

Big Data, Small Data: Considerations for Modeling and Inference in Neuroimaging

John Darrell Van Horn, Ph.D., M.Eng.

Department of Psychology and
School of Data Science

University of Virginia



Introduction

Find Counselling (City or Postcode)

Damian K. F. Pang M.Sc.
Consciousness and Beyond

NEUROSCIENCE

The Staggering Complexity of the Human Brain

Why our brains are the most complex structures in the known universe.

Posted September 2, 2023 | Reviewed by Ray Parker



KEY POINTS

- The human brain has been described as the most complex structure in the known universe.
- The brain contains around 86 billion neurons, 85 billion other cells, and over 100 trillion connections.
- Several hundred-million-dollar research projects failed to fully map the structure of the brain.
- The function of the brain is even more complex than its structure with consciousness as the greatest mystery.



While the Milky Way has more stars than the brain

The human brain has been described as the most complex known structure in the universe (Dolan, 2007). The numbers are indeed staggering: Ingenious research methods that examined the almost 1,000 brain regions, and extrapolated the expected

Sign in



Enter search keywords

News Features Newsletters Podcasts Video Comment Culture Crosswords | This week's magazine

Health Space Physics Technology Environment Mind Humans Life Mathematics Chemistry Earth Society

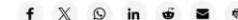
Mind

Is the human brain really the most complex object in the universe?

There are 86 billion neurons in your brain, roughly the same number as there are galaxies in the observable universe. Whether the mind is more complex than the cosmos, however, is up for debate

By Thomas Lewton

21 February 2024



The Crescent Nebula: more complex than the human brain?

Reinhold Wittich/Stocktrek Images/Alamy

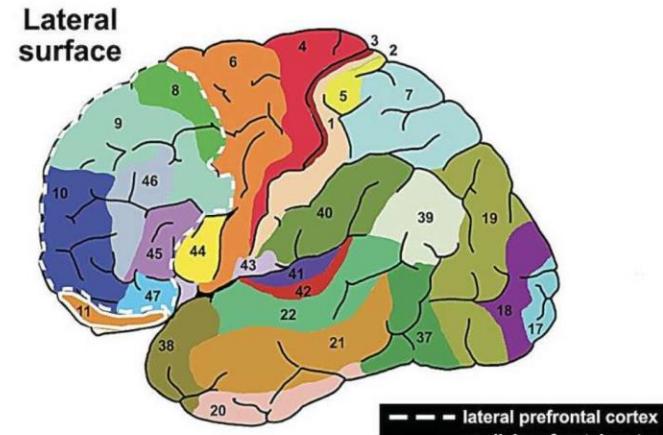
BACK in 2012, neuroscientist Christof Koch wrote in his book *Consciousness: Confessions of a romantic reductionist* that the human brain is “the most complex object in the known universe”. Given that there are about 86 billion neurons in a brain, connected up in ways that we are only beginning to unravel, this seems intuitive. But when I put it to David Wolpert at the Santa Fe Institute in New Mexico – created in the 1980s as a hub for the budding field of complexity science – he doesn’t see it that way. “It’s almost farcical to entertain that we are the most complex system in the universe,” he says. “The question is actually wrongheaded.”

Home / Medicine / Neurology /

The brain is the most complicated object in the universe—scientists seek to decode it and read people's minds

Updated on 2024-06-26 10:47:45 4 2

by Nicholas J. Kelley, Stephanie Sheir and Timo Istace, The Conversation

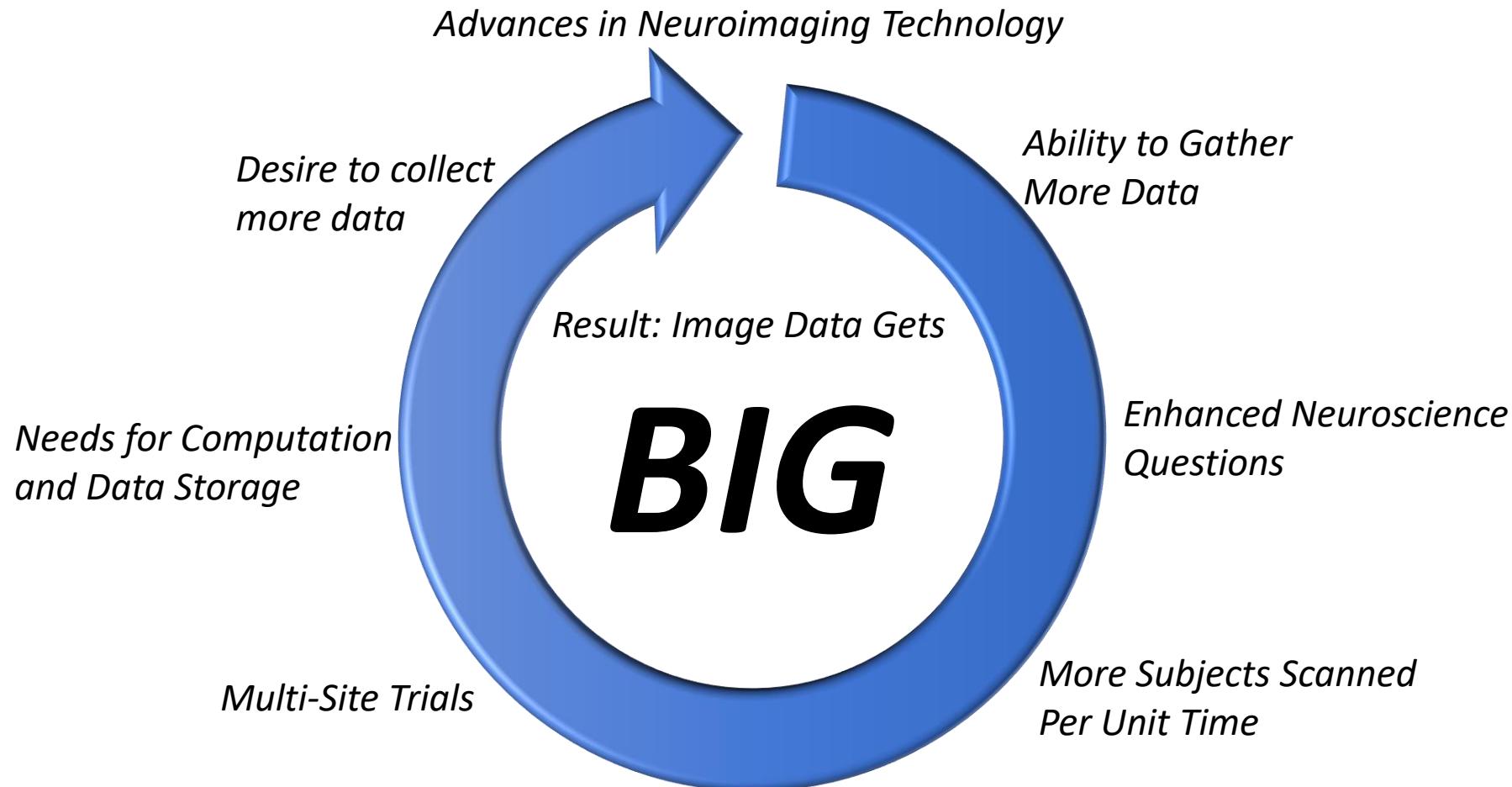


Brodman's brain map. Credit: Vysha/Wikimedia, CC BY-SA

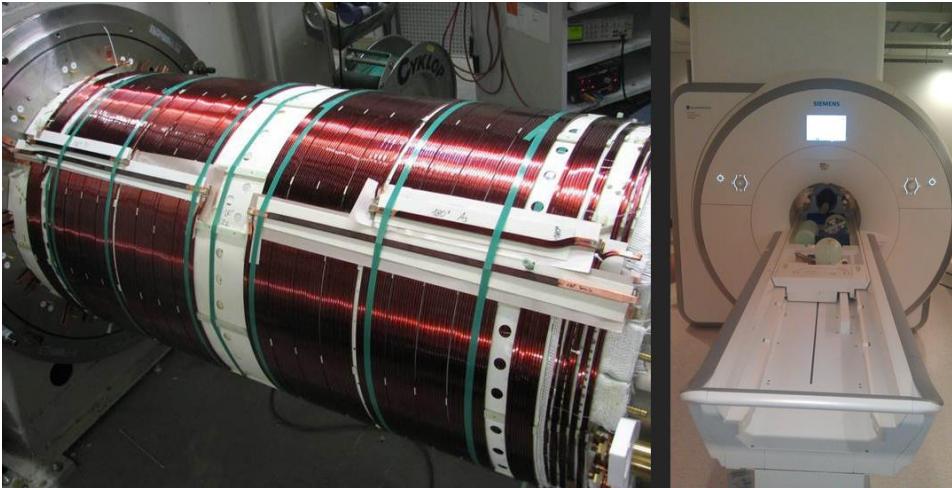
In the middle of 2023, a study conducted by the HuthLab at the University of Texas sent shockwaves through the realms of neuroscience and technology. For the first time, the thoughts and impressions of people unable to communicate with the outside world were translated into continuous natural language, using a combination of artificial intelligence (AI) and brain imaging technology.

This is the closest science has yet come to reading someone's mind. While advances in neuroimaging over the past two decades have enabled non-responsive and minimally conscious patients to control a computer cursor with their brain, HuthLab's research is a significant step closer towards accessing people's actual thoughts. As Alexander Huth, the neuroscientist who co-led the research, told the New York Times:

A Cycle of Data Growth in Neuroimaging



Neuroimaging as “Big Data”



MGH CONNECTOM 3T Scanner (Human Connectome Project)

“The methods by which these data are obtained are themselves contributing to this growth, involving finer spatial and temporal resolution as MR physicists push the limits of what is possible and as brain scientists then rush to meet those limits. It is safe to say that human neuroimaging is now, officially, a ‘big data’ science.”

Van Horn and Toga, 2014, *Brain Imaging and Behavior*

Fun Fact:



**Data expands to fill the space available for storage.
- *Parkinson's Law***

Jansen, Peter (2008). *IT-Service-Management Volgens ITIL*. Derde Editie. [Pearson Education](#). p. 179. [ISBN 978-90-430-1323-9](#).

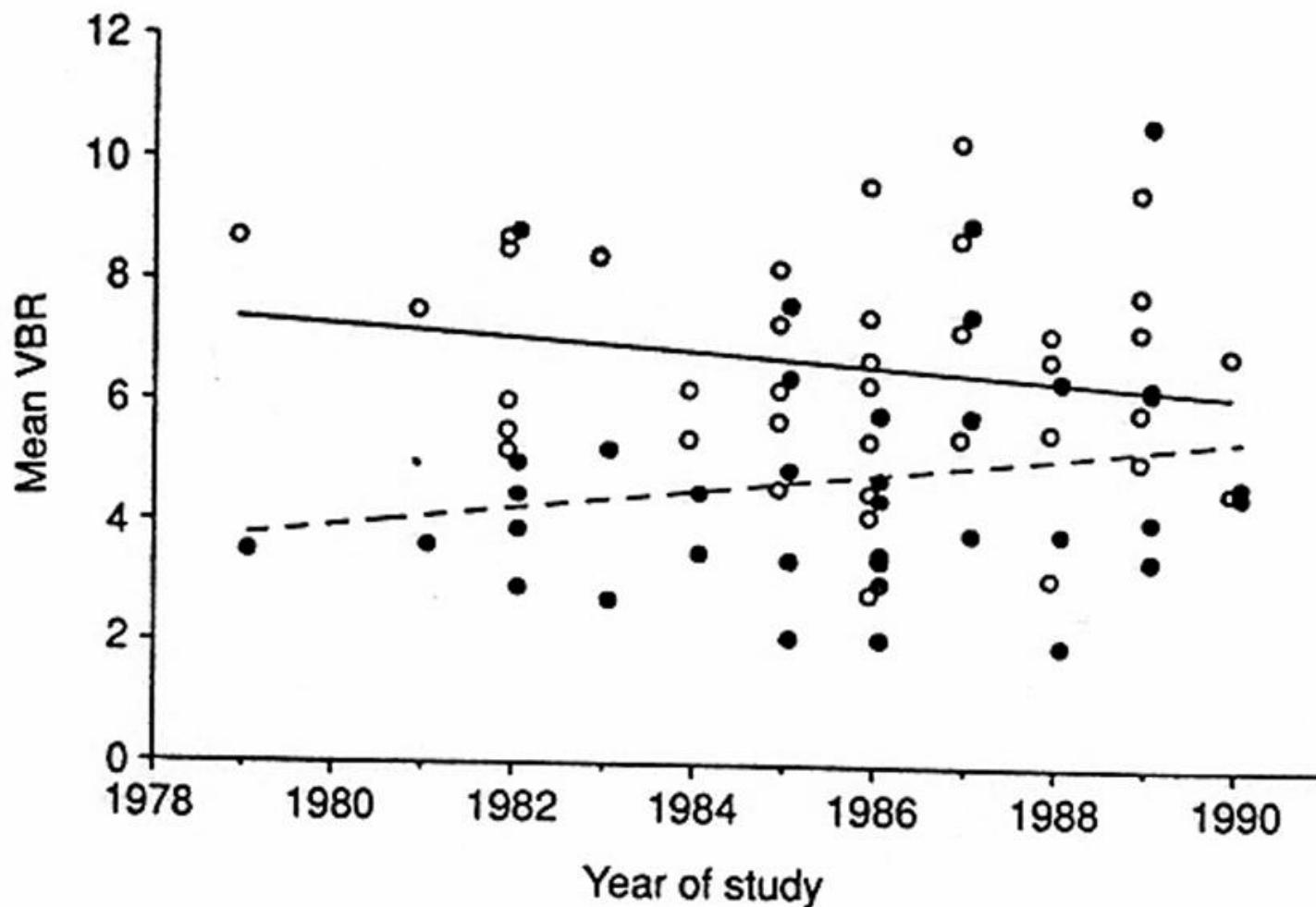
- Fish do not grow to the size of their fish tank.
- Fish in a natural setting continue to grow throughout their lifetimes.
- It is the tank, however, that constrains the size of the fish.

Shared Data Resources



Ventricular Enlargement in Schizophrenia A Meta-analysis of Studies of the Ventricle:Brain Ratio (VBR)

J. D. VAN HORN and I. C. McMANUS



Study factors influencing ventricular enlargement in schizophrenia: A 20 year follow-up meta-analysis

Angelo Sayo, Robin G. Jennings, John Darrell Van Horn *

Laboratory of Neuro Imaging (LONI), Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, 635 Charles E. Young Drive SW, Suite 225, Los Angeles, CA 90095-7334, USA

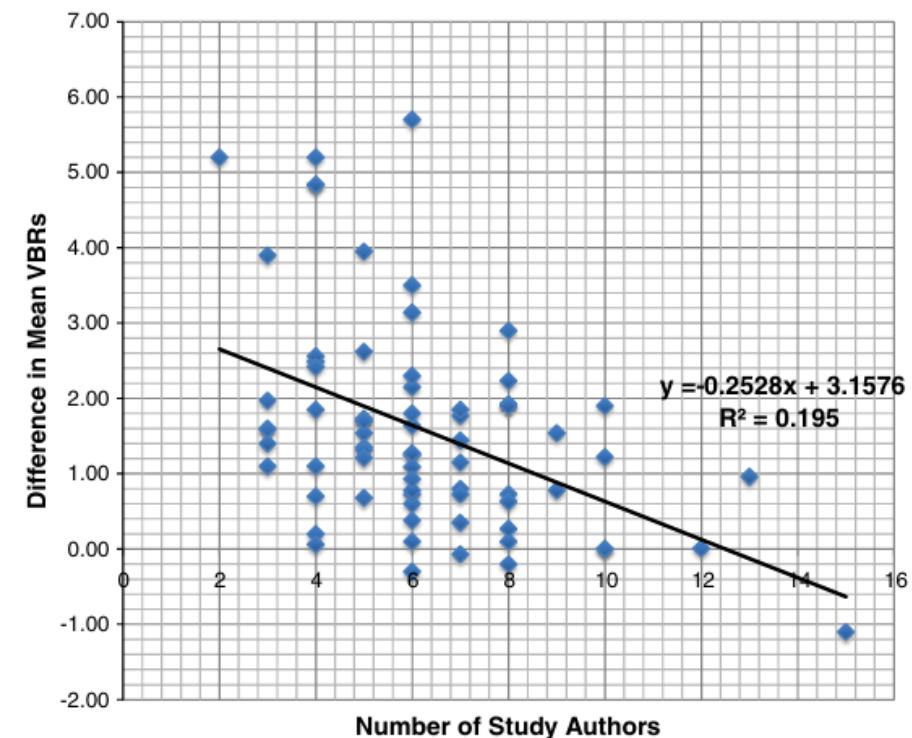
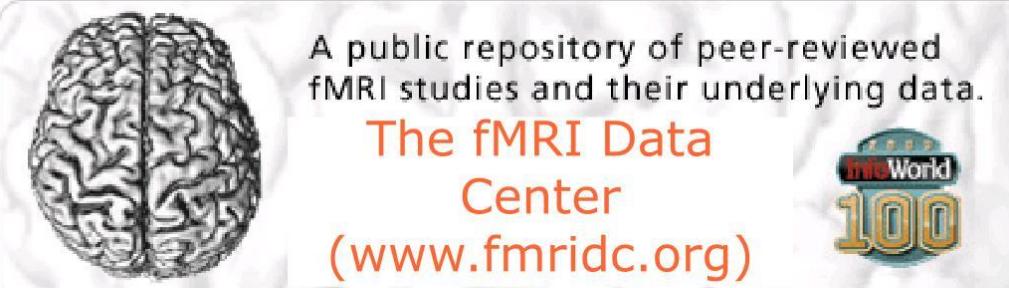


Fig. 2. The number of co-authors effect on the difference between schizophrenic and control sample mean VBRs. The correlation between the number of study authors and the VBR difference was $r = 0.442$, $p < 0.001$.

The screenshot shows the fMRI Data Center website (fMRI-DC) as it appeared in Microsoft Internet Explorer. At the top, there's a standard Windows-style menu bar with File, Edit, View, Favorites, Tools, Help, and a toolbar with icons for Back, Forward, Stop, Refresh, Home, Search, Favorites, and Print. Below the toolbar is a search bar with 'Google' and a search button, followed by a link to 'SnagIt'. The address bar shows the URL 'http://www.fmridc.org/fmridc'. The main content area has a header 'The fMRI Data Center fmRIDC' with a search bar for 'fmRIDC Database'. To the right are 'My Account' and 'Request List (Empty)' buttons. On the left is a sidebar with links for HOME, DATABASE, SUBMISSIONS, RESOURCES, DATA MANAGEMENT TOOL, HELP, and ABOUT US. Below this are links for Sitemap and Contact Us. The main content includes a large image of a brain, text about being a public repository of peer-reviewed fMRI studies, funding from the National Science Foundation, W.M. Keck Foundation, National Institutes of Mental Health, and Sun Center of Excellence for Neuroscience, and a 'InfoWorld 100' logo. It also features sections for INFORMATION (How do I get started?, Q&A about fmRIDC, Available Datasets), FMRI-DC NEWS (## FMRI-DC SYSTEM MAINTENANCE TODAY, NIH Neuroscience Blueprint seeks to advance cooperative neuroscience activities across NIH institutes and investigators, MRI Safety: Not to be taken for granted, 2004 Summer Workshop Video Presentations Now Online!, The New Perspectives in fMRI Research Award), and PROJECT STATISTICS (Registered users: 1630, Datasets available: 97, Dataset requests: 1531). There are also links for Special Collections (Data from special or rare populations of subjects), Summer Workshops, and the fMRI-DC Data Management Tool (Written in Java). At the bottom, there's a footer with copyright information and links to Dartmouth College Disclaimer and Privacy Statement, and credits to Paradigm Consulting Co. for web site design and development.



A public repository of peer-reviewed fMRI studies and their underlying data.

The fMRI Data Center

(www.fmridc.org)



- A typical study comprises
3 groups,
20 subjects/group,
5 runs/subject,
300 volumes/run
→ 90,000 volumes, 60 GB raw
→ 1.2 million files processed
 - 100s of such studies in total

PERSPECTIVES

OPINION

Databasing fMRI studies — towards a 'discovery science' of brain function

John D. Van Horn and Michael S. Gazzaniga

Enormous progress has been made over the past decades in the development of neuroimaging technologies to study the living brain function. But as we move into the genomic era, much of the raw functional imaging data that have been described in the literature have not been made available to other researchers. The data are scattered across numerous functional imaging data from peer-reviewed publications, making it very difficult to assess the validity of previous findings, test new methods and generate new hypotheses. This bold proposal is to develop a common language of understanding of complex cognitive processes and their study in the study of neuroscience.

Everybody agrees that *Genbank* (from The National Center for Biotechnology Information) is a great success. It has become the standard for sharing genetic information — stop shopping for genetic information — the central source for genetics, the study of genes and their products. It contains information on the genomes of assured organisms from numerous viewpoints, including *diagrammatic* representations of genes on disease, and on taxonomic, and these data have been used to support medical research, in research, in clinical applications and in education¹. So, just as genecists have gone beyond simply considering gene sequence and function, we believe that the complexity of brain function, neuroscience must reveal hidden patterns of brain function that can be used to support medical research, in research, in clinical applications and in education². Because, say, neuroscientists did not have access to the same kind of evidence to make progress towards the larger goal of understanding the brain and its variants functionally.

Genbank has also ushered into a new way of doing research. Genome information is now freely available to all researchers — for example, the *Handbook of Chemistry & Physics* is now freely available online. Once data is assembled, the semantic content of the information is still largely unknown. In order to make the most of this information — the identification of data-mining tools, and through homology searching, the identification of

coding regions and genes, and so on, that information about gene function will be discovered. And this led to the emergence of *Gene Ontology*, which attempts to categorize every gene and its products according to an object and you define all its elements and you create a database of information about that object. This is what we mean by hypothesis-driven view.

