cognition are related to changes in students' overall intellectual skills, but differently with respect to their classroom environment, casts further doubt on the notion of IQ or intelligence as a fixed, innate measure and, instead, highlights the role of environment and experience. Although differences in these relationships between female and male students support the substantial body of literature supporting this notion of sex differences in the biological representations of intelligence (Neubauer and Fink, 2009; Dunst et al., 2014; Haier et al., 2005; Jiang et al., 2019; Ryman et al., 2016; Satterthwaite et al., 2015; Schmithorst and Holland, 2007; Wu et al., 2013), here we suggest a potential sociological explanation. These findings indicate, too, neural support for intelligence exhibits domain-specific relations in the context of STEM education. Not only does cognitive context matter to the relations between intelligence and brain network organization, our data indicate that sex and learning environment matter, too.

3.4.4 Limitations and future directions

Here, we demonstrate increases in intellectual ability over a semester of physics instruction. While students' sex and classroom environment did not affect the extent of these increases, the data suggest meaningful consequences of classroom environment on the relations between intellectual ability and underlying brain network organization. The implication that the learning environment affects male and female students differently, both cognitively and neurobiologically, in a field as male-

dominated as physics demands attention and further study. However, as there were no observed sex differences in final course grade or in change in intellectual ability, any differences in classroom experience are not differentially affecting female and male students' academic performance in the course. Further work should assess whether differences in experience and associated brain function are linked to long-term success for male and female students. This assessment should consider factors beyond overt measures of success and focus on variables related to self-efficacy and in-classroom experiences, both previously been shown to affect men and women differently in physics education (Nissen and Shemwell, 2016).

While we have identified differences in physics-related brain organization and its relation to WAIS-IV scores based on class type, this study is unable to distinguish whether these differences are due to differences in pedagogy or social classroom environment, or to practice effects across a short assessment period. Future research should include in-classroom assessments of social climate and the possibility of gender differences in social interactions during physics instruction. Furthermore, control groups in (1) another, less male-dominated, domain and (2) an age- and sex-matched groups outside of university would provide insight into both sex differences in and the degree to which changes in intellectual ability are associated with STEM education.

3.5 Conclusion

While our data indicate clear relations between domain-specific components of intellectual ability and performance on assessments of physics conceptual reasoning and content knowledge, we found no association between academic outcomes and intelligence. Likewise, intelligence was related to functional brain connections dur-

ing these assessments, but none that explain its association with performance. Our multifaceted approach to studying the neurobiological underpinnings of intelligence did not uncover a single, robust aspect of brain network organization that was consistent across cognitive contexts and experiential influences. However, relations between intelligence and, separately, physics task performance and task-based functional connectivity were moderated by students' sex and classroom environment. Ultimately, both the magnitude and development of perceptual reasoning skills is meaningful over the course of a semester of physics instruction. As these data show significant increases in perceptual reasoning over a semester, we present an optimistic and more plastic view of intelligence, as a set of skills to be developed, rather than an innate capacity that students either have or don't have. Together, these data highlight the complicated nature of relations between intelligence and classroom successes, which vary with students' sex and domain-specific experiences.

CHAPTER 4

ENDOCRINE INFLUENCES ON WITHIN-INDIVIDUAL VARIABILITY IN BRAIN AND MOOD

4.1 Introduction

The brain is a critical node of the endocrine system that both directs hormone release throughout the body and is subject to direct action of those hormones, which impact both its structure and function (Albert et al., 2017; Beltz and Moser, 2020; Brown et al., 2015; Galea et al., 2017; Taylor et al., 2020). Cortisol, an end-product of "stress-response" HPA axis activity in humans, binds to glucocorticoid receptors distributed across the prefrontal cortex (PFC) and hippocampus (Barfield and Gourley, 2018; de Kloet et al., 2019), influencing social cognition (Smeets et al., 2009), affective processing (Wirth et al., 2011; Peters et al., 2016), attention, and memory (Vedhara et al., 2000). Ovarian hormones, estradiol and progesterone, are integral to the female reproductive system and fluctuate throughout the menstrual cycle and bind to receptors in the hypothalamus, hippocampus, and PFC (Rossetti et al., 2016). Estradiol influences dendritic spine density (Hao et al., 2006) and plays roles in learning and memory (Jacobs and D'Esposito, 2011; Jacobs et al., 2016; Taxier et al., 2020), as well as emotion regulation (Graham et al., 2017; van Wingen et al., 2011). Progesterone's role in cognition, however, is largely inconclusive (Henderson, 2018). Taken together, widespread impacts of direct hormone action in the brain and the well-documented fluctuations in hormone levels with respect to circadian rhythms and the menstrual cycle suggest that incorporating endocrine measures adds significant value to any neuroscientific study.

Menstrual cycle-related and circadian rhythms are intertwined in women with ovulatory menstrual cycles, influencing each other in myriad, bidirectional ways

(Baker and Driver, 2007). These associations are due in large part to biochemical interactions between the hypothalamic-pituitary-adrenal (HPA) axis and the female reproductive system (Chrousos et al., 1998; Kajantie and Phillips, 2006; Rabin et al., 1990; Stockhorst and Antov, 2015; Handa and Weiser, 2014; Roy et al., 1999) and are reflected by differences in sleep (Meers and Nowakowski, 2020), circulating cortisol (Hamidovic et al., 2020; Nepomnaschy et al., 2011) and stress (Leeners et al., 2019). Subjective sleep quality is notably poorer leading up to menses (Meers et al., 2019; Meers and Nowakowski, 2020; Baker and Lee, 2018; Becker et al., 2018), poor sleep quality influences female reproductive health (Willis et al., 2019), and HPA axis dysfunction can cause disrupted sleep (Buckley and Schatzberg, 2005). Furthermore, women are more likely as men to experience severe consequences of sleep dysregulation (Mong and Cusmano, 2016) and to be diagnosed with insomnia (Mallampalli and Carter, 2014; Krishnan and Collop, 2006; Suh et al., 2018; Zhang and Wing, 2006), which, in turn, increases their likelihood of developing depression (Miller, 2004). Similarly, not only do women report different stress levels (Leeners et al., 2019; Duchesne and Pruessner, 2013; Woods et al., 1998) and experience different neuroendocrine stress responses (Collins et al., 1985) across the menstrual cycle, there is overwhelming evidence that stress inhibits the reproductive system and disrupts the menstrual cycle (Nagma et al., 2015; Nillni et al., 2018; Schliep et al., 2015; Xiao and Ferin, 1997; Tarín et al., 2010; Dobson et al., 2003). As is the case with sleep, women are more likely than men to experience stress-related mental disorders (McLean et al., 2011; Bangasser and Valentino, 2014; Tolin and Foa, 2008), symptoms of which are exacerbated during the days preceding menses (Nillni et al., 2018), and may be more susceptible to stress-linked cardiovascular conditions (Vaccarino and Bremner, 2017). These effects are likely products of tangled interactions between ovarian hormones and the HPA axis (Bangasser and Valentino, 2014;

Handa and Weiser, 2014; Oyola and Handa, 2017; Nicolaides et al., 2014; Buckley and Schatzberg, 2005), which is implicated in both stress (Karin et al., 2020; Faravelli et al., 2012) and sleep dysregulation (Buckley and Schatzberg, 2005). Studies of both humans and non-human animals have demonstrated both ovarian hormones (e.g., progesterone and estrogen) influence on HPA axis function (Bangasser and Valentino, 2014; Handa and Weiser, 2014; Lalmansingh and Uht, 2008; Ochedalski et al., 2007; Phumsatitpong et al., 2020; Stephens et al., 2016; Roy et al., 1999; Wirth et al., 2007) and differences in HPA axis reactivity as ovarian hormones fluctuate across menstrual cycle phases (Baker and Driver, 2007; Parry et al., 2000; Stephens et al., 2016). This constellation of neuroendocrine actions and interactions has implications not only for reproduction and women's reproductive health, but for cardiovascular and mental health, as well (Vaccarino and Bremner, 2017).

Unfortunately, issues of women's health and endocrine interactions have received little attention in human neuroscience (Taylor et al., 2021). While sex differences have been studied at length (Sacher et al., 2013), specific questions of women's physiology are not. Human neuroimaging research on the menstrual cycle is incredibly limited, with vital questions about the roles of day-to-day hormone fluctuations largely unanswered. Together with the knowledge that ovarian hormones easily diffuse across the blood-brain barrier and bind to receptors throughout the central nervous system, it is surprising that the neurobiological and neurocognitive consequences of such ubiquitous hormonal phenomena as the menstrual cycle and neuroendocrine stress cascade (i.e., HPA axis activity) have received relatively little attention in human neuroscience research. Indeed, a recent review revealed that less than 8% of published human neuroimaging studies (across imaging modalities) of the menstrual cycle that directly assessed hormone levels assessed individuals at more than three time points and only 30% assessed individuals more than twice (Dubol

et al., 2021). This is problematic, as menstrual cycles vary within an individual as much as they do between individuals (Fehring et al., 2006) and hormone fluctuations are not limited to discrete changes between menstrual cycle phases. Thus, what little attention ovarian hormones and endocrine rhythms have received in human neuroimaging research is insufficient to characterize how hormone dynamics across the menstrual cycle interact with the brain. Given the expression of ovarian hormone receptors across the brain *and* the interactions of ovarian hormone metabolites with major neurotransmitter systems (Baulieu et al., 1996; Brinton et al., 2008), these regular fluctuations have greater consequences than are currently considered in neuroimaging research.

A recent trend in neuroimaging research has potential to address these gaps, moving toward "precision neuroscience" by densely-sampling a few individuals with repeated measurements over days, weeks, and months to better understand both between- and within-individual variability in the brain (Gordon et al., 2017; Poldrack et al., 2015; Poldrack, 2017; Pritschet et al., 2020). Historically, the study of individual variability in the brain has been limited by issues of reliability, validity, and statistical power, which has further limited our understanding of just how much the human brain varies in the course of everyday life. One significant limitation is the problem of "unmodeled noise" (Dubois and Adolphs, 2016; Van Horn et al., 2008), variability in the brain that introduces noise into fMRI research and potentially obscures effects of interest. This noise comes a host of sources, from physiological measures that affect the measurement of the BOLD signal (Handwerker et al., 2012; Liu, 2016, 2013; Murphy et al., 2013) to factors influencing cognition and both brain function and structure (Taylor et al., 2020; Pritschet et al., 2020; Stewart et al., 2003; Nyberg et al., 2006). As research continually uncovers sources of such noise, accounting for potential confounds becomes unweildy and im-

practical, and researchers are forced to prioritize the measures and confounds they include in their study. Without understanding the degrees to which each source impacts the variables and neural structures of interest, this prioritization is largely unguided. Dense-sampling approaches allow for better measurement and estimation of individual variability, including previously unmodeled sources of noise. Such research has already shown us how much variability in the brain comes from individual, group, and cognitive factors (Gratton et al., 2018), how large-scale brain networks vary between individuals and the stability of these networks across individuals (Seitzman et al., 2019), and how brain structure and function varies over the menstrual cycle (Fitzgerald et al., 2020; Pritschet et al., 2020; Taylor et al., 2020). While dense-sampling approaches offer a nearly unparalleled opportunity to better elucidate hormone-brain interactions, only a minority include endocrine assessments (Arélin et al., 2015; Barth et al., 2016; Filevich et al., 2017; Pritschet et al., 2020). As the study of individual differences in fMRI advances, with dense-sampling and precision approaches, it is imperative that women not be left behind.

To advance an understanding of women's health, we asked to what extent functional brain network organization varies with respect to lifestyle factors (i.e., stress, sleep) and ovarian hormones across the menstrual cycle. Here, we will use data from two dense-sampling datasets: the Dense Investigation of Variability in Affect (DIVA) Study and 28andMe, which both include resting-state fMRI data, hormone measurements, and behavioral assessments. Resting state functional connectivity matrices were computed per session for each individual. First, we assessed magnitudes of variability in functional brain networks due to site, individual, and birth control use and examined how they were distributed across the brain. Then, we assessed proportions of variability in functional brain network connectivity and topology associated with sleep quality, perceived stress, and ovarian hormone levels within

individuals over time with an ordinary least squares regression model, controlling for site, individual, and BC use. We expected that site and individual would account for more variability in functional connectivity than birth control use, in line with findings from Gratton et al. (2018), and that endocrine measures (i.e., levels of estradiol and progesterone) would account for more variability than lifestyle measures (i.e., stress and sleep), due to their direct biochemical mechanisms in the brain. Finally, we expected that functional connectivity of the prefrontal cortex and networks in to which it is connected would be specifically associated with hormonal changes over the course of the menstrual cycle. Learning more about the interplay between endocrine factors, lifestyle measures, and within-individual variability in brain function will help us better understand the brain on a day-to-day level, while specifically advancing knowledge of factors contributing to differential health outcomes for women.

4.2 Methods

Here, we leverage two, complementary datasets to assess how hormones and behavior are associated with brain network variability: the DIVA Study and 28andMe.

4.2.1 Participants and Study Design

DIVA

Data were collected from three premenopausal, female participants (age range = 26-31). Two participants were using combined oral contraceptives and the third was freely cycling, with a history of regular menstrual cycles who had not used hormonal contraceptives in the prior year. Participants completed behavioral assessments

and collected saliva samples twice a week, 3-4 days apart, and completed MRI scanning sessions once a week (on a behavioral + hormone collection day). Written, informed consent was obtained from each participant before data collection began, in accordance with Florida International University's Institutional Review Board approval.

28andMe

Data were collected from one premenopausal, female participant (age 23, right-handed) with no history of neurological or psychiatric diagnoses, a history of regular menstrual cycles, and no hormonal contraceptive use in the prior year. She completed daily hormone, behavioral, and MRI assessments for 28 days. Written, informed consent was obtained from this participant before data collection began, in accordance with the University of California, Santa Barbara Human Subjects Committee.

4.2.2 Hormone Data

DIVA

Endocrine measures include salivary estradiol and progesterone. Saliva samples were collected via passive drool into 2mL sterile cryovials shortly after waking twice a week (3-4 days apart) and stored at -20C until shipping. Samples were shipped following completion of the study to Salimetrics' SalivaLab (Carlsbad, CA), where they were assayed using the Salimetrics Salivary Estradiol Assay Kit (Cat. No. 1-3702) and the Salimetrics Salivary Progesterone Assay Kit (Cat. No. 1-1502), without modifications to the manufacturers' protocols. All samples were assayed in duplicate.

28andMe

Endocrine measures include serum estradiol and progesterone. Blood draws were completed daily at or within 30 minutes of 10:00AM local time.

4.2.3 Behavioral Data

DIVA

Participants completed semiweekly assessments via mobile device shortly after waking, following saliva collection. The data used here are a subset of a larger battery, and include the Perceived Stress Scale (Cohen et al., 1994) and and the Pittsburgh Sleep Quality Index (Buysse et al., 1989).

28andMe

The participant completed daily assessments shortly before having blood drawn for hormone assessments. The data used here are a subset of a larger battery, and include the Perceived Stress Scale and the Pittsburgh Sleep Quality Index.

4.2.4 Neuroimaging Data Preprocessing

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 20.2.1 (Esteban et al., 2018a,b); RRID:SCR_016216), which is based on *Nipype* 1.5.1 ((Gorgolewski et al., 2011, 2018, RRID:SCR_002502).

Anatomical data preprocessing

A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU)

with `N4BiasFieldCorrection` (Tustison et al., 2010), distributed with ANTs 2.3.3 (RRID:SCR_004757 Avants et al., 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast` (FSL 5.0.9, RRID:SCR_002823 Zhang et al., 2001). Brain surfaces were reconstructed using `recon-all` (FreeSurfer 6.0.1, RRID:SCR_001847 Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438 Klein et al., 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with `antsRegistration` (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* (RRID:SCR_008796; TemplateFlow ID: MNI152 NLin2009cAsym Fonov et al., 2009).

Functional data preprocessing

For each of the 3 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated by aligning and averaging the first echo of 4 single-band references (SBRefs). A B0-nonuniformity map (or *fieldmap*) was estimated based on two (or more) echo-planar imaging (EPI) references with opposing phase-encoding directions, with `3dQwarp` (Cox, 1996) (AFNI 20160207). Based on the estimated susceptibility distortion, a corrected EPI (echo-planar imaging) reference was calculated for

a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve and Fischl, 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (FSL 5.0.9 Jenkinson et al., 2002). BOLD runs were slice-time corrected using `3dTshift` from AFNI 20160207 (RRID:SCR_005927 Cox and Hyde, 1997). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. A T2* map was estimated from the preprocessed BOLD by fitting to a monoexponential signal decay model with nonlinear regression, using T2*/S0 estimates from a log-linear regression fit as initial values. For each voxel, the maximal number of echoes with reliable signal in that voxel were used to fit the model. The calculated T2* map was then used to optimally combine preprocessed BOLD across echoes following the method described in Posse et al. (1999). The optimally combined time series was carried forward as the *preprocessed BOLD*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD time-series were resampled onto the following surfaces (FreeSurfer reconstruction nomenclature): *fsnative*, *fsaverage5*. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise

displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, (Power et al., 2014)) and Jenkinson (relative root mean square displacement between affines (Jenkinson et al., 2002)). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, the aCompCor masks are subtracted a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The

confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each confound (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of *fMRIPrep* use *Nilearn* 0.6.2 (Abraham et al., 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in *fMRIPrep*'s documentation (https://fmriprep.readthedocs.io/en/latest/workflows.html).

Copyright Waiver

The above boilerplate text was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts *unchanged*. It is released under the CC0 license (https://creativecommons.org/publicdomain/zero/1.0/).

4.2.5 Neuroimaging Data Analysis

Following fMRI preprocessing, the Individual Differences in Connectivity (IDConn) pipeline was used to estimate connectivity per participant per session, from which measures of network topology were computed.

First, Nilearn was used to mask preprocessed fMRI data with two atlases in turn: a 400-region cortical parcellation parcel atlas (Schaefer et al., 2018) and a whole-brain parcellation based on clustered resting-state functional connectivity (resulting in 393 regions of interest) (Craddock et al., 2012). The rest of the data analysis and inferential statistics were performed in duplicate, on each atlas, separately, to prevent any conclusions drawn from further analysis from being tainted by parcellation artifacts.

After regressing out six motion parameters, CSF and white matter signals, and their temporal derivatives, the average residual BOLD signal was extracted per region and standardized. Each pair of mean BOLD signals were then correlated, to assemble an adjacency matrix per participant per session. Then, proportional thresholds, κ, were estimated at which (a) each matrix ceases to be node-connected (i.e., fragments) and (b) each matrix begins to exhibit scale-free degree distribution. These thresholds were to define the upper and lower limits (respectively) of the range across which networks would be thresholded (in increments of $\delta\kappa = 0.01$).

Graph measures calculated here include global and local estimates of each network integration and segregation. Global measures, modularity (Q) and characteristic path length (l_G), were calculated with each network across the range of thresholds. Then, the area under the curve (AUC) was calculated per measure, across the range of thresholds, for a single measure of each modularity and characteristic path length per subject per session that is less biased by choice of a single proportional threshold. Local measures, clustering coefficient (C_i,) and betweenness centrality

(g_i), were calculated for each node, i, of each network across the range of thresholds. The AUC for each of these measures, per region, were then calculated across the range of thresholds for a single vector of each clustering coefficient and betweenness centrality per participant per session.

4.2.6 Inferential Statistics

Analysis 1: Lifestyle and Endocrine Variables

First, we assessed distributions of and relationships between independent variables: hormone levels, stress, and sleep quality. Mean, median, standard deviation, skewness, and kurtosis were calculated per variable (Table 1) and Pearson's correlation coefficients were calculated between each pair of variables (Table 2).

Analysis 2: Connectome Similarity

The degree of similarity in functional connectivity across the whole brain, i.e, connectome similarity, due to each (1) group, (2) site, (3) individual, and (4) hormonal contraceptive use was quantified by computing a Pearson's correlation coefficient for each pair of networks, across individuals and sessions, Fisher transforming these correlations for normality, and averaging the resultant Z-scores within each of those four groupings for an overall estimate of whole-brain network similarity.

Analysis 3: Connectome Variability

Then, the degree to which values of each of levels of estradiol ($[E_2]$), progesterone ($[P_4]$), and the interaction thereof ($[E_2] \times [P_4]$), sleep, stress, and the interaction thereof are associated with functional connectivity was assessed using mass-univariate ordinary least squares regressions for each network edge (i.e., connection

between brain regions), controlling for the effects of site, individual, session, and hormonal contraceptive use (assessed above) and averaging the coefficient of determination (R^2) across edges for a single measure per session per individual. Effects were further assessed (a) region-wise, by averaging coefficients of determination for all of a region's edges and (b) system-wise, by averaging coefficients of determination for all edges of all regions within each of 17 canonical, large-scale functional brain networks (Thomas Yeo et al., 2011).

The magnitude of variability in functional network topology due to each (1) site, (2) individual, (3) hormonal contraceptive use, and (4) values of each of $[E_2]$, $[P_4]$, $[E_2] \times [P_4]$, sleep, stress, and the interaction thereof was quantified by running a series of OLS regressions of the form shown in Equation 4.2.6, and extracting the coefficient of determination (R^2) associated with the model at each edge, in addition to the proportion of variability associated with each predictor at each edge, using the following equation (from Cohen (1973)):

$$\omega^2 = \frac{SS_{effect} - df_{effect} \times MS_{error}}{MS_{error} + SS_{total}} \tag{4.1}$$

.

By doing so, we estimate what proportion of variability in brain network connectivity and topology is accounted for by each variable, beyond that which is already accounted for by the other variables.

$$\begin{aligned} connectivity = \beta_0 + \beta_1 site + \beta_2 participant + \beta_3 hormonal\ contraceptive\ use \\ + \beta_4[E_2] + \beta_5[P_4] + \beta_6[E_2] \times [P_4] + \beta_7 sleep + \beta_8 stress + \beta_9 sleep \times stress \\ + \beta_{10} head\ motion \end{aligned} \tag{4.2}$$

Analysis 4: Toplogical Variability

Using OLS regressions of the form shown in Equation 4.2.6, associations between brain network topology and each endocrine, lifestyle, and logistical variables were likewise assessed. Specifically, we regressed, separately, each modularity (Q) and characteristic path length (l_G) on: site, participant identity, HC use, $[E_2]$, $[P_4]$, $[E_2] \times [P_4]$, sleep, stress, and the interaction thereof, while controlling for head motion. These topological measures were selected because they represent opposing network properties: segregation and integration, respectively.

4.3 Results

Table 4.1

Descriptive statistics of participant and lifestyle variables across the sample

	$\bar{x}$	s	Q_1	Q_3
Age (years)	26.3	2.86		
Sleep quality	8.0	8.62	3.0	8.5
Stress	11.9	8.58	5.0	17.5
Head Motion	0.10	0.017	0.09	0.11

Note. Values of sleep quality are from the Pittsburgh Sleep Quality Index (Buysse et al., 1989), on which higher scores indicate poorer sleep quality, and values of stress are from the Perceived Stress Scale (Cohen et al., 1994), on which higher scores indicate greater perceived stress. Head motion is estimated from framewise displacement across fMRI scans included in data analysis.

Data were collected from 4 pre-menopausal women (ages 23 to 30), from 2 sites, over the course of three to four weeks, on a daily-to-weekly timescale. Questionnaire measures included the Pittsburgh Sleep Quality Index (Buysse et al., 1989) and Perceived Stress Scale (Cohen et al., 1994), endocrine measures included as-

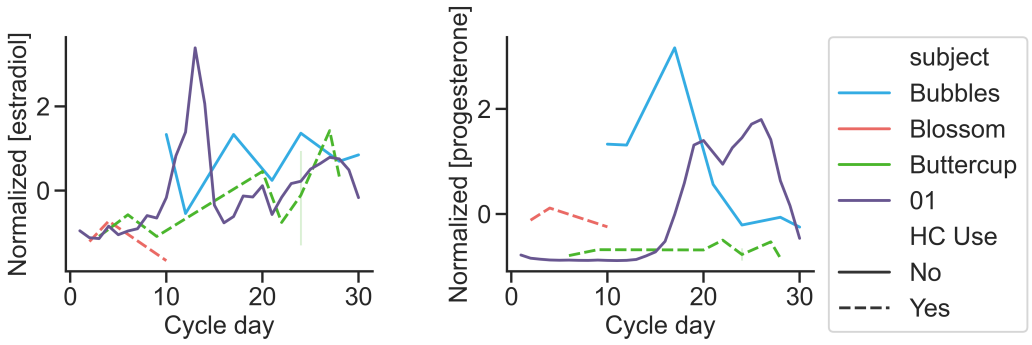

Figure 4.1. Normalized estradiol and progesterone concentrations.

says of salivary or serum estradiol and progesterone concentrations ($[E_2]$ and $[P_4]$, respectively), which were standardized to enable comparison across measurement technique, which differed by site. Hormone concentrations across the measurement period are displayed in Figure 4.1, which displays differences in cycle timing and measurement period across subjects. Here, "cycle day" refers to the number of days since the start of the participant's most recent menses. Finally, participants' head motion was well within the acceptable range for resting-state fMRI data, indicating that no participant's imaging data was sufficiently corrupted by motion.

4.3.1 Analysis 1: Lifestyle and Endocrine Collinearity

Pairwise Pearson's and point biserial (for dichotomous variables site and HC use) correlation coefficients and associated significance testing revealed significant associations between non-brain variables in this dataset. Site was significantly associated with hormonal contraceptive use, sleep quality, and perceived stress; hormonal contraceptive use was significantly associated with sleep quality and perceived stress; menstrual cycle day was significantly associated with $[E_2]$; and perceived stress was significantly associated with sleep quality (all $p < 0.01$. Due to this evidence of

Table 4.2

Correlations between hormone, lifestyle, and logistical variables

	Site	HC Use	Cycle Day	$[E_2]$	$[P_4]$	Sleep quality	Perceived Stress
Site	–	4.0×10^{-9}	0.49	1	1	3.7×10^{-7}	6.3×10^{-4}
HC Use	***0.71***	–	0.24	0.093	0.046	1.1×10^{-17}	2.5×10^{-6}
Cycle Day	0.098	-0.26	–	0.007	0.41	0.91	0.27
$[E_2]$	0	-0.24	***0.58***	–	0.15	0.63	0.64
$[P_4]$	0	*-0.29*	-0.21	0.21	–	0.030	0.50
Sleep quality	***0.64***	***0.88***	0.025	-0.070	*-0.32*	–	2.0×10^{-10}
Perceived stress	***0.46***	***0.60***	0.24	-0.068	-0.099	***0.75***	–

Note. Values above the diagonal are p-values to the corresponding Pearson's correlation coefficients shown below the diagonal. Significant correlations at p ¡ 0.05 are indicated in italics, while significant correlations at $p < 0.01$ are indicated in bold italics.

multicollinearity, these variables were orthogonalized before their inclusion in the following regression models.