It is against this backdrop that we have launched a new initiative to build a database for functional magnetic resonance imaging (fMRI) studies. The first order of business is to make sure that the data that will allow scientists easy access to raw data from published, peer-reviewed studies. This project, called *fmriretrieval.org* (fMRI) and the *fMRI Data Center* (fMRIdc) is up and running. It is currently in beta testing, and both challenging and exciting — is to enter the raw data into a database that will allow scientists to search for specific and voluminous fMRI data. This "data-mining" goal involves the use of information gathering techniques to identify and extract the huge amount of information that is contained in the fMRI images to give efficient and effective answers to questions such as "all study data that are close to the following?" or "all study data that are similar to the following?". So, just as genecists have gone beyond simply considering gene sequence and function, we believe that the complexity of brain function, neuroscience must reveal hidden patterns of brain function that can be used to support medical research, in research, in clinical applications and in education². Because, say, neuroscientists did not have access to the same kind of evidence to make progress towards the larger goal of understanding the brain and its variants functionally.

current effort we are clearly dealing with one, small aspect of neuroscience research. However, this is a critical aspect of our system so that, in the future, it can exchange information with other database systems. This will allow us to have a perhaps single resource for all of neuroscience. We review our progress in the following sections.

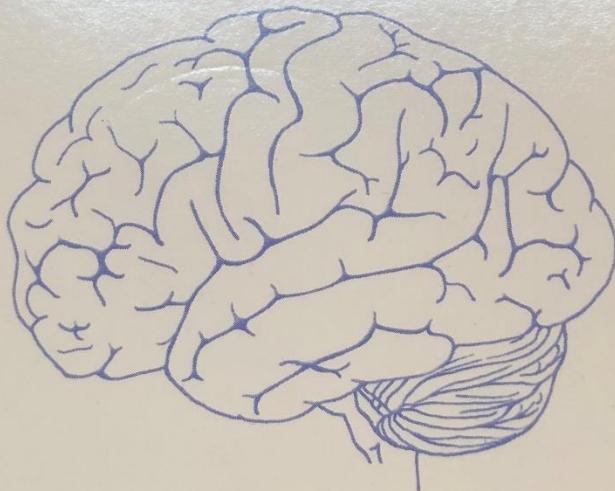
Progress to date and current usage

Data challenges are faced at building a database for the analysis of the multitude of the fMRI Data Center. We have discussed these technical issues in greater detail elsewhere³. In brief, the main challenge is how to store the data in a way that can be easily retrieved, given the amount that has recently gone to print⁴.

Subject selection and protection. In compliance with 7 CFR 17.2000, the *Code of Federal Regulations* human subject protection, we had to address the requirement that any and all information that can be used to identify a subject must be removed from the data, while maintaining its experimental integrity. This is accomplished in two ways. First, the data are de-identified using authors' identifiers before submitting the data to the *fMRI Data Center*. Then the *fMRI Data Center* identifies that the subjects may have been missed by the authors, while maintaining the original data and its subject descriptions. Beyond the obvious need to preserve the anonymity of subject data, neuroscientists must also consider the fact that reconstructed, high-resolution images of the brain in principle, can be used as identifiers. So, the *fMRI Data Center* must ensure that data can be reconstructed in three dimensions, a surface fit to the data and the context of the study. Second, the high-resolution images are reconstructed to ensure that the possibility that the identity of the subject may be determined by three-dimensional reconstruction of the data is removed. To aide the United States are encouraged to consider their functional imaging-data to the best of their ability, and to encourage others, carefully considering their country's policy on the sharing of data from human subjects.

Content and organization. Ensuring the quality of the content of this neuroimaging data is a critical concern. Communication between the *Data Center* personnel and the authors of the study, and the *fMRI Data Center* is critical to ensure that they own their study, thereby avoiding the possibility of misinterpretation by the *Data Center* personnel. In addition, the authors ask questions about appropriate parameter values that have been entered by the

The fMRI Data Center



DARTMOUTH COLLEGE

The fMRI Data Center • 6162 Moore Hall • Hanover, NH 03755-3569

nature neuroscience

A debate over fMRI data sharing



news

Prospect of data sharing gives brain mappers a headache

Peter Aldhous

An attempt to encourage the sharing of brain images has ignited a fierce controversy about neuroscientists' right to control their data. Some specialists in functional magnetic resonance imaging (fMRI) fear that the National fMRI Data Center, newly established at Dartmouth College in Hanover, New Hampshire, will take over control of their results. Others are concerned that the centre's operation might breach patient confidentiality.

Blood flow and energy consumption in the brain are mapped by fMRI, showing which brain structures are active when particular cognitive tasks are performed. But behind each image lies a mass of raw nuclear magnetic resonance data, statistical analyses and detailed anatomical records.

Having access to these underlying data could help fMRI specialists interpret one another's work. But brain mappers are divided over the merit of sharing primary data, and many of them object to the way in which Michael Gazzaniga, director of the National fMRI Data Center, has set about the task.

In mid-June, Gazzaniga wrote to fMRI specialists who had submitted manuscripts to the *Journal of Cognitive Neuroscience*. The journal would not accept fMRI papers to the data from the centre, he said.

Having been invited to speak at Neuroscience, Greek participants in the letter would add to the number of journals that would accept manuscripts, many of which was its implication.

This prompted a response from the University of Isabel University in Nashua, New Hampshire, dozen fMRI specialists and financial backer — and the editors of the letter to science, plus those on the publication.

NATURE | VOL 406 | 3 AUGUST 2000

from their own work," the letter argued. "The archive would allow scientists to spot questionable images more easily, he says.

An archive would allow scientists to spot questionable images more easily, he says.

nature

3 August 2000 Volume 406 Issue no 6795

Whose scans are they, anyway?

Raw data are useful for researchers wishing to replicate the results of an experiment. Care needs to be taken when, as with brain-imaging measurements, such data can be misused or misinterpreted.

Like motherhood and apple pie, the concept of sharing primary data is widely recognized among scientists as a good thing. The difficulty lies in putting this laudable aim into practice — and

from results they have worked hard to acquire. At the same time, they must devise robust procedures to protect confidentiality, for carelessly archived fMRI data could allow experimental subjects to be identified.

(which techniques for

dized. For instance,

asymmetries in brain

and

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

for delaying discussions

provides a powerfu

to artefact with

als. With different

serve their interests,

it's claims. And as

is, many of whom

ill grow. Ultimately,

the best solution

ing (OHBM) has

is watershed discus-

sions. For instance,

Creating a database

escriptions can be in

Publication Bias in Neuroimaging Research: Implications for Meta-Analyses

Robin G. Jennings · John D. Van Horn

Published online: 4 June 2011
© Springer Science+Business Media, LLC (outside the USA) 2011

Abstract Neuroimaging and the neurosciences have made notable advances in sharing activation results through detailed databases, making meta-analysis of the published research faster and easier. However, the effect of publication bias in these fields has not been previously addressed or accounted for in the developed meta-analytic methods. In this article, we examine publication bias in functional magnetic resonance imaging (fMRI) for tasks involving working memory in the frontal lobes (Brodmann Areas 4, 6, 8, 9, 10, 37, 45, 46, and 47). Seventy-four studies were selected from the literature and the effect of publication bias was examined using a number of regression-based techniques. Pearson's r correlation coefficient and Cohen's d effect size estimates were computed for the activation in each study and compared to the study sample size using Egger's regression, Macaskill's regression, and the 'Trim and Fill' method. Evidence for publication bias was identified in this body of literature ($p < 0.01$ for each test), generally, though was neither task- nor sub-region-dependent. While we focused our analysis on this subgroup of brain mapping studies, we believe our findings generalize to the brain imaging literature as a whole and databases

seeking to curate their collective results. While neuroimaging databases of summary effects are of enormous value to the community, the potential publication bias should be considered when performing meta-analyses based on database contents.

Keywords Brain imaging · fMRI · Databases · Meta-analysis · Publication bias

Introduction

In recent years meta-analyses have become increasingly popular in neuroimaging as large databases of structural and functional brain imaging data have been created and employed to aggregate results from across individual studies (Murphy et al. 2003; Neumann et al. 2008; Fusar-Poli et al. 2009). Meta-analytic methods to examine these data have become increasingly refined (Turkeltaub et al. 2002), and these techniques are rapidly becoming particularly important tools for understanding fundamental questions underlying patterns of cognitively induced activity.

The development of highly detailed neuroimaging databases of published results has made quantitative assessment of the available research much easier (Fox et al. 2005), and the ability to pool studies and sample sizes to make inferences about functional brain activity has become increasingly valuable in diagnostics (Peyron et al. 2000). These resources provide a useful means for combining the results of studies in specific research domains and have offered a unique solution for examining variation in reported activation foci (Nielsen and Hansen 2002).

However, while meta-analyses of functional imaging studies may provide invaluable insights, caution must be taken due to the potential for bias in the current literature,

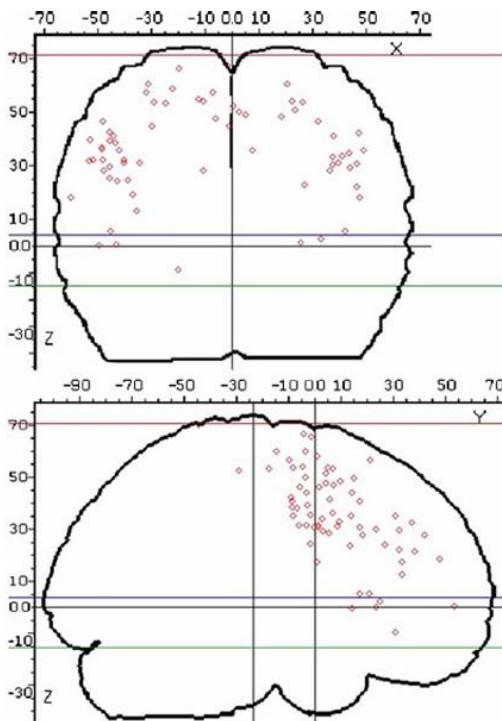


Fig. 1 Results for studies, plotted on a standard glass brain in Talairach space using BrainMap, showing each reported study local maxima located in the frontal lobes ($n=68$)

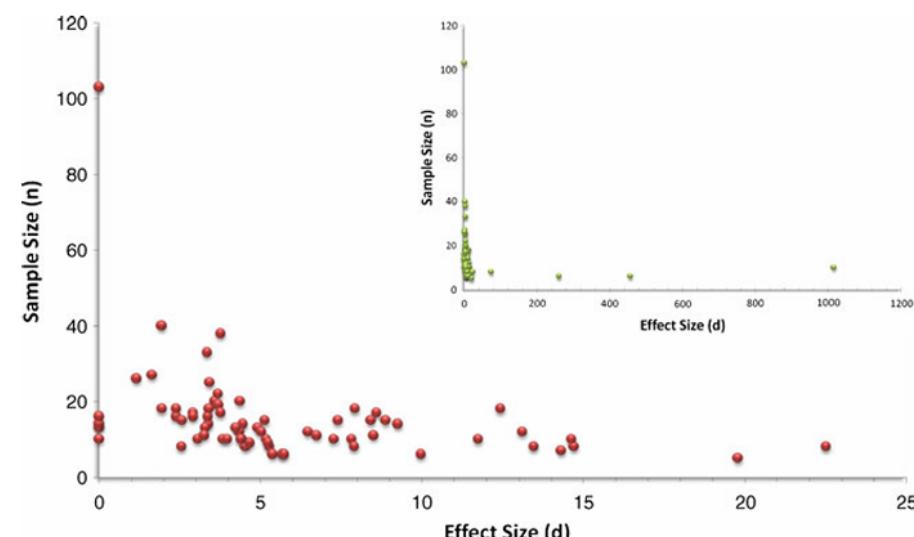


Fig. 3 Funnel plot of Cohen's d by sample size for studies without extreme values ($n=70$). While a 'large' Cohen's d value is usually $d>0.8$, most of our values are between 1 and 25, with funnel plot asymmetry due to the heavy right-tail evident here. **Inset:** Funnel plot of Cohen's d by sample size for each study ($n=74$), showing the four extreme outlier values

R. G. Jennings (✉)
Department of Biostatistics, University of California Los Angeles,
635 Charles Young Drive South, Suite 225,
Los Angeles, CA 90095, USA
e-mail: rgjennings@ucla.edu

R. G. Jennings · J. D. Van Horn
Laboratory of Neuro Imaging (LONI), Department of Neurology,
David Geffen School of Medicine,
University of California Los Angeles,
635 Charles Young Drive South, Suite 225,
Los Angeles, CA 90095, USA

J. D. Van Horn
e-mail: jvanhorn@loni.ucla.edu

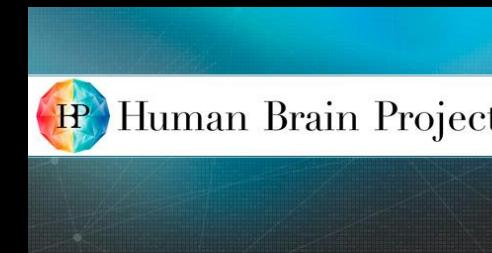
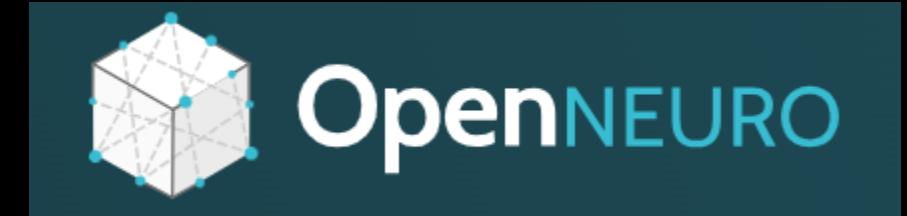
Sources of Publication Bias

- Rush to publish
 - Getting out some potentially interesting result before someone else
- Focusing on the least publishable unit
 - Quantity over quality
- Promotion and Tenure
 - Plumping up of one's CV in advance of P&T decisions
- Justification for a research funding proposal
 - To claim a preliminary result as a basis for getting a grant

Linden AH, Pollet TV, Hönekopp J. Publication bias in psychology: A closer look at the correlation between sample size and effect size. PLoS One. 2024 Feb 15;19(2):e0297075. doi: 10.1371/journal.pone.0297075. PMID: 38359021; PMCID: PMC10868788.

Kwee TC, Almaghrabi MT, Kwee RM. Scientific Fraud, Publication Bias, and Honorary Authorship in Nuclear Medicine. J Nucl Med. 2023 Feb;64(2):200-203. doi: 10.2967/jnumed.122.264679. Epub 2022 Sep 8. PMID: 36215567.

Online Brain Data Resources



Power Determined by Effect Size and N

“More subjects is good!”

- Achieve statistically significant activations
- Avoid publication biases
- Help ensure replication
- Optimize statistical power
- Publish in the top journals
- Etc
- Etc

A power calculation guide for fMRI studies

Jeanette A. Mumford

Department of Psychology, University of Texas at Austin, University Station A8000, Austin, TX 78712-0187, USA

In the past, power analyses were not that common for fMRI studies, but recent advances in power calculation techniques and software development are making power analyses much more accessible. As a result, power analyses are more commonly expected in grant applications proposing fMRI studies. Even though the software is somewhat automated, there are important decisions to be made when setting up and carrying out a power analysis. This guide provides tips on carrying out power analyses, including obtaining pilot data, defining a region of interest and other choices to help create reliable power calculations.



Keywords: functional magnetic resonance imaging, power analysis, fMRI

INTRODUCTION

When running a functional magnetic resonance experiment, we hope that our data has interest and, more importantly, that we detect this signal. The ability to detect referred to as statistical power and is 80% or higher. The interpretation of repeat our study 100 times, and the signal detected in 80 of the studies. Unlike other analysis needs to be performed prior to study planning tool. Most commonly, it is a grant proposals. Although different styles have existed since the early 2000s required lengthy simulations (Desmond complicated for a non-statistician to Mumford and Nichols, 2008). The recent age, fMRI Power (fmripower.org), allows (ROI) power calculations based on a simulation in Mumford and Nichols (2008). fMRI analysis software described in Joyce and Sibley for any investigator to perform power analysis are more common grant applications and can help reduce fMRI studies in the future.

The power analysis model described (2008) has the flexibility to calculate power for many subjects (even and how many) are presented in a run. Changing the generation of new first-level design automatically in a software package. Therefore power calculation as a function of the analysis assumes that the future data will contain the number of runs as the pilot analysis. This Matlab-based software uses the FSL (www.fmrib.ox.ac.uk/fsl) or SPSS software packages and automatically executes power calculation. fMRI Power can calculate two-sample and paired *t*-tests. The soft what type of analysis was originally run only need to specify the number of subjects.

Received: 13 April 2013; Accepted: 23 May 2013
Editorial Access publication: 26 May 2013
Correspondence should be addressed to Jeanette A. Mumford, Department of Psychology, University of Texas at Austin, University Station A8000 Austin, TX 78712-0187, USA.
© The Author (2013). Published by Oxford University Press.

RESEARCH ARTICLE

The relation between statistical power and inference in fMRI

Henk R. Creemers^{1*}, Tor D. Wager², Tal Yarkoni³

¹ Department of Clinical Psychology, University of Amsterdam, Amsterdam, Netherlands, ² Department of Psychology and Neuroscience, University of Colorado at Boulder, Boulder, Colorado, United States of America, ³ Department of Psychology, University of Texas at Austin, Austin, Texas, United States of America



The Statistical Analysis of fMRI Data

Martin A. Lindquist

OPEN ACCESS

Citation:

Creemers HR,

Wager TD,

Yarkoni T

(2017)

The relation between statistical power and inference in fMRI. PLoS ONE 12(1): e0184923.

<https://doi.org/10.1371/journal.pone.0184923>

Editor:

Sam Gilbert,

University College London,

UNITED KINGDOM

Received:

January 26, 2017

Accepted:

September 1, 2017

Published:

November 20, 2017

Copyright:

© 2017

Creemers et al.

This is an open

access article distributed under the terms of the

Creative Commons Attribution License,

which

permits unrestricted

use, distribution,

and

reproduction in any medium, provided the original

author and source are credited.

Data Availability Statement:

The third-party data

underlying this study are available from the Human

Connectome Project.

The data set can be accessed

in the same manner as the authors by registering

on <http://www.humanconnectome.org/data/>.

Funding:

The Data analyzed in section 3.2 were

provided (in part) by the Human Connectome

Project, WU-Minn Consortium (Principal

Investigators: David Van Essen and Kamil Ugurbil;

15USMHQ01657) funded by the 16 NIH Institutes

and Centers that support the NIH Blueprint for

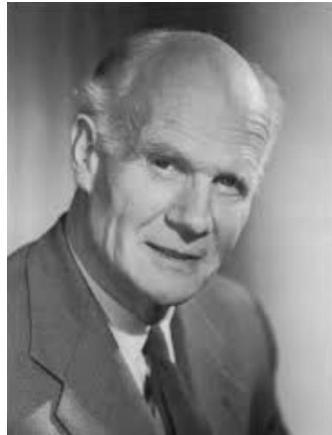
Neuroscience Research; and by the McDonnell

Center for Systems Neuroscience at Washington

The Statistical Power of a Test



Jerzy Neyman
(1894 – 1981)



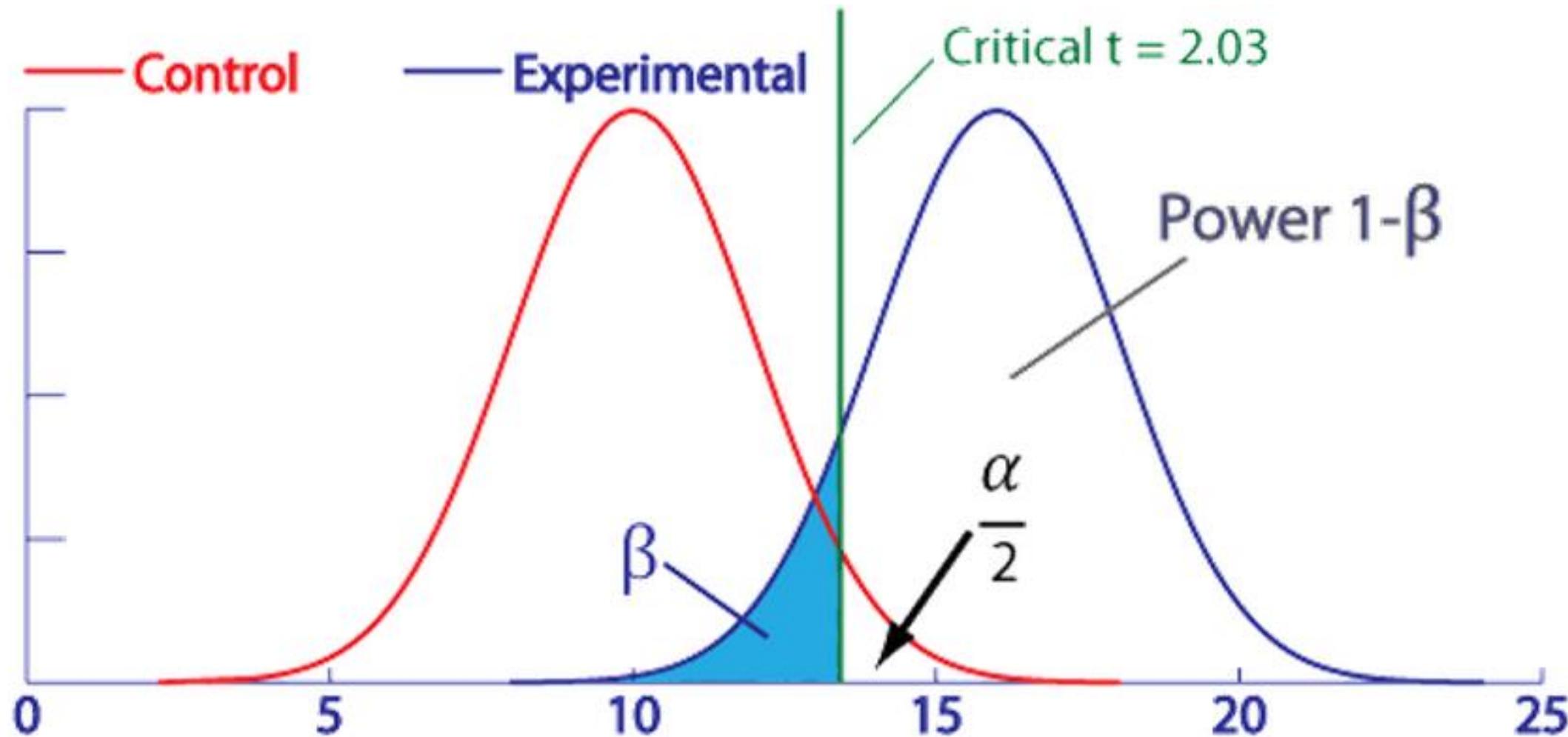
Egon Pearson
(1895-1980)

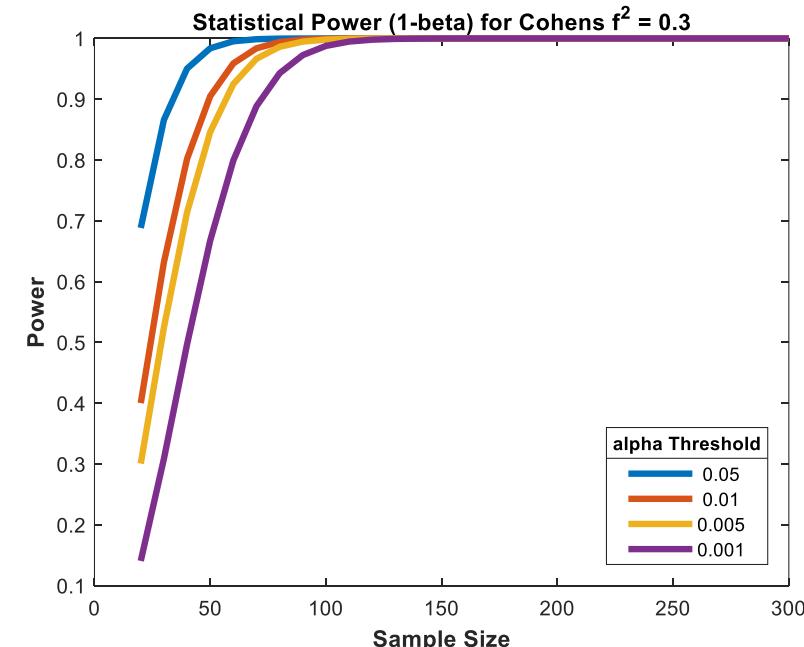
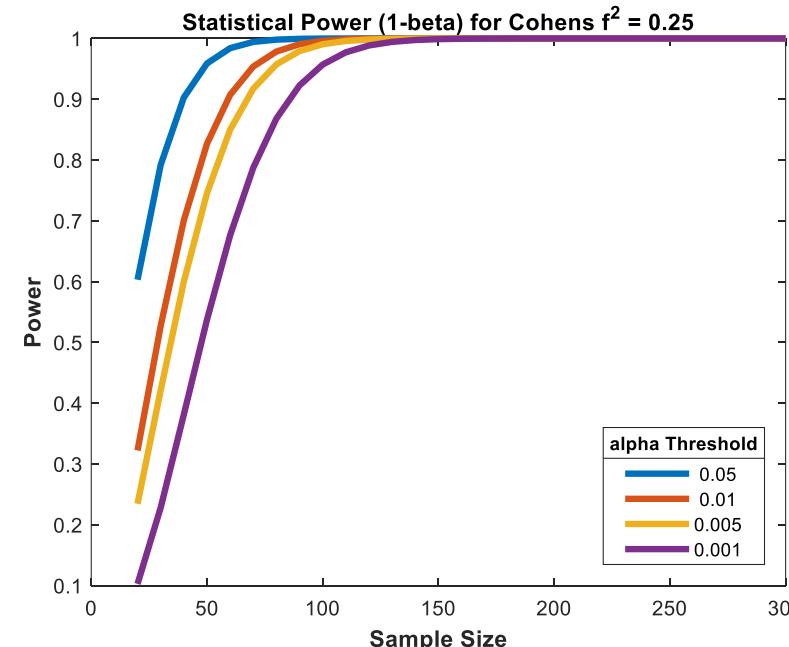
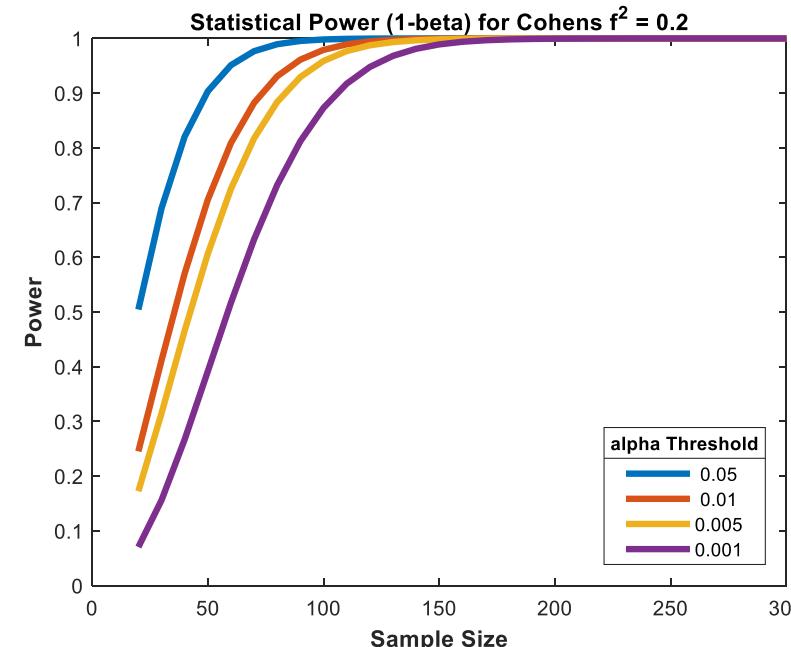
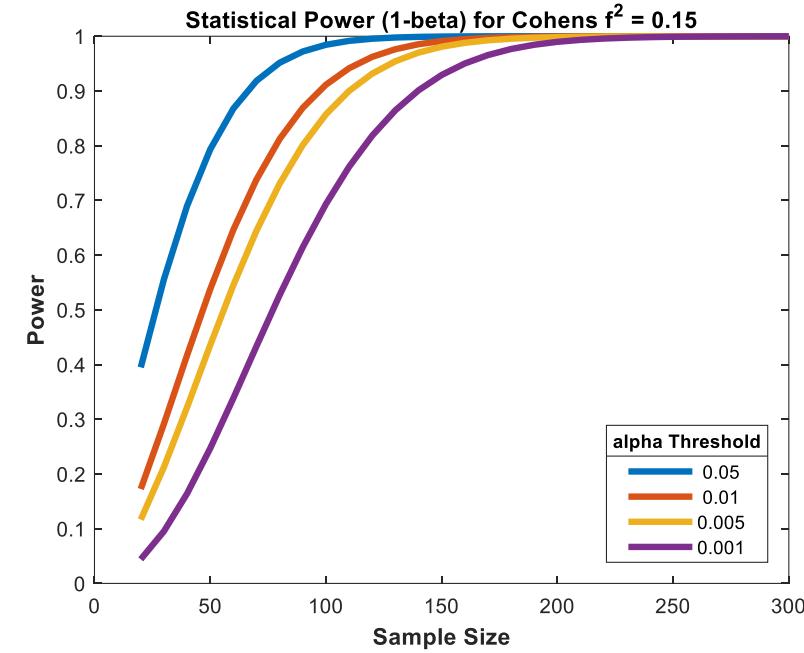
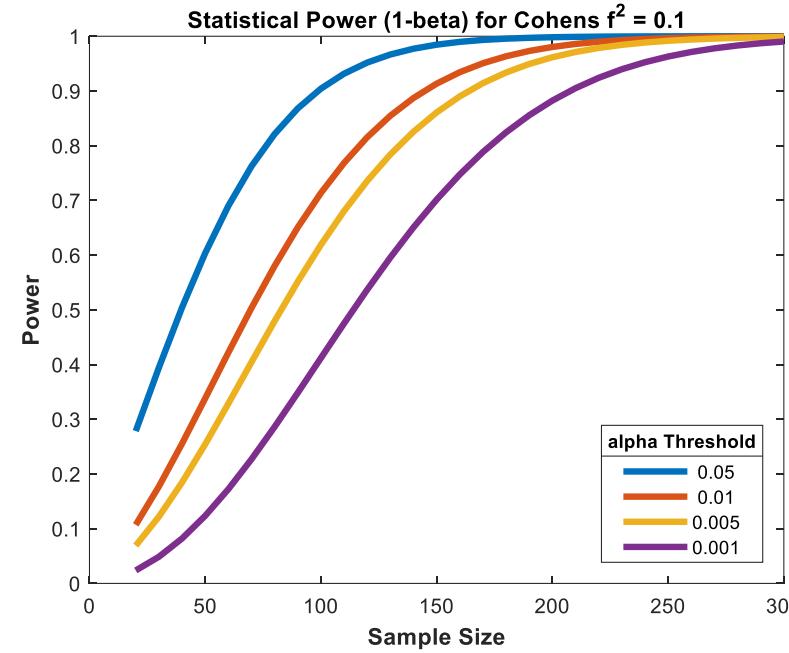
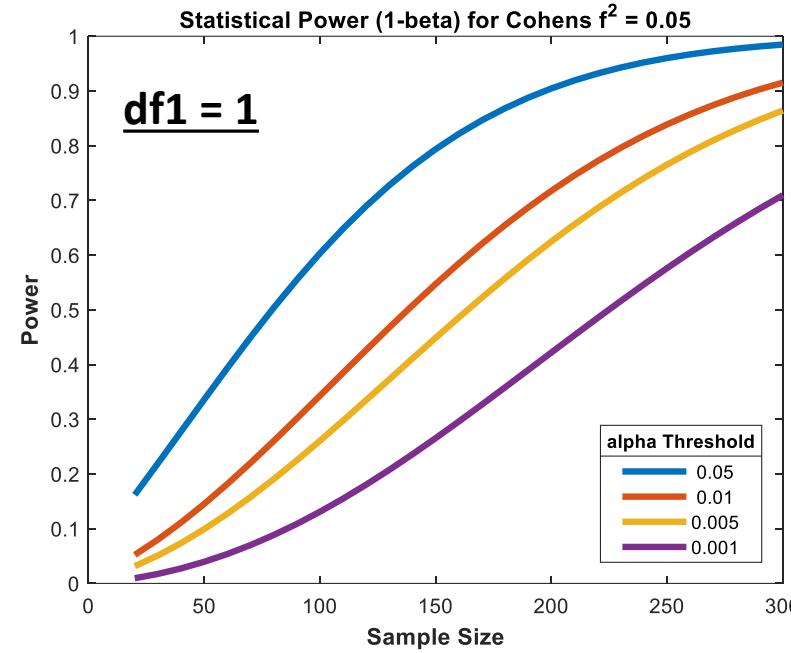
		Do not reject H_0	Reject H_0
H_0 is true	Correct Decision	Incorrect Decision: Type I error α	
	Incorrect Decision: Type II error β	Correct Decision “Power of the test”	
H_0 is false			

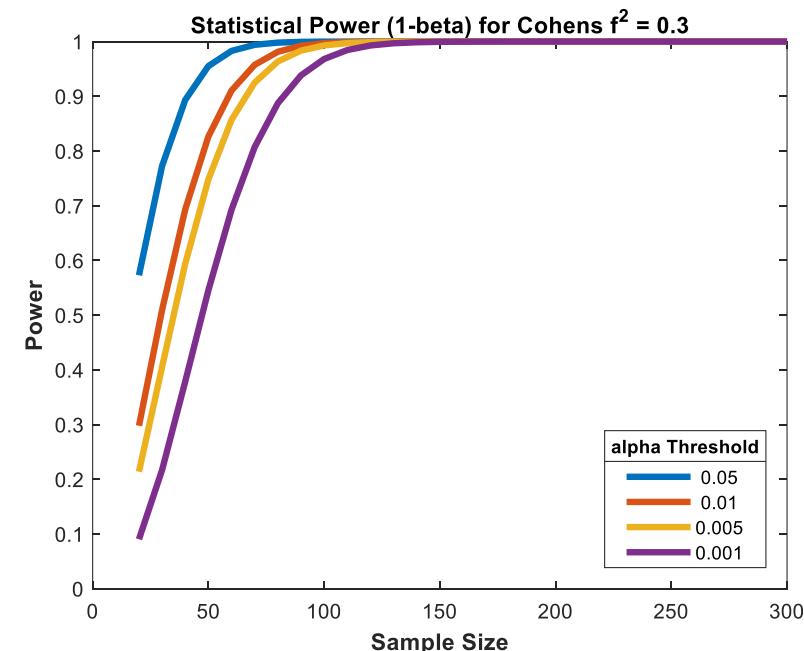
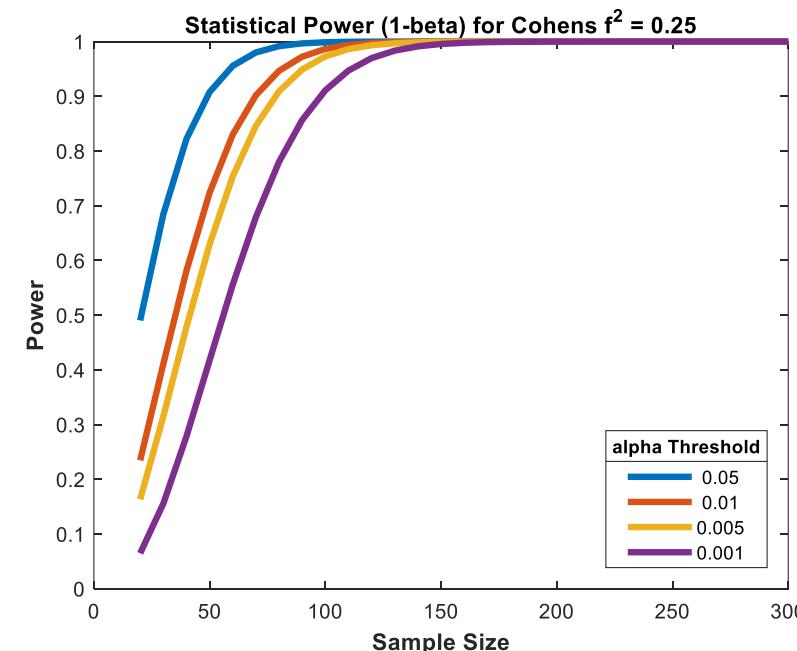
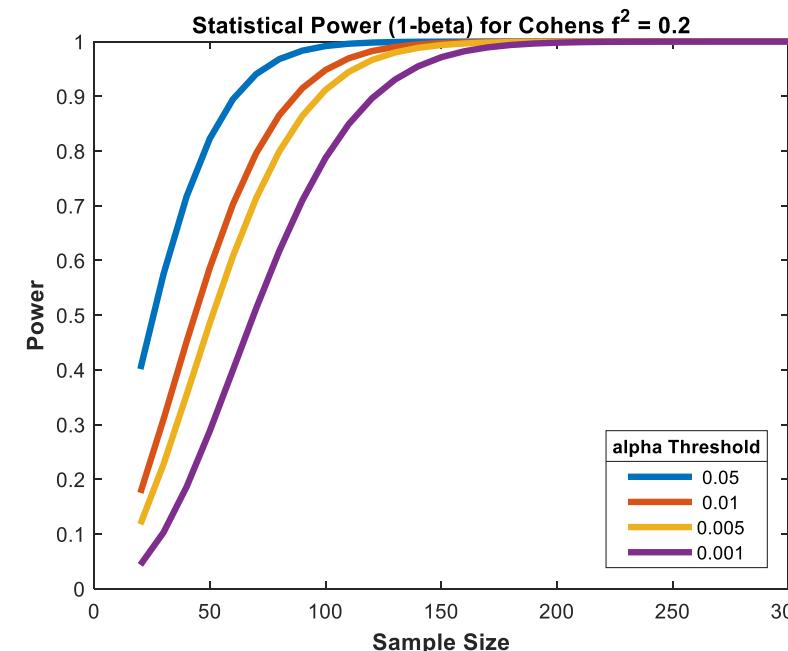
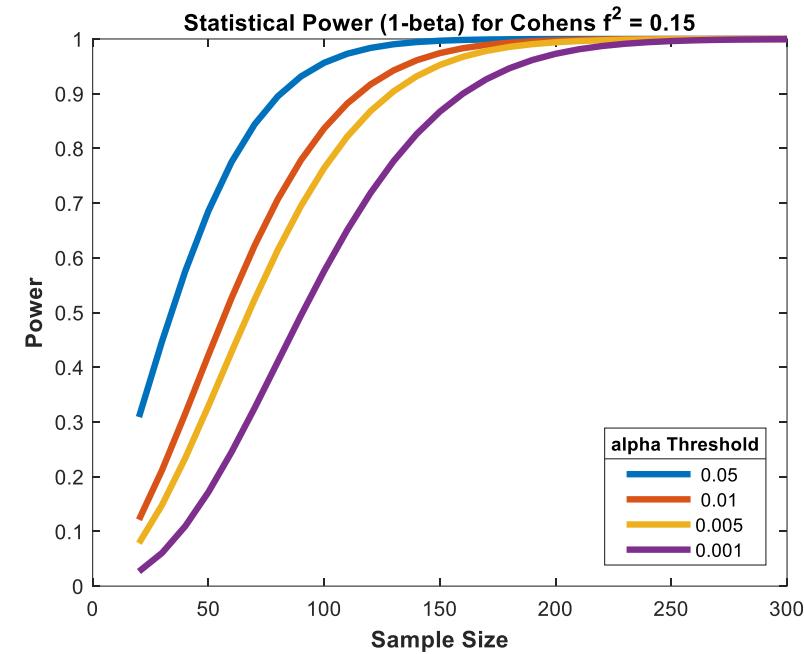
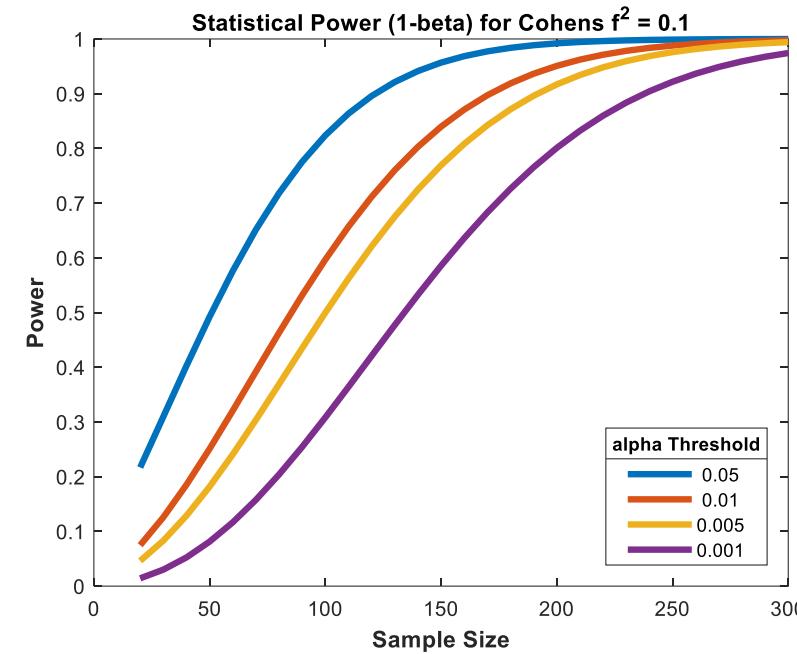
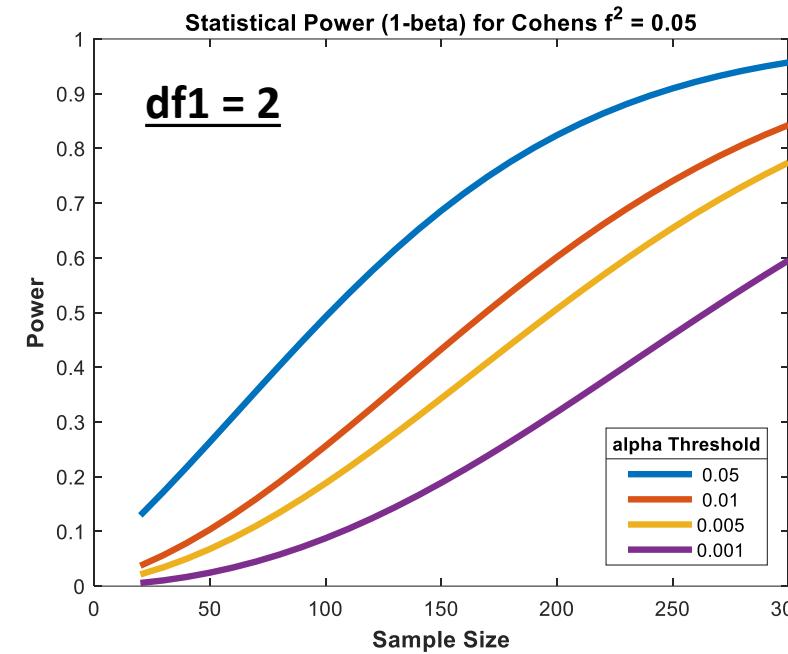
Statistical power is the probability that a statistical test will correctly reject a false null hypothesis. It reflects the test's ability to detect an effect or difference when one is truly exists. In frequentist statistics, ***power is a measure of the ability of an experimental design and hypothesis testing setup to detect a particular effect if it is truly present.***