4.3.2 Analysis 2: Connectome Similarity

First, the upper triangles of each connectivity matrix (i.e., per subject per session) were vectorized and Pearson's correlation coefficients were calculated for each pair of connectivity matrices (Figure 4.2). Then, network similarity due to each group, site, HC use, and individual was calculated as the average Fisher-transformed Pearson's correlations for each (1) the whole *group*, but different sites, individuals, HC use status, (2) the same *site*, but different individuals and HC use status, (3) the same *HC use* status, but different individuals, and (4) the same *individual*, but different sessions, respectively (Figure 4.3). In line with results from Gratton et al. (2018), we found that network similarity within each individual was more than twice the magnitude of similarity between individuals (i.e., across the group). Additionally,

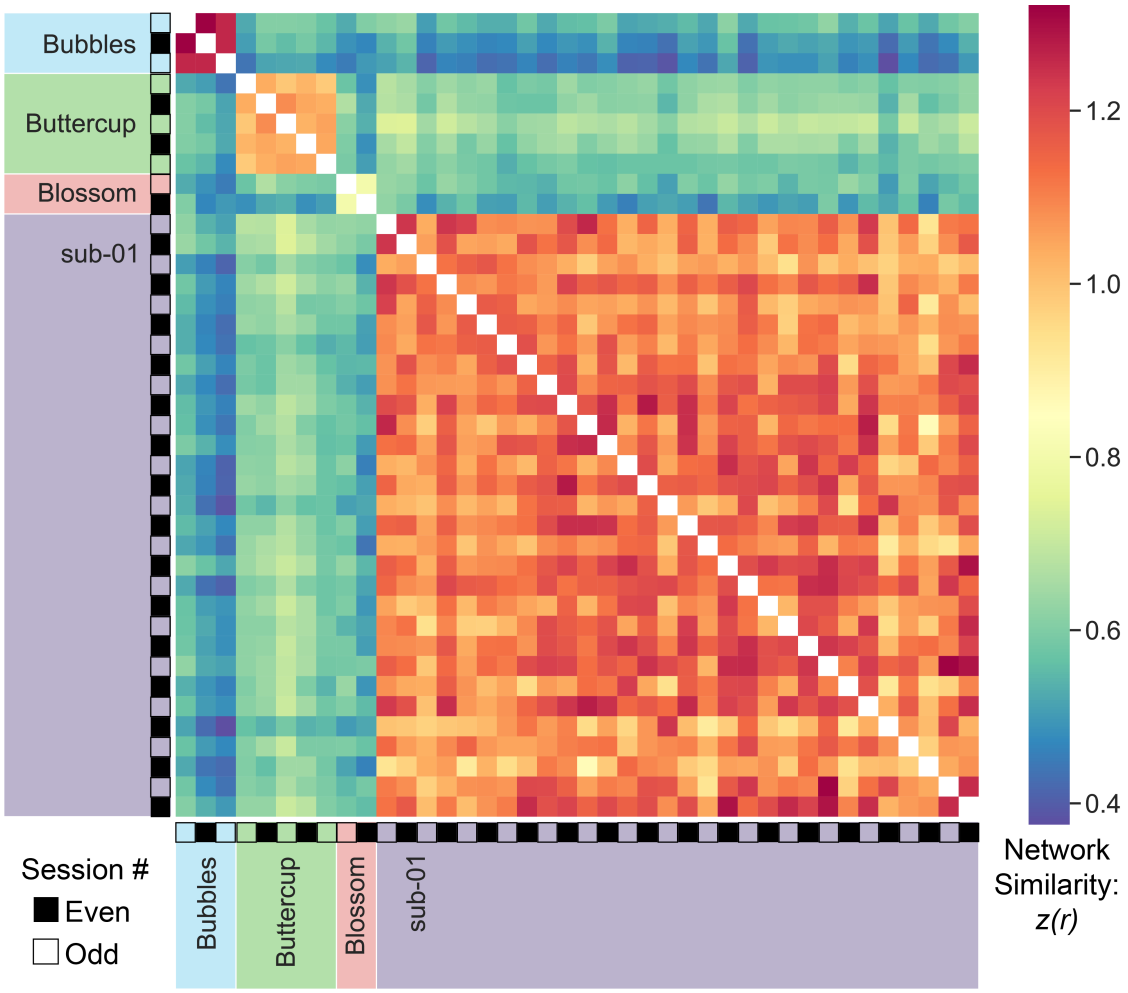

Figure 4.2. Resting-state functional connectome similarity across subjects and sessions.

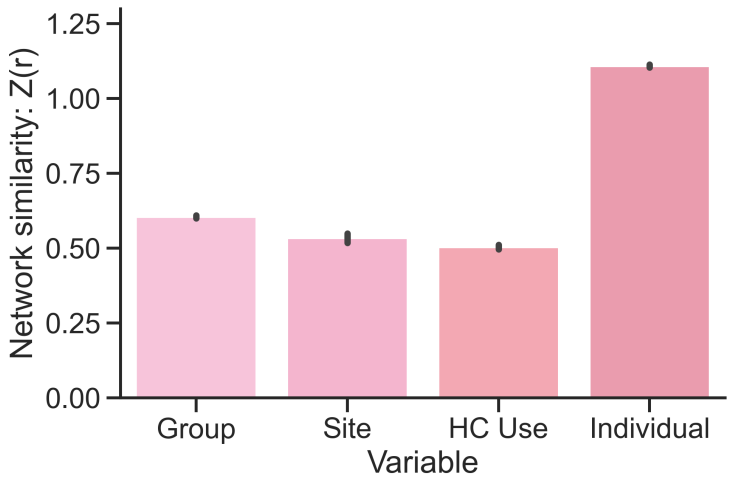

Figure 4.3. Resting-state brain network similarity across categorical variables.

we found that site and HC use status account for similar magnitudes of network similarity to group, though variability due to each of these variables was significantly different from that of the others ($p < 0.01$). Due to the significant correlation between HC use and site, these effects warrant further exploration (see Analysis 3 below).

4.3.3 Analysis 3: Connectome Variability

Each edge from the vectorized upper triangles of each connectivity matrix was regressed on site, HC use, $[E_2]$, $[P_4]$, $[E_2] \times [P_4]$, sleep, stress, sleep $\times$ stress, and head motion following the form shown in Equation 4.2.6. Proportions of variability in functional connectivity (i.e., per edge) were assessed for the whole model using R^2, adjusted for the number of predictors, and for each predictor using $\omega^2_{partial}$.

Table 4.3

Proportions of variance in functional connectivity explained by endocrine, lifestyle, and logistical variables

	$R^{\bar{2}}_{adj}$ *or* $\bar{\omega}^2$	s_{R^2}	Q_1	Q_3	95th %ile
Model	0.36	0.23	0.18	0.54	0.74
Site	0.035	0.060	0	0.045	0.15
Participant	0.069	0.10	5.0×10^{-3}	0.090	0.26
Cycle Day	2.7×10^{-3}	9.2×10^{-3}	0	7.0×10^{-4}	0.016
HC Use	2.9×10^{-3}	0.010	0	7.0×10^{-4}	0.016
$[E_2]$	4.8×10^{-3}	0.016	0	1.5×10^{-3}	0.27
$[P_4]$	2.4×10^{-3}	9.6×10^{-3}	0	0	0.014
$[E_2] \times [P_4]$	5.5×10^{-3}	0.018	0	2.7×10^{-3}	0.030
Sleep	3.2×10^{-3}	0.011	0	1.1×10^{-3}	0.019
Stress	4.1×10^{-3}	0.012	0	2.3×10^{-3}	0.023
Sleep × Stress	4.0×10^{-3}	0.013	0	1.3×10^{-3}	0.022
Head Motion	2.8×10^{-3}	9.6×10^{-3}	0	6.0×10^{-4}	0.016

Note. Model values reflect the coefficient of determination (R^2) adjusted for the number of predictors in the model, while individual predictor values reflect the proportion of variance in connectivity that is accounted for by the predictor itself (ω^2). Values are averaged across all edges in the 400-region functional connectivity matrix, representing an average of the effect of each predictor across the functional connectome.

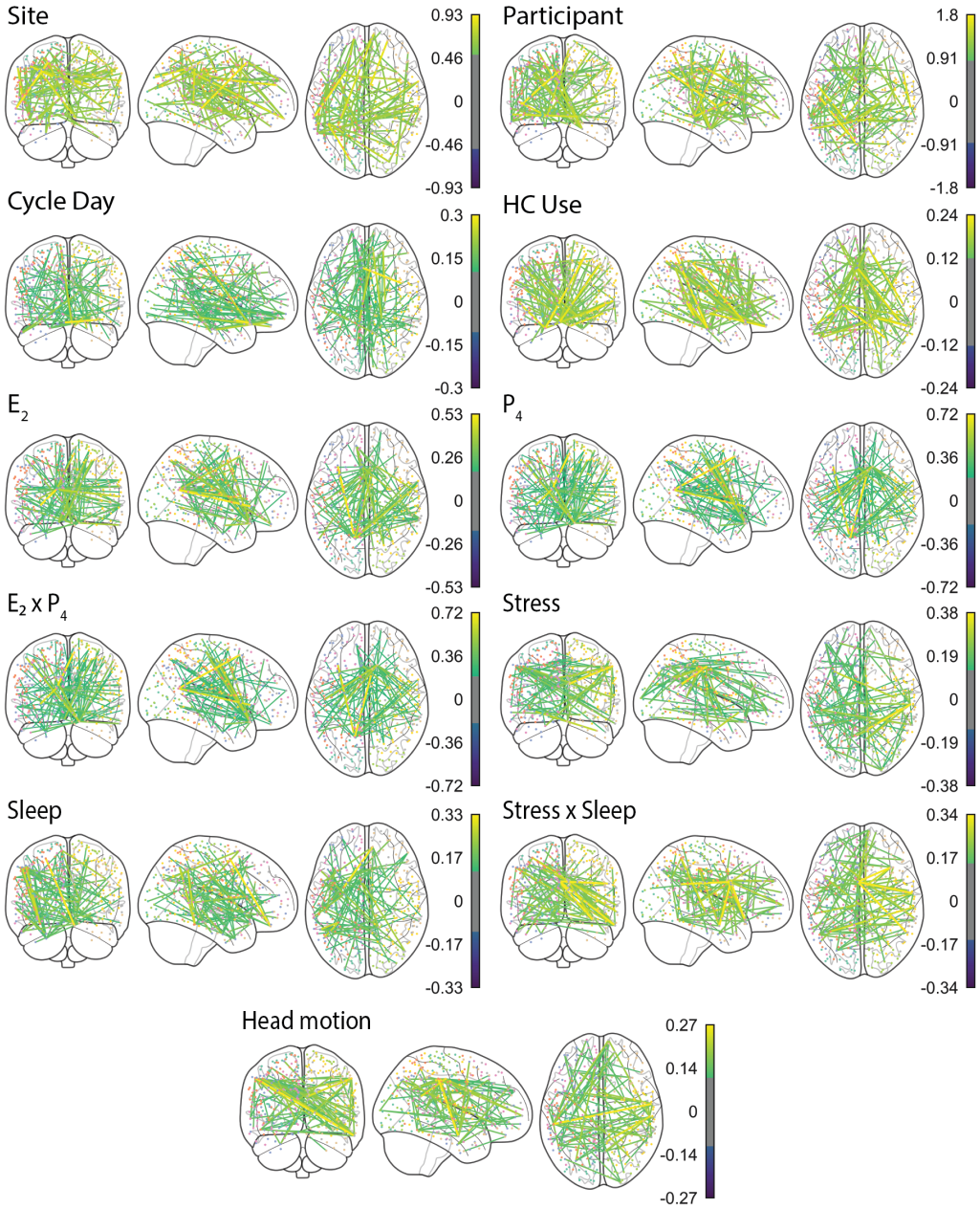

Figure 4.4. Resting-state functional connectome variance explained by predictor.

Edge-Wise Variability

Across the brain, the model shown in Equation 4.2.6 accounted for 36% of the variability in resting-state functional connectivity, adjusted for the number of terms in the model. Participant identity accounted for a greater proportion of of variability (ω^2) in connectivity (6.9%) than any other predictor, including site (3.4%) and HC use status (0.3%), in line with the results of the previously described whole-brain similarity analysis (Section 4.3.2). Values of ω^2 were compared between predictors, across connections, using paired t-tests, which confirmed that participant identity accounted for significantly more variance than any other predictor (average $t(80{,}199) = 176.9$, $p < 0.01$), followed by site (average $t = 123.1$, $p < 0.01$). The interaction $[E_2] \times [P_4]$ accounted for significantly more variability in connectivity than the next most predictive predictor, $[E_2]$ ($t = 24.6$), which accounted for significantly more variability than perceived stress ($t = 9.8$) and the interaction of stress and sleep ($t = 11.9$), which did not account for a significantly different proportion of variability from each other, but which both accounted for significantly more variability than sleep quality alone ($t_{stress} = 15.8$, $t_{stress \times sleep} = 12.5$), which accounted for significantly more variability than HC use ($t = 7.7$), cycle day ($t = 10.8$), and head motion ($t = 9.8$), each of which did not significantly differ in the amount of variability they accounted for, but all of which accounted for significantly more variability in connectivity than $[P_4]$ ($t_{HC} = 12.8$, $t_{cycle\ day} = 7.7$, $t_{head\ motion} = 8.3$, respectively), all while controlling the familywise error rate at $p_{FWE-corr}(t) < 0.01$, per the Šidak correction, adjusted for the number of effective comparisons between variables (Li and Ji, 2005; Šidák, 1967).

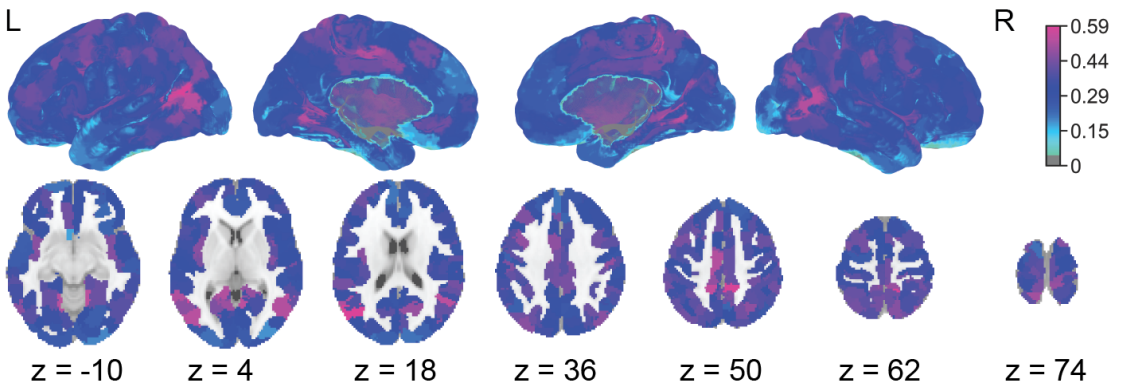

Figure 4.5. Overall resting-state functional connectome variance explained, region-wise.

Region-Wise Variability

Proportion of variance in all functional connections of each region was computed by averaging the model R^2_{adj} within each region (Figure 4.5). The model shown in Equation 4.2.6 accounted for more than half of the variance in functional connectivity of the temporooccipital junction (bilaterally), right medial frontal and parietal regions, left striate and extrastriate regions, left inferior parietal lobule, and left postcentral gyrus. On the other hand, the model accounted for less than 20% of the variance in functional connectivity of the orbitofrontal cortex (bilaterally), left temporal pole, and right extrastriate regions.

Proportions of variance in each region's functional connectivity was assessed, likewise, per predictor included in the model. Individual factors were associated with higher proportions of variance in medial temporal, inferior frontal, and temporal connectivity than in connectivity elsewhere in the brain. Site accounted for a higher proportion of variance in frontal and parietal connectivity. Cycle day accounted for higher proportions of variance in ventromedial prefrontal, medial temporal, and precuneus connectivity. HC use accounted for higher proportions of variance in ventromedial prefrontal and medial temporal connectivity. Estradiol levels ($[E_2]$), pro-

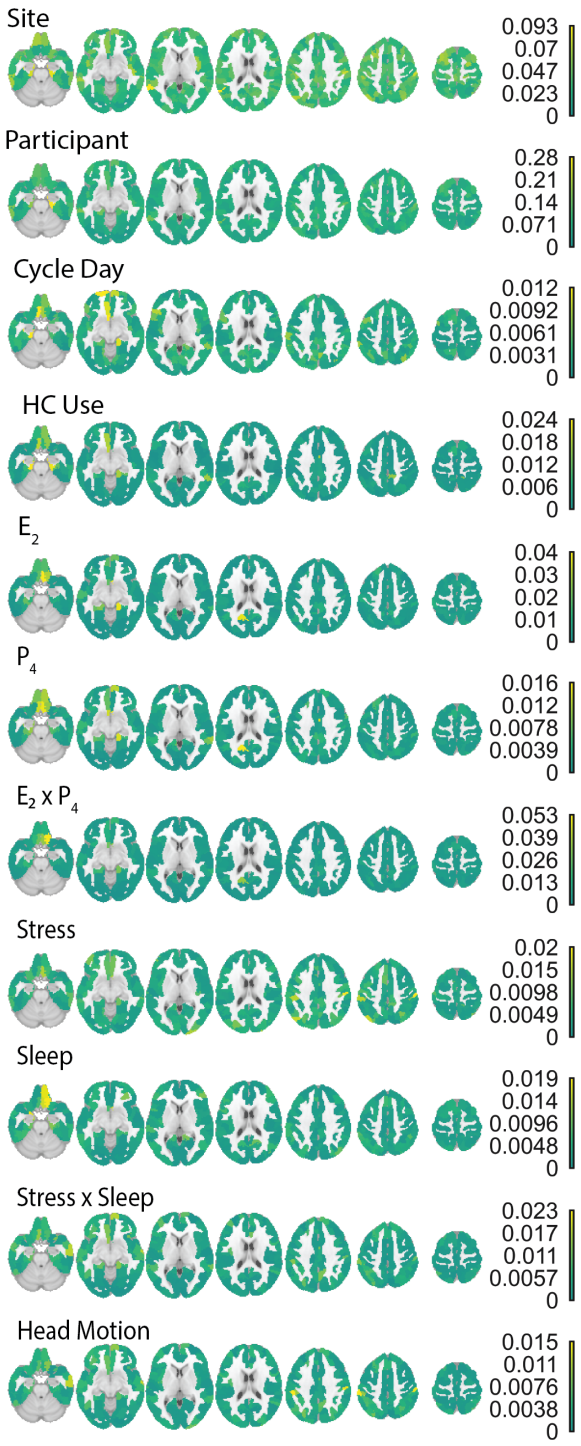

Figure 4.6. Resting-state functional connectivity variance explained by predictor, region-wise.

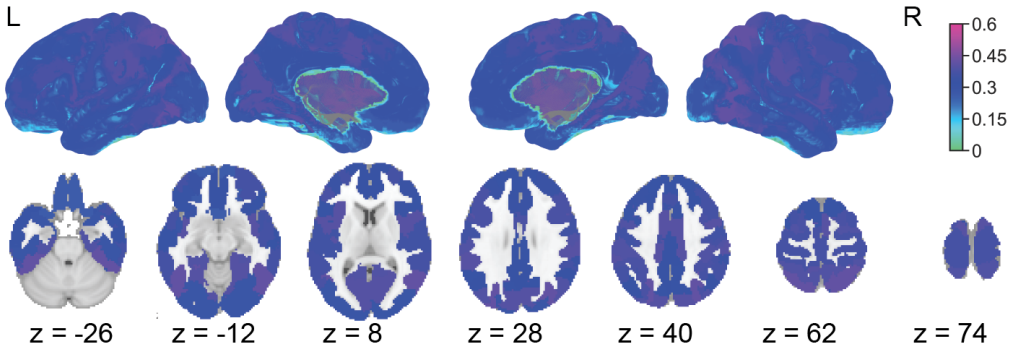

Figure 4.7. Overall resting-state functional connectome variance explained, system-wise.

gesterone levels ($[P_4]$), and the interaction thereof ($[E_2] \times [P_4]$) accounted for higher proportions of variance in ventromedial prefrontal, medial temporal, and precuneus connectivity. Stress accounted for higher proportions of variance in ventromedial prefrontal and parietal connectivity. Sleep quality accounted for higher proportions of variance in right orbitofrontal, middle frontal, and precuneus connectivity. The interaction of stress and sleep accounted for higher proportions of variance in anterior frontal, temporal, and parietal connectivity. Head motion accounted for higher proportions of variance in anterior frontal, temporal, and postcentral connectivity.

System-Wise Variability

Proportion of variance in all functional connections of each functional system was computed by averaging R^2_{adj} within each of 17 large-scale functional systems (or networks; Figure 4.7). The most variance in functional connectivity accounted for by the model shown in Equation 4.2.6 was in portions of the dorsal ($\bar{R}^2_{adj} = 0.42$) and ventral ($\bar{R}^2_{adj} = 0.40$) attention networks, peripheral visual network ($\bar{R}^2_{adj} = 0.40$), and portions of the default mode network ($\bar{R}^2_{adj} = 0.39$). The least variance in functional connectivity accounted was in the limbic networks ($\bar{R}^2_{adj} = 0.29$, $\bar{R}^2_{adj} = 0.26$), central visual network ($\bar{R}^2_{adj} = 0.29$), and in other portions of the default

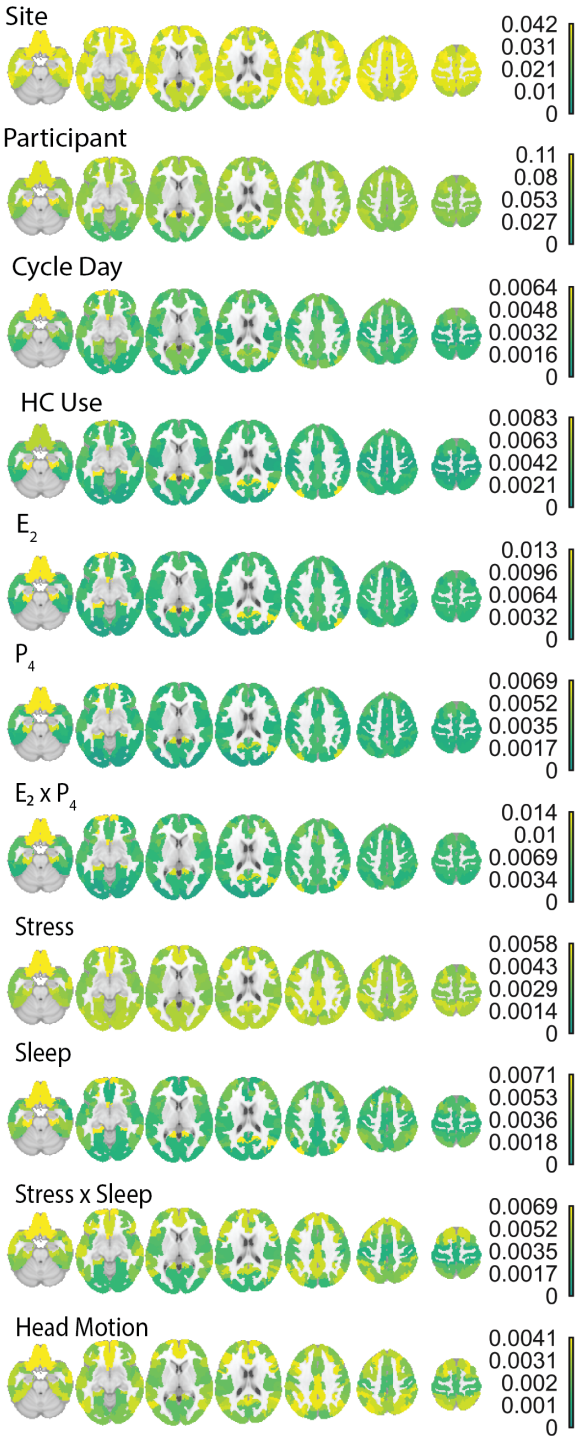

Figure 4.8. Resting-state functional connectivity variance explained by predictor, system-wise.

mode network (anterior prefrontal, temporal: $R^{\bar{2}}_{adj} = 0.31$; medial prefrontal, precuneus/posterior cingulate: $R^{\bar{2}}_{adj} = 0.32$), and in posterior portions of the control network ($R^{\bar{2}}_{adj} = 0.32$).

Proportions of variance in each network's functional connectivity was assessed, likewise, per predictor included in the model. Individual factors were associated with higher proportions of variance in default mode, limbic, and control network connectivity than in connectivity elsewhere in the brain. Site accounted for a higher proportion of variance in limbic and control network connectivity. Cycle day accounted for higher proportions of variance in limbic and default mode network connectivity. HC use accounted for higher proportions of variance in limbic and sparse default mode network connectivity. Estradiol levels ($[E_2]$) accounted for higher proportions of variance in widespread limbic and default mode connectivity. Progesterone levels ($[P_4]$) accounted for higher proportions of variance in limbic, default mode, and control network connectivity. The interaction of estradiol and progesterone ($[E_2] \times [P_4]$) accounted for higher proportions of variance in limbic, default mode, and salience/ventral attention network connectivity. Stress accounted for higher proportions of variance in limbic, default mode, and portions of control and dorsal attention network connectivity. Sleep quality accounted for higher proportions of variance in limbic, salience/ventral attention, and control network connectivity. The interaction of stress and sleep accounted for higher proportions of variance in limbic, control, and default mode network connectivity. Head motion accounted for higher proportions of variance in limbic and default mode network connectivity.

Table 4.4

Variability in network topology associated with endocrine, lifestyle, and logistical variables

	Modularity		Characteristic Path Length	
	Statistic	p-value	Statistic	p-value
R^2_{adj}	0.384	–	0.491	–
$F(df = 11)$	1.73	0.123	***13.14***	5.6×10^{-8}
Intercept	***0.045***	0.003	0.040	0.301
Site	***0.014***	0.00	-0.017	0.38
Participant				
Cycle Day	-3.4×10^{-5}	0.51	4.0×10^{-4}	0.29
HC Use	4.9×10^{-5}	0.75	0.0001	0.78
$[E_2]$	2.0×10^{-4}	0.35	1.7×10^{-3}	0.13
$[P_4]$	-2.0×10^{-4}	0.54	-7.0×10^{-4}	0.38
$[E_2] \times [P_4]$	-6.7×10^{-5}	0.49	-1.1×10^{-3}	.055
Sleep	1.87×10^{-5}	0.65	-2.0×10^{-4}	0.21
Stress	-1.0×10^{-4}	0.064	1.0×10^{-4}	0.56
Sleep × Stress	-6.6×10^{-5}	0.38	-9.6×10^{-5}	0.68
Head Motion	2.0×10^{-4}	0.25	9.3×10^{-5}	0.80

Note. Standard errors are robust to heteroscedasticity (HC1). Statistics that are significant at $\alpha < 0.05$ are italicized; $\alpha < 0.01$, italicized and bolded.