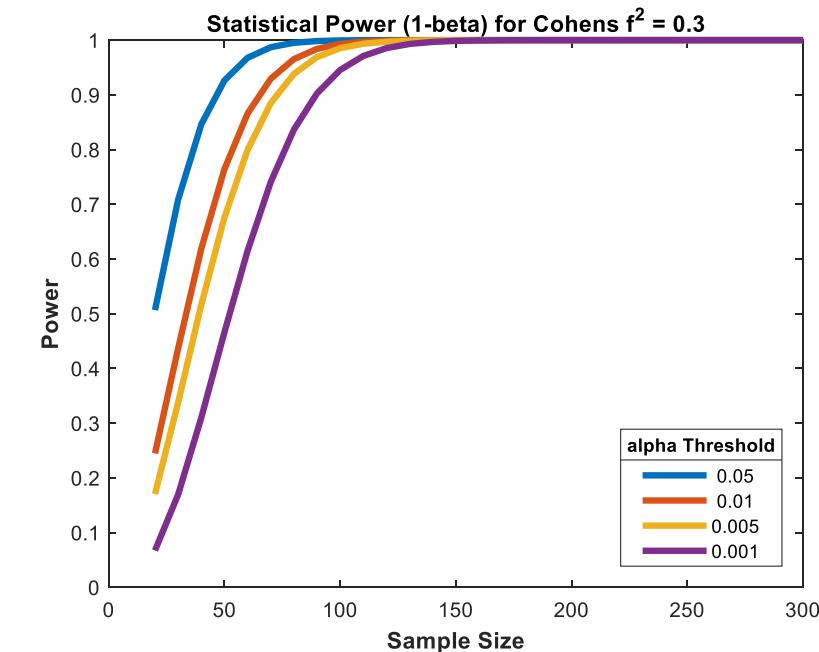
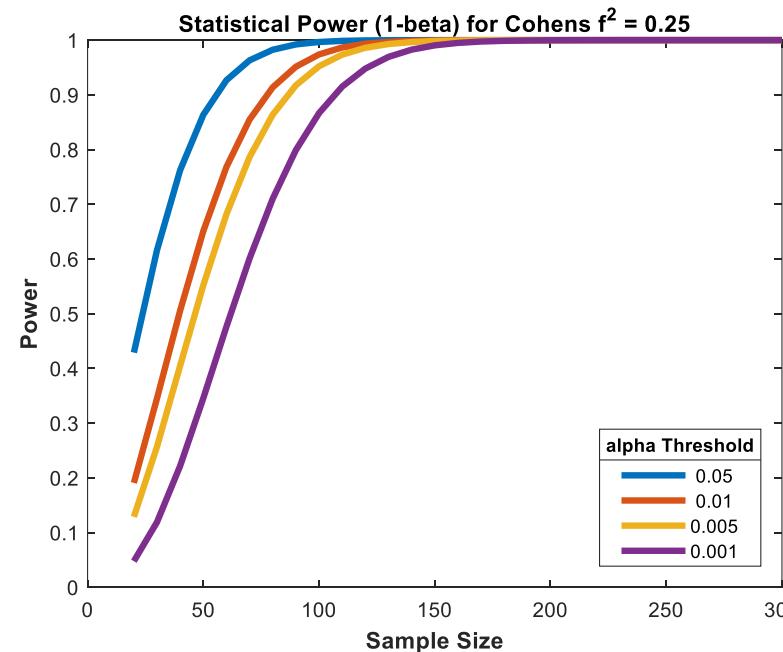
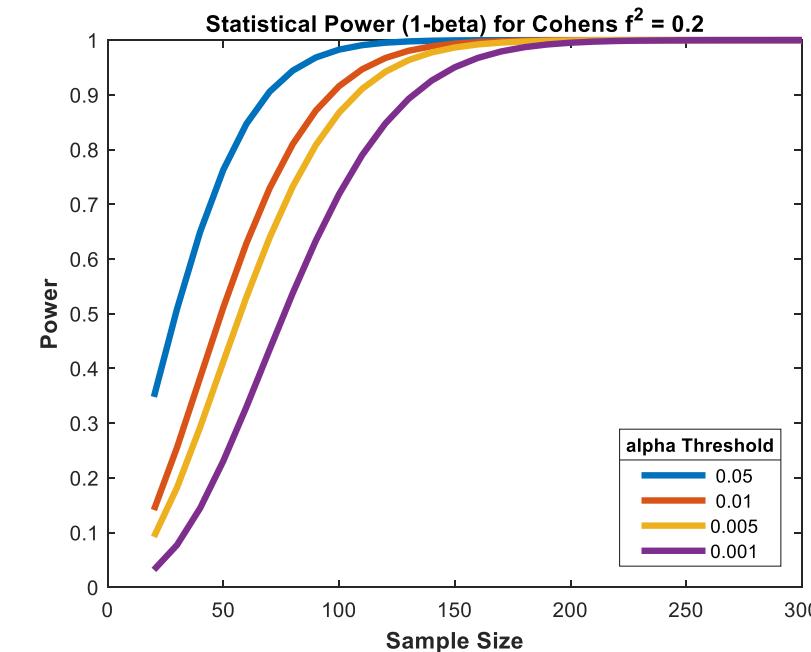
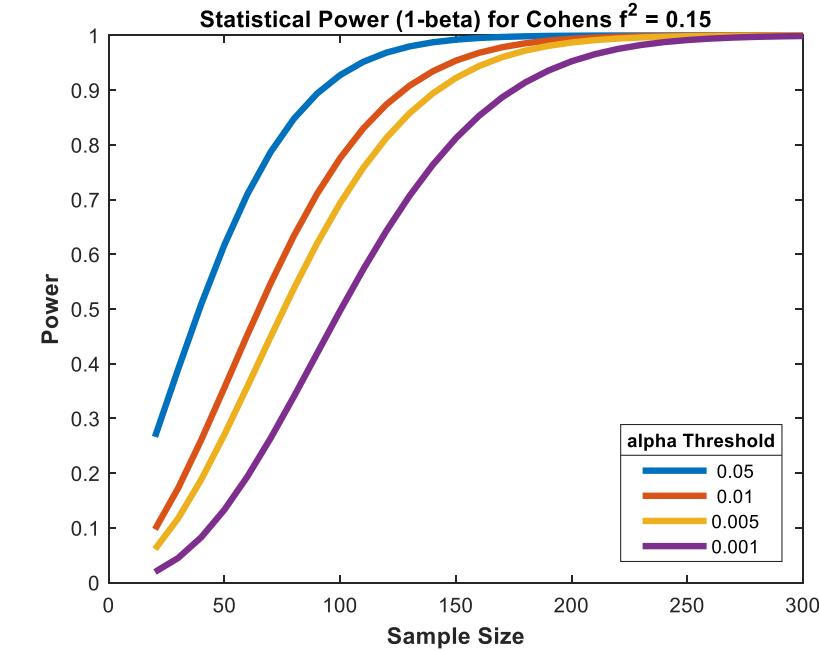
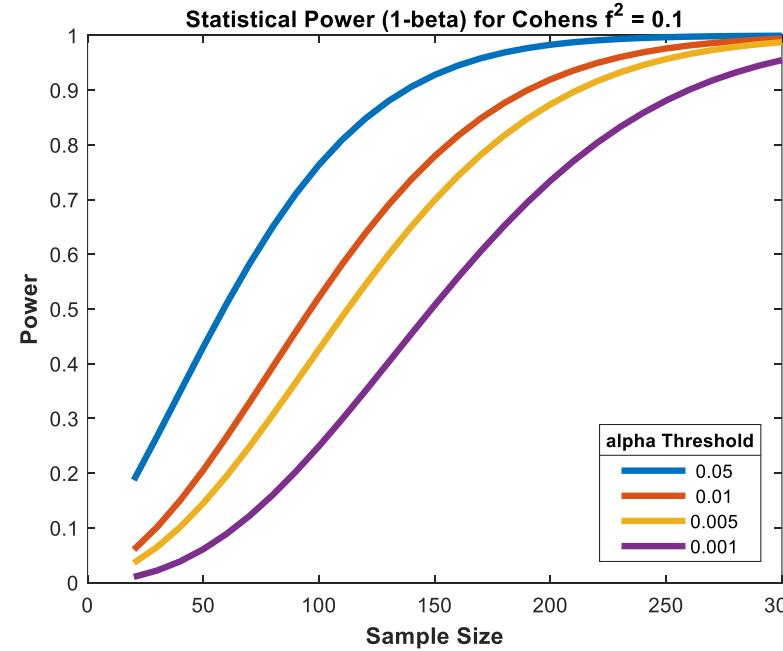
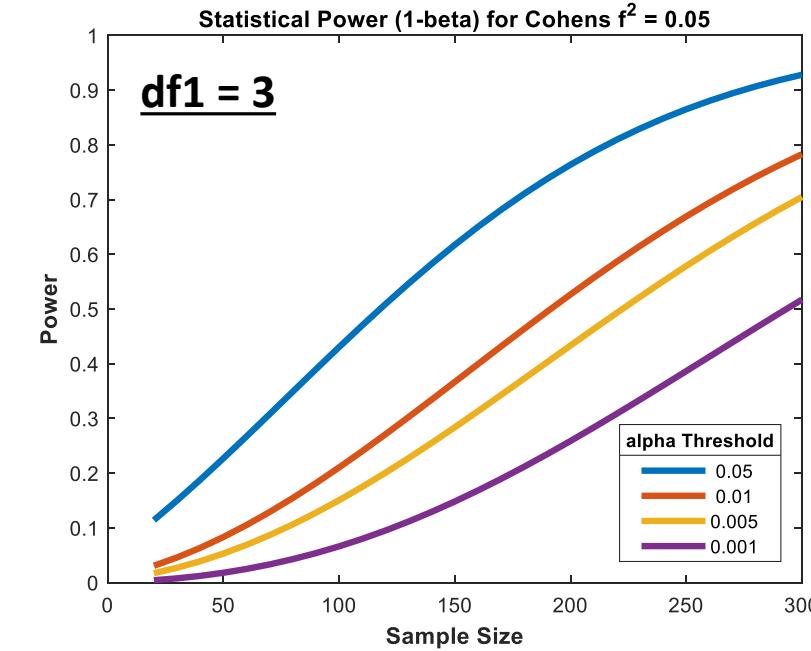
$$\text{Power} = 1 - \beta$$

The Statistical Power of a Test

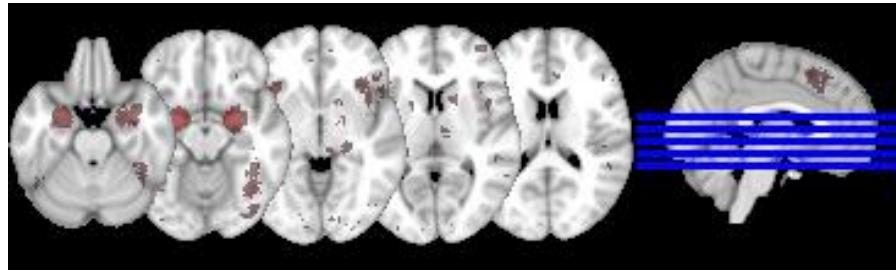








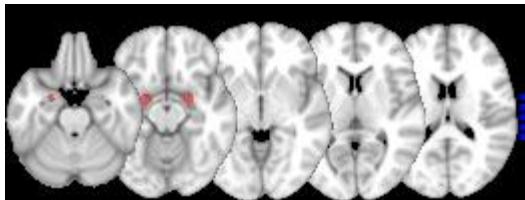
META-ANALYTIC ACTIVATION MAP FOR EMOTIONAL FACES



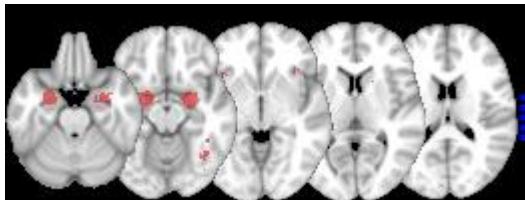
[Neurosynth: emotional faces](#)

Cohen's $f^2 = 0.30$

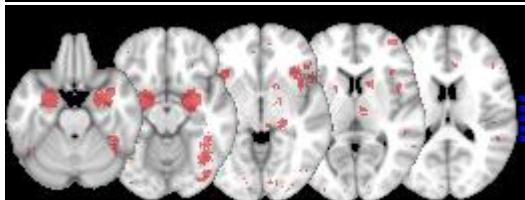
Power $\geq 0.80, p \leq 0.05$



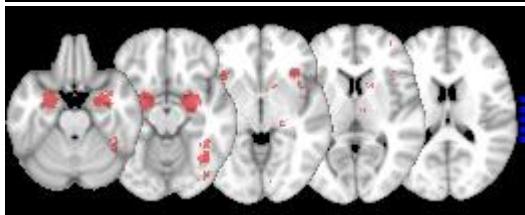
Power $\geq 0.85, p \leq 0.05$



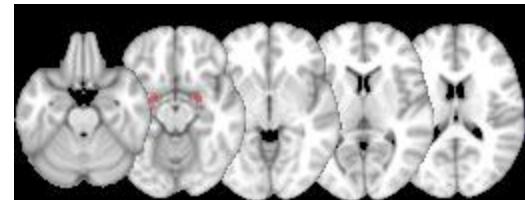
Power $\geq 0.90, p \leq 0.05$



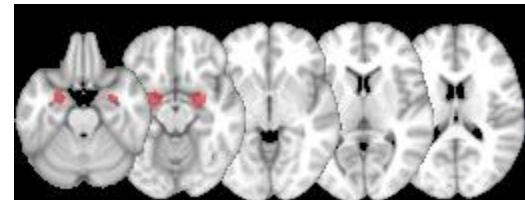
Power $\geq 0.95, p \leq 0.05$



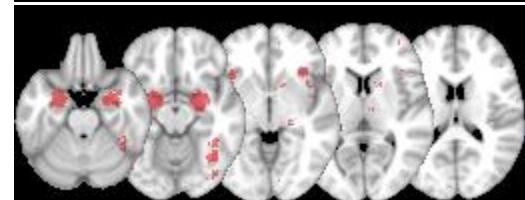
Power $\geq 0.90, p \leq 0.05$



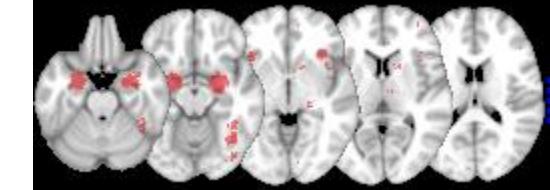
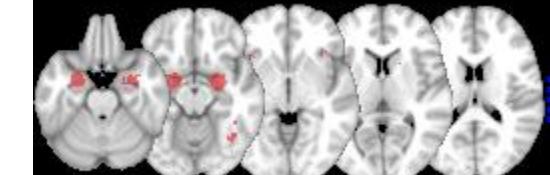
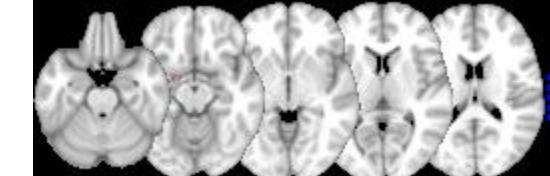
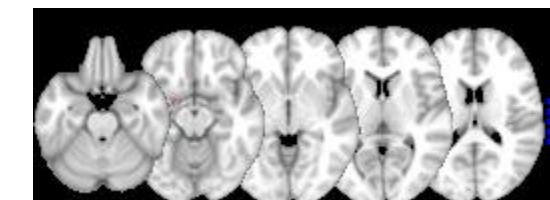
Power $\geq 0.85, p \leq 0.05$



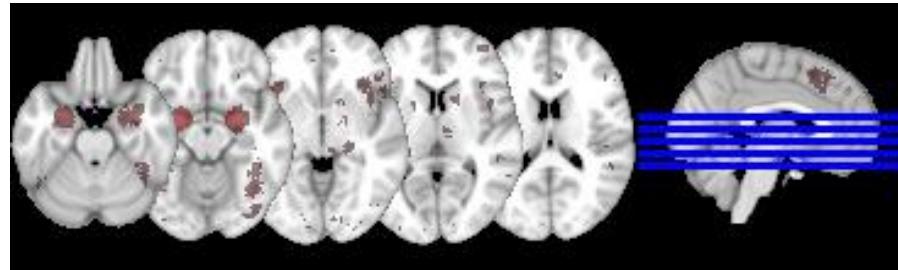
Power $\geq 0.90, p \leq 0.05$



Power $\geq 0.95, p \leq 0.05$



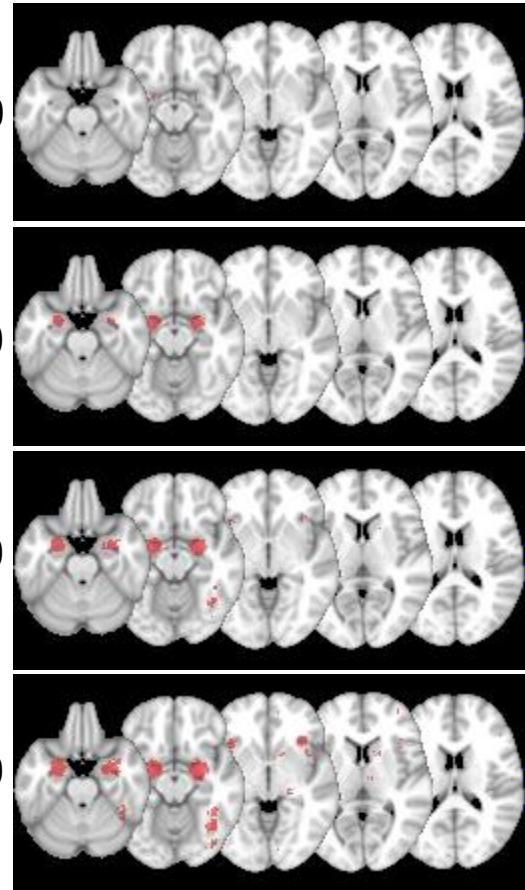
META-ANALYTIC ACTIVATION MAP FOR EMOTIONAL FACES



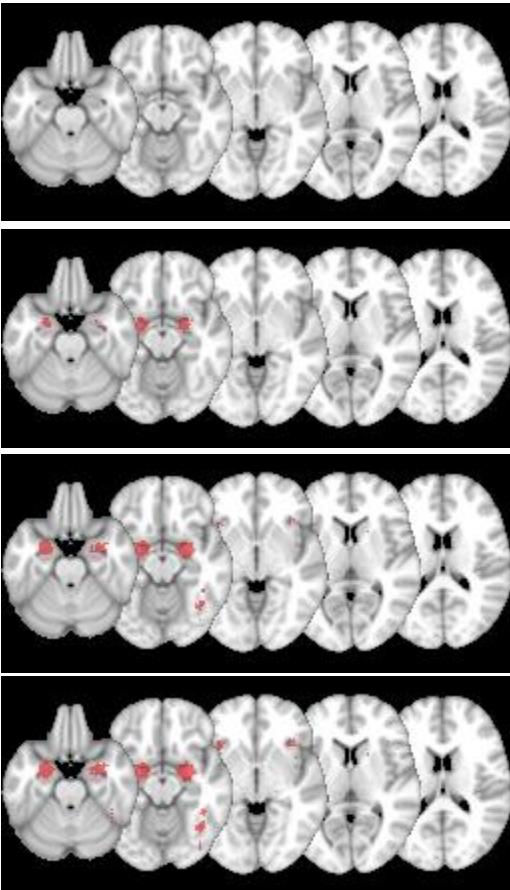
[Neurosynth: emotional faces](#)

Cohen's $f^2 = 0.20$

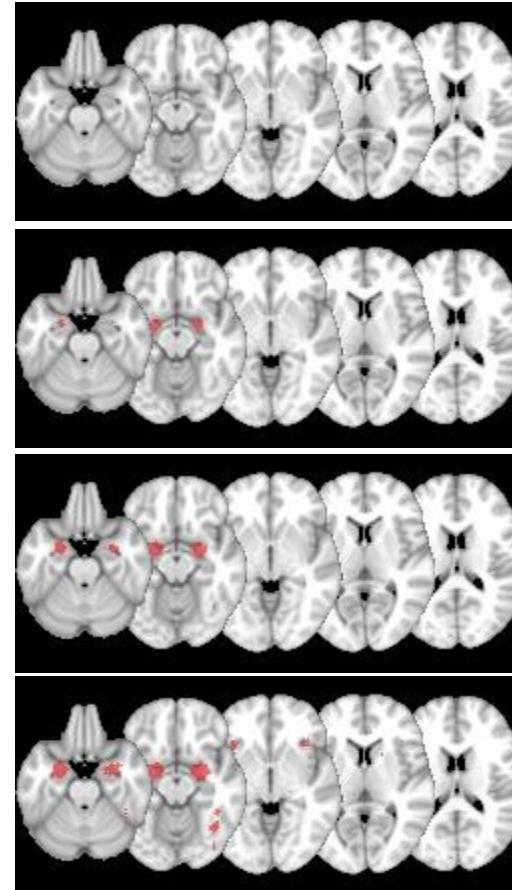
Power $\geq 0.80, p \leq 0.05$



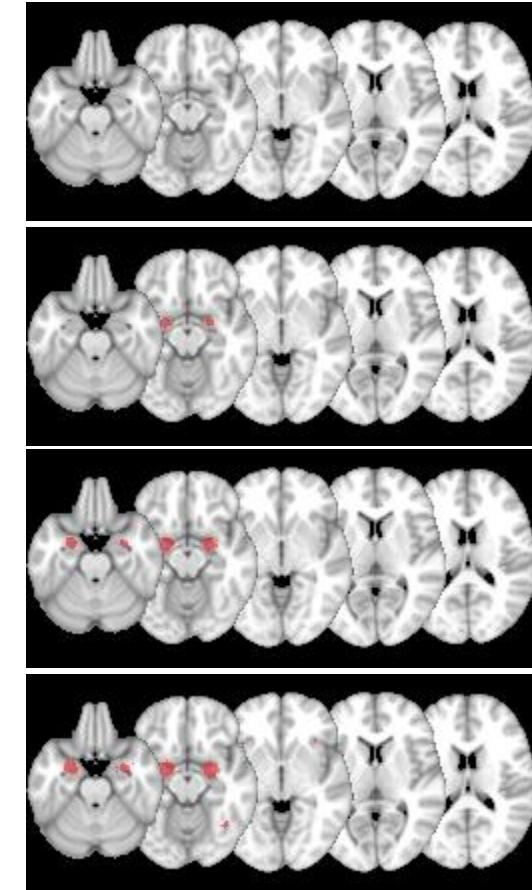
Power $\geq 0.85, p \leq 0.05$



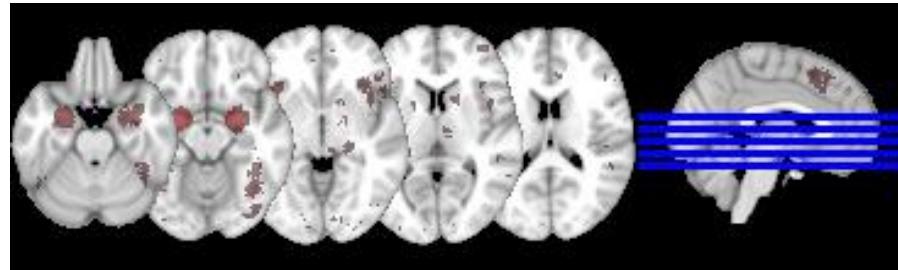
Power $\geq 0.90, p \leq 0.05$



Power $\geq 0.95, p \leq 0.05$



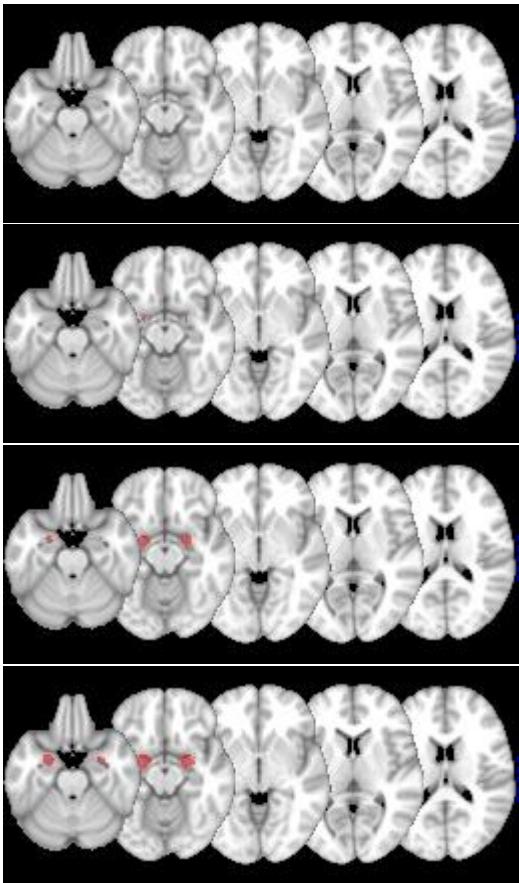
META-ANALYTIC ACTIVATION MAP FOR EMOTIONAL FACES



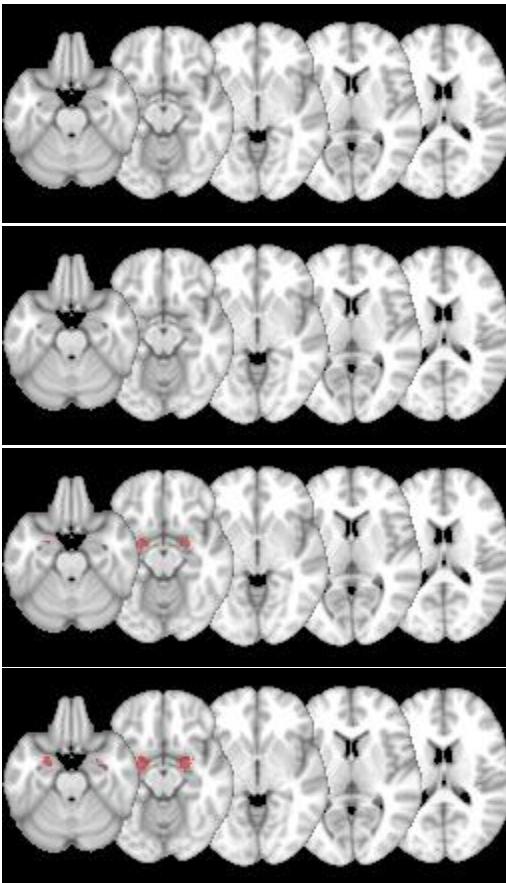
[Neurosynth: emotional faces](#)