4.3.4 Analysis 4: Topological Variability

The model shown in Equation 4.2.6 was used to investigate brain network topology (i.e., modularity and characteristic path length), as well. These regressions revealed that this constellation of endocrine, lifestyle, and logistical factors do not significantly predict modularity ($F(11) = 1.73, p < 0.123$), they do significantly predict characteristic path length ($F(11) = 13.1, p < 0.01$), though no single endocrine or lifestyle variable accounts for a significant proportion of variability in either modularity or characteristic path length (see Table 4.4).

4.4 Discussion

Here, we combined two datasets that densely sampled a few individuals to better elucidate within-individual variability in the brain and replicated previous findings that group and individual factors account for the largest proportions of variability in functional brain connectivity, while extending these findings to address relative contributions of scanning site and hormonal contraceptive use. Overall, we found that, individually, day-to-day variations in lifestyle factors affecting women's health (i.e., sleep quality, stress) and ovarian hormone levels explain little variance in resting-state functional connectivity (on average $< 1\%$ each). Within this group of factors, ovarian hormone levels However, when considered together, site, HC use, hormone levels, cycle day, perceived stress, sleep quality, and head motion accounted for more than a third of day-to-day variability in functional connectivity. After participant identity and site, the interaction of estradiol and progesterone levels explained the next most connectome variance, followed by estradiol levels alone, stress, the interaction of stress and sleep, sleep quality, HC use, cycle day, head motion, and progesterone levels. These results further the current understanding of individual

connectome variability by incorporating continuous endocrine and lifestyle factors, in addition to discrete individual, group, and site factors.

4.4.1 Menstrual cycle, sleep, and stress

Across the sample we did not observe significant associations between ovarian hormone levels and either sleep quality or perceived stress. We did, however, observe significant associations between HC use, perceived stress, and sleep quality within the sample, which cannot be generalized to a larger population effect due to the small sample size. As only individuals from one site used HCs, collinearity between site and HC use makes it difficult to disentangle their effects in this sample. Further, there was no observed association between menstrual cycle day and either subjective sleep quality or perceived stress (Table 4.2), contrary to prior research (Meers et al., 2019; Meers and Nowakowski, 2020; Leeners et al., 2019; Duchesne and Pruessner, 2013). On the other hand, perceived stress levels reported by freely-cycling individuals in the sample ranged from low to moderate, below thresholds that might cause reproductive system dysfunction.

4.4.2 Logistical contributions to connectome variability.

In line with previous work by Gratton et al. (2018), we found that day-to-day variability in the resting-state functional connectivity was dominated by stable, individual factors. Not only were the relative magnitudes of variation due to group and individual factors revealed here in line with those previously reported, our analyses found commensurate network similarity values to those reported by Gratton et al. (2018): connectome similarity due to group found here was $z(\bar{r}) = 0.60$, similar

to the previously reported $z(\bar{r}) = 0.56$; connectome similarity due to individual factors found here was $z(\bar{r}) = 1.11$, similar to the previously reported $z(\bar{r}) = 1.08$.

In building a reliable, sufficiently powered study of individual variability in the brain, researchers are increasingly moving to consortium science (Thompson et al., 2014; Van Horn and Toga, 2009; Mackey et al., 2016) and multi-site imaging endeavors to increase sample size and participant variability. Here, we build on previous work quantifying sources of variability to include an assessment of site effects ($z(\bar{r}) = 0.53$; $\bar{\omega}^2 = 0.034$, when accounting for additional variables). These results suggest the effect of site is commensurate with variability due to group factors and nearly half the effect of individual factors and an order of magnitude larger than the endocrine and lifestyle effects seen here. Furthermore, the effect of site is not uniformly distributed across the brain For researchers studying individual differences with fMRI data collected at multiple sites (e.g., with data from the ABCD Study℠, UK Biobank), the effect of site must be considered, adding to the existing literature on accounting for site in multi-site neuroimaging studies (Li et al., 2020; Marek et al., 2019).

4.4.3 Endocrine contributions to connectome variability.

Prior research focused on quantifying sources of variability found that within- individual variability over time was relatively minor, compared with individual and group contributions to connectome variability. While the analyses performed here do not explicitly assess the effect of time, they do consider measures of time-varying biological phenomena on daily to weekly timescales. Estradiol, progesterone, and cycle day provide estimates of the effects of the hormonal milieu associated with the menstrual cycle. After individual and site, endocrine factors explained the largest

proportions of variability in functional connectivity. While individual factors accounted for an average of 7% of the variance in connectivity across brain regions, estradiol fluctuations accounted for 0.5%, progesterone fluctuations accounted for 0.24%, and the effect of estradiol fluctuations that differ based on on progesterone levels, 0.55%. Menstrual cycle day accounted for 0.27% of the variance, on average. Together, these data suggest a few considerations for future study of menstrual cycle effects on functional connectivity. First, the effects of estradiol and progesterone alone (i.e., while statistically holding the other constant) account for less variability in functional connectivity than do the effects of estradiol that depend on progesterone levels. Second, menstrual cycle day effect sizes are less than half the magnitudes of either estradiol alone or the interaction of estradiol and progesterone. Together with prior research showing the extent of within-individual variation in the menstrual cycle (Fehring et al., 2006), this suggests that neuroimaging investigations of the menstrual cycle greatly benefit from direct estimation of estradiol and progestone. Further, this adds to extant work suggesting that cycle staging should not rely on estimations based on participant report of the number of days since the start of a previous cycle (Wideman et al., 2013; Allen et al., 2016).

This finding of differences in each estradiol's and progesterone's association with functional connectivity builds on by prior work with one of the datasets included here by Pritschet et al. (2020). These data highlight associations between estradiol and default mode connectivity, which is characterized by trait-like variations of functional network connectivity between individuals (Gratton et al., 2018; Seitzman et al., 2019). Furthermore, the default mode network is characterized by connectivity of medial prefrontal cortex, which our results have associated with estradiol levels and in which prior research with humans and nonhuman animals has identified particularly high concentrations of estrogen receptors (Maki, 2005; Wang et al.,

2010). While our results cannot indicate causality or biochemical mechanisms, when informed by prior research from different fields of neuroscience the suggest that not only do estradiol's actions in the brain preferentially occur in the prefrontal cortex, they influence functional connectivity in brain networks that demonstrate significant variability within individuals. The default mode network is broadly associated with psychopathology (Menon, 2011; Whitfield-Gabrieli and Ford, 2012), particularly with depressive symptoms, which are more common in women, worldwide (Weissman et al., 1993) and in which, estrogen has been broadly implicated (Hernández-Hernández et al., 2019; Shariff et al., 1995; Young et al., 2000). These results provide additional information for neuroimaging researchers studying either or both of depression and default mode network function and variability, and characterize findings of individual variability in default mode connectivity.

Additionally, these results add to previous reports that medial temporal lobe and hippocampal volume changes dramatically over the course of the menstrual cycle, associated with fluctuations in progesterone levels (Taylor et al., 2020; Protopopescu et al., 2008). Our results add that as medial temporal lobe volume changes, so to does its connectivity across the cortex. These findings have profound implications for the study of learning and memory, which focus on hippocampal structure, function, and connectivity. While there is a preponderance of evidence changes in learning and memory with respect to the menstrual cycle (Ikarashi et al., 2020; Pompili et al., 2012; Taxier et al., 2020), endocrine measures are often left out of such research. This and previous work, however, make it clear that individual differences in learning and memory using functional neuroimaging techniques should be aware of the degree to which their results may be influenced by the menstrual cycle and act accordingly.

4.4.4 Lifestyle and endocrine overlap in connectome variability.

Lifestyle variables, stress and sleep, accounted for significantly less connectome variance than did estradiol and the interaction of estradiol and progesterone, but more than HC use, cycle day, and progesterone. The region- and network-wise distributions of connectome variability explained by stress and sleep demonstrates notable overlap with those of endocrine variables. On one hand, this overlap reflects a complex constellation of menstrual and circadian rhythms that interact via the HPA axis, with implications for stress and sleep (Chrousos et al., 1998; Bangasser and Valentino, 2014; Meers and Nowakowski, 2020). Not only does this overlap hint at a functional neuroanatomy of these interactions in the cortex, it potentially complicates the separate studies of sleep, stress, and the menstrual cycle. As changes in each of these variables is reflected by changes in the functional connectivity of the ventromedial prefrontal cortex, medial temporal lobe, and precuneus, isolating the effects of any one of them compels the measurement and consideration of the other two. As the proportion of connectivity variance in these regions accounted for by estradiol levels is up to twice as large as those accounted for by sleep and stress, studies of the latter are more subject to the confounding effects of endocrine fluctuations. This issue is reflected in proportions of network-wise connectivity variance explained by stress, sleep, and endocrine variables, and subject to the same differences in magnitudes of effect between estradiol, and each sleep and stress.

4.4.5 Limitations

Inherent to small sample dense-sampling datasets is the limitation of generalizability. While these data provide insight into how much variability in functional connectiv-

ity across the brain is attributable to endocrine, lifestyle, and logistical variables, they do not allow population-level inference regarding the effects of these variables on brain function. Furthermore, the DIVA Study dataset only includes a handful of timepoints per participant due to premature termination of the project due to the onset of the coronavirus pandemic in early 2020, leaving room for additional phenotyping of these individuals. Furthermore, the differences in measurement timing between the two datasets included here preclude an investigation of autoregressive or time-lagged effects of hormones, sleep quality, and perceived stress on brain network variability, as demonstrated by Pritschet et al. (2020).

4.4.6 Advancing women's health

Here, we present an exploration of several factors specifically related to women's health in an effort to both, (1) inform some implications of menstrual and circadian interactions for brain function and (2) inform future research on issues of women's health in neuroscience. Not only do our results help to characterize findings of brain network variability from prior research (Gratton et al., 2018; Seitzman et al., 2019), they provide guidance to researchers focusing on individual differences in brain function and connectivity, regarding potential confounds in studies of each default mode network function, learning and memory, and the menstrual cycle (i.e., which of the sources of previously "unmodeled noise" to consider). With no shortage of variables to measure and account for, this important for better understanding day-to-day fluctuations in the brain over the course of everyday life and for understanding which lifestyle changes are associated with larger alterations in brain connectivity. Ultimately, building knowledge about the relative influence of endocrine and lifestyle

variables like those included here will hopefully influence future study design, for a more efficient study of women's health.

CHAPTER 5

CONCURRENT FMRI-PHYSIOLOGICAL STUDY OF PSYCHOPHYSIOLOGICAL VARIABILITY

5.1 Introduction

Physiological recordings collected simultaneously during functional magnetic resonance imaging (fMRI) can add valuable information about a participant's physical state and provide quantitative assessment of psychological phenomena. Furthermore, they offer the opportunity to study relations between the central and autonomic nervous systems (CNS and ANS, respectively) that underlie cognition and behavior. For example, physiological arousal has been assessed during fMRI using heart rate, via electrocardiogram (ECG) recordings and skin conductance, via electrodermal activity (EDA) recordings. Inclusion of such measures may enhance interpretation of studies examining decision making (reviewed in (Wong et al., 2011), typical and disordered affective processing (Shokri-Kojori et al., 2018; Mulcahy et al., 2019; Kraus et al., 2007; Goldstein et al., 2005), pain (Perlaki et al., 2015; Mobascher et al., 2010, 2009), autonomic regulation (Napadow et al., 2008; Valenza et al., 2014; Sclocco et al., 2019; Lane et al., 2013; Kuoppa et al., 2012), and fMRI denoising (Hu et al., 2019; Abreu et al., 2017).

Collecting electrophysiological recordings in the MR environment adds MR-induced artifacts to the recordings. Often, the magnitude of these MR artifacts is much larger than that of the phenomena of interest, necessitating additional data cleaning steps before such data can be used to assess psychophysiological phenomena. Single-echo MRI sequences (Figure 5.1A: radio frequency (RF) pulse in the first row, time-varying gradient fields in the following three rows) that measure the blood-oxygenation level dependent (BOLD) signal are the norm in fMRI research.

However, recent advances in MR technology and denoising approaches are prompting researchers to increasingly turn to multi-echo sequences (Figure 5.1B: radio frequency (RF) pulse in the first row, time-varying gradient fields in the following three rows; (Auerbach et al., 2013; Cohen et al., 2017)), which offer better differentiation between BOLD signal and non-neural noise for improved estimates of brain activation (Kundu et al., 2017; Boyacioğlu et al., 2015; Olafsson et al., 2015). While these sequences arguably offer better quality fMRI data, they require more complex radio frequency (RF) and gradient pulses that introduce added artifacts to simultaneously collected electrophysiological recordings. Furthermore, adding echoes to an MRI sequence introduces further limitations on the temporal resolution of the sequence, requiring a longer repetition time (TR). Using multiband or simultaneous multi-slice (SMS) excitation allows researchers to reduce the amount of time required to acquire a single volume, by acquiring several slices simultaneously, and effectively minimizing the multi-echo temporal constraints on TR. This is an important consideration for human fMRI studies, in which a study's power to detect an effect is linearly related to the number of time points in a scan (Desmond and Glover, 2002).

The present study assessed MR-artifact removal strategies from electrophysiological (i.e., ECG and EDA) data collected during single- and multi-echo multiband EPI scans. Currently, there are few commercial manufacturers of MR-compatible equipment for collecting and filtering electrophysiological data during MRI scans for research purposes. Of these, BIOPAC Systems, Inc. (biopac.com) is a popular choice among neuroimaging researchers and among the only manufacturers to offer recommendations for filtering MR-noise out of concurrently collected electrophysiological data. Our goals were to (1) compare MR-related noise from single- and multi-echo EPI sequences, (2) assess current filtering recommendations in multi-

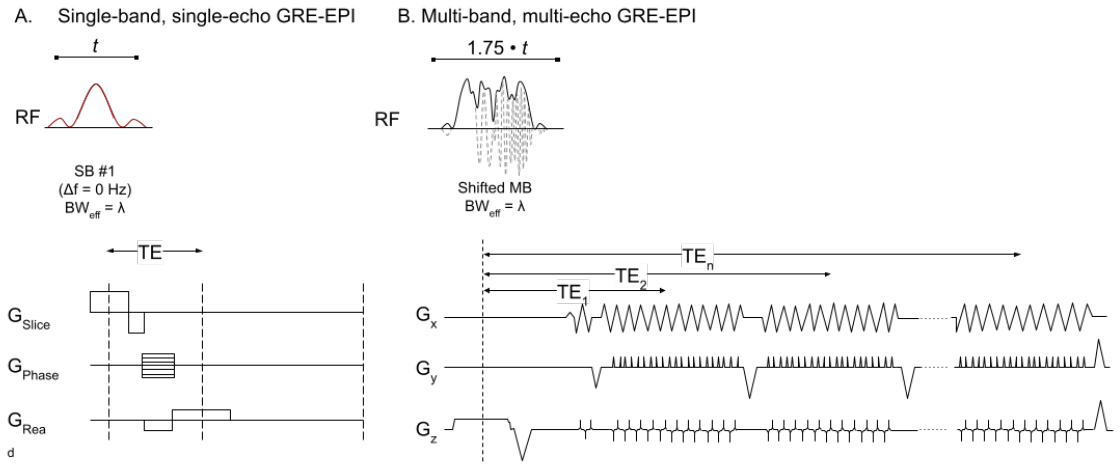

Figure 5.1. Single-band, single-echo vs. multiband, multi-echo echo planar imaging pulse sequences.

band and multi-echo contexts, and (3) if current filtering recommendations appeared insufficient for removing MR-related artifacts from concurrent electrophysiological data, to redefine these recommendations accordingly. To achieve this, we used ECG and EDA data collected during both single- and multi-echo multiband BOLD EPI sequences, from 5 participants across 700+ minutes of scanning. First, data were Fourier transformed to identify MR-artifact frequencies, then digital filters were applied, and cleaned data were compared to data collected in the absence of MR sequences, both visually and quantitatively. We anticipated that MR-artifacts would be greatest during multiband, multi-echo EPI sequences and while current filtering recommendations would mitigate MR-artifacts, adaptations may be necessary for multiband and multi-echo pulse sequences. We expected that adaptations to the slice collection frequency would be necessary, to account for slices collected in parallel. Finally, we make these findings openly available both as interactive online code and easy-to-use command-line software, sharing updated recommendations for removing MR-artifacts from these physiological data in multiband and multi-echo contexts.

5.2 Methods

5.2.1 Physiological recordings

Physiological data were collected using MRI-compatible modules, leads, and electrodes from BIOPAC Systems. Data were acquired using a BIOPAC MP150 system, connected to subject leads by two standard MEC-MRI cables that passed through the MRI patch panel via MRI-RFIF filters and ran without loops to the bore, then parallel with the subject. Electrocardiogram (ECG) recordings were collected using radiotranslucent EL508 electrodes with GEL100 and LEAD108 leads, with an ECG100C-MRI amplifier. Electrodes were placed in a 3-lead bipolar monitoring configuration, 6 to 8 inches apart diagonally across the heart from right clavicle to left rib cage, with the ground placed 6 to 8 inches away on the right rib cage. Electrodermal activity (EDA) recordings were collected using radiotranslucent EL509 electrodes with GEL101 and LEAD108 leads, with an EDA100C-MRI amplifier. Leads were placed on the palm of the participant's non-dominant hand, on the thenar and hypothenar eminences. All physiological data were collected at a rate of 2000 Hz, with ECG and EDA collected concurrently from all participants. Physiological data collection began once participants were loaded on the scanner bed and continued until the scanner bed exited the bore after the scanning session, including several minutes per participant of data collected in the absence of an MR pulse sequence.

5.2.2 BOLD EPI Sequences

Physiological recordings were acquired in the bore of a whole-body 3-Tesla Siemens MAGNETOM Prisma with a 32-channel head/neck coil, during both a multiband,

single-echo (MBSE) blood-oxygenation-level-dependent (BOLD) echo planar imaging (EPI) sequence and a multiband, multi-echo (MBME) BOLD EPI sequence.

Multiband, single-echo BOLD EPI sequence.

The MBSE sequence used here is the one used by the Adolescent Brain Cognitive Development (ABCD)℠ Study (Casey et al., 2018). In brief, this sequence acquired 60 transverse slices, with an anterior to posterior phase encoding direction, using a single echo (TE = 30ms) with TR = 800ms, a multiband acceleration factor of 6, interleaved acquisition, in-plane GRAPPA acceleration, and a 52º flip angle. More information about the scan protocols is available with the curated ABCD data via the NIMH Data Archive (NDA; https://abcdstudy.org/scientists/protocols/). Participants in the current study (n=5, all female, aged 26-39) completed four 6-minute runs of an emotion regulation task (Blair et al., 2007; Ochsner et al., 2002) and two 5-minute runs of rest.

Multiband, multi-echo BOLD EPI sequence.

The MBME BOLD EPI sequence used here is from the distribution of multi-band accelerated EPI sequences (Moeller et al., 2010) developed by the Center for Magnetic Resonance Research at the University of Minnesota. The MBME GRE-EPI sequence acquired 48 slices at a 30º transverse-to-coronal orientation with anterior-to-posterior phase encoding direction at each of 4 echoes (TE1 = 11.80ms, TE2 = 28.04ms, TE3 = 44.28ms, TE4 = 60.52ms) with TR = 1500ms, a multiband acceleration factor of 3, interleaved acquisition, in-plane GRAPPA acceleration, a 77º flip angle, and an excite pulse duration of 2560µs (Figure 1B). Participants completed six runs, 6 to 11 minutes each, of film watching (Duffer and Duffer, 2017), two runs of the same emotion regulation task (28), two runs of a probabilistic selection task,

one 6 minutes and the other 9 minutes (Frank et al., 2004), and two runs of 5 minutes of rest. The full parameters and fMRI data are available on OpenNeuro.org (https://openneuro.org/datasets/ds002278/versions/1.0.1).

5.2.3 Software tools

All code used to create and apply these filters was written and run using Python 3.7.3. and is available on GitHub:(https://github.com/62442katieb/mbme-physio-denoising). The bioread library (https://github.com/uwmadison-chm/bioread, v. 1.0.4) was used to read in physiological recordings stored in AcqKnowledge format, data were manipulated using Pandas (pandas.pydata.org, v. 1.0.3), digital filters were created and applied using SciPy (scipy.org, v. 1.4.1; (Virtanen et al., 2020)), and fast Fourier transforms implemented in NumPy (numpy.org, v. 1.18.2 Harris et al., 2020).

5.2.4 Denoising electrophysiological recordings

Fourier transform was applied to ECG and EDA data collected both in the presence and absence of MR pulse sequences to identify the frequencies of MR-related artifacts. Then, in parallel processing streams for each the ECG and EDA data, we applied digital filters to mitigate the effects of these artifacts on the recordings. First, we applied the manufacturer (i.e., BIOPAC) recommendation for single-band, single-echo sequences: comb band-stop filters at the slice collection frequency and its harmonics up to the Nyquist frequency, and then Fourier transformed the results to assess how these filters mitigated artifacts. This slice collection frequency is defined as:

$$\nu_{slice\ collection} = number\ of\ slices \div TR$$

And here, comb band-stop filters were implemented as a series of infinite impulse response (IIR) notch filters to account for the fundamental frequency and its harmonics.

Then, we adjusted the frequencies of these filters to account for additional MR-related artifacts identified in frequency spectra, applied the adjusted filters, and Fourier transformed the results.

Finally, physiological data were compared across steps and to data collected in the absence of MR pulse sequences, using magnitude squared coherence to assess linear dependence across the frequency band in which physiologically-relevant signals were found: $0.5 - 50Hz$ for ECG and $< 0.5Hz$ for EDA.

5.3 Results

5.3.1 Frequencies of MR-related artifacts

We applied Fourier transforms to ECG recordings in the absence (Figure 5.2, first row) and presence (Figure 5.2, second row) of both single- (Figure 5.2, left column) and multi-echo (Figure 5.2, right column) BOLD EPI sequences to identify frequencies of confounding MR-related noise. This revealed MR-related noise in frequencies corresponding to the TR (denoted by blue-green circles in Figure 5.2D, L) and slice acquisition (denoted by pink triangles in Figure 5.2D, L). The presence and relative impacts of this noise is visually apparent in the difference between recordings collected before and during the two BOLD EPI sequences in Figure 5.2 (MBSE-ECG

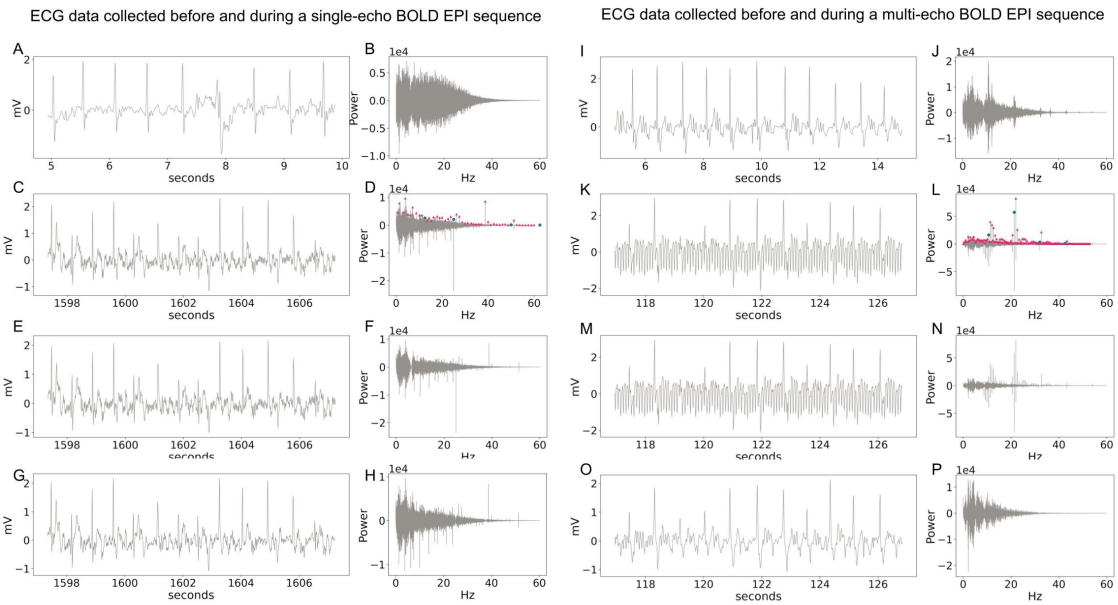

Figure 5.2. Electrocardiogram recordings through the denoising process.

in Figure 5.2A vs. C; MBME-ECG in Figure 5.2I vs. K) and evidenced by greater power in the frequencies corresponding with TR and slice acquisition (MBSE-ECG in Figure 5.2B vs. D; MBME-ECG, 5.2J vs. L). These artifacts occur at frequencies equal to (a) the slice frequency which is equal to the number of slices divided by the multiband factor per TR (indicated by circular, blue-green markers), (b) the TR frequency (indicated by triangular, pink markers), and the harmonics of these frequencies. Furthermore, the power of these confounding signals was much greater in MBME-ECG recordings than in MBSE-ECG recordings (Figure 5.2D vs. L), resulting in greater corruption of the ECG signal (Figure 5.2C vs K).

5.3.2 Denoising electrocardiogram recordings

We applied BIOPAC-recommended filtering to ECG recordings collected during MBSE and MBME BOLD EPI sequences (MBSE-ECG and MBME-ECG, respectively), via IIR notch filters centered at the slice frequency (Figure 5.2E-F, M-N).

This resulted in an incomplete mitigation of MR-related artifacts, which are still clearly present in the frequency spectra (Figure 5.2F, N).

Based on the identified artifact frequencies, we then adjusted these recommendations to account for the multiband factor of each sequence, with IIR notch filters centered at the frequency corresponding to the number of slices divided by the multiband factor per TR (see equation below) and applied the resultant filter (again, including harmonics) to MBSE-ECG recordings (Figure 5.2G-H).

$$\nu_{MB\ slice\ collection} = number\ of\ slices \div TR \div MBfactor$$

Finally, we applied these adjusted recommendations and additional IIR notch filters centered at the TR frequency to mitigate the effects of additional confounding frequencies present in MBME-ECG recordings, demonstrating the additional need for filtering beyond that which MBSE-ECG requires, in order to obtain MR-denoised ECG recordings simultaneous with MBME BOLD EPI data (Figure 5.2O-P).