Cohen's $f^2 = 0.10$

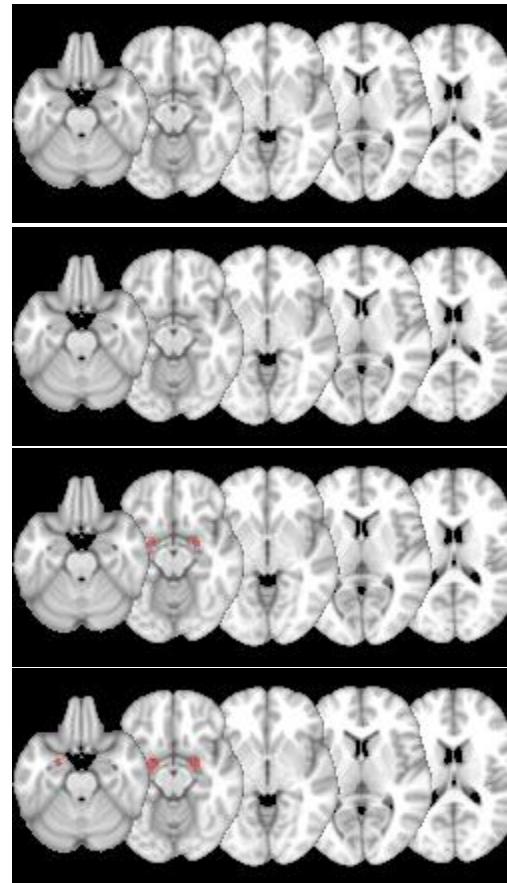
Power $\geq 0.80, p \leq 0.05$



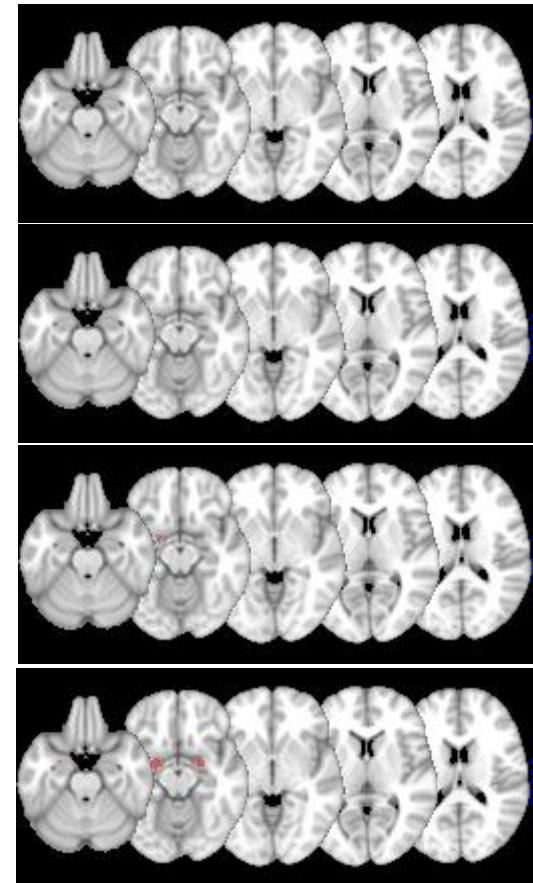
Power $\geq 0.85, p \leq 0.05$



Power $\geq 0.90, p \leq 0.05$



Power $\geq 0.95, p \leq 0.05$



N = 50

N = 100

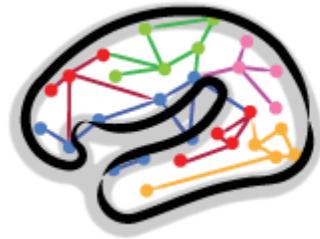
N = 150

N = 200

Sweeping (and possibly wrong) Conclusions about power in an Emotional Faces Task

- At an omnibus peak effect size of $f^2 = 0.1$ and a p-value of 0.05, one would need a sample size of ~150 or greater (75/group) to have 80% power to detect robust activations in the amygdala.
- For an omnibus peak effect size of $f^2 = 0.2$ and a p-value of 0.05, one would need a sample size of 100 or greater (50/group) to have 80% power to detect robust activations in the FFA.
- With omnibus peak effect size of $f^2 = 0.3$ and a p-value of 0.05, one would need a sample size of ~100 (50/group) or greater to have 80% power to detect robust activations in **cortical (e.g. frontal) brain regions**.
- Caveat emptor: The FFA is a highly robust effect in the cognitive neuroscience literature on face processing. Relatively low N fMRI studies spearheaded and have driven this literature (e.g. Courtney et al, *Science*, 1998 (N=11); Kosakowski et al., *eNeuro*, 2024 (N=65 combining two datasets)). The N's obtained here are theoretical and for our use in justifying our proposed sample size based on published results images. No “real” data was collected! BUT, these total sample sizes would certainly do the trick!

Multi-Site Studies



CONNECTOME
COORDINATION FACILITY



uk
biobank

Neuroimaging Parameters

	Matrix	Slices	FOV	% FOV phase	Resolution (mm)	TR (ms)	TE (ms)	TI (ms)	Flip Angle (deg)	Parallel Imaging	MultiBand Acceleration	Phase partial Fourier	Diffusion Directions	b-values	Acquisition Time
Siemens															
T1	256 x 256	176	256 x 256	100%	1.0 x 1.0 x 1.0	2500	2.88	1060	8	2x	Off	Off	N/A	N/A	7:12
T2	256 x 256	176	256 x 256	100%	1.0 x 1.0 x 1.0	3200	565	N/A	Variable	2x	Off	Off	N/A	N/A	6:35
Diffusion fMRI	140 x 140	81	240 x 240	100%	1.7 x 1.7 x 1.7	4100	88	N/A	90	Off	3	6/8	96	3000 (60-dirs)	7:31
	90 x 90	60	216 x 216	100%	2.4 x 2.4 x 2.4	800	30	N/A	52	Off	6	Off	N/A	N/A	
Philips															
T1	256 x 256	225	256 x 240	93.75%	1.0 x 1.0 x 1.0	6.31	2.9	1060	8	1.5 x 2.2	Off	N/A	N/A	N/A	5:38
T2	256 x 256	256	256 x 256	100%	1.0 x 1.0 x 1.0	2500	251.6	N/A	90	1.5 x 2.0	Off	N/A	N/A	N/A	2:53
Diffusion fMRI	140 x 140	81	240 x 240	100%	1.7 x 1.7 x 1.7	5300	89	N/A	78	Off	3	0.6	96	3000 (60-dirs)	9:14
	90 x 90	60	216 x 216	100%	2.4 x 2.4 x 2.4	800	30	N/A	52	Off	6	0.9	N/A	N/A	
GE															
T1	256 x 256	208	256 x 256	100%	1.0 x 1.0 x 1.0	2500	2	1060	8	2x	Off	Off	N/A	N/A	6:09
T2	256 x 256	208	256 x 256	100%	1.0 x 1.0 x 1.0	3200	60	N/A	Variable	2x	Off	Off	N/A	N/A	5:50
Diffusion fMRI	140 x 140	81	240 x 240	100%	1.7 x 1.7 x 1.7	4100	81.9	N/A	77	Off	3	5.5/8	96	3000 (60-dirs)	7:30
	90 x 90	60	216 x 216	100%	2.4 x 2.4 x 2.4	800	30	N/A	52	Off	6	Off	N/A	N/A	

https://abcdstudy.org/images/Protocol_Imaging_Sequences.pdf

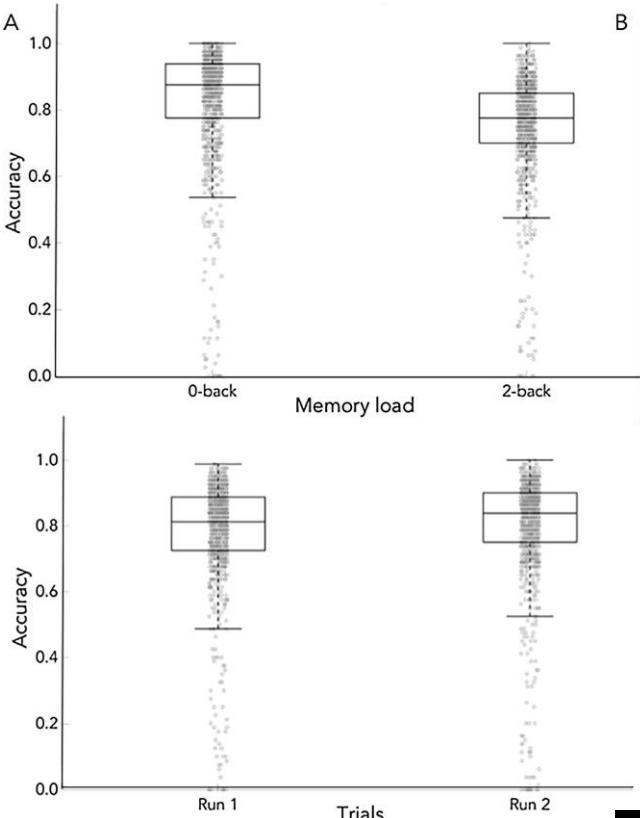
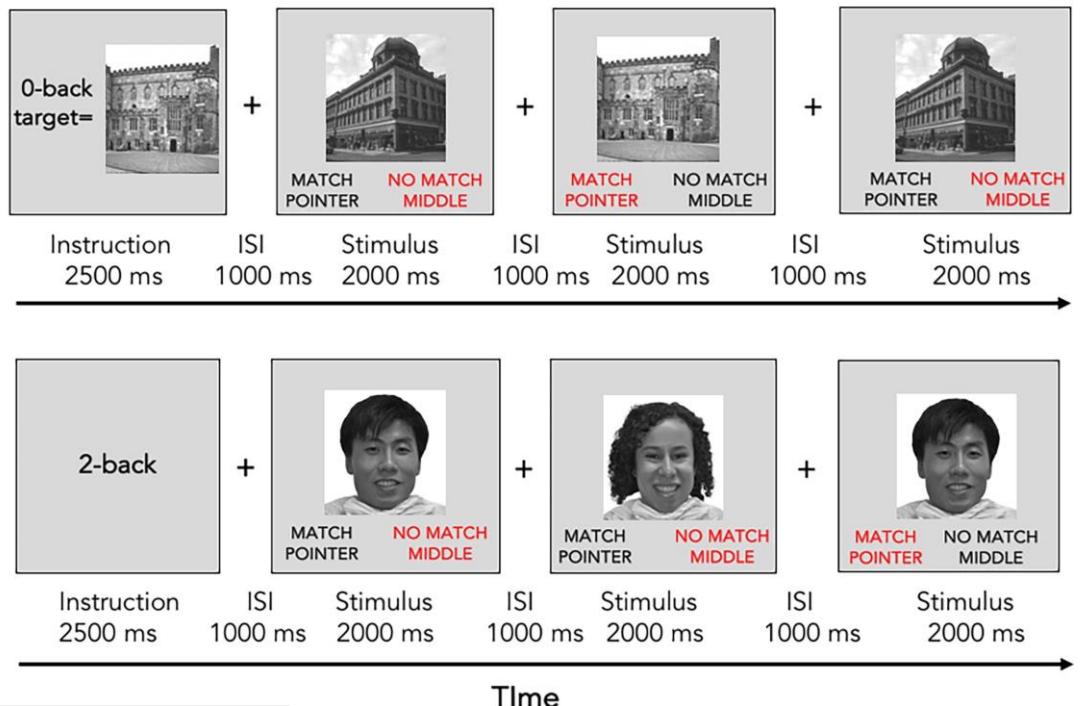


Adolescent Brain Cognitive Development®
Teen Brains. Today's Science. Brighter Future.

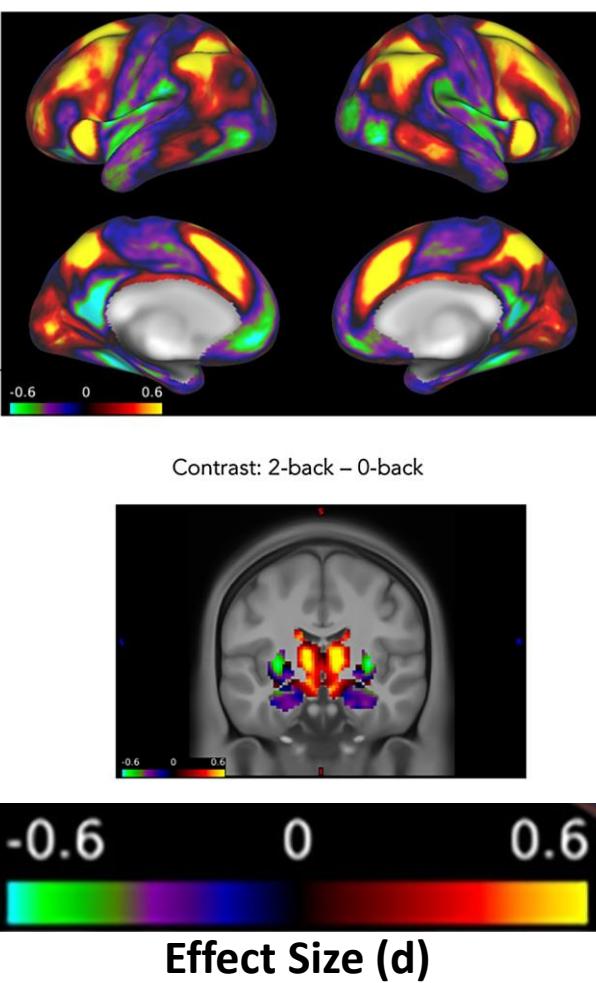
ABCDStudy.org

Emotional N-Back Task

O-back condition



n = 517

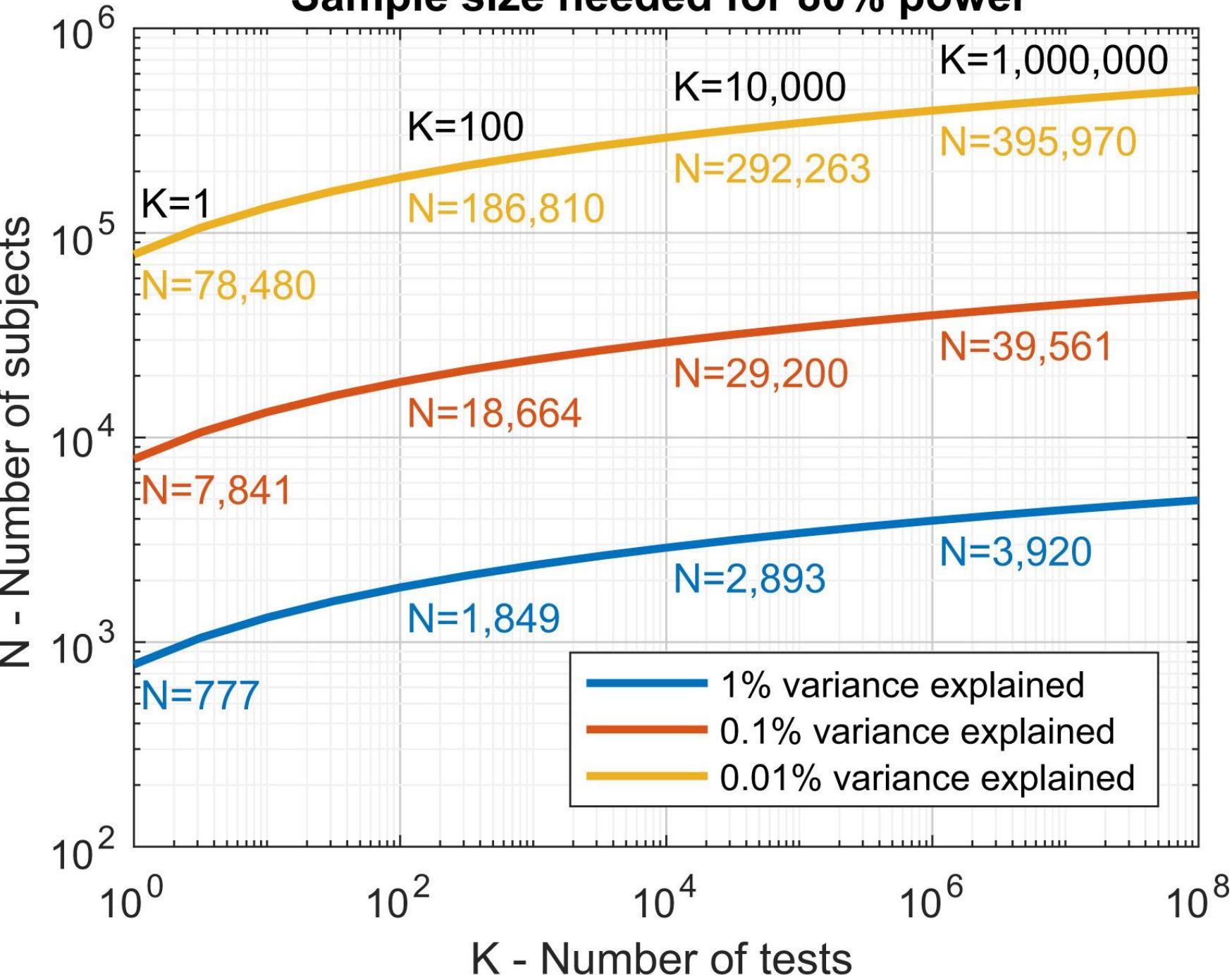


	Cohen's d Effect Size			
	0.1	0.2	0.3	0.6
Non-Centrality Parameter	3.61	3.62	3.64	3.75
Critical t	1.962	1.967	1.967	2.02
df	1301	326	146	38
Sample Size Requirement	1302	327	147	39
Power	0.95	0.95	0.95	0.95

Assuming α (uncorrected) = 0.05



Sample size needed for 80% power



Statistical Challenges in "Big Data" Human Neuroimaging

Stephen M. Smith^{1,*} and Thomas E. Nichols^{1,2,3}

¹Wellcome Trust Centre for Integrative Neuroimaging (WIN-FMRI), University of Oxford, Oxford, UK

²Big Data Institute, University of Oxford, Oxford, UK

³Department of Statistics, University of Warwick, Coventry, UK

*Correspondence: steve@fmri.ox.ac.uk

<https://doi.org/10.1016/j.neuron.2017.12.018>

Smith and Nichols discuss "big data" human neuroimaging studies, with very large subject numbers and amounts of data. These studies provide great opportunities for making new discoveries about the brain but raise many new analytical challenges and interpretational risks.