5.3.3 Denoising electrodermal activity recordings

Prior research on EDA recordings collected during single-band, single-echo BOLD sequences has shown minimal MR-related artifacts in EDA data (Robinson et al., 2014). As such, EDA recordings collected simultaneously with fMRI data do not typically require MR-specific denoising. However, a Fourier transform of EDA recordings acquired during the multiband, single-echo BOLD EPI sequence (hereafter MBSE-EDA) in question revealed noise in sequence-specific frequency bands (Figure 5.3D, L) corresponding to the harmonics of the TR frequency (pink triangles) and, to a lesser extent, of the slice collection frequency (blue-green circles). To re-

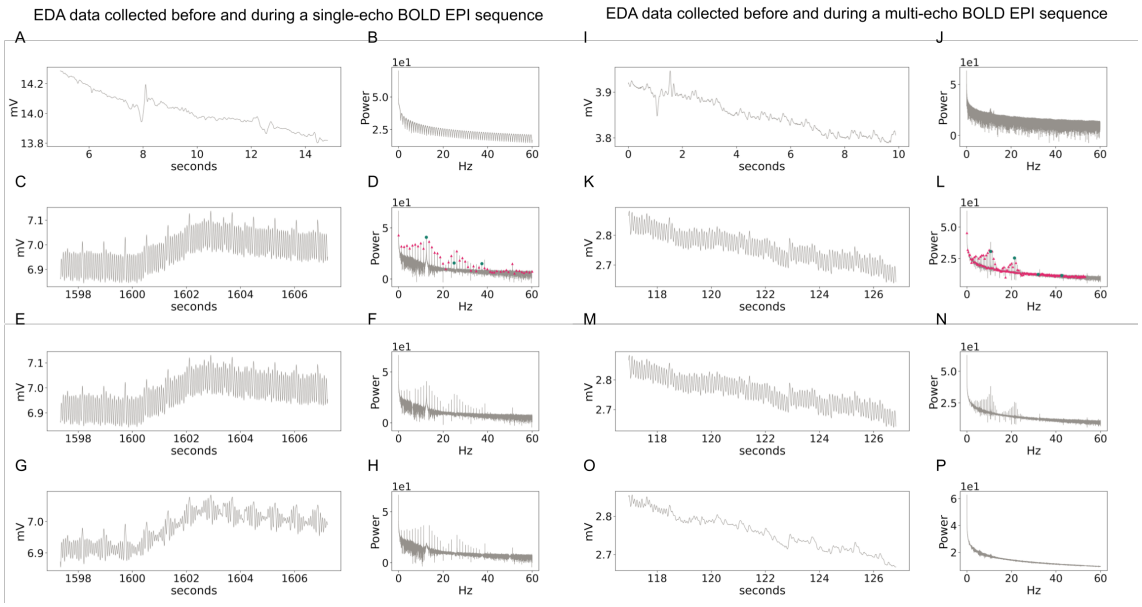

Figure 5.3. Electrodermal activity recordings through the denoising process.

move MR-related noise from these EDA recordings, we followed the same filtering procedures as mentioned above (see Denoising electrocardiogram recordings).

5.3.4 Quantitative assessments of denoised data

Finally, magnitude squared coherence was computed between each pair of MBSE-ECG recordings across filtering approaches and, separately, between each pair of MBME-ECG recordings. This allowed a direct comparison of frequency spectra between (i) filtering approaches, between (ii) filtered recordings and unfiltered recordings, and between (iii) filtered recordings and ECG recordings in the absence of EPI sequences.

Linear dependence of ECG recordings across the denoising approaches described above was assessed using pairwise magnitude squared coherence of the cleaned ECG recordings across filtering approaches, the raw ECG recordings collected during BOLD EPI sequences, and the no-MR recordings (Figure 5.4). These comparisons

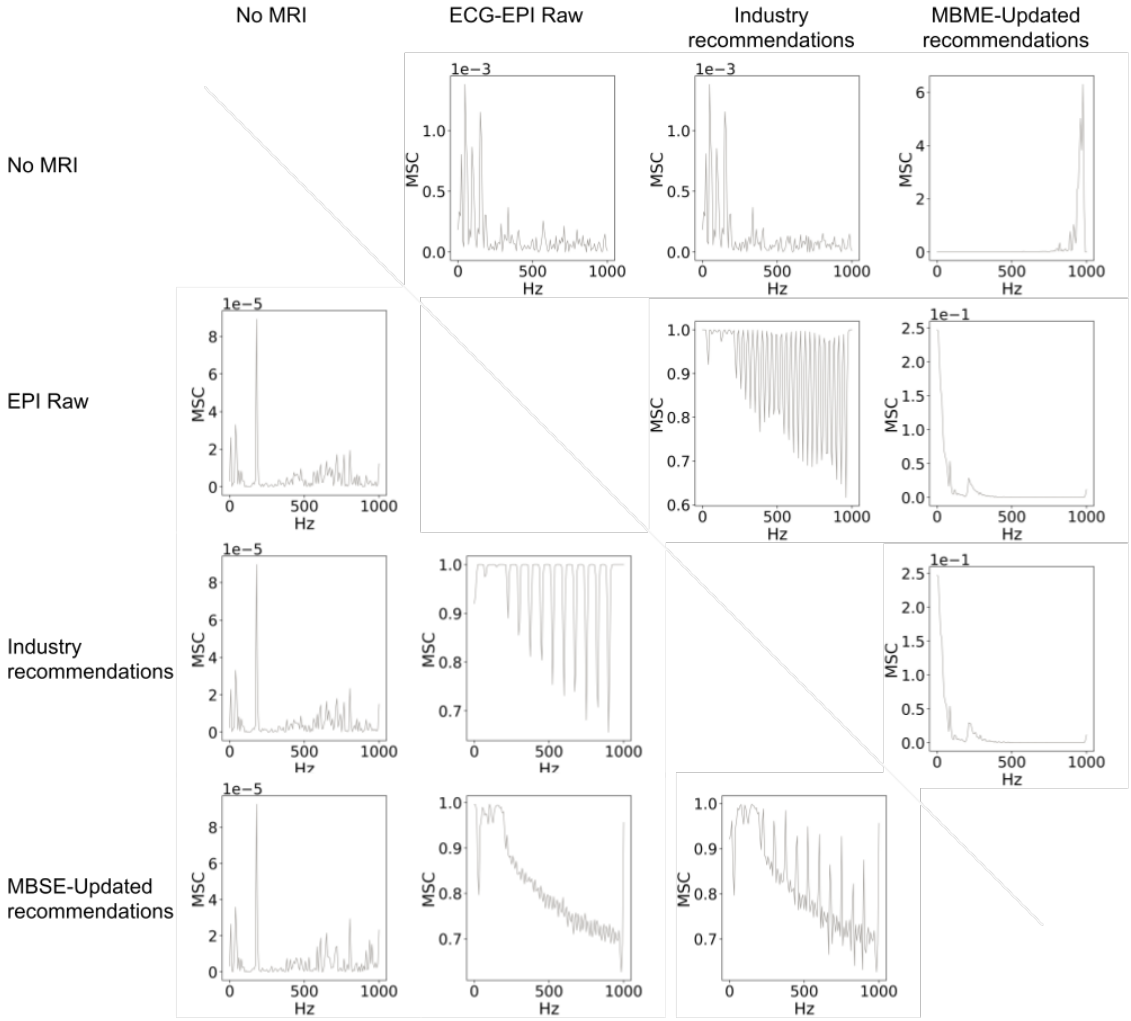

Figure 5.4. Magnitude squared coherence between raw and cleaned ECG recordings.

between recordings across the denoising process clearly demonstrate the differences in filtering between BIOPAC-recommended and updated approaches for data collected during multiband and multi-echo pulse sequences. MBSE-ECG recordings (Figure 5.4, lower triangle) retained the same frequency spectra in common with ECG recordings collected in the absence of an MR sequence, regardless of filters applied (BIOPAC-recommended or its multiband equivalent; Figure 5.4, column 1). Between cleaned MBSE-ECG recordings (Figure 5.4, upper triangle) and ECG recordings collected in the absence of an MRI sequence (Figure 5.4, row 1), manufacturer recommendations did not change the frequency spectra in common, when compared with unfiltered MBME-ECG data, but the thoroughly filtered (i.e., with multiband-updated slice frequency filtered and the additional TR frequency filtering) there are more high frequencies in common. Linear dependence between these recordings is lower than between any other pair of recordings. Between raw and cleaned MBSE- and MBME-ECG recordings (Figure 5.4, second and third columns and rows), there remained more low frequencies ($< 250Hz$) in common. However, this may be due to the relatively higher power of lower frequencies in these recordings in general, regardless of filtering approach (see Figure 5.2).

Within the frequency band in which physiologically-relevant signals are found in ECG (0.5 - 50Hz), we assessed the correlations between power spectra across filtering approaches. Average correlation between MBSE-ECG recordings filtered with MB-updated manufacturer recommendations ($r = 0.633$) was slightly higher than that of MBSE-ECG recordings filtered with BIOPAC-recommended filters ($r = 0.619$), both of which were higher than the average correlation between power spectra of MBSE-ECG and no-MR-ECG recordings ($r = 0.573$) (Supplemental Table 1, Figure 5.5). Average correlation between power spectra of MBME-ECG recordings filtered with our recommended slice- and TR-frequency filters (r = 0.659) was higher than

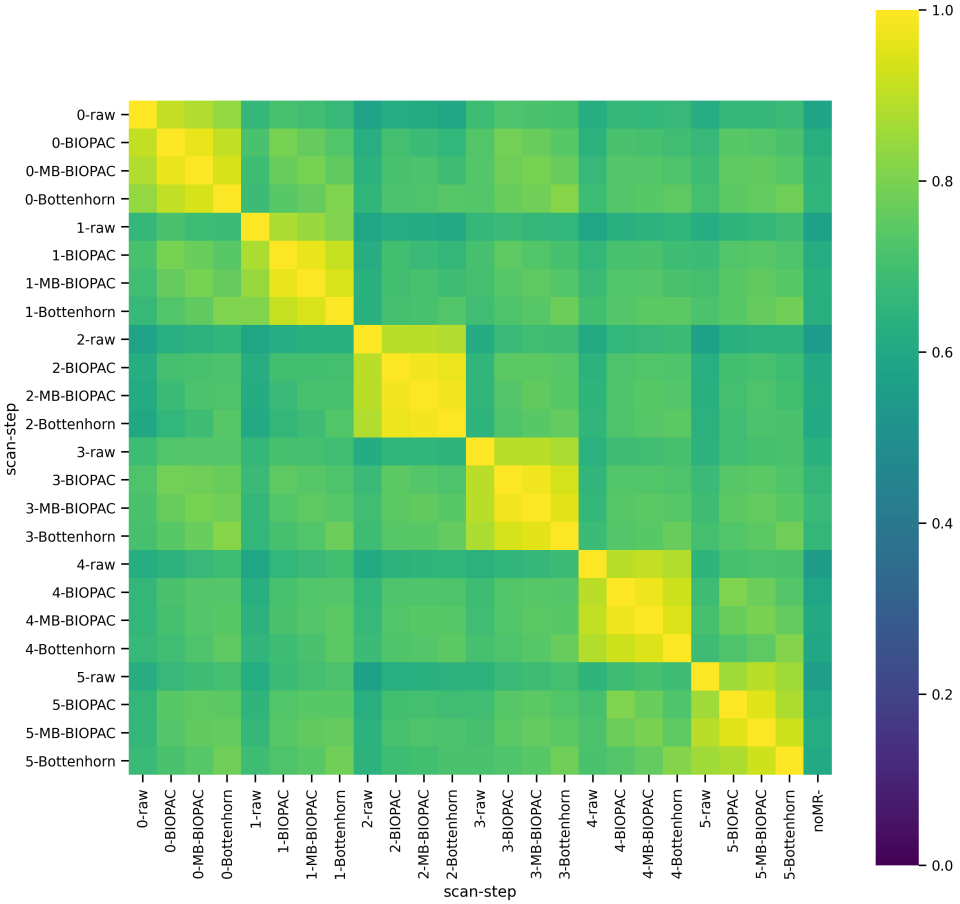

Figure 5.5. Average correlation between multiband, single-echo-concurrent electrocardiogram recordings.

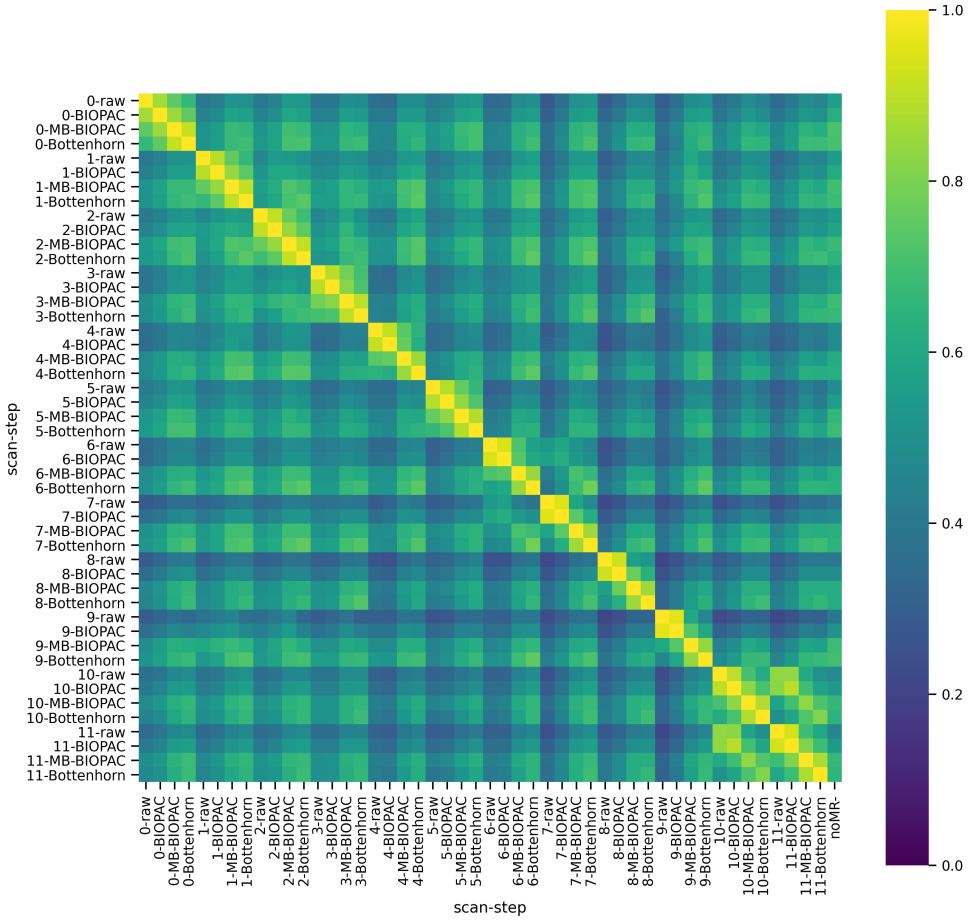

Figure 5.6. Average correlation between multiband, multi-echo-concurrent electrocardiogram recordings.

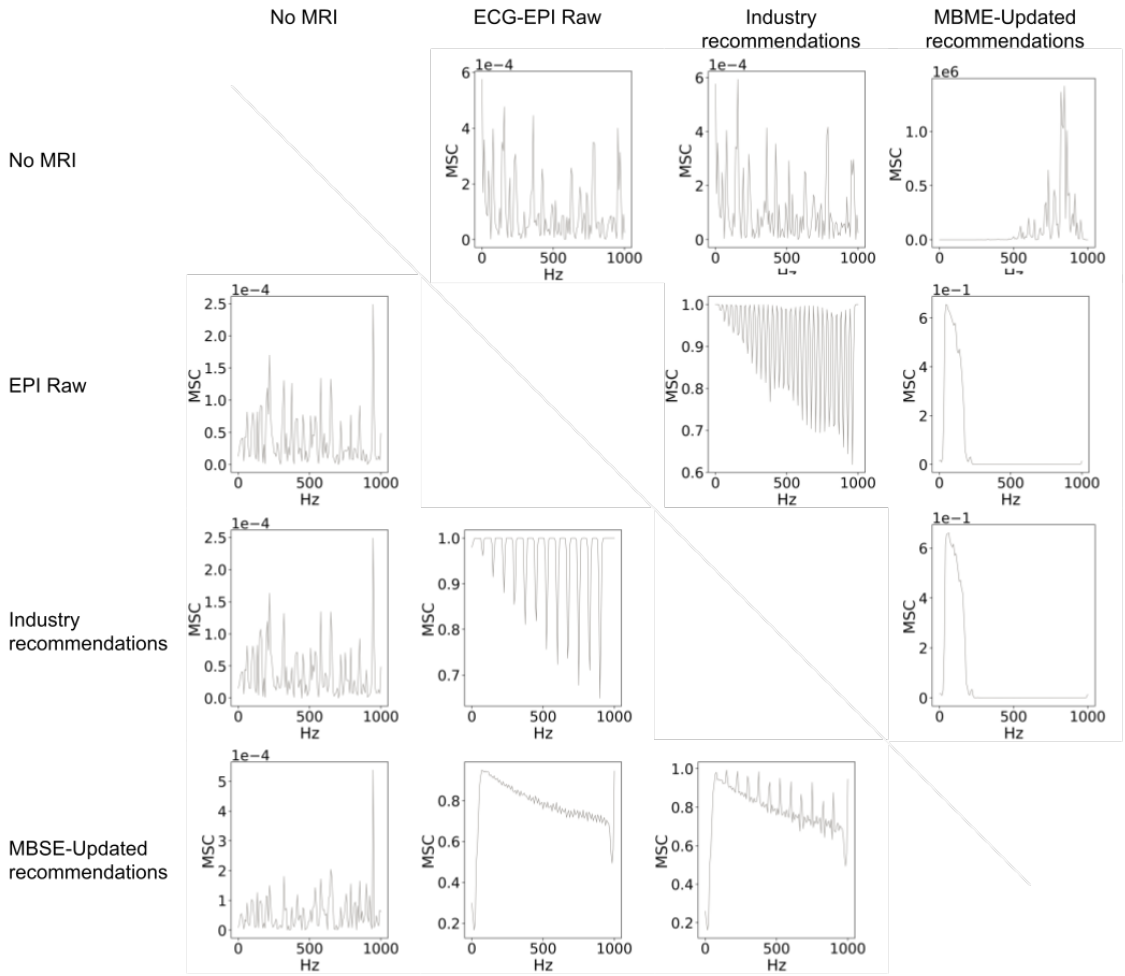

Figure 5.7. Magnitude squared coherence between raw and cleaned EDA recordings.

the average correlation between no-MR-ECG and MBME-ECG recordings denoised with BIOPAC-recommended filters ($r = 0.539$), both of which were higher than the average correlation between unfiltered MBME-ECG recordings and no-MR- ECG recordings ($r = 0.480$) (Supplemental Table 2, Figure 5.6).

This approach was repeated for each pair of MBSE-EDA recordings across filtering approaches and, separately, between each pair of MBME-EDA recordings.

The magnitude squared coherence of EDA recordings collected before and during BOLD EPI sequences, and across filtering approaches, show linear dependence (i.e., similarity) between pairs of recordings across frequencies. These indicate that, as

with MBSE-ECG recordings, MBSE-EDA recordings' similarity to EDA recordings in the absence of an MR sequence do not differ greatly across the cleaning process (Figure 5.7, column 1). Likewise, MBME-EDA recordings showed greater similarity to EDA recordings in the absence of an MR sequence in higher frequency bands following the slice- and TR-frequency filtering (Figure 5.7, row 1, MBME-Updated Recommendation).

Within the frequency band in which physiologically-relevant signals are found in EDA (<0.5Hz), we assessed the correlations between power spectra across filtering approaches. Average correlations of MBSE- and MBME-EDA power spectra and no-MR-EDA power spectra were much higher than their -ECG counterparts ($r_{MBSE} \geq 0.99$, $r_{MBME} \geq 0.999$) with very little variability ($\sigma_r(MBSE) = 0.001$, $\sigma_r(MBME) = 1.84 \times 10^{-5}$) (Supplemental Tables 4 and 5, Figures 5.5 and 5.9). Overall, this supports the claim in prior research that the impacts of MR-related artifacts on simultaneously-collected EDA recordings are minimal, although not nonexistent.

5.3.5 Research Products

The workflows used to clean these data and create the associated figures are available as a command-line Python script and in interactive Jupyter Notebooks available at: https://github.com/62442katieb/mbme-physio-denoising/. These notebooks are additionally available, interactively, at: https://mybinder.org/v2/gh/62442katieb/mbme-physio-denoising/binder-live.

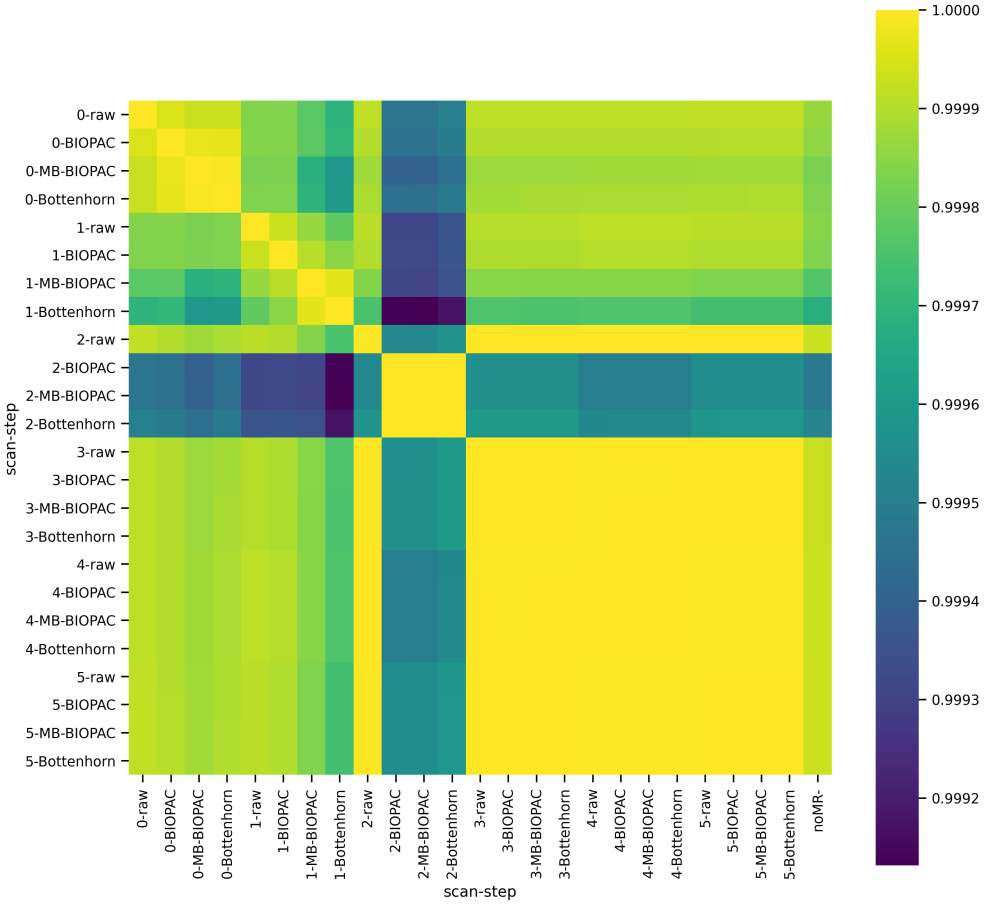

Figure 5.8. Average correlation between multiband, cingle-echo-concurrent electrodermal activity recordings.

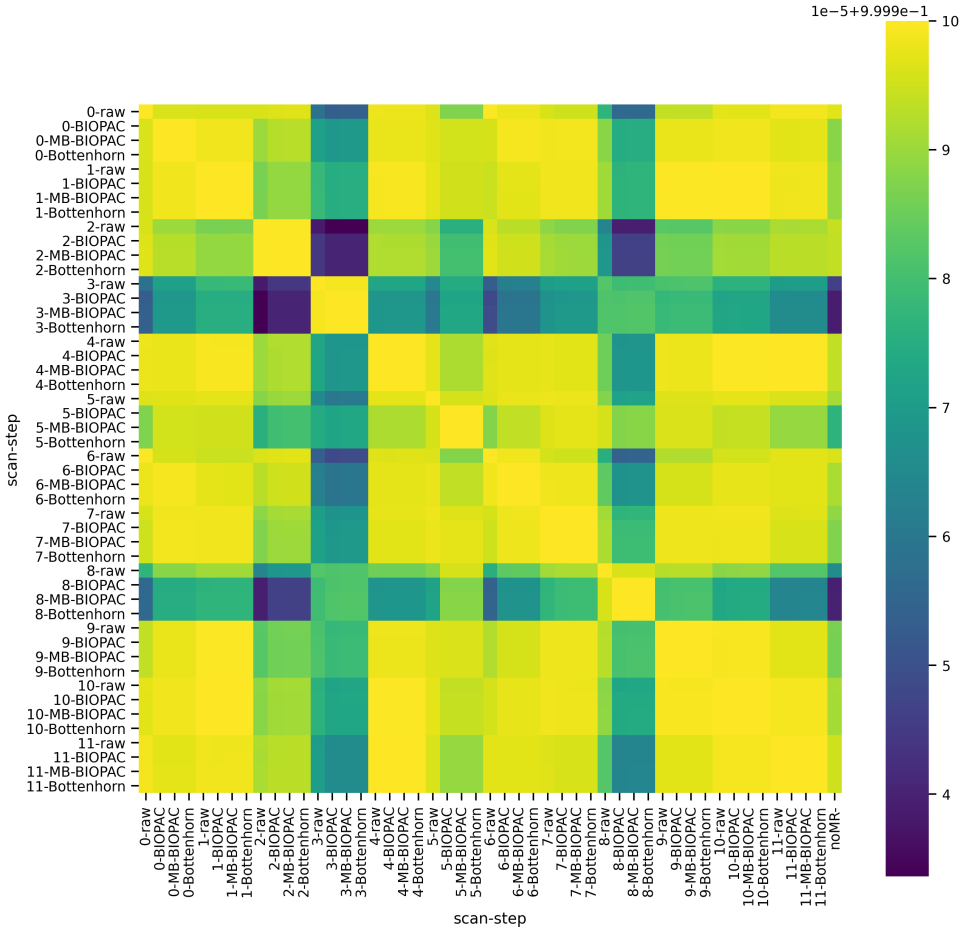

Figure 5.9. Average correlation between multiband, multi-echo-concurrent electrodermal activity recordings.