Introduction

Over the last 30 years, structural and functional brain imaging has become a powerful and widespread tool for clinical and basic neuroscience. However, it is only now that some brain imaging studies have started to join the ranks of "big data" science. Whereas most neuroimaging studies continue to have modest sample sizes (number of subjects, $N < 50$) and modest amounts of data collected per subject, a small number of studies are now starting to embrace "big" imaging in a number of ways—with subject numbers in tens of thousands, or taking advantage of huge increases in the quantity of imaging and non-imaging data collected on each subject. For example, the Human Connectome Project (HCP) (Van Essen et al., 2013) has recently completed imaging of >1,000 young adults, with an impressive 4 hr of scanning per subject, and utilizing vast improvements in the spatial and temporal resolutions of the acquired data. In a complementary manner, UK Biobank (UKB) (Miller et al., 2016) is acquiring more modest amounts of imaging data per subject but is scanning 100,000 volunteers, and this imaging is just part of the much larger overall UKB project that includes genetics, biological and lifestyle measures, and health outcomes from 500,000 subjects. In yet another approach to big data imaging, the ENIGMA consortium is amassing imaging and genetic data from tens of thousands of subjects by pooling many existing studies, using advanced meta-analysis techniques to overcome restrictions on sharing individual-subject data (Bearden and Thompson 2017).

In this brief NeuroView, we highlight some of the challenges that "big data neuroimaging" brings. While there is no hard definition, we may consider an imaging study "big" if it has 1,000 or more subjects and/or collects a substantially larger set of measurements than usual or expands on traditional measurements (e.g., with greater resolution or duration than typical). We focus on MRI-based brain imaging, though many of the points made are also relevant for valuable complementary techniques such as EEG, MEG, and PET. All aspects of neuroimaging are rapidly becoming "bigger." Spatial resolution and number of images are constantly increasing, enabled by enhancements to hardware (higher field strength and more coil channels) and software (advanced acquisition and reconstruction techniques). Additionally, the proliferation of research scanners means that study designs are becoming more ambitious, with longer scan durations, multiple scan sessions, and large subject numbers. For example, if one considers the increases in state-of-the-art fMRI imaging spatial resolution, temporal resolution, and receive-coil numbers, data size has increased quite in line with Moore's law (a doubling every 2 years, i.e., a 1,000-fold increase in the last 20 years).

These numbers immediately illustrate perhaps the most obvious difficulty with big data neuroimaging—data sizes can quickly become hard to manage, both in terms of long-term storage and compute speed and memory. However this is just one of the challenges.

Big Sample Sizes: Small Effect Sizes

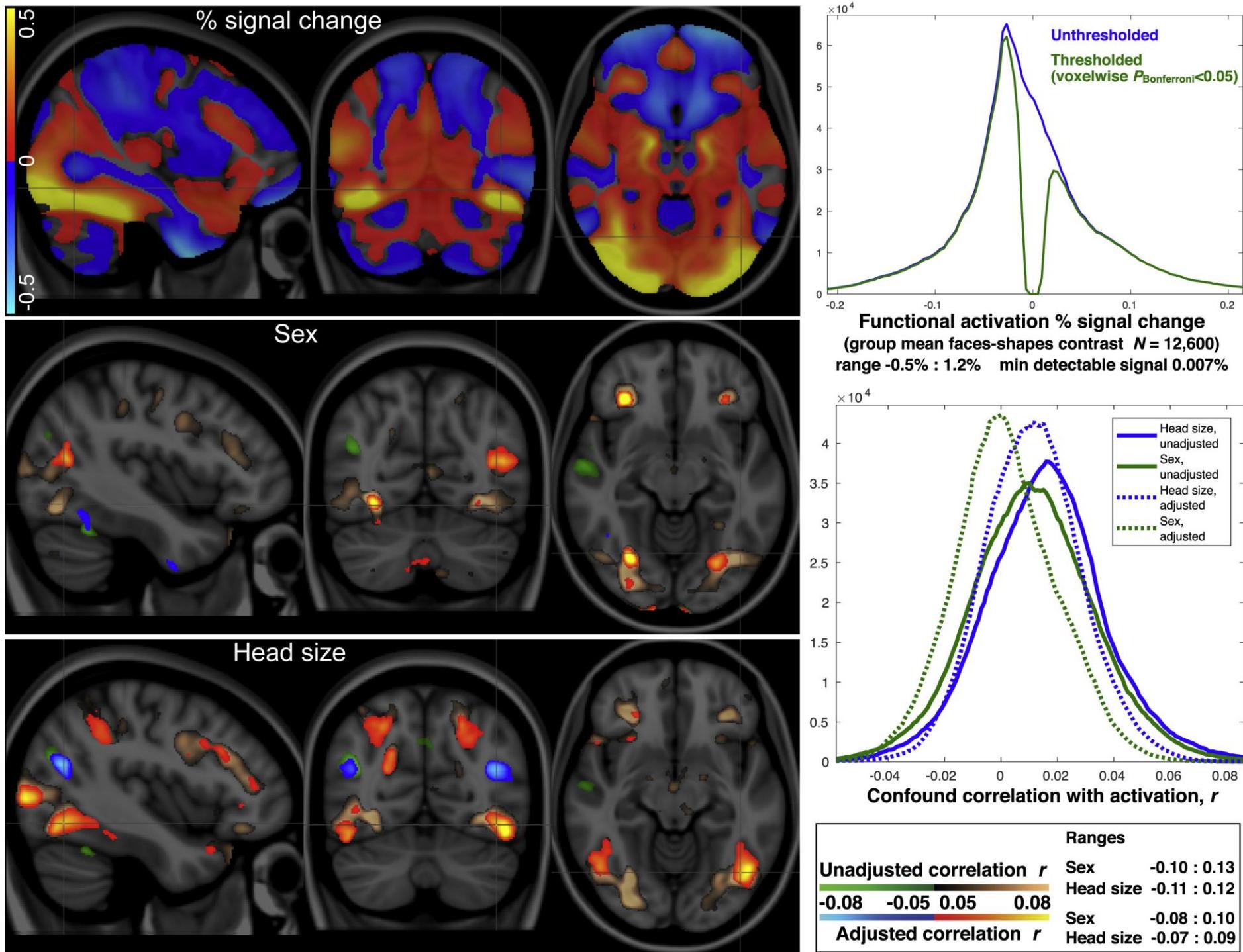
A primary reason that prospective epidemiological health studies (such as UKB) need very large subject numbers is to be able to acquire data on subjects in advance of later disease development, hence allowing researchers to learn the earliest markers of (and possibly mechanisms for) disease. The large subject numbers are needed as there is no single

Voxelwise Analyses of the *Faces-Shapes* Contrast in the [UK Biobank](#) Task-fMRI Data, from 12,600 Subjects

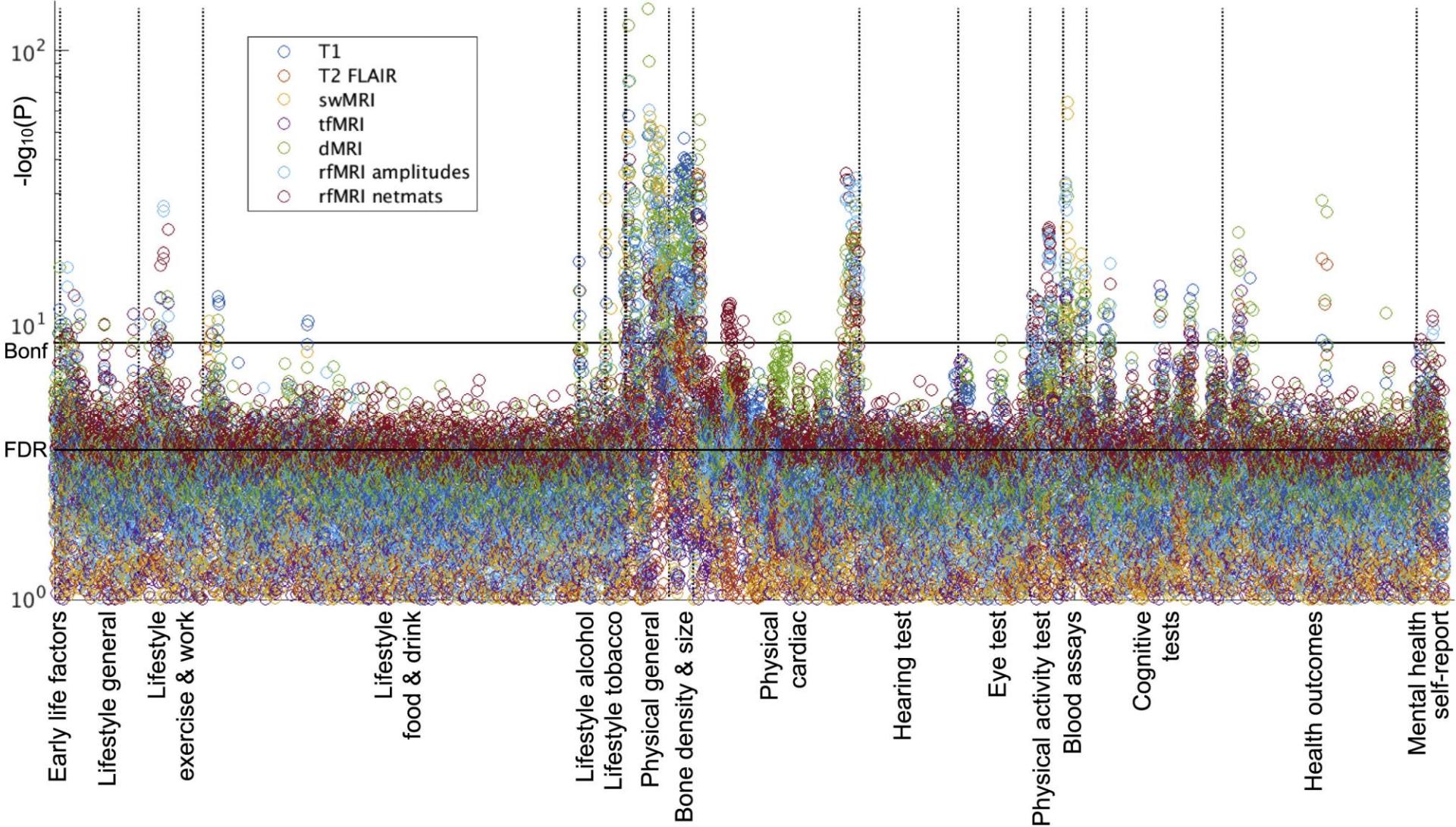
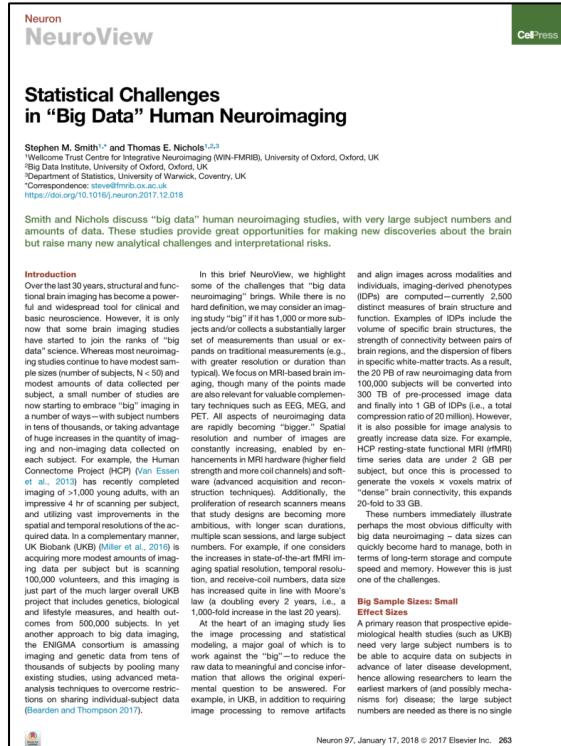
The *% signal change* color overlay shows the fMRI signal change associated with the faces-shapes contrast, masked by significant [voxels](#) from mixed-effects modeling of the group-average signal (all maps shown here conservatively corrected for [multiple comparisons](#) across 1.8 million voxels using Bonferroni correction, $p<0.05$).

Thresholding excludes only 19% of voxels, i.e., showing a significant response to the task in most of the [brain](#).

The maximum statistical effect size (Cohen's d) is 1.57, equivalent to a one-group T [statistic](#) of 176.2.

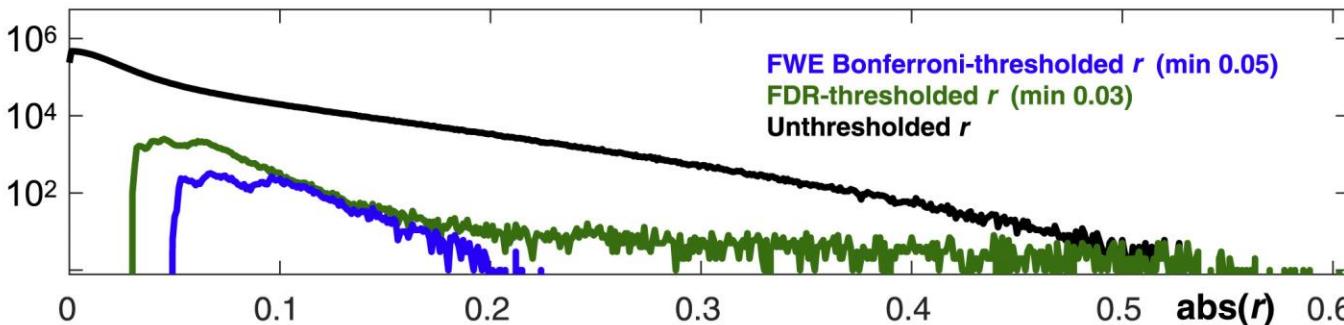


14 Million Univariate Association Tests between IDPs and Non-Brain-Imaging Variables in UK Biobank, 14,500 Subjects



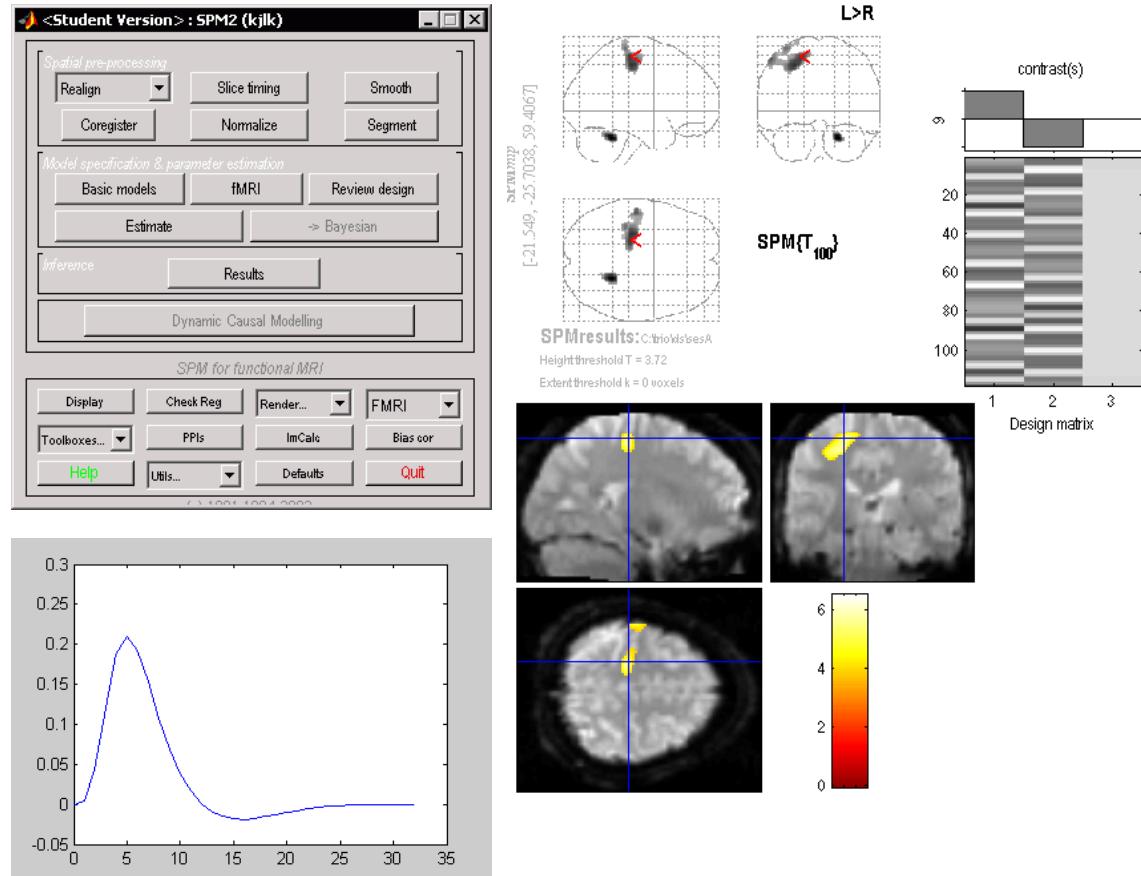
PWAS: Phenome-Wide Association Study

Approximately 100,000 associations over 5,456 variables are FDR significant, and 15,000 are Bonferroni significant. The histograms show the distributions of correlation size (across all 14 million tests); depending on [thresholding](#) method, the minimum detectable r is 0.03–0.05, meaning that for FDR thresholding an association with 0.1% variance explained is detectable.

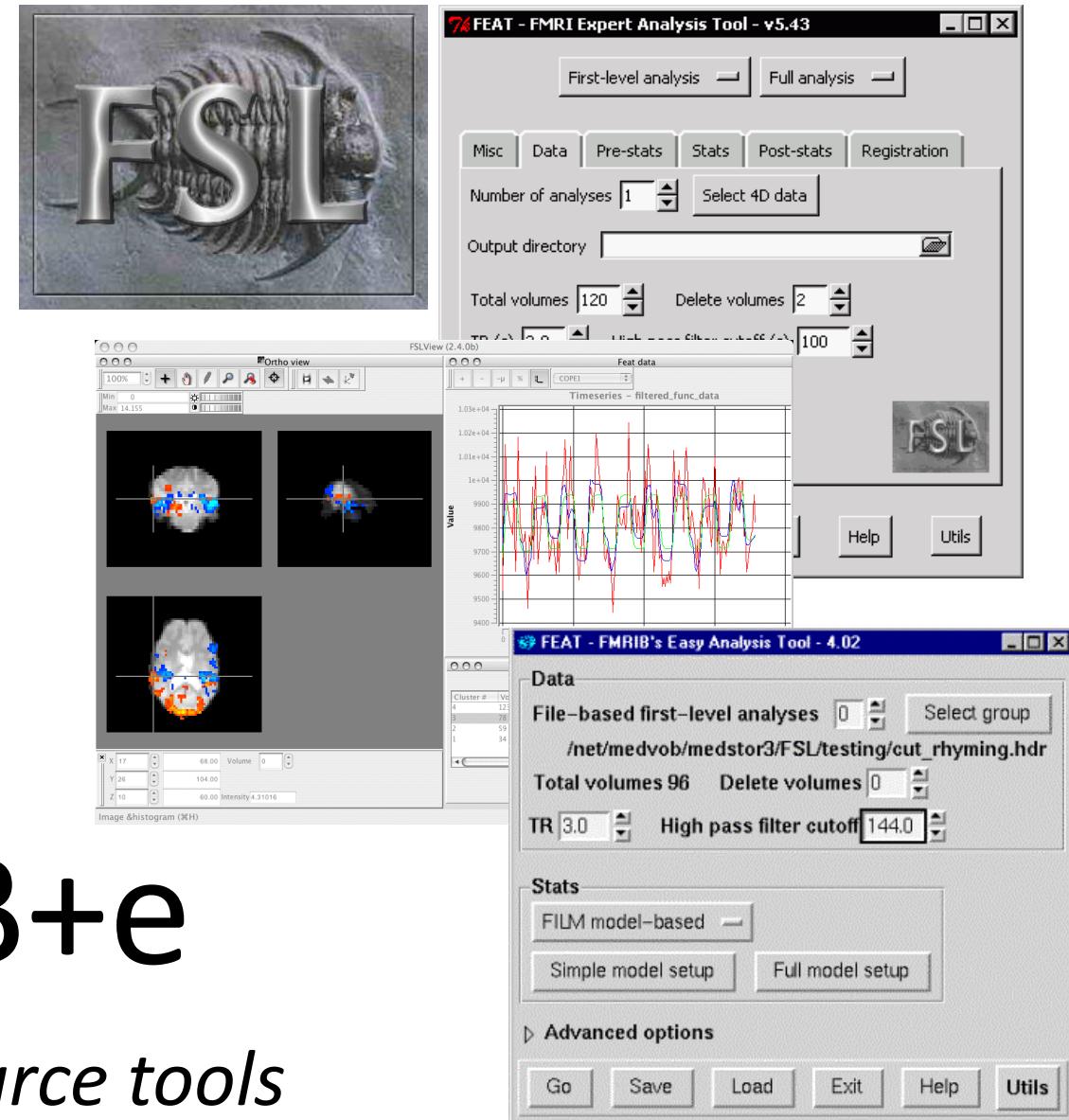


SGPV

SPM (1990'ish to Present)



FSL (2000'ish to Present)

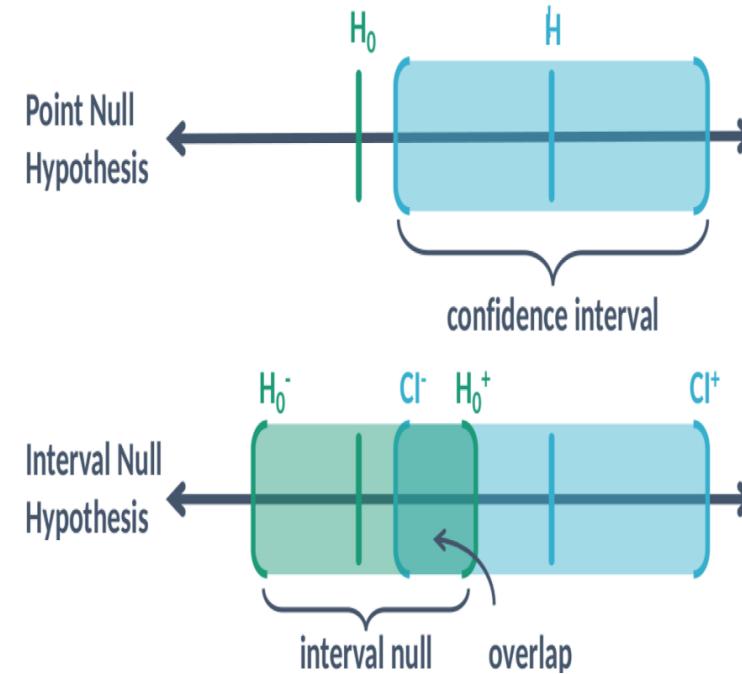
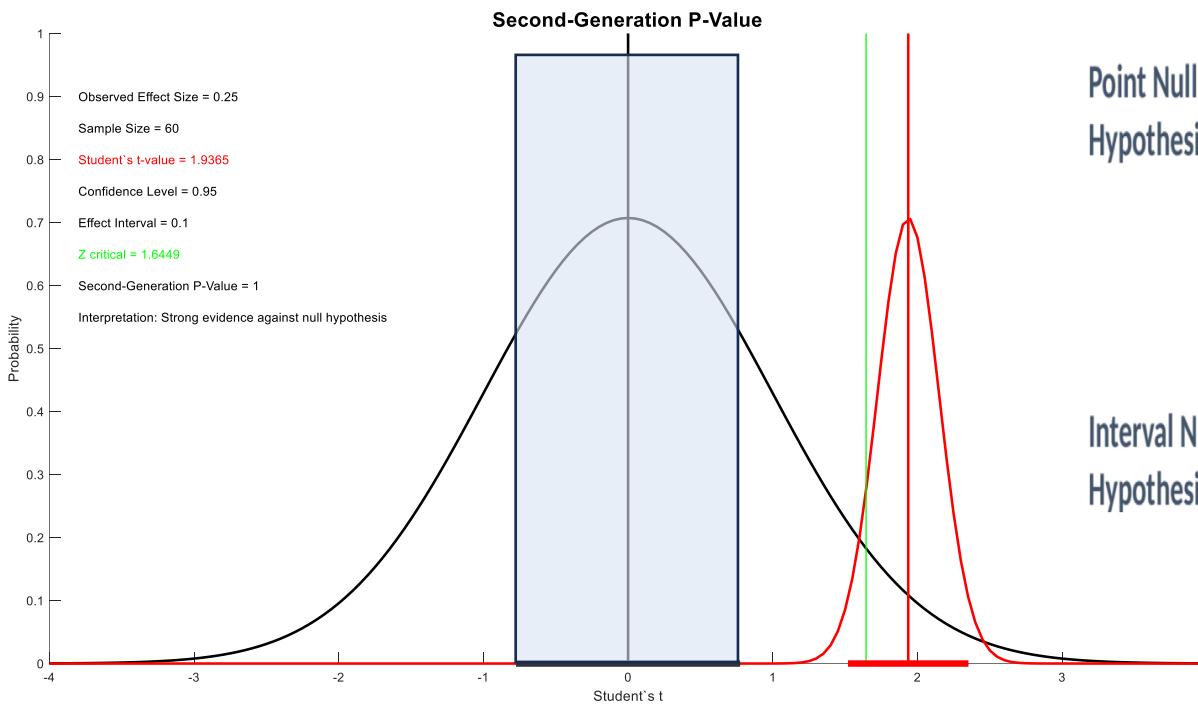


$$Y = XB + e$$

AFNI, too!