5.4 Discussion

Here, we assessed the confounding influence of both multiband, single-echo and a multiband, multi-echo BOLD MRI sequences on simultaneously acquired peripheral physiological recordings (e.g., ECG and EDA). These artifacts were demonstrated in recordings collected over the course of several MRI scans, using an MBSE BOLD EPI sequence with a multiband factor of 6 and a MBME BOLD EPI sequence that acquired 4 volumes per RF excitation with a multiband factor of 3. Two fundamental confounding frequencies were identified, corresponding with the slice frequency and the repetition time of the MRI sequence. Applying a series of notch filters centered at frequencies corresponding to the sequence's TR and slice collection frequency, mimicking a comb band-stop filter (per manufacturer (i.e., BIOPAC) recommendations) provided marked decrease of confounding signals. Based on this, we present an updated set of recommendations for mitigation of pulse sequence-related artifacts in ECG and EDA recordings collected during multiband, multi-echo BOLD MRI scans. These recommendations make it easier for researchers to include physiological recordings during functional MRI studies that capitalize on the improved temporal signal-to-noise ratio (tSNR) of multi-echo pulse sequences and the improvements to temporal resolution made possible by simultaneous multi-slice acquisition. While we did not test these recommendations across a range of pulse sequences with different numbers of echoes and multiband factors, it is likely that our recommendations will generalize across MBME BOLD EPI sequences due to the linear relationship between confounding frequency bands and the sequence's TR and multiband factor.

Building on prior research, we found MRI sequence artifacts in simultaneously collected ECG recordings in frequency bands correspond to the number of asynch-

ronously-collected slices (i.e., the number of slices divided by the multiband factor). Furthermore, these confounding frequencies were of greater power in recordings collected during multi-echo BOLD EPI scans than during single-echo. The impact of these corrupting frequencies can be removed with a series of IIR notch filters corresponding to the TR and slice collection frequencies and their harmonics up to the Nyquist frequency. Further processing is needed in order to distinguish moment-to-moment heart rate and data derived therefrom, but the data have been cleaned of the confounding MR-related artifacts.

Contrary to prior research, we found MRI sequence artifacts in simultaneously collected EDA recordings, both during single- and multi-echo BOLD EPI sequences (Figure 5.3D, L). These artifacts corresponded with the TR frequency, likely related to the transmission of RF excitation pulses, and the gradient pulses during slice collection. However, the relative power of these confounding frequencies did not differ between MBSE-EDA and MBME-EDA, in contrast to those in the ECG signals (see Figure2). Here, we demonstrate that these artifacts can be removed in the same manner as from MBSE- and MBME-ECG recordings with a series of IIR notch filters, mimicking a comb band-stop filter that removes a frequency band and its harmonics up to the Nyquist frequency.

Confounding frequencies corresponding with the TR of the sequence were detected via comparison of the power spectra of ECG and EDA recordings before (Figures 5.2B and 5.3B, respectively) and during a MBME BOLD EPI sequence (Figures 5.2D and 5.3D, respectively). This frequency is not often mentioned in the simultaneous physiology-fMRI literature as RF pulse-related artifacts are either of a much lesser amplitude than other MR-artifacts or they are filtered out entirely by MRI-specific amplifiers (see Figure 1 for comparison) (Negishi et al., 2007; Allen et al., 2000). However, the RF excitation that precedes slice collection in multiband

MRI pulse sequences has a higher amplitude and/or greater total power than that of a single-band sequence. This increased power may explain why the artifact is seen here, in frequency bands corresponding with the sequence TR as shown in Figures 5.2F and 5.3, but not usually seen in physiological recordings collected simultaneously with single-band BOLD sequences and is not mentioned in prior simultaneous ECG- or EDA-fMRI research or the associated manufacturer recommendations for signal cleaning.

The confounding fundamental frequency identified here corresponds with the slice collection or slice repetition frequency (Allen et al., 2000). This artifact is more commonly seen in ECG recordings collected during fMRI scans, though not in EDA recordings (Robinson et al., 2014). The literature on simultaneous EEG-fMRI acquisition and data cleaning suggests that the magnitude of artifacts due to electromotive force caused by time-varying magnetic field gradients during slice acquisition far surpasses that of the RF excitation pulse (Allen et al., 2000). While this artifact is seen in physiological recordings acquired during single-band BOLD sequences, as well, the power of the harmonics of this confounding frequency are much greater in data collected during multiband BOLD sequences. Although slice collection in multi-echo GRE-EPI sequences is more prolonged over the course of a timepoint of data acquisition, due to the acquisition of multiple volumes of data per RF excitation pulse, the duration of slice collection is short (¡75ms) compared to the repetition time (1500ms). As such, the confounding frequency associated with time-varying gradients is centered on the slice frequency (slices / MB factor / TR) and confounding frequencies associated with individual echoes were not observed. Notch filters centered at the slice frequency and its harmonics sufficiently removed the artifact caused by shifting gradient fields.

5.4.1 Limitations and Considerations

The temporal resolutions of each electrophysiological ($1000 - 5000Hz$) and fMRI ($0.5 - 1.5Hz$) data complicate psychophysiological analyses. First, in relating physiological processes to BOLD signal fluctuations, accounting for differences in the timing of individual slice collection, typically performed in the beginning of fMRI preprocessing (Jones et al., 2008; Parker et al., 2017; Sladky et al., 2011), becomes crucial. Second, physiological data should be downsampled for such investigations. fMRI data are collected with a TR between 500ms and 3s and while multiband acquisition can shorten TRs, multi-echo acquisition often lengthens TRs. When TRs exceed 2 seconds, the temporal resolution of fMRI data becomes low enough to induce aliasing in biologically-relevant frequency bands of physiological data downsampled to match. Researchers should proceed with caution when this is the case. On another note, researchers collecting data regarding heart rate and cardiac pulsations can avoid MR artifacts entirely by using a photoplethysmograph, which collects optical instead of electrical measurements.

5.4.2 Conclusions

While MBSE and MBME pulse sequences introduce more complicated artifacts into simultaneously acquired electrophysiological recordings than can be addressed with current manufacturer recommendations, the data presented here suggest that these artifacts are predictable and their effects can be greatly mitigated with notch filters centered at their fundamental frequencies and harmonics. By targeting the slice acquisition frequency, updated to account for multiband factor, and, especially in the case of MBME-simultaneous recordings, TR frequency, researchers should be able to remove significant MR-related artifacts from ECG and EDA data collected during

fMRI scans. Recommendations such as those demonstrated here allow researchers to capitalize on the improved SNR afforded by MBSE and MBME BOLD sequences, while including rich information concerning a participant's peripheral, visceral state.

CHAPTER 6

CONCLUSION

Here, I present a body of work aimed at furthering our understanding of individual variability in large-scale brain networks and both the sources and consequences thereof.

IDConn, a Python pipeline for assessing individual differences in brain connectivity, applies techniques from neuroimaging, network science, statistics, and data science with adherence to cutting-edge data standards for transparent, reproducible analysis of both connectivity and topology of functional brain networks. This will allow cognitive neuroscientists and psychologists to perform robust studies of functional brain networks without the hurdle of amassing knowledge across those disciplines. Because IDConn is open-source, its code is available for anyone to view, test, use, and contribute to, which not only furthers reproducible science, it allows researchers a better understanding of what occurs "under the hood", which is not an option with most widely-used functional neuroimaging software. This work demonstrates IDConn's flexibility and applicability in two disparate use cases:

- Chapter 3: a cross-sectional study of associations between physics cognition and functional brain networks following a semester of physics instruction in undergraduate students
- Chapter 4: a dense-sampling, longitudinal study of associations between endocrine and lifestyle variables with functional brain networks across the menstrual cycle

Learning-related changes in the brain and behavior are a part of every day life. One source of these changes is formal education, a nearly ubiquitous experience for young people across the United States. In studying how skill acquisition and

physics cognition are associated with large-scale functional network connectivity in the brain, this work provides a window into individual differences are associated with a common external intervention (formal education). Not only was skill acquisition associated with physics-related cognition, these associations were dependent on the pedagogical style in which students were instructed and on students' sex, which were reflected by individual differences in students' brain network connectivity while they were solving physics problems. Together, these findings not only describe factors associated with student performance in school, they highlight the degree to which students' experience impact their learning on a behavioral and biological level.

Menstruation is, was, or will be an aspect of everyday life for nearly 50% of people and has at least indirectly affected the lives of the other 50%, as well. However, as functional magnetic resonance imaging (fMRI) has grown in popularity as a method for *in vivo* assessment of brain function, this crucial aspect of life has received relatively little attention, as have many other issues associated with women's health. In studying how endocrine and lifestyle factors explain changes in women's brains over the course of the menstrual cycle, we contribute not only to the neural understanding of factors disproportionately impacting women's health but to future study of individual variability in brain network connectivity across studies. Only by understanding the degree to which aspects of everyday life affect brain function can researchers make informed decisions regarding study design and analytic techniques. As the study of individual variability in the brain grows, identification of confounding variables becomes increasingly important. Here, we demonstrate that endocrine factors account for more variance in brain network connectivity than do stress, sleep, and head motion, and that they account for comparatively more variance in default mode and limbic connectivity than in connectivity elsewhere in the brain, while sleep and stress account for more diffuse connectivity. As large-scale

brain networks are differently implicated in psychopathology and associated with different aspects of cognition, researchers should consider not only the magnitude of confounding variables, but the topography of their associations with the brain, as well.

Finally, BOLD fMRI measures are effected by differences in individual physiology, as well (Handwerker et al., 2012). As MR technology improves, the prospect of measuring sources of potential physiological confounds becomes more daunting. While altering BOLD fMRI sequences to collect more data within a single scan can provide significant improvements in estimating the BOLD signal and, thus, brain function, these alterations impact significant artifacts on simultaneously collected electrophysiological data. The final chapter of this work details this problem and presents a solution that applies digital filters to significantly diminish the impacts of these artifacts on concurrently collected electrocardiogram and electrodermal activity recordings. Not only is this approach effective, it presents the code used to implement the approach and provides a command-line program that applies the filtering approach, allowing users to easily apply these solutions to their own data. Thus, this work promotes technological and methodological improvements to the robust study of individual variability in the brain and moves us closer to the goal of "precision neuroscience."

Overall, this work makes several strides toward an individualized understanding of brain function and how it changes over the course of everyday life.

BIBLIOGRAPHY

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., and Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8:14.

Abreu, R., Nunes, S., Leal, A., and Figueiredo, P. (2017). Physiological noise correction using ECG-derived respiratory signals for enhanced mapping of spontaneous neuronal activity with simultaneous EEG-fMRI. *NeuroImage*, 154:115–127.

Albert, K., Hiscox, J., Boyd, B., Dumas, J., Taylor, W., and Newhouse, P. (2017). Estrogen enhances hippocampal gray-matter volume in young and older postmenopausal women: a prospective dose-response study. *Neurobiology of Aging*, 56:1–6.

Alexander, L. M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N., Alexander, A., Kovacs, M., Litke, S., O'Hagan, B., Andersen, J., Bronstein, B., Bui, A., Bushey, M., Butler, H., Castagna, V., Camacho, N., Chan, E., Citera, D., Clucas, J., Cohen, S., Dufek, S., Eaves, M., Fradera, B., Gardner, J., Grant-Villegas, N., Green, G., Gregory, C., Hart, E., Harris, S., Horton, M., Kahn, D., Kabotyanski, K., Karmel, B., Kelly, S. P., Kleinman, K., Koo, B., Kramer, E., Lennon, E., Lord, C., Mantello, G., Margolis, A., Merikangas, K. R., Milham, J., Minniti, G., Neuhaus, R., Levine, A., Osman, Y., Parra, L. C., Pugh, K. R., Racanello, A., Restrepo, A., Saltzman, T., Septimus, B., Tobe, R., Waltz, R., Williams, A., Yeo, A., Castellanos, F. X., Klein, A., Paus, T., Leventhal, B. L., Craddock, R. C., Koplewicz, H. S., and Milham, M. P. (2017). An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*, 4(1):170181. Number: 1 Publisher: Nature Publishing Group.

Allen, A. M., McRae-Clark, A. L., Carlson, S., Saladin, M. E., Gray, K. M., Wetherington, C. L., McKee, S. A., and Allen, S. S. (2016). Determining Menstrual Phase in Human Biobehavioral Research: A Review with Recommendations. *Experimental and clinical psychopharmacology*, 24(1):1–11.

Allen, N. E., Sudlow, C., Peakman, T., Collins, R., and Biobank, o. b. o. U. (2014). UK Biobank Data: Come and Get It. *Science Translational Medicine*, 6(224):224ed4–224ed4. Publisher: American Association for the Advancement of Science Section: Editorial.

Allen, P. J., Josephs, O., and Turner, R. (2000). A method for removing imaging artifact from continuous EEG recorded during functional MRI. *NeuroImage*, 12(2):230–239.

Alloway, T. P. and Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, 106(1):20–29.

Anderson, M. J. and Robinson, J. (2001). Permutation Tests for Linear Models. *Australian & New Zealand Journal of Statistics*, 43(1):75–88.

Andersson, J. L. R., Jenkinson, M., and Smith, S. (2007). Non-linear registration, aka spatial normalization. Technical Report TR07JA2, FMRIB Centre, Oxford, UK.

Andreassen, R. and Bråten, I. (2010). Examining the prediction of reading comprehension on different multiple-choice tests. *Journal of Research in Reading*, 33(3):263–283. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9817.2009.01413.x.

Arélin, K., Mueller, K., Barth, C., Rekkas, P. V., Kratzsch, J., Burmann, I., Villringer, A., and Sacher, J. (2015). Progesterone mediates brain functional connectivity changes during the menstrual cycle—a pilot resting state MRI study. *Frontiers in Neuroscience*, 9.

Ashburner, J. (2012). SPM: a history. *NeuroImage*, 62(2):791–800.

Auerbach, E. J., Xu, J., Yacoub, E., Moeller, S., and Uğurbil, K. (2013). Multiband accelerated spin-echo echo planar imaging with reduced peak RF power using time-shifted RF pulses. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 69(5):1261–1267.

Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41.

Baird, C. L. (2018). Male-dominated stem disciplines: How do we make them more attractive to women? *IEEE Instrumentation Measurement Magazine*, 21(3):4–14.

Baker, F. C. and Driver, H. S. (2007). Circadian rhythms, sleep, and the menstrual cycle. *Sleep Medicine*, 8(6):613–622.

Baker, F. C. and Lee, K. A. (2018). Menstrual cycle effects on sleep. *Sleep medicine clinics*, 13(3):283–294. ISBN: 1556-407X Publisher: Elsevier.

Bangasser, D. A. and Valentino, R. J. (2014). Sex differences in stress-related psychiatric disorders: Neurobiological perspectives. *Frontiers in Neuroendocrinology*, 35(3):303–319.

Barfield, E. T. and Gourley, S. L. (2018). Prefrontal cortical trkB, glucocorticoids, and their interactions in stress and developmental contexts. *Neuroscience and biobehavioral reviews*, 95:535–558.

Barone, C. (2011). Some Things Never Change: Gender Segregation in Higher Education across Eight Nations and Three Decades. *Sociology of Education*, 84(2):157–176.

Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., and Williams, S. C. R. (1999). Social intelligence in the normal and autistic brain: an fMRI study. *European Journal of Neuroscience*, 11(6):1891–1898. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1460-9568.1999.00621.x.

Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M., and Boomsma, D. I. (2002). Genetic and environmental influences on the development of intelligence. *Behavior Genetics*, 32(4):237–249.

Barth, C., Steele, C. J., Mueller, K., Rekkas, V. P., Arélin, K., Pampel, A., Burmann, I., Kratzsch, J., Villringer, A., and Sacher, J. (2016). In-vivo Dynamics of the Human Hippocampus across the Menstrual Cycle. *Scientific Reports*, 6(1):32833. Number: 1 Publisher: Nature Publishing Group.

Bartley, J. E., Riedel, M. C., Salo, T., Boeving, E. R., Bottenhorn, K. L., Bravo, E. I., Odean, R., Nazareth, A., Laird, R. W., Sutherland, M. T., Pruden, S. M., Brewe, E., and Laird, A. R. (2019). Brain activity links performance in science reasoning with conceptual approach. *npj Science of Learning*, 4(1):1–8.

Bassett, D. S. and Bullmore, E. (2006). Small-World Brain Networks. *The Neuroscientist*, 12(6):512–523. Publisher: SAGE Publications Inc STM.

Bassett, D. S., Meyer-Lindenberg, A., Achard, S., Duke, T., and Bullmore, E. (2006). Adaptive reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National Academy of Sciences*, 103(51):19518.

Bassett, D. S., Nelson, B. G., Mueller, B. A., Camchong, J., and Lim, K. O. (2012). Altered resting state complexity in schizophrenia. *NeuroImage*, 59(3):2196–2207.

Baulieu, E. E., Schumacher, M., Koenig, H., Jung-Testas, I., and Akwa, Y. (1996). Progesterone as a neurosteroid: Actions within the nervous system. *Cellular and Molecular Neurobiology*, 16(2):143–154.

Becker, S. P., Jarrett, M. A., Luebbe, A. M., Garner, A. A., Burns, G. L., and Kofler, M. J. (2018). Sleep in a large, multi-university sample of college students: sleep problem prevalence, sex differences, and mental health correlates. *Sleep Health*, 4(2):174–181.

Begovac, I., Begovac, B., Majić, G., and Vidović, V. (2009). LONGITUDINAL STUDIES OF IQ STABILITY IN CHILDREN WITH CHILDHOOD AUTISM – LITERATURE SURVEY. *Psychiatria Danubina*, 21(3):310–319. Publisher: Medicinska naklada.

Behzadi, Y., Restom, K., Liau, J., and Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1):90–101.

Beilock, S. L. and DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(6):983–998. Place: US Publisher: American Psychological Association.

Beltz, A. M. and Moser, J. S. (2020). Ovarian hormones: a long overlooked but critical contributor to cognitive brain structures and function. *Annals of the New York Academy of Sciences*, 1464(1):156–180.

Bergold, S. and Steinmayr, R. (2018). Personality and Intelligence Interact in the Prediction of Academic Achievement. *Journal of Intelligence*, 6(2):27. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

Betzel, R. F., Satterthwaite, T. D., Gold, J. I., and Bassett, D. S. (2017). Positive affect, surprise, and fatigue are correlates of network flexibility. *Scientific Reports*, 7(1):520.

Bicer, A., Nite, S. B., Capraro, R. M., Barroso, L. R., Capraro, M. M., and Lee, Y. (2017). Moving from STEM to STEAM: The effects of informal STEM learning on students' creativity and problem solving skills with 3D printing. In *2017 IEEE Frontiers in Education Conference (FIE)*, pages 1–6.

Blair, K. S., Smith, B. W., Mitchell, D. G. V., Morton, J., Vythilingam, M., Pessoa, L., Fridberg, D., Zametkin, A., Nelson, E. E., Drevets, W. C., Pine, D. S., Martin, A., and Blair, R. J. R. (2007). Modulation of emotion by cognition and cognition by emotion. *NeuroImage*, 35(1):430–440.

Bookheimer, S. Y., Wang, A. T., Scott, A., Sigman, M., and Dapretto, M. (2008). Frontal contributions to face processing differences in autism: Evidence from fMRI of inverted face processing. *Journal of the International Neuropsychological Society*, 14(6):922–932. Publisher: Cambridge University Press.

Bottenhorn, K. L., Bartley, J. E., Riedel, M. C., Salo, T., Bravo, E. I., Odean, R., Nazareth, A., Laird, R. W., Musser, E. D., Pruden, S. M., Brewe, E., Sutherland, M. T., and Laird, A. R. (2021). Intelligence and academic performance: Is it all in your head? *bioRxiv*, page 2021.01.23.427928. Publisher: Cold Spring Harbor Laboratory Section: New Results.

Boyacioğlu, R., Schulz, J., Koopmans, P. J., Barth, M., and Norris, D. G. (2015). Improved sensitivity and specificity for resting state and task fMRI with multiband multi-echo EPI compared to multi-echo EPI at 7 T. *NeuroImage*, 119:352–361.

Brinton, R. D., Thompson, R. F., Foy, M. R., Baudry, M., Wang, J., Finch, C. E., Morgan, T. E., Pike, C. J., Mack, W. J., Stanczyk, F. Z., and Nilsen, J. (2008). Progesterone receptors: Form and function in brain. *Frontiers in Neuroendocrinology*, 29(2):313–339.

Brown, E. S., Jeon-Slaughter, H., Lu, H., Jamadar, R., Issac, S., Shad, M., Denniston, D., Tamminga, C., Nakamura, A., and Thomas, B. P. (2015). Hippocampal Volume in Healthy Controls Given 3-Day Stress Doses of Hydrocortisone. *Neuropsychopharmacology*, 40(5):1216–1221.

Buckley, T. M. and Schatzberg, A. F. (2005). On the interactions of the hypothalamic-pituitary-adrenal (HPA) axis and sleep: normal HPA axis activity and circadian rhythm, exemplary sleep disorders. *The Journal of Clinical Endocrinology and Metabolism*, 90(5):3106–3114.

Buffler, A., Lubben, F., Ibrahim, B., and Pillay, S. (2008). A model-based framework for understanding the role of visualization in physics education. In *Proceedings of the 16th Annual Meeting of the Southern African Association for Research in Mathematics, Science and Technology Education, Maseru, Lesotho*, pages 435–441.

Bullmore, E. and Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336–349. Number: 5 Publisher: Nature Publishing Group.

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., and Gabrieli, J. D. E. (2002). Immature Frontal Lobe Contributions to Cognitive Control in Children: Evidence from fMRI. *Neuron*, 33(2):301–311.

Busato, V. V., Prins, F. J., Elshout, J. J., and Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences*, 29(6):1057–1068.

Bush, G., Frazier, J. A., Rauch, S. L., Seidman, L. J., Whalen, P. J., Jenike, M. A., Rosen, B. R., and Biederman, J. (1999). Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the counting stroop. *Biological Psychiatry*, 45(12):1542–1552.

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry research*, 28(2):193–213.

Canivez, G. L. and Watkins, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children—Third Edition. *Psychological Assessment*, 10(3):285–291.

Cao, H., Plichta, M. M., Schäfer, A., Haddad, L., Grimm, O., Schneider, M., Esslinger, C., Kirsch, P., Meyer-Lindenberg, A., and Tost, H. (2014). Test–retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. *NeuroImage*, 84:888–900.

Casey, B., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. M., Chaarani, B., Mejia, M. H., Hagler, D. J., Daniela Cornejo, M., Sicat, C. S., Harms, M. P., Dosenbach, N. U., Rosenberg, M., Earl, E., Bartsch, H., Watts, R., Polimeni, J. R., Kuperman, J. M., Fair, D. A., and Dale, A. M. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, 32:43–54.

Cassels, J. R. T. and Johnstone, A. H. (1984). The effect of language on student performance on multiple choice tests in chemistry. *Journal of Chemical Education*, 61(7):613. Publisher: American Chemical Society.

Chamberland, M., Girard, G., Bernier, M., Fortin, D., Descoteaux, M., and Whittingstall, K. (2017). On the Origin of Individual Functional Connectivity Variability: The Role of White Matter Architecture. *Brain Connectivity*, 7(8):491–503.

Charles, M. and Bradley, K. (2009). Indulging Our Gendered Selves? Sex Segregation by Field of Study in 44 Countries. *American Journal of Sociology*, 114(4):924–976.

Chen, J. K., Johnston, K. M., Frey, S., Petrides, M., Worsley, K., and Ptito, A. (2004). Functional abnormalities in symptomatic concussed athletes: an fMRI study. *NeuroImage*, 22(1):68–82.

Chen, W.-C. and Whitehead, R. (2009). Understanding physics in relation to working memory. *Research in Science & Technological Education*, 27(2):151–160. Publisher: Routledge _eprint: https://doi.org/10.1080/02635140902853624.

Chong, M., Bhushan, C., Joshi, A. A., Choi, S., Haldar, J. P., Shattuck, D. W., Spreng, R. N., and Leahy, R. M. (2017). Individual parcellation of resting fMRI with a group functional connectivity prior. *NeuroImage*, 156:87–100.

Chrousos, G. P., Torpy, D. J., and Gold, P. W. (1998). Interactions between the hypothalamic-pituitary-adrenal axis and the female reproductive system: clinical implications. *Annals of Internal Medicine*, 129(3):229–240.

Cohen, A. D., Nencka, A. S., Lebel, R. M., and Wang, Y. (2017). Multiband multi-echo imaging of simultaneous oxygenation and flow timeseries for resting state connectivity. *PLOS ONE*, 12(3):e0169253. Publisher: Public Library of Science.

Cohen, J. (1973). Eta-Squared and Partial Eta-Squared in Fixed Factor Anova Designs. *Educational and Psychological Measurement*, 33(1):107–112. Publisher: SAGE Publications Inc.

Cohen, S., Kamarck, T., and Mermelstein, R. (1994). Perceived stress scale. *Measuring stress: A guide for health and social scientists*, 10.

Collins, A., Eneroth, P., and Landgren, B.-M. (1985). Psychoneuroendocrine stress responses and mood as related to the menstrual cycle. *Psychosomatic Medicine*, 47(6):512–527. Place: US Publisher: Lippincott Williams & Wilkins.

Cox, R. W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*, 29(3):162–173.

Cox, R. W. and Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, 10(4-5):171–178.

Craddock, R. C., James, G. A., Holtzheimer, P. E., Hu, X. P., and Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, 33(8):1914–1928.

Cullum, C. M. and Larrabee, G. J. (2010). CHAPTER 6 - WAIS-IV Use in Neuropsychological Assessment. In Weiss, L. G., Saklofske, D. H., Coalson, D. L., and Raiford, S. E., editors, *WAIS-IV Clinical Use and Interpretation*, Practical Resources for the Mental Health Professional, pages 167–187. Academic Press, San Diego.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, 9(2):179–194.

Daneman, M. and Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology. General*, 130(2):208–223.

Day, J., Stang, J. B., Holmes, N., Kumar, D., and Bonn, D. (2016). Gender gaps and gendered action in a first-year physics laboratory. *Physical Review Physics Education Research*, 12(2):020104.

Day, L., Hanson, K., Maltby, J., Proctor, C., and Wood, A. (2010). Hope uniquely predicts objective academic achievement above intelligence, personality, and previous academic achievement. *Journal of Research in Personality*, 44(4):550–553.

de Kloet, E. R., de Kloet, S. F., de Kloet, C. S., and de Kloet, A. D. (2019). Top-down and bottom-up control of stress-coping. *Journal of Neuroendocrinology*, 31(3).

Deary, I. J., Strand, S., Smith, P., and Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35(1):13–21.

Desmond, J. E. and Glover, G. H. (2002). Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. *Journal of Neuroscience Methods*, 118(2):115–128.

D'Esposito, M. (2019). Are individual differences in human brain organization measured with functional MRI meaningful? *Proceedings of the National Academy of Sciences of the United States of America*, 116(45):22432–22434.

Ding, L. (2014). Verification of causal influences of reasoning skills and epistemology on physics conceptual learning. *Physical Review Special Topics - Physics Education Research*, 10(2):023101.

Dobson, H., Ghuman, S., Prabhakar, S., and Smith, R. (2003). A conceptual model of the influence of stress on female reproduction. *Reproduction (Cambridge, England)*, 125(2):151–163.

Docktor, J. and Heller, K. (2008). Gender Differences in Both Force Concept Inventory and Introductory Physics Performance. *AIP Conference Proceedings*, 1064(1):15–18.

DoE, U. (2019). Status and Trends on the Education of Racial and Ethnic Groups, Indicator 26: STEM Degrees. Technical Report NCES 2019-038, U.S. Department of Education, National Center for Education Statistics.

Dubois, J. and Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends in cognitive sciences*, 20(6):425.

Dubois, J., Galdi, P., Paul, L. K., and Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 373(1756).

Dubol, M., Epperson, C. N., Sacher, J., Pletzer, B., Derntl, B., Lanzenberger, R., Sundström-Poromaa, I., and Comasco, E. (2021). Neuroimaging the menstrual cycle: A multimodal systematic review. *Frontiers in Neuroendocrinology*, 60:100878.

Duchesne, A. and Pruessner, J. C. (2013). Association between subjective and cortisol stress response depends on the menstrual cycle phase. *Psychoneuroendocrinology*, 38(12):3155–3159.

Duffer, M. and Duffer, R. (2017). Stranger things.

Dunst, B., Benedek, M., Koschutnig, K., Jauk, E., and Neubauer, A. C. (2014). Sex differences in the IQ-white matter microstructure relationship: A DTI study. *Brain and Cognition*, 91:71–78.

Dzirasa, K. (2017). A brilliant approach to study the basis of intelligence? *Science Translational Medicine*, 9(401):eaao0978.

Eaton, P., Vavruska, K., and Willoughby, S. (2019). Exploring the preinstruction and postinstruction non-Newtonian world views as measured by the Force Concept Inventory. *Physical Review Physics Education Research*, 15(1):010123.

Eddy, S. L. and Brownell, S. E. (2016). Beneath the numbers: A review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines. *Physical Review Physics Education Research*, 12(2):020106.

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, James D. andGoncalves, M., DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., and Gorgolewski, K. J. (2018a). fMRIPrep. *Software*. Publisher: Zenodo.

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R., and Gorgolewski, K. J. (2018b). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*.

Esteban, O., Markiewicz, C. J., Burns, C., Goncalves, M., Jarecka, D., Ziegler, E., Berleant, S., Ellis, D. G., Pinsard, B., Madison, C., Waskom, M., Notter, M. P., Clark, D., Manhães-Savio, A., Clark, D., Jordan, K., Dayan, M., Halchenko, Y. O., Loney, F., Salo, T., Dewey, B. E., Johnson, H., Bougacha, S., Keshavan, A., Yvernault, B., Hamalainen, C., Christian, H., Ćirić, R., Dubois, M., Joseph, M., Cipollini, B., Tilley II, S., Visconti di Oleggio Castello, M., Wong, J., De La Vega, A., Kaczmarzyk, J., Huntenburg, J. M., Clark, M. G., Benderoff, E., Erickson, D., Kent, J. D., Hanke, M., Giavasis, S., Moloney, B., Nichols, B. N., Tungaraza, R., Frohlich, C., Wassermann, D., de Hollander, G., Eshaghi, A., Millman, J., Mancini, M., Nielson, D. M., Varoquaux, G., Watanabe, A., Mordom, D., Guillon, J., Koudoro, S., Chetverikov, A., Rokem, A., Acland, B., Forbes, J., Markello, R., Gillman, A., Kong, X.-Z., Geisler, D., Salvatore, J., Gramfort, A., Doll, A., Buchanan, C., DuPre, E., Liu, S., Schaefer, A., Kleesiek, J., Sikka, S., Schwartz, Y., Lee, J. A., Mattfeld, A., Richie-Halford, A., Liem, F., Perez-Guevara, M. F., Heinsfeld, A. S., Haselgrove, C., Durnez, J., Lampe, L., Poldrack, R., Glatard, T., Tabas, A., Cumba, C., Pérez-García, F., Blair, R., Iqbal, S., Welch, D., Triplett, W., Ghayoor, A., Craddock, R. C., Correa, C., Papadopoulos Orfanos, D., Stadler, J., Warner, J., Sisk, L. M., Falkiewicz, M., Sharp, P., Rothmei, S., Kim, S., Weinstein, A., Kahn, A. E., Kastman, E., Bottenhorn, K., Grignard, M., Perkins, L. N., Contier, O., Zhou, D., Bielievtsov, D., Cooper, G., Stojic, H., Linkersdörfer, J., Waller, L., Renfro, M., Hinds, O., Stanley, O., Küttner, R., Pauli, W. M., Glen, D., Kimbler, A., Meyers, B., Tarbert, C., Ginsburg, D., Haehn, D., Margulies, D. S., Ma, F., Malone, I. B., Snoek, L., Brett, M., Cieslak, M., Hallquist, M., Molina-Romero, M., Bilgel, M., Lee, N., Inati, S., Gerhard, S., Mathotaarachchi, S., Saase, V., Van, A., Steele, C. J., Ort, E., Condamine, E., Lerma-Usabiaga, G., Schwabacher, I., Arias, J., Lai, J., Pellman, J., Huguet, J., WEN, J., Leinweber, K., Chawla, K., Weninger, L., Modat, M., Harms, R.,

Andberg, S. K., Baratz, Z., Matsubara, K., González Orozco, A. A., Marina, A., Davison, A., Floren, A., Park, A., Cheung, B., McDermottroe, C., McNamee, D., Shachnev, D., Flandin, G., Gonzalez, I., Varada, J., Schlamp, K., Podranski, K., Huang, L., Noel, M., Crusoe, M. R., Pannetier, N., Khanuja, R., Urchs, S., Nickson, T., Huang, L., Broderick, W., Tambini, A., Mihai, P. G., Gorgolewski, K. J., and Ghosh, S. (2020). nipy/nipype: 1.5.1.

Estevis, E., Basso, M. R., and Combs, D. (2012). Effects of Practice on the Wechsler Adult Intelligence Scale-IV Across 3- and 6-Month Intervals. *The Clinical Neuropsychologist*, 26(2):239–254.

Etkina, E., Murthy, S., and Zou, X. (2006). Using introductory labs to engage students in experimental design. *American Journal of Physics*, 74(11):979–986.

Faravelli, C., Lo Sauro, C., Lelli, L., Pietrini, F., Lazzeretti, L., Godini, L., Benni, L., Fioravanti, G., Alina Talamba, G., and Castellini, G. (2012). The role of life events and HPA axis in anxiety disorders: a review. *Current pharmaceutical design*, 18(35):5663. ISBN: 1381-6128.

Fehring, R. J., Schneider, M., and Raviele, K. (2006). Variability in the Phases of the Menstrual Cycle. *Journal of Obstetric, Gynecologic & Neonatal Nursing*, 35(3):376–384.

Filevich, E., Lisofsky, N., Becker, M., Butler, O., Lochstet, M., Martensson, J., Wenger, E., Lindenberger, U., and Kühn, S. (2017). Day2day: investigating daily variability of magnetic resonance imaging measures over half a year. *BMC Neuroscience*, 18(1):65.

Finkel, D. and Reynolds, C. A., editors (2014). *Behavior Genetics of Cognition Across the Lifespan*. Advances in Behavior Genetics. Springer-Verlag, New York.

Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., and Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *NeuroImage*, 160:140–151.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., and Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11):1664–1671.

Fischer, F. U., Wolf, D., Scheurich, A., and Fellgiebel, A. (2014). Association of Structural Global Brain Network Properties with Intelligence in Normal Aging. *PLoS ONE*, 9(1):e86258.

Fitzgerald, M., Pritschet, L., Santander, T., Grafton, S. T., and Jacobs, E. G. (2020). Cerebellar network organization across the human menstrual cycle. *Scientific Reports*, 10(1):20732. Number: 1 Publisher: Nature Publishing Group.

Fonov, V., Evans, A., McKinstry, R., Almli, C., and Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47:S102.

Frank, M. J., Seeberger, L. C., and O'Reilly, R. C. (2004). By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism. *Science*, 306(5703):1940–1943.

Freedman, D. and Lane, D. (1983). A Nonstochastic Interpretation of Reported Significance Levels. *Journal of Business & Economic Statistics*, 1(4):292.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., and Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23):8410.

Frisby, C. L. and Henry, B. (2016). Science, politics, and best practice: 35 years after Larry P. *Contemporary School Psychology*, 20(1):46–62. Place: Germany Publisher: Springer.

Friston, K. J., Frith, C. D., Dolan, R. J., Price, C. J., Zeki, S., Ashburner, J. T., and Penny, W. D. (2004). *Human Brain Function*. Elsevier. Google-Books-ID: AoWD2S8759kC.

Friston, K. J., Holmes, A. P., Price, C. J., Büchel, C., and Worsley, K. J. (1999). Multisubject fMRI Studies and Conjunction Analyses. *NeuroImage*, 10(4):385–396.

Gabel, D. L. and Sherwood, R. D. (1983). Facilitating problem solving in high school chemistry. *Journal of Research in Science Teaching*, 20(2):163–177.

Galea, L. A. M., Frick, K. M., Hampson, E., Sohrabji, F., and Choleris, E. (2017). Why estrogens matter for behavior and brain health. *Neuroscience & Biobehavioral Reviews*, 76:363–379.

Geerligs, L., Rubinov, M., Cam-Can, and Henson, R. N. (2015). State and Trait Components of Functional Connectivity: Individual Differences Vary with Mental State. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 35(41):13949–61.

Gewin, V. (2011). Gender divide in physics spans globe. *Nature*, 473(7348):547–547.

Ginestet, C. E. and Simmons, A. (2011). Statistical parametric network analysis of functional connectivity dynamics during a working memory task. *NeuroImage*, 55(2):688–704.

Goebel, R. (2012). BrainVoyager–past, present, future. *NeuroImage*, 62(2):748–756.

Goldstein, J. M., Jerram, M., Poldrack, R., Ahern, T., Kennedy, D. N., Seidman, L. J., and Makris, N. (2005). Hormonal cycle modulates arousal circuitry in women using functional magnetic resonance imaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(40):9309–9316.

Goldstein-Piekarski, A. N., Holt-Gosselin, B., O'Hora, K., and Williams, L. M. (2020). Integrating sleep, neuroimaging, and computational approaches for precision psychiatry. *Neuropsychopharmacology*, 45(1):192–204. Number: 1 Publisher: Nature Publishing Group.

Gong, G., He, Y., Concha, L., Lebel, C., Gross, D. W., Evans, A. C., and Beaulieu, C. (2009). Mapping Anatomical Connectivity Patterns of Human Cerebral Cortex Using In Vivo Diffusion Tensor Imaging Tractography. *Cerebral Cortex*, 19(3):524–536. Publisher: Oxford Academic.

Gonzalez, A. A., Bottenhorn, K. L., Bartley, J. E., Hayes, T., Riedel, M. C., Salo, T., Bravo, E. I., Odean, R., Nazareth, A., Laird, R. W., Sutherland, M. T., Brewe, E., Pruden, S. M., and Laird, A. R. (2019). Sex differences in brain correlates of STEM anxiety. *npj Science of Learning*, 4(1):1–10.

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J. M., Coalson, R. S., Nguyen, A. L., McDermott, K. B., Shimony, J. S., Snyder, A. Z., Schlaggar, B. L., Petersen, S. E., Nelson, S. M., and Dosenbach, N. U. F. (2017). Precision Functional Mapping of Individual Human Brains. *Neuron*, 95(4):791–807.e7.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., Poline, J.-B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J. A., Varoquaux, G., and Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3:160044.

Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., and Ghosh, S. S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics*, 5:13.

Gorgolewski, K. J., Esteban, O., Ellis, D. G., Notter, M. P., Ziegler, E., Johnson, H., Hamalainen, C., Yvernault, B., Burns, C., Manhães-Savio, A., Jarecka, D., Markiewicz, C. J., Salo, T., Clark, D., Waskom, M., Wong, J., Modat, M., Dewey, B. E., Clark, M. G., Dayan, M., Loney, F., Madison, C., Gramfort, A., Keshavan, A., Berleant, S., Pinsard, B., Goncalves, M., Clark, D., Cipollini, B., Varoquaux, G., Wassermann, D., Rokem, A., Halchenko, Y. O., Forbes, J., Moloney, B., Malone, I. B., Hanke, M., Mordom, D., Buchanan, C., Pauli, W. M., Huntenburg, J. M., Horea, C., Schwartz, Y., Tungaraza, R., Iqbal, S., Kleesiek, J., Sikka, S., Frohlich, C., Kent, J., Perez-Guevara, M., Watanabe, A., Welch, D., Cumba, C., Ginsburg, D., Eshaghi, A., Kastman, E., Bougacha, S., Blair, R., Acland, B., Gillman, A., Schaefer, A., Nichols, B. N., Giavasis, S., Erickson, D., Correa, C., Ghayoor, A., Küttner, R., Haselgrove, C., Zhou, D., Craddock, R. C., Haehn, D., Lampe, L., Millman, J., Lai, J., Renfro, M., Liu, S., Stadler, J., Glatard, T., Kahn, A. E., Kong, X.-Z., Triplett, W., Park, A., McDermottroe, C., Hallquist, M., Poldrack, R., Perkins, L. N., Noel, M., Gerhard, S., Salvatore, J., Mertz, F., Broderick, W., Inati, S., Hinds, O., Brett, M., Durnez, J., Tambini, A., Rothmei, S., Andberg, S. K., Cooper, G., Marina, A., Mattfeld, A., Urchs, S., Sharp, P., Matsubara, K., Geisler, D., Cheung, B., Floren, A., Nickson, T., Pannetier, N., Weinstein, A., Dubois, M., Arias, J., Tarbert, C., Schlamp, K., Jordan, K., Liem, F., Saase, V., Harms, R., Khanuja, R., Podranski, K., Flandin, G., Papadopoulos Orfanos, D., Schwabacher, I., McNamee, D., Falkiewicz, M., Pellman, J., Linkersdörfer, J., Varada, J., Pérez-García, F., Davison, A., Shachnev, D., and

Ghosh, S. (2017). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. 0.13.1.

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D., Wong, J., Loney, F., Modat, M., Dewey, B. E., Madison, C., Visconti di Oleggio Castello, M., Clark, M. G., Dayan, M., Clark, D., Keshavan, A., Pinsard, B., Gramfort, A., Berleant, S., Nielson, D. M., Bougacha, S., Varoquaux, G., Cipollini, B., Markello, R., Rokem, A., Moloney, B., Halchenko, Y. O., Wassermann, D., Hanke, M., Horea, C., Kaczmarzyk, J., Hollander, G. d., DuPre, E., Gillman, A., Mordom, D., Buchanan, C., Tungaraza, R., Pauli, W. M., Iqbal, S., Sikka, S., Mancini, M., Schwartz, Y., Malone, I. B., Dubois, M., Frohlich, C., Welch, D., Forbes, J., Kent, J., Watanabe, A., Cumba, C., Huntenburg, J. M., Kastman, E., Nichols, B. N., Eshaghi, A., Ginsburg, D., Schaefer, A., Acland, B., Giavasis, S., Kleesiek, J., Erickson, D., Küttner, R., Haselgrove, C., Correa, C., Ghayoor, A., Liem, F., Millman, J., Haehn, D., Lai, J., Zhou, D., Blair, R., Glatard, T., Renfro, M., Liu, S., Kahn, A. E., Pérez-García, F., Triplett, W., Lampe, L., Stadler, J., Kong, X.-Z., Hallquist, M., Chetverikov, A., Salvatore, J., Park, A., Poldrack, R., Craddock, R. C., Inati, S., Hinds, O., Cooper, G., Perkins, L. N., Marina, A., Mattfeld, A., Noel, M., Snoek, L., Matsubara, K., Cheung, B., Rothmei, S., Urchs, S., Durnez, J., Mertz, F., Geisler, D., Floren, A., Gerhard, S., Sharp, P., Molina-Romero, M., Weinstein, A., Broderick, W., Saase, V., Andberg, S. K., Harms, R., Schlamp, K., Arias, J., Papadopoulos Orfanos, D., Tarbert, C., Tambini, A., De La Vega, A., Nickson, T., Brett, M., Falkiewicz, M., Podranski, K., Linkersdörfer, J., Flandin, G., Ort, E., Shachnev, D., McNamee, D., Davison, A., Varada, J., Schwabacher, I., Pellman, J., Perez-Guevara, M., Khanuja, R., Pannetier, N., McDermottroe, C., and Ghosh, S. (2018). Nipype. *Software*. Publisher: Zenodo.

Gottfredson, L. S. (1998). The General Intelligence Factor. *Scientific American Presents: Exploring Intelligence*, 9(4):24–30.

Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance*, 15(1-2):25–46.

Graham, B. M., Ash, C., and Den, M. L. (2017). High endogenous estradiol is associated with enhanced cognitive emotion regulation of physiological conditioned fear responses in women. *Psychoneuroendocrinology*, 80:7–14.

Graham, S., Jiang, J., Manning, V., Nejad, A. B., Zhisheng, K., Salleh, S. R., Golay, X., Berne, Y. I., and McKenna, P. J. (2010). IQ-related fMRI differences during cognitive set shifting. *Cerebral cortex (New York, N.Y. : 1991)*, 20(3):641–649.

Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., Nelson, S. M., Coalson, R. S., Snyder, A. Z., Schlaggar, B. L., Dosenbach, N. U. F., and Petersen, S. E. (2018). Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron*, 98(2):439–452.e5.

Greene, A. S., Gao, S., Scheinost, D., and Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, 9(1):2807.

Greve, D. N. and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1):63–72.

Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., and Alkire, M. T. (2005). The neuroanatomy of general intelligence: sex matters. *NeuroImage*, 25(1):320–327.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1):64–74.

Halloun, I. A. and Hestenes, D. (1985). Common sense concepts about motion. *American Journal of Physics*, 53(11):1056–1065.

Hamidovic, A., Karapetyan, K., Serdarevic, F., Choi, S. H., Eisenlohr-Moul, T., and Pinna, G. (2020). Higher Circulating Cortisol in the Follicular vs. Luteal Phase of the Menstrual Cycle: A Meta-Analysis. *Frontiers in Endocrinology*, 11:311.

Handa, R. J. and Weiser, M. J. (2014). Gonadal Steroid Hormones and the Hypothalamo-Pituitary-Adrenal Axis. *Frontiers in neuroendocrinology*, 35(2):197–220.

Handwerker, D. A., Gonzalez-Castillo, J., D'Esposito, M., and Bandettini, P. A. (2012). The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *NeuroImage*, 62(2):1017–1023.

Hao, J., Rapp, P. R., Leffler, A. E., Leffler, S. R., Janssen, W. G., Lou, W., McKay, H., Roberts, J. A., Wearne, S. L., and Hof, P. R. (2006). Estrogen alters spine number and morphology in prefrontal cortex of aged female rhesus monkeys. *Journal of Neuroscience*, 26(9):2571–2578. ISBN: 0270-6474 Publisher: Soc Neuroscience.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362. Number: 7825 Publisher: Nature Publishing Group.

Harris, D. (1940). Factors affecting college grades: a review of the literature, 1930-1937. *Psychological Bulletin*, 37(3):125–166. Place: US Publisher: American Psychological Association.

Hawco, C., Dickie, E. W., Jacobs, G., Daskalakis, Z. J., and Voineskos, A. N. (2021). Moving Beyond the Mean: Subgroups and dimensions of brain activity and cognitive performance across domains. *NeuroImage*, page 117823.

Hayasaka, S. and Laurienti, P. J. (2010). Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data. *NeuroImage*, 50(2):499–508.

He, Y., Chen, Z. J., and Evans, A. C. (2007). Small-World Anatomical Networks in the Human Brain Revealed by Cortical Thickness from MRI. *Cerebral Cortex*, 17(10):2407–2419. Publisher: Oxford Academic.

Hearne, L. J., Mattingley, J. B., and Cocchi, L. (2016). Functional brain networks related to individual differences in human intelligence at rest. *Scientific Reports*, 6:32328.

Hegarty, M. (2014). Spatial Thinking in Undergraduate Science Education. *Spatial Cognition & Computation*, 14(2):142–167.

Hegarty, M., Stieff, M., and Dixon, B. L. (2013). Cognitive change in mental models with experience in the domain of organic chemistry. *Journal of Cognitive Psychology*, 25(2):220–228. Publisher: Routledge _eprint: https://doi.org/10.1080/20445911.2012.725044.

Helwig, R., Rozek-tedesco, M. A., Tindal, G., Heath, B., and Almond, P. J. (1999). Reading as an Access to Mathematics Problem Solving on Multiple-Choice Tests for Sixth-Grade Students. *The Journal of Educational Research*, 93(2):113–125. Publisher: Routledge _eprint: https://doi.org/10.1080/00220679909597635.

Henderson, V. W. (2018). Progesterone and human cognition. *Climacteric : the journal of the International Menopause Society*, 21(4):333–340.

Hernández-Hernández, O. T., Martínez-Mota, L., Herrera-Pérez, J. J., and Jiménez-Rubio, G. (2019). Role of Estradiol in the Expression of Genes Involved in Serotonin Neurotransmission: Implications for Female Depression. *Current Neuropharmacology*, 17(5):459–471.

Herrnstein, R. J. and Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press. Google-Books-ID: s4CKqxi6yWIC.

Hertzog, C. and Schaie, K. W. (1986). Stability and change in adult intelligence: 1. Analysis of longitudinal covariance structures. *Psychology and Aging*, 1(2):159–171.

Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3):141–158.

Heuvel, M. P., Stam, C. J., Kahn, R. S., and Hulshoff Pol, H. E. (2009). Efficiency of functional brain networks and intellectual performance. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(23):7619–24.

Hilger, K., Ekman, M., Fiebach, C. J., and Basten, U. (2017a). Efficient hubs in the intelligent brain: Nodal efficiency of hub regions in the salience network is associated with general intelligence. *Intelligence*, 60:10–25.

Hilger, K., Ekman, M., Fiebach, C. J., and Basten, U. (2017b). Intelligence is associated with the modular structure of intrinsic brain networks. *Scientific Reports*, 7(1):16088.

Holdnack, J. A., Drozdick, L. W., Iverson, G. L., and Chelune, G. J. (2013). Chapter 6 - Serial Assessment with WAIS–IV and WMS–IV. In Holdnack, J. A., Drozdick, L. W., Weiss, L. G., and Iverson, G. L., editors, *WAIS-IV, WMS-IV, and ACS*, pages 279–329. Academic Press, San Diego.

Howe, M. J. A. (1997). *IQ in Question: The Truth about Intelligence*. SAGE. Google-Books-ID: p7HWNjU8dDoC.

Hu, C., Tokoglu, F., Scheinost, D., Qiu, M., Shen, X., Peters, D. C., Galiana, G., and Constable, R. T. (2019). Dynamic-flip-angle ECG-gating with nuisance signal regression improves resting-state BOLD functional connectivity mapping by reducing cardiogenic noise. *Magnetic Resonance in Medicine*, 82(3):911–923.

Hunt, E. (1997). The Concept and Utility of Intelligence. In Devlin, B., Fienberg, S. E., Resnick, D. P., and Roeder, K., editors, *Intelligence, Genes, and Success: Scientists Respond to The Bell Curve*, pages 157–176. Springer New York, New York, NY.

Hunter, J. E. and Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96(1):72–98.

Igolkina, A. A. and Meshcheryakov, G. (2020). semopy: A Python Package for Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(6):952–963. Publisher: Routledge _eprint: https://doi.org/10.1080/10705511.2019.1704289.

Ikarashi, K., Sato, D., Iguchi, K., Baba, Y., and Yamashiro, K. (2020). Menstrual Cycle Modulates Motor Learning and Memory Consolidation in Humans. *Brain Sciences*, 10(10).

Jackson, J., Dukerich, L., and Hestenes, D. (2008). Modeling Instruction: An Effective Model for Science Education. *Science Educator*, 17(1):10–17.

Jacobs, E. and D'Esposito, M. (2011). Estrogen Shapes Dopamine-Dependent Cognitive Processes: Implications for Women's Health. *Journal of Neuroscience*, 31(14):5286–5293. Publisher: Society for Neuroscience Section: Articles.

Jacobs, E. G., Weiss, B. K., Makris, N., Whitfield-Gabrieli, S., Buka, S. L., Klibanski, A., and Goldstein, J. M. (2016). Impact of sex and menopausal status on episodic memory circuitry in early midlife. *Journal of Neuroscience*, 36(39):10163–10173. ISBN: 0270-6474 Publisher: Soc Neuroscience.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825–841.

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *NeuroImage*, 62(2):782–790.

Jiang, R., Calhoun, V. D., Fan, L., Zuo, N., Jung, R., Qi, S., Lin, D., Li, J., Zhuo, C., Song, M., Fu, Z., Jiang, T., and Sui, J. (2019). Gender Differences in Connectome-based Predictions of Individualized Intelligence Quotient and Sub-domain Scores. *Cerebral Cortex*, 30(3):888–900.

Johnstone, B. and Wilhelm, K. L. (1996). The longitudinal stability of the WRAT-R Reading subtest: Is it an appropriate estimate of premorbid intelligence? *Journal of the International Neuropsychological Society*, 2(4):282–285. Publisher: Cambridge University Press.

Jones, T. B., Bandettini, P. A., and Birn, R. M. (2008). Integration of motion correction and physiological noise regression in fMRI. *NeuroImage*, 42(2):582–590.

Jones, T. B., Bandettini, P. A., Kenworthy, L., Case, L. K., Milleville, S. C., Martin, A., and Birn, R. M. (2010). Sources of group differences in functional connectivity: An investigation applied to autism spectrum disorder. *NeuroImage*, 49(1):401–414.

Jung, R. E. and Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *The Behavioral and Brain Sciences*, 30(2):135–154; discussion 154–187.

Kajantie, E. and Phillips, D. I. W. (2006). The effects of sex and hormonal status on the physiological response to acute psychosocial stress. *Psychoneuroendocrinology*, 31(2):151–178.

Karin, O., Raz, M., Tendler, A., Bar, A., Korem Kohanim, Y., Milo, T., and Alon, U. (2020). A new model for the HPA axis explains dysregulation of stress hormones on the timescale of weeks. *Molecular Systems Biology*, 16(7).

Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology*, 77(2):323–338.

Keller, J. and Dauenheimer, D. (2003). Stereotype Threat in the Classroom: Dejection Mediates the Disrupting Threat Effect on Women's Math Performance. *Personality and Social Psychology Bulletin*, 29(3):371–381.

Kelly, A. M. (2016). Social cognitive perspective of gender disparities in undergraduate physics. *Physical Review Physics Education Research*, 12(2):020116.