And an increasing number of open-source tools

Second Generation p-Values



PLOS ONE

RESEARCH ARTICLE

Second-generation p-values: Improved rigor, reproducibility, & transparency in statistical analyses

Jeffrey D. Blume^{1*}, Lucy D'Agostino McGowan², William D. Dupont³, Robert A. Greevy, Jr.¹

¹ Associate Professor, Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, ² PhD Candidate, Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, ³ Professor, Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America

* j.blume@vanderbilt.edu

Abstract

Verifying that a statistically significant result is scientifically meaningful is not only good scientific practice, it is a natural way to control the Type I error rate. Here we introduce a novel extension of the p-value—a second-generation p-value (p_2)—that formally accounts for scientific relevance and leverages this natural Type I Error control. The approach relies on a pre-specified interval null hypothesis that represents the collection of effect sizes that are scientifically uninteresting or are practically null. The second-generation p-value is the proportion of data-supported hypotheses that are also null hypotheses. As such, second-generation p-values indicate when the data are compatible with null hypotheses ($p_2 = 1$), or with alternative hypotheses ($p_2 = 0$), or when the data are inconclusive ($0 < p_2 < 1$). Moreover, second-generation p-values provide a proper scientific adjustment for multiple comparisons and reduce false discovery rates. This is an advance for environments rich in data, where traditional p-value adjustments are needlessly punitive. Second-generation p-values promote transparency, rigor and reproducibility of scientific results by *a priori* specifying which competing hypotheses are practically meaningful and by providing a more reliable statistical summary of when the data are compatible with alternative or null hypotheses.

OPEN ACCESS

Citation: Blume JD, D'Agostino McGowan L, Dupont WD, Greevy RA, Jr. (2018) Second-generation p-values: Improved rigor, reproducibility, & transparency in statistical analyses. PLoS ONE 13(3): e0188299. <https://doi.org/10.1371/journal.pone.0188299>

Editor: Neil R. Smalheiser, University of Illinois-Chicago, UNITED STATES

Received: May 25, 2017
Accepted: November 4, 2017
Published: March 22, 2018

Copyright: © 2018 Blume et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The raw data are publicly available online via the “*globeEts*” package (<http://www.datamocracy.org/globe/>) and GitHub (<https://github.com/BlumeJD/globeEts>). You can also access the data via <http://biostat.vanderbilt.edu/publications/globe/globe.html>.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

PLOS ONE | <https://doi.org/10.1371/journal.pone.0188299> March 22, 2018

1 / 17

Second-generation p-values: Improved rigor, reproducibility, & transparency in statistical analyses

Jeffrey D. Blume, Lucy D'Agostino McGowan, William D. Dupont, Robert A. Greevy Jr, PLOS ONE, Published: March 22, 2018: <https://doi.org/10.1371/journal.pone.0188299>

Need a better idea of what the **Null Distribution** of the data happens to be!
But just collecting more data may not be helpful.



A public repository of unthresholded statistical maps, parcellations, and atlases of the brain.

spmT 0001

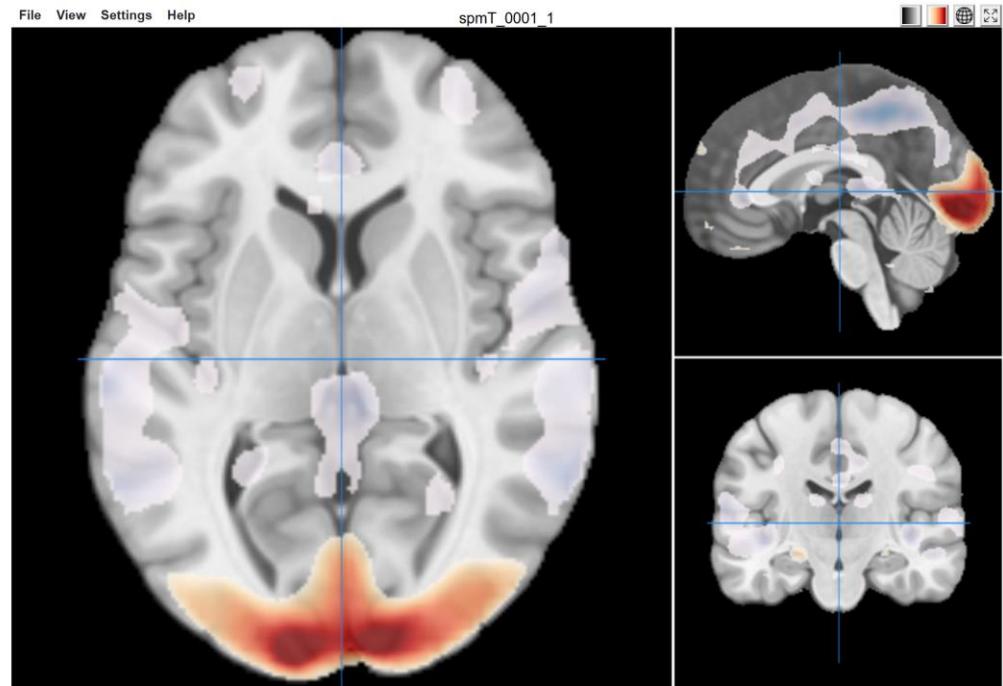
Contributed by jflournoy on Jan. 29, 2016

Collection: Neural Reactivity to Emotional Faces May Mediate the Relationship Between Childhood Empathy and Adolescent Prosocial Behavior

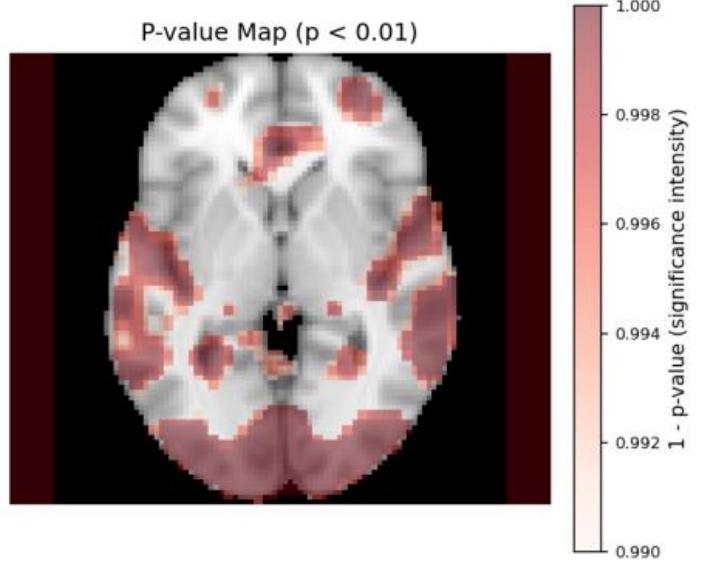
Description: SPM($T_{[53.0]}$) - contrast 1: Mean

Task View 3D View Download ▾ Analysis ▾

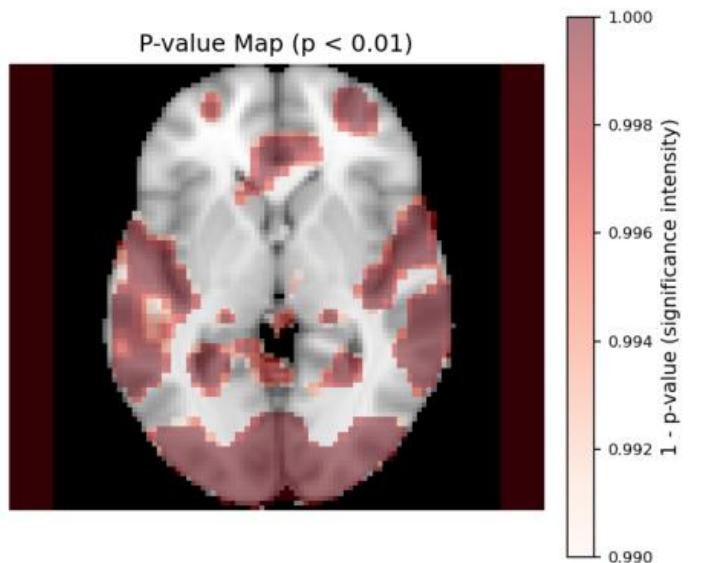
Warning: This map is missing some mandatory metadata!



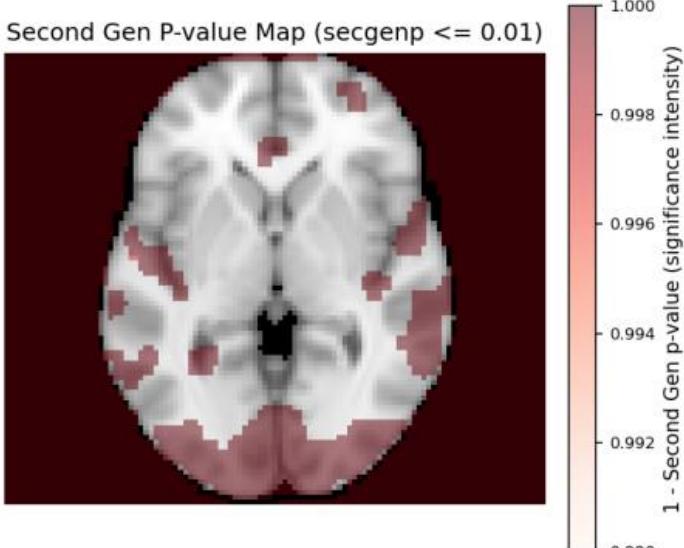
Stat P (Subjects = 20, alpha = 0.01)



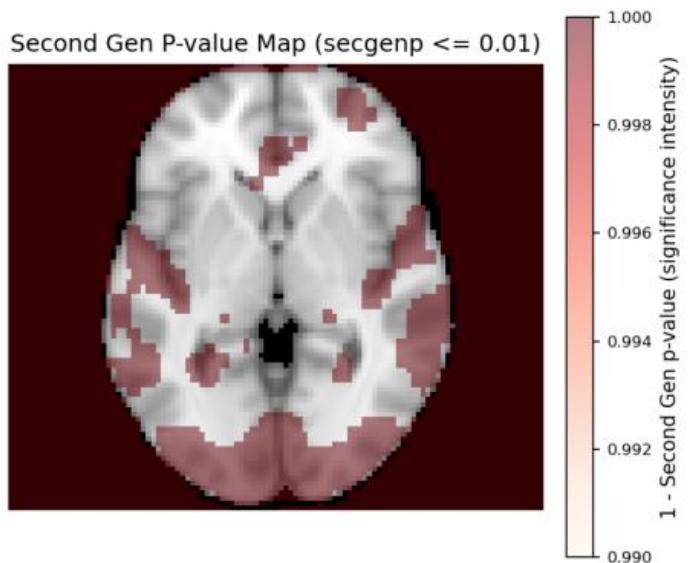
Stat P (Subjects = 2000, alpha = 0.01)



Second Gen P (Subjects = 20, alpha = 0.01)



Second Gen P (Subjects = 2000, alpha = 0.01)



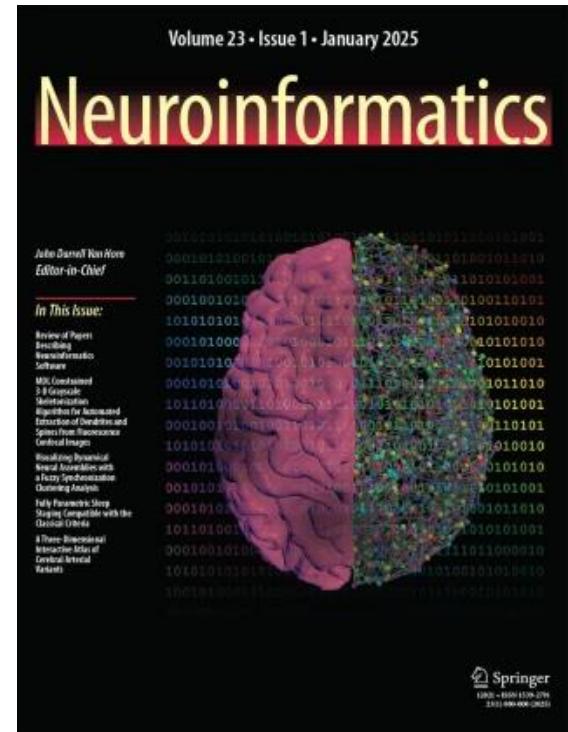
Is fMRI Alone Enough?

- fMRI seeks to describe task induced or resting signal changes over time
- The GLM has been the workhorse in cognitive neuroscience for ~30 years
- Functional connectivity studies depend upon region-to-region correlations
- BUT, like causality, correlations do not imply connection, *per se*
- Attempts to correlate functional networks with structural networks are kind of unsatisfying (don't you think?)
- Wildly popular, is ML/DL/AI the answer?
- Can we draw inspiration on how best to model neural dynamics from other branches of science?

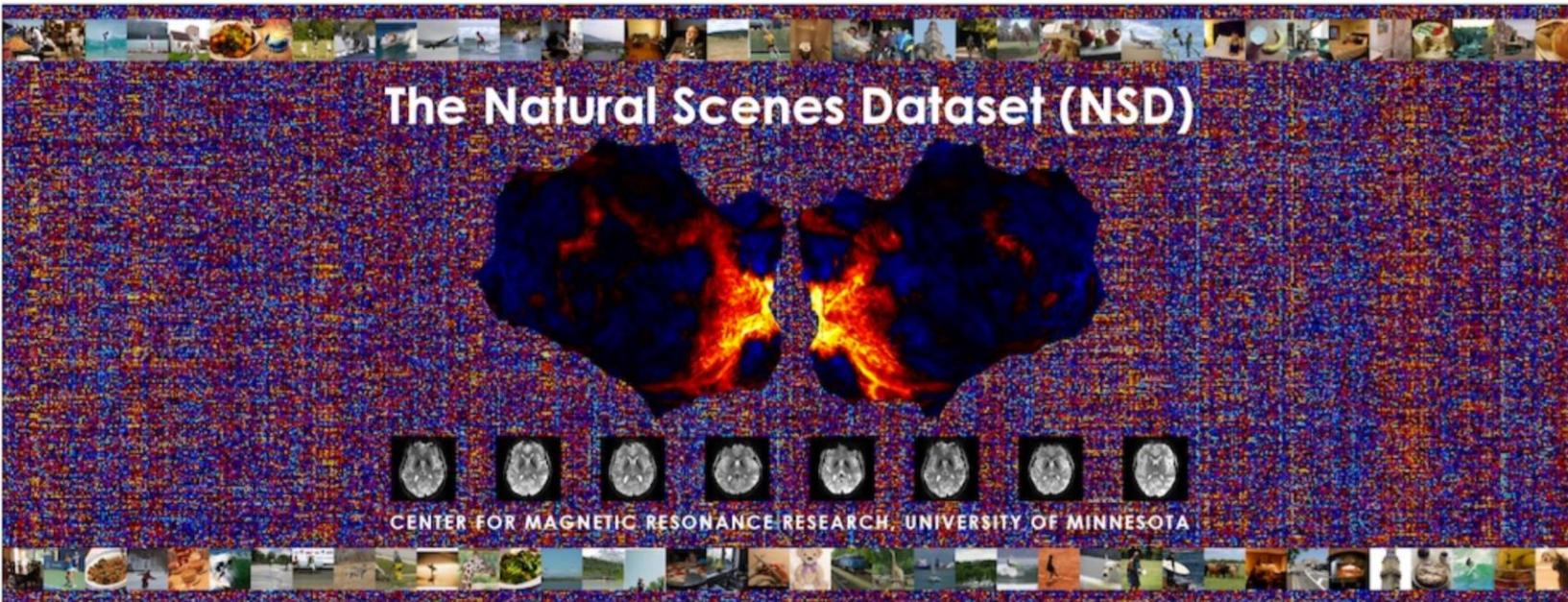
A Comment on ML

The Typical ML/DL/AI Study Submitted to Neuroinformatics

- Disease X is a worldwide burden on patients and clinicians
- Early detection is vital
- We downloaded data on disease X from the internet
- We propose that a machine/deep learning/AI model is needed
- Insert highly complicated and difficult to follow model here
- AUC and other performance metrics are near perfect
- Thus, our model correctly classifies disease X cases
- These results indicate that this may be useful in early diagnosis of X

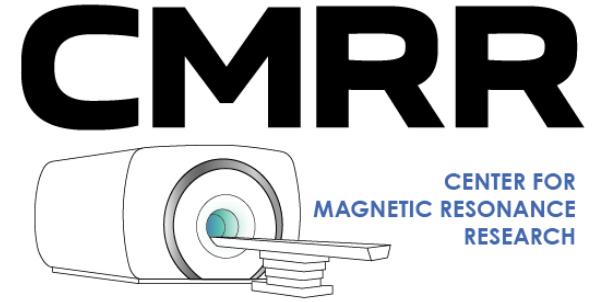


Natural Scenes Dataset (NSD)



News:

- March 11, 2025 - The NSD synthetic data (one additional 7T fMRI scan session) have now been publicly released.
- April 2, 2024: Take the [NSD / large-scale neuroimaging dataset anonymous survey!](#) Deadline May 15, 2024.
- January 16, 2023: Announcing that NSD data are used as part of the [2023 Algonauts Challenge](#)!
- January 13, 2023: A list of papers and pre-prints using NSD data added below.
- December 16, 2021: The [NSD data paper](#) is now published.
- September 3, 2021: The [NSD dataset](#) is now publicly available.



The Natural Scenes Dataset (NSD) is a large-scale fMRI dataset conducted at ultra-high-field (7T) strength at the [Center of Magnetic Resonance Research \(CMRR\)](#), at the University of Minnesota. The dataset consists of whole-brain, high-resolution (1.8-mm isotropic, 1.6-s sampling rate) fMRI measurements of 8 healthy adult subjects while they viewed thousands of color natural scenes over the course of 30–40 scan sessions. While viewing these images, subjects were engaged in a continuous recognition task in which they reported whether they had seen each given image at any point in the experiment. These data constitute a massive benchmark dataset for computational models of visual representation and cognition, and can support a wide range of scientific inquiry.

For a formal description of the dataset, please see the NSD data paper:

“High-resolution image reconstruction with latent diffusion models from human brain activity” (Takagi and Nishimoto, 2023, *bioRxiv*)

<https://www.biorxiv.org/content/10.1101/2022.11.18.517004v2.full.pdf>



Figure 1. Presented images (red box, top row) and images reconstructed from fMRI signals (gray box, bottom row) for one subject (subj01).