Kirilina, E., Lutti, A., Poser, B. A., Blankenburg, F., and Weiskopf, N. (2016). The quest for the best: The impact of different EPI sequences on the sensitivity of random effect fMRI group analyses. *NeuroImage*, 126:49–59.

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., and Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, 13(2):e1005350. Publisher: Public Library of Science.

Kleinhans, N. M., Richards, T., Johnson, L. C., Weaver, K. E., Greenson, J., Dawson, G., and Aylward, E. (2011). fMRI evidence of neural abnormalities in the subcortical face processing system in ASD. *NeuroImage*, 54(1):697–704.

Knaus, T. A., Silver, A. M., Lindgren, K. A., Hadjikhani, N., and Tager-Flusberg, H. (2008). fMRI activation during a language task in adolescents with ASD. *Journal of the International Neuropsychological Society*, 14(6):967–979. Publisher: Cambridge University Press.

Konrad, K., Neufang, S., Thiel, C. M., Specht, K., Hanisch, C., Fan, J., Herpertz-Dahlmann, B., and Fink, G. R. (2005). Development of attentional networks: An fMRI study with children and adults. *NeuroImage*, 28(2):429–439.

Kozhevnikov, M., Hegarty, M., and Mayer, R. E. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and instruction*, 20(1):47–77.

Kozhevnikov, M. and Thornton, R. (2006). Real-Time Data Display, Spatial Visualization Ability, and Learning Force and Motion Concepts. *Journal of Science Education and Technology*, 15(1):111.

Kraus, T., Hösl, K., Kiess, O., Schanze, A., Kornhuber, J., and Forster, C. (2007). BOLD fMRI deactivation of limbic and temporal brain structures and mood enhancing effect by transcutaneous vagus nerve stimulation. *Journal of Neural Transmission (Vienna, Austria: 1996)*, 114(11):1485–1493.

Krishnan, V. and Collop, N. A. (2006). Gender differences in sleep disorders. *Current opinion in pulmonary medicine*, 12(6):383–389. ISBN: 1070-5287 Publisher: LWW.

Kundu, P., Voon, V., Balchandani, P., Lombardo, M. V., Poser, B. A., and Bandettini, P. A. (2017). Multi-echo fMRI: A review of applications in fMRI denoising and analysis of BOLD signals. *NeuroImage*, 154:59–80.

Kuoppa, P., Niskanen, E., Karjalainen, P., and Tarvainen, M. P. (2012). Cerebral cortex and sub-cortex lateralization in cardiovascular regulation: correlations of BOLD fMRI and heart rate variability. *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2012:3412–3415.

Kwong, K. K. (2012). Record of a Single fMRI Experiment in May of 1991. *Neuroimage*, 62(2):610–612.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., and Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12):5675–5679. ISBN: 0027-8424 Publisher: National Acad Sciences.

Laird, A. R., Fox, P. M., Eickhoff, S. B., Turner, J. A., Ray, K. L., McKay, D. R., Glahn, D. C., Beckmann, C. F., Smith, S. M., and Fox, P. T. (2011). Behavioral interpretations of intrinsic connectivity networks. *Journal of Cognitive Neuroscience*, 23(12):4022–4037.

Lalmansingh, A. S. and Uht, R. M. (2008). Estradiol regulates corticotropin-releasing hormone gene (crh) expression in a rapid and phasic manner that parallels estrogen receptor-alpha and -beta recruitment to a 3',5'-cyclic adenosine 5'-monophosphate regulatory region of the proximal crh promoter. *Endocrinology*, 149(1):346–357.

Lanczos, C. (1964). Evaluation of Noisy Data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, 1(1):76–85.

Lane, R. D., Weidenbacher, H., Smith, R., Fort, C., Thayer, J. F., and Allen, J. J. B. (2013). Subgenual anterior cingulate cortex activity covariation with cardiac vagal control is altered in depression. *Journal of Affective Disorders*, 150(2):565–570.

Langer, N., Pedroni, A., Gianotti, L. R., Hänggi, J., Knoch, D., and Jäncke, L. (2012). Functional brain network efficiency predicts intelligence. *Human Brain Mapping*, 33(6):1393–1406.

Lee, K. H., Choi, Y. Y., Gray, J. R., Cho, S. H., Chae, J.-H., Lee, S., and Kim, K. (2006). Neural correlates of superior intelligence: Stronger recruitment of posterior parietal cortex. *NeuroImage*, 29(2):578–586.

Leeners, B., Krüger, T. H. C., Geraedts, K., Tronci, E., Mancini, T., Egli, M., Röblitz, S., Saleh, L., Spanaus, K., Schippert, C., Zhang, Y., and Ille, F. (2019). Associations Between Natural Physiological and Supraphysiological Estradiol Levels and Stress Perception. *Frontiers in Psychology*, 10:1296.

Li, H., Smith, S. M., Gruber, S., Lukas, S. E., Silveri, M. M., Hill, K. P., Killgore, W. D. S., and Nickerson, L. D. (2020). Denoising scanner effects from multimodal MRI data using linked independent component analysis. *NeuroImage*, 208:116388.

Li, J. and Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*, 95(3):221–227.

Li, Y., Liu, Y., Li, J., Qin, W., Li, K., Yu, C., and Jiang, T. (2009). Brain Anatomical Network and Intelligence. *PLoS Computational Biology*, 5(5):e1000395. not individual differences, separated into high- and low-IQ groups.

Liu, T. T. (2013). Neurovascular factors in resting-state functional MRI. *NeuroImage*, 80:339–348.

Liu, T. T. (2016). Noise contributions to the fMRI signal: An overview. *NeuroImage*, 143:141–151.

Lorenzo, M., Crouch, C. H., and Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics*, 74(2):118–122.

Lynall, M.-E., Bassett, D. S., Kerwin, R., McKenna, P. J., Kitzbichler, M., Muller, U., and Bullmore, E. (2010). Functional Connectivity and Brain Networks in Schizophrenia. *Journal of Neuroscience*, 30(28):9477–9487. Publisher: Society for Neuroscience Section: Articles.

Lynch, C. J., Power, J. D., Scult, M. A., Dubin, M., Gunning, F. M., and Liston, C. (2020). Rapid Precision Functional Mapping of Individuals Using Multi-Echo fMRI. *Cell Reports*, 33(12):108540.

Mackey, S., Kan, K.-J., Chaarani, B., Alia-Klein, N., Batalla, A., Brooks, S., Cousijn, J., Dagher, A., de Ruiter, M., Desrivieres, S., Feldstein Ewing, S. W., Goldstein, R. Z., Goudriaan, A. E., Heitzeg, M. M., Hutchison, K., Li, C.-S. R., London, E. D., Lorenzetti, V., Luijten, M., Martin-Santos, R., Morales, A. M., Paulus, M. P., Paus, T., Pearlson, G., Schluter, R., Momenan, R., Schmaal, L., Schumann, G., Sinha, R., Sjoerds, Z., Stein, D. J., Stein, E. A., Solowij, N., Tapert, S., Uhlmann, A., Veltman, D., van Holst, R., Walter, H., Wright, M. J., Yucel, M., Yurgelun-Todd, D., Hibar, D. P., Jahanshad, N., Thompson, P. M., Glahn, D. C., Garavan, H., and Conrod, P. (2016). Chapter 10 - Genetic imaging consortium for addiction medicine: From neuroimaging to genes. In Ekhtiari, H. and Paulus, M. P., editors, *Progress in Brain Research*, volume 224 of *Neuroscience for Addiction Medicine: From Prevention to Rehabilitation - Methods and Interventions*, pages 203–223. Elsevier.

Madsen, A., McKagan, S. B., and Sayre, E. C. (2013). Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap? *Physical Review Special Topics - Physics Education Research*, 9(2):020121.

Madsen, A. M., Larson, A. M., Loschky, L. C., and Rebello, N. S. (2012). Differences in visual attention between those who correctly and incorrectly answer physics problems. *Physical Review Special Topics - Physics Education Research*, 8(1):010122.

Maki, P. M. (2005). Estrogen effects on the hippocampus and frontal lobes. *International Journal of Fertility and Women's Medicine*, 50(2):67–71.

Mallampalli, M. P. and Carter, C. L. (2014). Exploring Sex and Gender Differences in Sleep Health: A Society for Women's Health Research Report. *Journal of Women's Health*, 23(7):553–562. Publisher: Mary Ann Liebert, Inc., publishers.

Marek, S., Tervo-Clemmens, B., Nielsen, A. N., Wheelock, M. D., Miller, R. L., Laumann, T. O., Earl, E., Foran, W. W., Cordova, M., Doyle, O., Perrone, A., Miranda-Dominguez, O., Feczko, E., Sturgeon, D., Graham, A., Hermosillo, R., Snider, K., Galassi, A., Nagel, B. J., Ewing, S. W. F., Eggebrecht, A. T., Garavan, H., Dale, A. M., Greene, D. J., Barch, D. M., Fair, D. A., Luna, B., and Dosenbach, N. U. (2019). Identifying reproducible individual differences in childhood functional brain networks: An ABCD study. *Developmental Cognitive Neuroscience*, 40.

McAllister, T. W., Sparling, M. B., Flashman, L. A., Guerin, S. J., Mamourian, A. C., and Saykin, A. J. (2001). Differential Working Memory Load Effects after Mild Traumatic Brain Injury. *NeuroImage*, 14(5):1004–1012.

McCullough, L. (2013). Gender, Context, and Physics Assessment. *Journal of International Women's Studies*, 5(4):20–30.

McGonigle, D. J. (2012). Test–retest reliability in fMRI: Or how I learned to stop worrying and love the variability. *NeuroImage*, 62(2):1116–1120.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In Walt, S. v. d. and Millman, J., editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61.

McLean, C. P., Asnaani, A., Litz, B. T., and Hofmann, S. G. (2011). Gender Differences in Anxiety Disorders: Prevalence, Course of Illness, Comorbidity and Burden of Illness. *Journal of psychiatric research*, 45(8):1027–1035.

Mears, D. and Pollard, H. B. (2016). Network science and the human brain: Using graph theory to understand the brain and one of its hubs, the amygdala, in health and disease. *Journal of Neuroscience Research*, 94(6):590–605. An interesting future study would be to determine whether the nodal properties of the amygdala within this network are altered in patients with mood or anxiety disorders. Neuroticism positively correlated with the BC of the bilateral amygdala, right precentral gyrus, olfactory cortex, and caudate. The authors suggest that this result is evidence of a greater level of activity in limbic regions, including the bilateral amygdala, in resting brain networks of subjects with high neuroticism

scores (Gao et al., 2013). wever, in terms of graph theory, the positive correlation between amygdalar BC and neuroticism score indicates only that the amygdala has a greater influence over information transfer in resting networks of individuals with high neuroticism.

Meers, J., Stout-Aguilar, J., and Nowakowski, S. (2019). Chapter 3 - Sex differences in sleep health. In Grandner, M. A., editor, *Sleep and Health*, pages 21–29. Academic Press.

Meers, J. M. and Nowakowski, S. (2020). Sleep, premenstrual mood disorder, and women's health. *Current Opinion in Psychology*, 34:43–49.

Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. *Trends in Cognitive Sciences*, 15(10):483–506.

Menon, V. and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, 214(5):655–667.

Miller, E. H. (2004). Women and insomnia. *Clinical Cornerstone*, 6 Suppl 1B:S8–18.

Millman, K. J. and Brett, M. (2007). Analysis of Functional Magnetic Resonance Imaging in Python. *Computing in Science Engineering*, 9(3):52–55.

Mobascher, A., Brinkmeyer, J., Warbrick, T., Musso, F., Schlemper, V., Wittsack, H. J., Saleh, A., Schnitzler, A., and Winterer, G. (2010). Brain activation patterns underlying fast habituation to painful laser stimuli. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 75(1):16–24.

Mobascher, A., Brinkmeyer, J., Warbrick, T., Musso, F., Wittsack, H. J., Stoermer, R., Saleh, A., Schnitzler, A., and Winterer, G. (2009). Fluctuations in electrodermal activity reveal variations in single trial brain responses to painful laser stimuli–a fMRI/EEG study. *NeuroImage*, 44(3):1081–1092.

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., and Uğurbil, K. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic Resonance in Medicine*, 63(5):1144–1153.

Moffitt, T. E., Caspi, A., Harkness, A. R., and Silva, P. A. (1993). The Natural History of Change to Intellectual Performance: Who Changes? How Much? Is it Meaningful? *Journal of Child Psychology and Psychiatry*, 34(4):455–506. _eprint: https://acamh.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-7610.1993.tb01031.x.

Mong, J. A. and Cusmano, D. M. (2016). Sex differences in sleep: impact of biological sex and sex steroids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1688):20150110. Publisher: Royal Society.

Mulcahy, J. S., Larsson, D. E. O., Garfinkel, S. N., and Critchley, H. D. (2019). Heart rate variability as a biomarker in health and affective disorders: A perspective on neuroimaging studies. *NeuroImage*, 202:116072.

Murphy, K., Birn, R. M., and Bandettini, P. A. (2013). Resting-state fMRI confounds and cleanup. *NeuroImage*, 80:349–359.

Murray, C. A. (1998). *Income Inequality and IQ*. AEI Press. Google-Books-ID: rrdNL4dpijwC.

Nagma, S., Kapoor, G., Bharti, R., Batra, A., Batra, A., Aggarwal, A., and Sablok, A. (2015). To evaluate the effect of perceived stress on menstrual function. *Journal of clinical and diagnostic research: JCDR*, 9(3):QC01–03.

Napadow, V., Dhond, R., Conti, G., Makris, N., Brown, E. N., and Barbieri, R. (2008). Brain correlates of autonomic modulation: Combining heart rate variability with fMRI. *NeuroImage*, 42(1):169–177.

Negishi, M., Pinus, B. I., Pinus, A. B., and Constable, R. T. (2007). Origin of the Radio Frequency Pulse Artifact in Simultaneous EEG-fMRI Recording: Rectification at the Carbon-Metal Interface. *IEEE transactions on bio-medical engineering*, 54(9):1725–1727.

Neidorf, T., Arora, A., Erberber, E., Tsokodayi, Y., and Mai, T. (2020). Review of Research into Misconceptions and Misunderstandings in Physics and Mathematics. In Neidorf, T., Arora, A., Erberber, E., Tsokodayi, Y., and Mai, T., editors, *Student Misconceptions and Errors in Physics and Mathematics: Exploring Data from TIMSS and TIMSS Advanced*, IEA Research for Education, pages 11–20. Springer International Publishing, Cham.

Neisser, U., Boodoo, G., Bouchard Jr., T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., and Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2):77–101.

Nepomnaschy, P. A., Altman, R. M., Watterson, R., Co, C., McConnell, D. S., and England, B. G. (2011). Is cortisol excretion independent of menstrual cycle day? A longitudinal evaluation of first morning urinary specimens. *PloS One*, 6(3):e18242.

Neubauer, A. C. and Fink, A. (2009). Intelligence and neural efficiency: Measures of brain activation versus measures of functional connectivity in the brain. *Intelligence*, 37(2):223–229.

Nicolaides, N. C., Charmandari, E., Chrousos, G. P., and Kino, T. (2014). Circadian endocrine rhythms: the hypothalamic–pituitary–adrenal axis and its actions. *Annals of the New York Academy of Sciences*, 1318:71–80.

Nikolaidis, A., Solon Heinsfeld, A., Xu, T., Bellec, P., Vogelstein, J., and Milham, M. (2020). Bagging improves reproducibility of functional parcellation of the human brain. *NeuroImage*, 214:116678.

Nillni, Y. I., Wesselink, A. K., Hatch, E. E., Mikkelsen, E. M., Gradus, J. L., Rothman, K. J., and Wise, L. A. (2018). Mental health, psychotropic medication use, and menstrual cycle characteristics. *Clinical Epidemiology*, 10:1073–1082.

Nissen, J. M. and Shemwell, J. T. (2016). Gender, experience, and self-efficacy in introductory physics. *Physical Review Physics Education Research*, 12(2):020105.

Nosek, B. A., Banaji, M. R., and Greenwald, A. G. (2002). Math = male, me = female, therefore math not = me. *Journal of personality and social psychology*, 83(1):44–59.

Nyberg, L., Eriksson, J., Larsson, A., and Marklund, P. (2006). Learning by doing versus learning by thinking: An fMRI study of motor and mental training. *Neuropsychologia*, 44(5):711–717.

Ochedalski, T., Subburaju, S., Wynn, P. C., and Aguilera, G. (2007). Interaction Between Oestrogen and Oxytocin on Hypothalamic-Pituitary-Adrenal Axis Activity. *Journal of Neuroendocrinology*, 19(3):189–197. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2826.2006.01525.x.

Ochsner, K. N., Bunge, S. A., Gross, J. J., and Gabrieli, J. D. E. (2002). Rethinking Feelings: An fMRI Study of the Cognitive Regulation of Emotion. *Journal of Cognitive Neuroscience*, 14(8):1215–1229.

Ogawa, S., Lee, T.-M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *proceedings of the National Academy of Sciences*, 87(24):9868–9872. ISBN: 0027-8424 Publisher: National Acad Sciences.

Olafsson, V., Kundu, P., Wong, E. C., Bandettini, P. A., and Liu, T. T. (2015). Enhanced Identification of BOLD-like Components with Multi-Echo Simultaneous Multi-Slice (MESMS) fMRI and Multi-Echo ICA. *NeuroImage*, 112:43–51.

Oliphant, T. E. (2007). Python for Scientific Computing. *Computing in Science Engineering*, 9(3):10–20. Conference Name: Computing in Science Engineering.

Oswald, T. M., Beck, J. S., Iosif, A.-M., McCauley, J. B., Gilhooly, L. J., Matter, J. C., and Solomon, M. (2016). Clinical and Cognitive Characteristics Associated with Mathematics Problem Solving in Adolescents with Autism Spectrum Disorder. *Autism Research*, 9(4):480–490.

Oyola, M. G. and Handa, R. J. (2017). Hypothalamic–pituitary–adrenal and hypothalamic–pituitary–gonadal axes: sex differences in regulation of stress responsivity. *Stress (Amsterdam, Netherlands)*, 20(5):476–494.

Pallrand, G. J. and Seeber, F. (1984). Spatial ability and achievement in introductory physics. *Journal of Research in Science Teaching*, 21(5):507–516.

Parker, D., Liu, X., and Razlighi, Q. R. (2017). Optimal slice timing correction and its interaction with fMRI parameters and artifacts. *Medical Image Analysis*, 35:434–445.

Parry, B. L., Javeed, S., Laughlin, G. A., Hauger, R., and Clopton, P. (2000). Cortisol circadian rhythms during the menstrual cycle and with sleep deprivation in premenstrual dysphoric disorder and normal control subjects. *Biological Psychiatry*, 48(9):920–931. ISBN: 0006-3223 Publisher: Elsevier.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(Oct):2825–2830.

Perlaki, G., Orsi, G., Schwarcz, A., Bodi, P., Plozer, E., Biczo, K., Aradi, M., Doczi, T., Komoly, S., Hejjel, L., Kovacs, N., and Janszky, J. (2015). Pain-related autonomic response is modulated by the medial prefrontal cortex: An ECG-fMRI study in men. *Journal of the Neurological Sciences*, 349(1-2):202–208.

Peters, A. T., Van Meter, A., Pruitt, P. J., Briceño, E. M., Ryan, K. A., Hagan, M., Weldon, A. L., Kassel, M. T., Vederman, A., Zubieta, J.-K., McInnis, M., Weisenbach, S. L., and Langenecker, S. A. (2016). Acute cortisol reactivity attenuates engagement of fronto-parietal and striatal regions during emotion processing in negative mood disorders. *Psychoneuroendocrinology*, 73:67–78.

Phumsatitpong, C., De Guzman, R. M., Zuloaga, D. G., and Moenter, S. M. (2020). A CRH Receptor Type 1 Agonist Increases GABA Transmission to GnRH Neurons in a Circulating-Estradiol-Dependent Manner. *Endocrinology*, 161(11).

Piacente, T. (2012). Verbal comprehension of university students. *Orientación y Sociedad*, no. 12.

Pietschnig, J. and Voracek, M. (2015). One Century of Global IQ Gains: A Formal Meta-Analysis of the Flynn Effect (1909–2013). *Perspectives on Psychological Science*, 10(3):282–306. Publisher: SAGE Publications Inc.

Plomin, R. and von Stumm, S. (2018). The new genetics of intelligence. *Nature reviews. Genetics*, 19(3):148–159.

Poldrack, R. A. (2017). Precision Neuroscience: Dense Sampling of Individual Brains. *Neuron*, 95(4):727–729.

Poldrack, R. A., Laumann, T. O., Koyejo, O., Gregory, B., Hover, A., Chen, M.-Y., Gorgolewski, K. J., Luci, J., Joo, S. J., Boyd, R. L., Hunicke-Smith, S., Simpson, Z. B., Caven, T., Sochat, V., Shine, J. M., Gordon, E., Snyder, A. Z., Adeyemo, B., Petersen, S. E., Glahn, D. C., Reese Mckay, D., Curran, J. E., Göring, H. H. H., Carless, M. A., Blangero, J., Dougherty, R., Leemans, A., Handwerker, D. A., Frick, L., Marcotte, E. M., and Mumford, J. A. (2015). Long-term neural and physiological phenotyping of a single human. *Nature Communications*, 6:8885.

Pollock, S. J., Finkelstein, N. D., and Kost, L. E. (2007). Reducing the gender gap in the physics classroom: How sufficient is interactive engagement? *Physical Review Special Topics - Physics Education Research*, 3(1):010107.

Pompili, A., Arnone, B., and Gasbarri, A. (2012). Estrogens and memory in physiological and neuropathological conditions. *Psychoneuroendocrinology*, 37(9):1379–1396.

Posse, S., Wiese, S., Gembris, D., Mathiak, K., Kessler, C., Grosse-Ruyken, M.-L., Elghahwagi, B., Richards, T., Dager, S. R., and Kiselev, V. G. (1999). Enhancement of BOLD-contrast sensitivity by single-shot multi-echo functional MR imaging. *Magnetic Resonance in Medicine*, 42(1):87–97.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, 84:320–341.

Prescott, J., Gavrilescu, M., Cunnington, R., O'Boyle, M. W., and Egan, G. F. (2010). Enhanced brain connectivity in math-gifted adolescents: An fMRI study using mental rotation. *Cognitive Neuroscience*, 1(4):277–288. Publisher: Routledge _eprint: https://doi.org/10.1080/17588928.2010.506951.

Pritschet, L., Santander, T., Taylor, C. M., Layher, E., Yu, S., Miller, M. B., Grafton, S. T., and Jacobs, E. G. (2020). Functional reorganization of brain networks across the human menstrual cycle. *NeuroImage*, 220:117091.

Protopopescu, X., Butler, T., Pan, H., Root, J., Altemus, M., Polanecsky, M., McEwen, B., Silbersweig, D., and Stern, E. (2008). Hippocampal structural changes across the menstrual cycle. *Hippocampus*, 18(10):985–988. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hipo.20468.

Rabin, D. S., Johnson, E. O., Brandon, D. D., Liapi, C., and Chrousos, G. P. (1990). Glucocorticoids inhibit estradiol-mediated uterine growth: possible role of the uterine estradiol receptor. *Biology of Reproduction*, 42(1):74–80.

Rauch, S. L., Whalen, P. J., Shin, L. M., McInerney, S. C., Macklin, M. L., Lasko, N. B., Orr, S. P., and Pitman, R. K. (2000). Exaggerated amygdala response to masked facial stimuli in posttraumatic stress disorder: a functional MRI study. *Biological Psychiatry*, 47(9):769–776.

Reddy, A. (2007). The eugenic origins of IQ testing: Implications for post-Atkins litigation. *DePaul L. Rev.*, 57:667. Publisher: HeinOnline.

Reynolds, C. A., Finkel, D., McArdle, J. J., Gatz, M., Berg, S., and Pedersen, N. L. (2005). Quantitative Genetic Analysis of Latent Growth Curve Models of Cognitive Abilities in Adulthood. *Developmental Psychology*, 41(1):3–16. Place: US Publisher: American Psychological Association.

Reynolds, C. R. and Suzuki, L. A. (2012). Bias in Psychological Assessment. In *Handbook of Psychology, Second Edition*. American Cancer Society. Section: 4 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118133880.hop210004.

Richardson, K. (2002). What IQ Tests Test. *Theory & Psychology*, 12(3):283–314.

Richman, L. S., vanDellen, M., and Wood, W. (2011). How Women Cope: Being a Numerical Minority in a Male-Dominated Profession. *Journal of Social Issues*, 67(3):492–509.

Ritchie, S. J. and Tucker-Drob, E. M. (2018). How Much Does Education Improve Intelligence? A Meta-Analysis. *Psychological Science*, 29(8):1358–1369.

Roberts, J. D., While, A. E., and Fitzpatrick, J. M. (1993). Problem solving in nursing practice: application, process, skill acquisition and measurement. *Journal of Advanced Nursing*, 18(6):886–891.

Robinson, J. L., Miller, M. W., Beyers, R., Grand, K., Kirby, L. A., Macy, A., and Graap, K. (2014). NEUROPSYCHOPHYSIOLOGICAL MAPPING: CONCOMMITANT PSYCHOPHYSIOLOGICAL RECORDING AND SUBMILLIMETER FUNCTIONAL MAGNETIC RESONANCE IMAGING (FMRI) AT 7T. volume 51, pages S78–S78. Wiley-Blackwell.

Rosen, B. R. and Savoy, R. L. (2012). fMRI at 20: Has it changed the world? *NeuroImage*, 62(2):1316–1324.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(1):1–36. Number: 1.

Rossetti, M. F., Cambiasso, M. J., Holschbach, M. A., and Cabrera, R. (2016). Oestrogens and Progestagens: Synthesis and Action in the Brain. *Journal of Neuroendocrinology*, 28(7). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jne.12402.

Roy, B. N., Reid, R. L., and Van Vugt, D. A. (1999). The effects of estrogen and progesterone on corticotropin-releasing hormone and arginine vasopressin messenger ribonucleic acid levels in the paraventricular nucleus and supraoptic nucleus of the rhesus monkey. *Endocrinology*, 140(5):2191–2198.

Rubia, K., Overmeyer, S., Taylor, E., Brammer, M., Williams, S. C. R., Simmons, A., Andrew, C., and Bullmore, E. T. (2000). Functional frontalisation with age: mapping neurodevelopmental trajectories with fMRI. *Neuroscience & Biobehavioral Reviews*, 24(1):13–19.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069.

Ryman, S. G., Yeo, R. A., Witkiewitz, K., Vakhtin, A. A., Heuvel, M. v. d., Reus, M. d., Flores, R. A., Wertz, C. R., and Jung, R. E. (2016). Fronto-Parietal gray matter and white matter efficiency differentially predict intelligence in males and females. *Human Brain Mapping*, 37(11):4006–4016.

Sacher, J., Neumann, J., Okon-Singer, H., Gotowiec, S., and Villringer, A. (2013). Sexual dimorphism in the human brain: evidence from neuroimaging. *Magnetic Resonance Imaging*, 31(3):366–375.

Salehi, M., Karbasi, A., Shen, X., Scheinost, D., and Constable, R. T. (2018). An exemplar-based approach to individualized parcellation reveals the need for sex specific functional networks. *NeuroImage*, 170:54–67.

Santarnecchi, E., Galli, G., Polizzotto, N. R., Rossi, A., and Rossi, S. (2014). Efficiency of weak brain connections support general cognitive functioning. *Human Brain Mapping*, 35(9):4566–4582.

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., and Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, 64(1):240–256.

Satterthwaite, T. D., Wolf, D. H., Roalf, D. R., Ruparel, K., Erus, G., Vandekar, S., Gennatas, E. D., Elliott, M. A., Smith, A., Hakonarson, H., Verma, R., Davatzikos, C., Gur, R. E., and Gur, R. C. (2015). Linked Sex Differences in Cognition and Functional Connectivity in Youth. *Cerebral Cortex*, 25(9):2383–2394.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., and Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral cortex*, 28(9):3095–3114. ISBN: 1047-3211 Publisher: Oxford University Press.

Schliep, K. C., Mumford, S. L., Vladutiu, C. J., Ahrens, K. A., Perkins, N. J., Sjaarda, L. A., Kissell, K. A., Prasad, A., Wactawski-Wende, J., and Schisterman, E. F. (2015). Perceived stress, reproductive hormones, and ovulatory function: a prospective cohort study. *Epidemiology (Cambridge, Mass.)*, 26(2):177–184.

Schmithorst, V. J. and Holland, S. K. (2007). Sex differences in the development of neuroanatomical functional connectivity underlying intelligence found using Bayesian connectivity analysis. *NeuroImage*, 35(1):406–419.

Schneider, W., Niklas, F., and Schmiedeler, S. (2014). Intellectual development from early childhood to early adulthood: The impact of early IQ differences on stability and change over time. *Learning and Individual Differences*, 32:156–162.

Schuerger, J. M. and Witt, A. C. (1989). The temporal stability of individually tested intelligence. *Journal of Clinical Psychology*, 45(2):294–302. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1097-4679%28198903%2945%3A2%3C294%3A%3AAID-JCLP2270450218%3E3.0.CO%3B2-N.

Schönemann, P. H. (1983). Do IQ tests really measure intelligence? *Behavioral and Brain Sciences*, 6(2):311–313.

Sclocco, R., Garcia, R. G., Kettner, N. W., Isenburg, K., Fisher, H. P., Hubbard, C. S., Ay, I., Polimeni, J. R., Goldstein, J., Makris, N., Toschi, N., Barbieri, R., and Napadow, V. (2019). The influence of respiration on brainstem and cardiovagal response to auricular vagus nerve stimulation: A multimodal ultrahigh-field (7T) fMRI study. *Brain Stimulation*, 12(4):911–921.

Seitzman, B. A., Gratton, C., Laumann, T. O., Gordon, E. M., Adeyemo, B., Dworetsky, A., Kraus, B. T., Gilmore, A. W., Berg, J. J., Ortega, M., Nguyen, A., Greene, D. J., McDermott, K. B., Nelson, S. M., Lessov-Schlaggar, C. N., Schlaggar, B. L., Dosenbach, N. U. F., and Petersen, S. E. (2019). Trait-like variants in human functional brain networks. *Proceedings of the National Academy of Sciences*, 116(45):22851–22861.

Shariff, S., Cumming, C. E., Lees, A., Handman, M., and Cumming, D. C. (1995). Mood disorder in women with early breast cancer taking tamoxifen, an estradiol receptor antagonist: An expected or unexpected effect? *Annals of the New York Academy of Sciences*, 761(1):365–368. ISBN: 0077-8923 Publisher: Wiley Online Library.

Shen, X., Tokoglu, F., Papademetris, X., and Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, 82:403–415.

Shokri-Kojori, E., Tomasi, D., and Volkow, N. D. (2018). An Autonomic Network: Synchrony Between Slow Rhythms of Pulse and Brain Resting State Is Associated with Personality and Emotions. *Cerebral Cortex (New York, NY)*, 28(9):3356–3371.

Siegel, J. S., Power, J. D., Dubis, J. W., Vogel, A. C., Church, J. A., Schlaggar, B. L., and Petersen, S. E. (2014). Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Human Brain Mapping*, 35(5):1981–1996.

Simmons, F. R., Willis, C., and Adams, A.-M. (2012). Different components of working memory have different relationships with different mathematical skills. *Journal of Experimental Child Psychology*, 111(2):139–155.

Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., and Windischberger, C. (2011). Slice-timing effects and their correction in functional MRI. *NeuroImage*, 58(2):588–594.

Smeets, T., Dziobek, I., and Wolf, O. T. (2009). Social cognition under stress: Differential effects of stress-induced cortisol elevations in healthy young men and women. *Hormones and Behavior*, 55(4):507–513.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., and Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31):13040–5.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., and Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23 Suppl 1:S208–219.

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, 35(1):4–28.

Sporns, O. (2012). From simple graphs to the connectome: Networks in neuroimaging. *NeuroImage*, 62(2):881–886.

Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, 8(9):418–425.

Stanley, M. L., Dagenbach, D., Lyday, R. G., Burdette, J. H., and Laurienti, P. J. (2014). Changes in global and regional modularity associated with increasing working memory load. *Frontiers in Human Neuroscience*, 8.

Staver, J. R. and Jacks, T. (1988). The influence of cognitive reasoning level, cognitive restructuring ability, disembedding ability, working memory capacity, and prior knowledge on students' performance on balancing equations by inspection. *Journal of Research in Science Teaching*, 25(9):763–775. Publisher: John Wiley & Sons, Ltd.

Stephens, M. A. C., Mahon, P. B., McCaul, M. E., and Wand, G. S. (2016). Hypothalamic-pituitary-adrenal axis response to acute psychosocial stress: Effects of biological sex and circulating sex hormones. *Psychoneuroendocrinology*, 66:47–55.

STERNBERG, R. J. (1984). What Should Intelligence Tests Test? Implications of a Triarchic Theory of Intelligence for Intelligence Testing. *Educational Researcher*, 13(1):5–15.

Stevens, A. A., Tappon, S. C., Garg, A., and Fair, D. A. (2012). Functional Brain Network Modularity Captures Inter- and Intra-Individual Variation in Working Memory Capacity. *PLOS ONE*, 7(1):e30468.

Stewart, L., Henson, R., Kampe, K., Walsh, V., Turner, R., and Frith, U. (2003). Brain changes after learning to read and play music. *NeuroImage*, 20(1):71–83.

Stieff, M. and Uttal, D. (2015). How Much Can Spatial Training Improve STEM Achievement? *Educational Psychology Review*, 27(4):607–615.

Stockhorst, U. and Antov, M. I. (2015). Modulation of Fear Extinction by Stress, Stress Hormones and Estradiol: A Review. *Frontiers in Behavioral Neuroscience*, 9:359.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, 35(5):401–426.

Suh, S., Cho, N., and Zhang, J. (2018). Sex Differences in Insomnia: from Epidemiology and Etiology to Intervention. *Current Psychiatry Reports*, 20(9):69.

Szucs, D. and Ioannidis, J. P. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage*, 221:117164.

Taasoobshirazi, G. and Sinatra, G. M. (2011). A structural equation model of conceptual change in physics. *Journal of Research in Science Teaching*, 48(8):901–918.

Tarín, J. J., Hamatani, T., and Cano, A. (2010). Acute stress may induce ovulation in women. *Reproductive Biology and Endocrinology*, 8(1):53.

Taxier, L. R., Gross, K. S., and Frick, K. M. (2020). Oestradiol as a neuromodulator of learning and memory. *Nature Reviews Neuroscience*, 21(10):535–550. Number: 10 Publisher: Nature Publishing Group.

Taylor, C. M., Pritschet, L., and Jacobs, E. G. (2021). The scientific body of knowledge – Whose body does it serve? A spotlight on oral contraceptives and women's health factors in neuroimaging. *Frontiers in Neuroendocrinology*, 60:100874.

Taylor, C. M., Pritschet, L., Olsen, R. K., Layher, E., Santander, T., Grafton, S. T., and Jacobs, E. G. (2020). Progesterone shapes medial temporal lobe volume across the human menstrual cycle. *NeuroImage*, 220:117125.

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

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., Wright, M. J., Martin, N. G., Agartz, I., Alda, M., Alhusaini, S., Almasy, L., Almeida, J., Alpert, K., Andreasen, N. C., Andreassen, O. A., Apostolova, L. G., Appel, K., Armstrong, N. J., Aribisala, B., Bastin, M. E., Bauer, M., Bearden, C. E., Bergmann, O., Binder, E. B., Blangero, J., Bockholt, H. J., Bøen, E., Bois, C., Boomsma, D. I., Booth, T., Bowman, I. J., Bralten, J., Brouwer, R. M., Brunner, H. G., Brohawn, D. G., Buckner, R. L., Buitelaar, J., Bulayeva, K., Bustillo, J. R., Calhoun, V. D., Cannon, D. M., Cantor, R. M., Carless, M. A., Caseras, X., Cavalleri, G. L., Chakravarty, M. M., Chang, K. D., Ching, C. R. K., Christoforou, A., Cichon, S., Clark, V. P., Conrod, P., Coppola, G., Crespo-Facorro, B., Curran, J. E., Czisch, M., Deary, I. J., de Geus, E. J. C., den Braber, A., Delvecchio, G., Depondt, C., de Haan, L., de Zubicaray, G. I., Dima, D., Dimitrova, R., Djurovic, S., Dong, H., Donohoe, G., Duggirala, R., Dyer, T. D., Ehrlich, S., Ekman, C. J., Elvsåshagen, T., Emsell, L., Erk, S., Espeseth, T., Fagerness, J., Fears, S., Fedko, I., Fernández, G., Fisher, S. E., Foroud, T., Fox, P. T., Francks, C., Frangou, S., Frey, E. M., Frodl, T., Frouin, V., Garavan, H., Giddaluru, S., Glahn, D. C., Godlewska, B., Goldstein, R. Z., Gollub, R. L., Grabe, H. J., Grimm, O., Gruber, O., Guadalupe, T., Gur, R. E., Gur, R. C., Göring, H. H. H., Hagenaars, S., Hajek, T., Hall, G. B., Hall, J., Hardy, J., Hartman, C. A., Hass, J., Hatton, S. N., Haukvik, U. K., Hegenscheid, K., Heinz, A., Hickie, I. B., Ho, B.-C., Hoehn, D., Hoekstra, P. J., Hollinshead, M., Holmes, A. J., Homuth, G., Hoogman, M., Hong, L. E., Hosten, N., Hottenga, J.-J., Hulshoff Pol, H. E., Hwang, K. S., Jack, C. R., Jenkinson, M., Johnston, C., Jönsson, E. G., Kahn, R. S., Kasperaviciute, D., Kelly, S., Kim, S., Kochunov, P., Koenders, L., Krämer, B., Kwok, J. B. J., Lagopoulos, J., Laje, G., Landen, M., Landman, B. A., Lauriello, J., Lawrie, S. M., Lee, P. H., Le Hellard, S., Lemaître, H., Leonardo, C. D., Li, C.-S., Liberg, B., Liewald, D. C., Liu, X., Lopez, L. M., Loth, E., Lourdusamy, A., Luciano, M., Macciardi, F., Machielsen, M. W. J., Macqueen, G. M., Malt, U. F., Mandl, R., Manoach, D. S., Martinot, J.-L., Matarin, M., Mather, K. A., Mattheisen, M., Mattingsdal, M., Meyer-Lindenberg, A., McDonald, C., McIntosh, A. M., McMahon, F. J., McMahon, K. L., Meisenzahl, E., Melle, I., Milaneschi, Y., Mohnke, S., Montgomery, G. W., Morris, D. W., Moses, E. K., Mueller, B. A., Muñoz Maniega, S., Mühleisen, T. W., Müller-Myhsok, B., Mwangi, B., Nauck, M., Nho, K., Nichols, T. E., Nilsson, L.-G., Nugent, A. C., Nyberg, L., Olvera, R. L., Oosterlaan, J., Ophoff, R. A., Pandolfo, M., Papalampropoulou-Tsiridou, M., Papmeyer, M., Paus, T., Pausova, Z., Pearlson, G. D., Penninx, B. W., Peterson, C. P., Pfennig, A., Phillips, M., Pike, G. B., Poline, J.-B., Potkin, S. G., Pütz, B., Ramasamy, A., Rasmussen, J., Rietschel, M., Rijpkema, M., Risacher, S. L., Roffman, J. L., Roiz-Santiañez, R., Romanczuk-Seiferth, N., Rose, E. J., Royle, N. A., Rujescu, D., Ryten, M., Sachdev, P. S., Salami, A., Satterthwaite, T. D., Savitz, J., Saykin, A. J., Scanlon, C., Schmaal, L., Schnack, H. G., Schork, A. J., Schulz, S. C., Schür, R., Seidman, L., Shen, L., Shoemaker, J. M., Simmons, A., Sisodiya, S. M., Smith, C., Smoller, J. W., Soares, J. C., Sponheim, S. R., Sprooten, E., Starr, J. M., Steen, V. M., Strakowski, S., Strike, L., Sussmann, J., Sämann, P. G., Teumer, A., Toga, A. W., Tordesillas-Gutierrez, D., Trabzuni, D., Trost, S., Turner, J., Van den Heuvel, M., van der Wee, N. J., van Eijk, K., van Erp, T. G. M., van Haren, N. E. M., van 't Ent, D., van Tol, M.-J., Valdés Hernández, M. C., Veltman, D. J., Versace, A., Völzke, H., Walker, R., Walter, H., Wang, L., Wardlaw, J. M., Weale, M. E., Weiner, M. W., Wen, W., Westlye, L. T., Whalley, H. C., Whelan, C. D., White, T., Winkler, A. M.,

Wittfeld, K., Woldehawariat, G., Wolf, C., Zilles, D., Zwiers, M. P., Thalamuthu, A., Schofield, P. R., Freimer, N. B., Lawrence, N. S., Drevets, W., and Alzheimer's Disease Neuroimaging Initiative, EPIGEN Consortium, IMAGEN Consortium, Saguenay Youth Study (SYS) Group (2014). The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior*, 8(2):153–182.

Tolin, D. F. and Foa, E. B. (2008). Sex differences in trauma and posttraumatic stress disorder: a quantitative review of 25 years of research. ISBN: 1942-969X Publisher: Educational Publishing Foundation.

Tompson, S. H., Falk, E. B., Vettel, J. M., and Bassett, D. S. (2018). Network Approaches to Understand Individual Differences in Brain Connectivity: Opportunities for Personality Neuroscience. *Personality Neuroscience*, 1:e5.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, 17(6):520–525. Publisher: Oxford Academic.

Tuma, J. M. and Appelbaum, A. S. (1980). Reliability and Practice Effects of Wisc-R Iq Estimates in a Normal Population. *Educational and Psychological Measurement*, 40(3):671–678. Publisher: SAGE Publications Inc.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320. Conference Name: IEEE Transactions on Medical Imaging.

Vaccarino, V. and Bremner, J. D. (2017). Behavioral, emotional and neurobiological determinants of coronary heart disease risk in women. *Neuroscience and Biobehavioral Reviews*, 74(Pt B):297–309.

Valenza, G., Nardelli, M., Lanatà, A., Gentili, C., Bertschy, G., Paradiso, R., and Scilingo, E. P. (2014). Wearable Monitoring for Mood Recognition in Bipolar Disorder Based on History-Dependent Long-Term Heart Rate Variability Analysis. *IEEE Journal of Biomedical and Health Informatics*, 18(5):1625–1635.

Van Horn, J. D., Grafton, S. T., and Miller, M. B. (2008). Individual Variability in Brain Activity: A Nuisance or an Opportunity? *Brain Imaging and Behavior*, 2(4):327–334.

Van Horn, J. D. and Toga, A. W. (2009). Multisite neuroimaging trials. *Current Opinion in Neurology*, 22(4):370–378.

van Wingen, G. A., Ossewaarde, L., Bäckström, T., Hermans, E. J., and Fernández, G. (2011). Gonadal hormone regulation of the emotion circuitry in humans. *Neuroscience*, 191:38–45.

Vanderwal, T., Eilbott, J., Finn, E. S., Craddock, R. C., Turnbull, A., and Castellanos, F. X. (2017). Individual differences in functional connectivity during naturalistic viewing conditions. *NeuroImage*, 157:521–530.

Vedhara, K., Hyde, J., Gilchrist, I. D., Tytherleigh, M., and Plummer, S. (2000). Acute stress, memory, attention and cortisol. *Psychoneuroendocrinology*, 25(6):535–549.

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, 18(2):332–339. Place: US Publisher: American Psychological Association.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., and van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272. Number: 3 Publisher: Nature Publishing Group.

von Stumm, S. (2017). Socioeconomic status amplifies the achievement gap throughout compulsory education independent of intelligence. *Intelligence*, 60:57–62.

Walker, A. J., Batchelor, J., and Shores, A. (2009). Effects of education and cultural background on performance on WAIS-III, WMS-III, WAIS-R and WMS-R measures: Systematic review. *Australian Psychologist*, 44(4):216–223. _eprint: https://aps.onlinelibrary.wiley.com/doi/pdf/10.1080/00050060902833469.

Wang, A. C. J., Hara, Y., Janssen, W. G. M., Rapp, P. R., and Morrison, J. H. (2010). Synaptic estrogen receptor-alpha levels in prefrontal cortex in female rhesus monkeys and their correlation with cognitive performance. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(38):12770–6.

Wechsler, D., Coalson, D. L., and Raiford, S. E. (2008). *WAIS-IV: Wechsler adult intelligence scale*. Pearson San Antonio, TX.

Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist*, 44(2):98–104.

Weiss, L. G., Chen, H., Harris, J. G., Holdnack, J. A., and Saklofske, D. H. (2010). CHAPTER 4 - WAIS-IV Use in Societal Context. In Weiss, L. G., Saklofske, D. H., Coalson, D. L., and Raiford, S. E., editors, *WAIS-IV Clinical Use and Interpretation*, Practical Resources for the Mental Health Professional, pages 97–139. Academic Press, San Diego.

Weissman, M. M., Bland, R., Joyce, P. R., Newman, S., Wells, J. E., and Wittchen, H.-U. (1993). Sex differences in rates of depression: cross-national perspectives. *Journal of Affective Disorders*, 29(2):77–84.

Welton, T., Kent, D. A., Auer, D. P., and Dineen, R. A. (2014). Reproducibility of Graph-Theoretic Brain Network Metrics: A Systematic Review. *Brain Connectivity*, 5(4):193–202. Publisher: Mary Ann Liebert, Inc., publishers.

Whitfield-Gabrieli, S. and Ford, J. M. (2012). Default mode network activity and connectivity in psychopathology. *Annual Review of Clinical Psychology*, 8:49–76.

Wideman, L., Montgomery, M. M., Levine, B. J., Beynnon, B. D., and Shultz, S. J. (2013). Accuracy of Calendar-Based Methods for Assigning Menstrual Cycle Phase in Women. *Sports Health*, 5(2):143–149.

Willis, S. K., Hatch, E. E., and Wise, L. A. (2019). Sleep and female reproduction. *Current Opinion in Obstetrics and Gynecology*, 31(4):222–227.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, 92:381–397.

Wirth, M. M., Meier, E. A., Fredrickson, B. L., and Schultheiss, O. C. (2007). Relationship between salivary cortisol and progesterone levels in humans. *Biological Psychology*, 74(1):104–107.

Wirth, M. M., Scherer, S. M., Hoks, R. M., and Abercrombie, H. C. (2011). The effect of cortisol on emotional responses depends on order of cortisol and placebo administration in a within-subjects design. *Psychoneuroendocrinology*, 36(7):945–954.

Wong, S. W. H., Xue, G., and Bechara, A. (2011). Integrating fMRI with psychophysiological measurements in the study of decision making. *Journal of Neuroscience, Psychology, and Economics*, 4(2):85–94.

Woods, N. F., Lentz, M. J., Mitchell, E. S., Heitkemper, M., Shaver, J., and Henker, R. (1998). Perceived stress, physiologic stress arousal, and premenstrual symptoms: Group differences and intra-individual patterns. *Research in Nursing & Health*, 21(6):511–523. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291098-240X%28199812%2921%3A6%3C511%3A%3AAID-NUR5%3E3.0.CO%3B2-W.

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45(1 Suppl):S173–186.

Wu, K., Taki, Y., Sato, K., Hashizume, H., Sassa, Y., Takeuchi, H., Thyreau, B., He, Y., Evans, A. C., Li, X., Kawashima, R., and Fukuda, H. (2013). Topological Organization of Functional Brain Networks in Healthy Children: Differences in Relation to Age, Sex, and Intelligence. *PLOS ONE*, 8(2):e55347.

Xiao, E. and Ferin, M. (1997). Stress-related disturbances of the menstrual cycle. *Annals of Medicine*, 29(3):215–219.

Yang, P., Jong, Y.-J., Hsu, H.-Y., and Lung, F.-W. (2011). Role of assessment tests in the stability of intelligence scoring of pre-school children with uneven/delayed cognitive profile. *Journal of Intellectual Disability Research*, 55(5):453–461. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2788.2011.01396.x.

Yarkoni, T. (2009). Big Correlations in Little Studies: Inflated fMRI Correlations Reflect Low Statistical Power—Commentary on Vul et al. (2009). *Perspectives on Psychological Science*, 4(3):294–298. Publisher: SAGE Publications Inc.

Young, E. A., Midgley, A. R., Carlson, N. E., and Brown, M. B. (2000). Alteration in the hypothalamic-pituitary-ovarian axis in depressed women. *Archives of general psychiatry*, 57(12):1157–1162. ISBN: 0003-990X Publisher: American Medical Association.

Yu, H., McCoach, D., Gottfried, A., and Gottfried, A. (2018). Stability of intelligence from infancy through adolescence: An autoregressive latent variable model. *Intelligence*, 69:8–15.

Zax, J. S. and Rees, D. I. (2002). IQ, Academic Performance, Environment, and Earnings. *The Review of Economics and Statistics*, 84(4):600–616.

Zhang, B. and Wing, Y.-K. (2006). Sex Differences in Insomnia: A Meta-Analysis. *Sleep*, 29(1):85–93.

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, 20(1):45–57.

Zhao, W., Palmer, C. E., Thompson, W. K., Chaarani, B., Garavan, H. P., Casey, B. J., Jernigan, T. L., Dale, A. M., and Fan, C. C. (2021). Individual Differences in Cognitive Performance Are Better Predicted by Global Rather Than Localized BOLD Activity Patterns Across the Cortex. *Cerebral Cortex*, 31(3):1478–1488.

Zilles, K. and Amunts, K. (2013). Individual variability is not noise. *Trends in cognitive sciences*, 17(4):153–155. ISBN: 1364-6613 Publisher: Elsevier.

Zuffianò, A., Alessandri, G., Gerbino, M., Luengo Kanacri, B. P., Di Giunta, L., Milioni, M., and Caprara, G. V. (2013). Academic achievement: The unique contribution of self-efficacy beliefs in self-regulated learning beyond intelligence, personality traits, and self-esteem. *Learning and Individual Differences*, 23:158–162.

Šidák, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*, 62(318):626–633.

VITA

KATHERINE L. BOTTENHORN

November 3, 1992	Born, Rockville, Maryland
2015	B.A., Chemistry Auburn University Auburn, AL
2015	B.A., Psychology Auburn University Auburn, AL
2018	M.S., Psychology Florida International University Miami, Florida

PUBLICATIONS AND PRESENTATIONS

*Gau R, *Noble S, *Heuer K, *Bottenhorn, KL, ... the Brainhack Community. (2021). Brainhack: developing a culture of open, inclusive, community-driven neuroscience. *Neuron.*

Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M, . . . , Bottenhorn KL, . . . Schonberg T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature.*

*Bielczyk NZ, *Ando A, *Badhwar A, *Caldinell C, *Gao M, *Haugg A, *Hernandez LM, *Ito K, *Kessler D, *Lurie D, *Makary MM, *Nikolaidis A, * Veldsman M,. . . Bottenhorn KL, . . . OHBM Student and Postdoc Special Interest Group (2020). Effective self-management for early career researchers in the natural and life sciences. *Neuron.*

*Waller R, *Hawes SW, Byrd AL, Dick AS, Sutherland MT, Riedel MC, Tobia MJ, Bottenhorn KL, Laird AR, Gonzalez R. (2020). Disruptive Behavior Problems, Callous-Unemotional Traits, and Regional Gray Matter Volume in the ABCD Study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging.*

Flannery JS, Riedel MC, Bottenhorn KL Poudel R, Salo T, Hill-Bowen LD, Laird AR, Sutherland MT (2020). Meta-analytic clustering dissociates brain activity and behavior profiles across reward processing paradigms. *Cogn Affect Behav Neurosci.*

*Gonzalez AA, *Bottenhorn KL, Bartley JE, Hayes T, Riedel MC, Salo T, ... Laird AR (2019). Sex differences in brain correlates of STEM anxiety. *Npj Science of Learning*.

Bartley JE, Riedel MC, Salo T, Boeving ER, Bottenhorn KL, Bravo EI, ... Laird AR (2019). Brain activity links performance in science reasoning with conceptual approach. *Npj Science of Learning*.
Bartley JE, Riedel MC, Salo T, Boeving ER, Bottenhorn KL, Bravo EI, ... Laird AR (2018). Meta-analytic evidence for a core problem solving network across multiple representational domains. *Neuroscience and Biobehavioral Reviews*.

Brewe E, Bartley JE, Riedel MC, Sawtelle V, Salo T, Boeving ER, Bravo EI, Odean R, Nazareth A, Bottenhorn KL, Laird RW, Sutherland MT, Pruden SM, Laird AR (2018). Toward a neurobiological basis for understanding learning in University Modeling Instruction physics courses. *Frontiers in ICT*.

Bottenhorn KL, Flannery JS, Boeving ER, Riedel MC, Eickhoff SB, Sutherland MT, Laird AR (2018). Cooperating yet distinct brain networks engaged during naturalistic paradigms: A meta-analysis of functional neuroimaging data. *Network Neuroscience*.

Robinson JL, Barron DS, Kirby LAJ, Bottenhorn KL, Hill AC, Murphy JE, ... Fox PT. (2015). Neurofunctional topography of the human hippocampus. *Human Brain Mapping*, 36(12), 5018–37.

*Indicates authors contributed equally to this work.