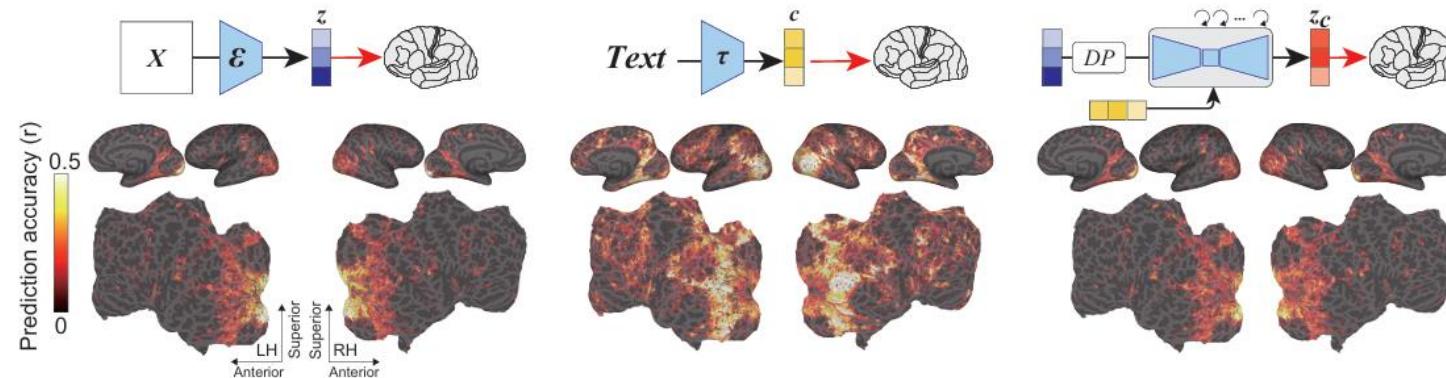
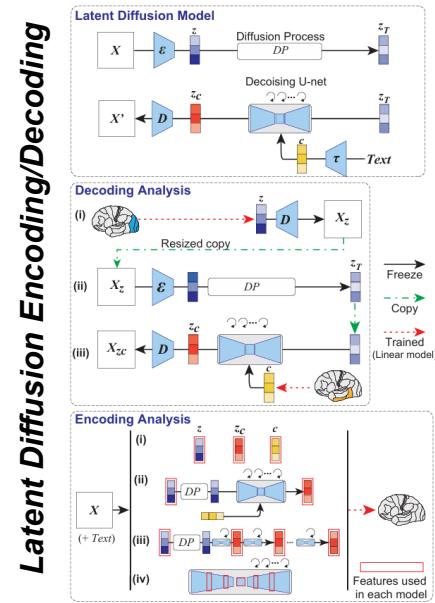
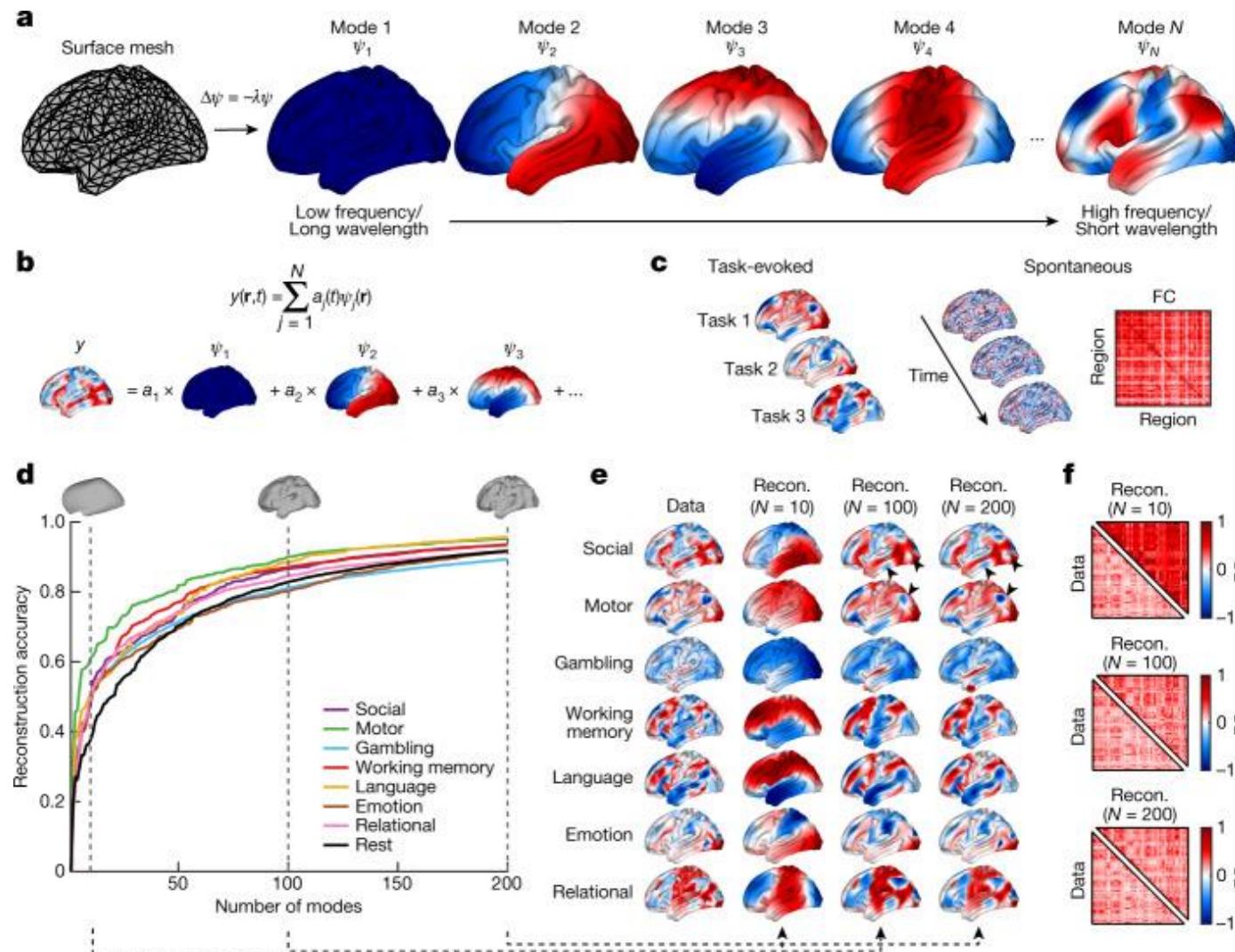


Figure 6. Prediction performance (measured using Pearson's correlation coefficients) for the voxel-wise encoding model applied to held-out test images in a single subject (subj01), projected onto the inflated (top, lateral and medial views) and flattened cortical surface (bottom, occipital areas are at the center), for both left and right hemispheres. Brain regions with significant accuracy are colored (all colored voxels $P < 0.05$, FDR corrected).

Developing Mathematical Models

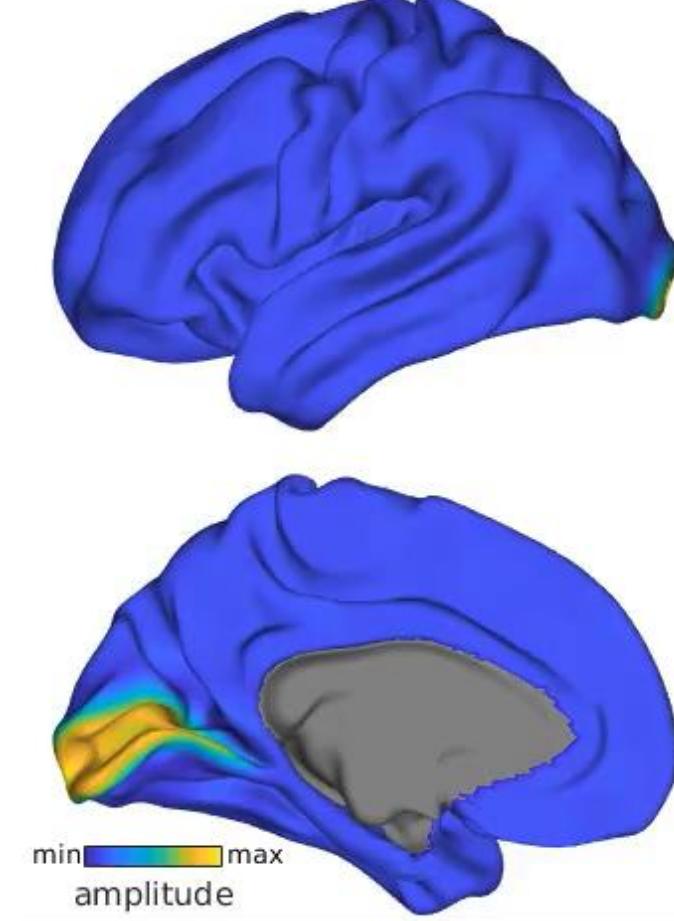
Functional Eigenmodes



Helmholtz Equation

$$\nabla^2 \psi = \Delta \psi = -\lambda \psi,$$

$t = 1.0 \text{ ms}$



The Deconstructed Standard Model Equation of Particle Physics

The Whole Thing

This version of the Standard Model is written in the Lagrangian form. The Lagrangian is a fancy way of writing an equation to determine the state of a changing system and explain the maximum possible energy the system can maintain. Technically, the Standard Model can be written in several different formulations, but, despite appearances, the Lagrangian is one of the easiest and most compact ways of presenting the theory.

$$\begin{aligned}
 & 1 -\frac{1}{2}\partial_\nu g_\mu^a \partial_\nu g_\mu^a - g_s f^{abc} \partial_\mu g_\nu^b g_\mu^c - \frac{1}{4}g_s^2 f^{abc} f^{ade} g_\mu^b g_\nu^c g_\mu^d g_\nu^e + \\
 & \quad \frac{1}{2}ig_s^2 (\bar{q}_i^\sigma \gamma^\mu q_j^\sigma) g_\mu^a + \bar{G}^a \partial^2 G^a + g_s f^{abc} \partial_\mu \bar{G}^a G^b g_\mu^c - \partial_\nu W_\mu^+ \partial_\mu W_\mu^- - \\
 & 2 M^2 W_\mu^+ W_\mu^- - \frac{1}{2} \partial_\nu Z_\mu^0 \partial_\nu Z_\mu^0 - \frac{1}{2c_w^2} M^2 Z_\mu^0 Z_\mu^0 - \frac{1}{2} \partial_\mu A_\nu \partial_\mu A_\nu - \frac{1}{2} \partial_\mu H \partial_\mu H - \\
 & \quad \frac{1}{2} m_h^2 H^2 - \partial_\mu \phi^+ \partial_\mu \phi^- - M^2 \phi^+ \phi^- - \frac{1}{2} \partial_\mu \phi^0 \partial_\mu \phi^0 - \frac{1}{2c_w^2} M \phi^0 \phi^0 - \beta_h [\frac{2M^2}{g^2} + \\
 & \quad \frac{2M}{g} H + \frac{1}{2}(H^2 + \phi^0 \phi^0 + 2\phi^+ \phi^-)] + \frac{2M^4}{g^2} \alpha_h - ig c_w [\partial_\nu Z_\mu^0 (W_\mu^+ W_\nu^- - \\
 & \quad W_\nu^+ W_\mu^-) - Z_\nu^0 (W_\mu^+ \partial_\nu W_\mu^- - W_\mu^- \partial_\nu W_\mu^+) + Z_\mu^0 (W_\nu^+ \partial_\nu W_\mu^- - \\
 & \quad W_\nu^- \partial_\nu W_\mu^+) - ig s_w [\partial_\nu A_\mu (W_\mu^+ W_\nu^- - W_\nu^+ W_\mu^-) - A_\nu (W_\mu^+ \partial_\nu W_\mu^- - \\
 & \quad W_\mu^- \partial_\nu W_\mu^+) + A_\mu (W_\nu^+ \partial_\nu W_\mu^- - W_\nu^- \partial_\nu W_\mu^+)] - \frac{1}{2} g^2 W_\mu^+ W_\mu^- W_\nu^+ W_\nu^- + \\
 & \quad \frac{1}{2} g^2 W_\mu^+ W_\nu^- W_\mu^- W_\nu^+ + g^2 c_w^2 (Z_\mu^0 W_\mu^+ Z_\nu^0 W_\nu^- - Z_\mu^0 Z_\mu^0 W_\nu^+ W_\nu^-) + \\
 & \quad g^2 s_w^2 (A_\mu W_\mu^+ A_\nu W_\nu^- - A_\mu A_\nu W_\mu^+ W_\nu^-) + g^2 s_w c_w [A_\mu Z_\nu^0 (W_\mu^+ W_\nu^- - \\
 & \quad W_\nu^+ W_\mu^-) - 2 A_\mu Z_\mu^0 W_\nu^+ W_\nu^-] - g \alpha [H^3 + H \phi^0 \phi^0 + 2 H \phi^+ \phi^-] - \\
 & \quad \frac{1}{8} g^2 \alpha_h [H^4 + (\phi^0)^4 + 4(\phi^+ \phi^-)^2 + 4(\phi^0)^2 \phi^+ \phi^- + 4 H^2 \phi^+ \phi^- + 2(\phi^0)^2 H^2] - \\
 & \quad g M W_\mu^+ W_\mu^- H - \frac{1}{2} g \frac{M}{c_w^2} Z_\mu^0 Z_\mu^0 H - \frac{1}{2} ig [W_\mu^+ (\phi^0 \partial_\mu \phi^- - \phi^- \partial_\mu \phi^0) - \\
 & \quad W_\mu^- (\phi^0 \partial_\mu \phi^+ - \phi^+ \partial_\mu \phi^0)] + \frac{1}{2} g [W_\mu^+ (H \partial_\mu \phi^- - \phi^- \partial_\mu H) - W_\mu^- (H \partial_\mu \phi^+ - \\
 & \quad \phi^+ \partial_\mu H)] + \frac{1}{2} g \frac{1}{c_w} (Z_\mu^0 (H \partial_\mu \phi^0 - \phi^0 \partial_\mu H) - ig \frac{s_w^2}{c_w} M Z_\mu^0 (W_\mu^+ \phi^- - W_\mu^- \phi^+) + \\
 & \quad ig s_w M A_\mu (W_\mu^+ \phi^- - W_\mu^- \phi^+) - ig \frac{1-2c_w^2}{2c_w} Z_\mu^0 (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) + \\
 & \quad ig s_w A_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) - \frac{1}{4} g^2 W_\mu^+ W_\mu^- [H^2 + (\phi^0)^2 + 2\phi^+ \phi^-] - \\
 & \quad \frac{1}{4} g^2 \frac{1}{c_w^2} Z_\mu^0 Z_\mu^0 [H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+ \phi^-] - \frac{1}{2} g^2 \frac{s_w^2}{c_w} Z_\mu^0 \phi^0 (W_\mu^+ \phi^- + \\
 & \quad W_\mu^- \phi^+) - \frac{1}{2} ig^2 \frac{s_w^2}{c_w} Z_\mu^0 H (W_\mu^+ \phi^- - W_\mu^- \phi^+) + \frac{1}{2} g^2 s_w A_\mu \phi^0 (W_\mu^+ \phi^- + \\
 & \quad W_\mu^- \phi^+) + \frac{1}{2} ig^2 s_w A_\mu H (W_\mu^+ \phi^- - W_\mu^- \phi^+) - g^2 \frac{s_w}{c_w} (2c_w^2 - 1) Z_\mu^0 A_\mu \phi^+ \phi^- - \\
 & \quad g^2 s_w^2 A_\mu \phi^+ \phi^- - \bar{e}^\lambda (\gamma \partial + m_e^\lambda) e^\lambda - \bar{\nu}^\lambda \gamma \partial \nu^\lambda - \bar{u}_j^\lambda (\gamma \partial + m_u^\lambda) u_j^\lambda - \\
 & 3 d_j^\lambda (\gamma \partial + m_d^\lambda) d_j^\lambda + ig s_w A_\mu [-(\bar{e}^\lambda \gamma^\mu e^\lambda) + \frac{2}{3} (\bar{u}_j^\lambda \gamma^\mu u_j^\lambda) - \frac{1}{3} (\bar{d}_j^\lambda \gamma^\mu d_j^\lambda)] + \\
 & \quad \frac{ig}{4c_w} Z_\mu^0 [(\bar{\nu}^\lambda \gamma^\mu (1 + \gamma^5) \nu^\lambda) + (\bar{e}^\lambda \gamma^\mu (4s_w^2 - 1 - \gamma^5) e^\lambda) + (\bar{u}_j^\lambda \gamma^\mu (\frac{4}{3}s_w^2 - \\
 & \quad 1 - \gamma^5) u_j^\lambda) + (\bar{d}_j^\lambda \gamma^\mu (1 - \frac{8}{3}s_w^2 - \gamma^5) d_j^\lambda)] + \frac{ig}{2\sqrt{2}} W_\mu^+ [(\bar{\nu}^\lambda \gamma^\mu (1 + \gamma^5) e^\lambda) + \\
 & \quad (\bar{u}_j^\lambda \gamma^\mu (1 + \gamma^5) C_{\lambda\kappa} d_j^\kappa)] + \frac{ig}{2\sqrt{2}} W_\mu^- [(\bar{e}^\lambda \gamma^\mu (1 + \gamma^5) \nu^\lambda) + (\bar{d}_j^\kappa C_{\lambda\kappa}^\dagger \gamma^\mu (1 + \\
 & \quad \gamma^5) u_j^\lambda)] + \frac{ig}{2\sqrt{2}} \frac{m_e^\lambda}{M} [-\phi^+ (\bar{\nu}^\lambda (1 - \gamma^5) e^\lambda) + \phi^- (\bar{e}^\lambda (1 + \gamma^5) \nu^\lambda)] - \\
 & 4 \frac{g}{2} \frac{m_e^\lambda}{M} [H(\bar{e}^\lambda e^\lambda) + i\phi^0(\bar{e}^\lambda \gamma^5 e^\lambda)] + \frac{ig}{2M\sqrt{2}} \phi^+ [-m_d^\kappa (\bar{u}_j^\lambda C_{\lambda\kappa} (1 - \gamma^5) d_j^\kappa) + \\
 & \quad m_u^\lambda (\bar{u}_j^\lambda C_{\lambda\kappa} (1 + \gamma^5) d_j^\kappa)] + \frac{ig}{2M\sqrt{2}} \phi^- [m_d^\lambda (\bar{d}_j^\lambda C_{\lambda\kappa}^\dagger (1 + \gamma^5) u_j^\kappa) - m_u^\kappa (\bar{d}_j^\lambda C_{\lambda\kappa}^\dagger (1 - \\
 & \quad \gamma^5) u_j^\kappa)] - \frac{g}{2} \frac{m_d^\lambda}{M} H(\bar{u}_j^\lambda u_j^\lambda) - \frac{g}{2} \frac{m_u^\lambda}{M} H(\bar{d}_j^\lambda d_j^\lambda) + \frac{ig}{2} \frac{m_d^\lambda}{M} \phi^0 (\bar{u}_j^\lambda \gamma^5 u_j^\lambda) - \\
 & \quad \frac{ig}{2} \frac{m_u^\lambda}{M} \phi^0 (\bar{d}_j^\lambda \gamma^5 d_j^\lambda) + \bar{X}^+(\partial^2 - M^2) X^+ + \bar{X}^-(\partial^2 - M^2) X^- + \bar{X}^0(\partial^2 - \\
 & 5 \frac{M^2}{c_w^2}) X^0 + \bar{Y} \partial^2 Y + ig c_w W_\mu^+ (\partial_\mu \bar{X}^0 X^- - \partial_\mu \bar{X}^+ X^0) + ig s_w W_\mu^+ (\partial_\mu \bar{Y} X^- - \\
 & \quad \partial_\mu \bar{X}^+ Y) + ig c_w W_\mu^- (\partial_\mu \bar{X}^- X^0 - \partial_\mu \bar{X}^0 X^+) + ig s_w W_\mu^- (\partial_\mu \bar{Y} X^- - \\
 & \quad \partial_\mu \bar{Y} X^+) + ig c_w Z_\mu^0 (\partial_\mu \bar{X}^+ X^+ - \partial_\mu \bar{X}^- X^-) + ig s_w A_\mu (\partial_\mu \bar{X}^+ X^+ - \\
 & \quad \partial_\mu \bar{X}^- X^-) - \frac{1}{2} g M [\bar{X}^+ X^+ H + \bar{X}^- X^- H + \frac{1}{c_w^2} \bar{X}^0 X^0 H] + \\
 & \quad \frac{1-2c_w^2}{2c_w} ig M [\bar{X}^+ X^0 \phi^+ - \bar{X}^- X^0 \phi^-] + \frac{1}{2c_w} ig M [\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + \\
 & \quad ig M s_w [\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + \frac{1}{2} ig M [\bar{X}^+ X^+ \phi^0 - \bar{X}^- X^- \phi^0]
 \end{aligned}$$

Section 1

These three lines in the Standard Model are ultraspecific to the gluon, the boson that carries the strong force. Gluons come in eight types, interact among themselves and have what's called a color charge.

Section 2

Almost half of this equation is dedicated to explaining interactions between bosons, particularly W and Z bosons. Bosons are force-carrying particles, and there are four species of bosons that interact with other particles using three fundamental forces. Photons carry electromagnetism, gluons carry the strong force and W and Z bosons carry the weak force. The most recently discovered boson, the Higgs boson, is a bit different; its interactions appear in the next part of the equation.

Section 3

This part of the equation describes how elementary matter particles interact with the weak force. According to this formulation, matter particles come in three generations, each with different masses. The weak force helps massive matter particles decay into less massive matter particles. This section also includes basic interactions with the Higgs field, from which some elementary particles receive their mass. Intriguingly, this part of the equation makes an assumption that contradicts discoveries made by physicists in recent years. It incorrectly assumes that particles called neutrinos have no mass.

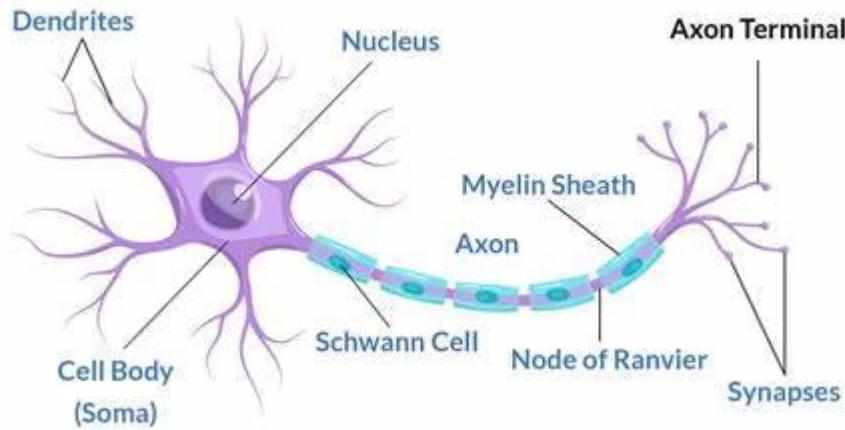
Section 4

In quantum mechanics, there is no single path or trajectory a particle can take, which means that sometimes redundancies appear in this type of mathematical formulation. To clean up these redundancies, theorists use virtual particles they call ghosts. This part of the equation describes how matter particles interact with Higgs ghosts, virtual artifacts from the Higgs field.

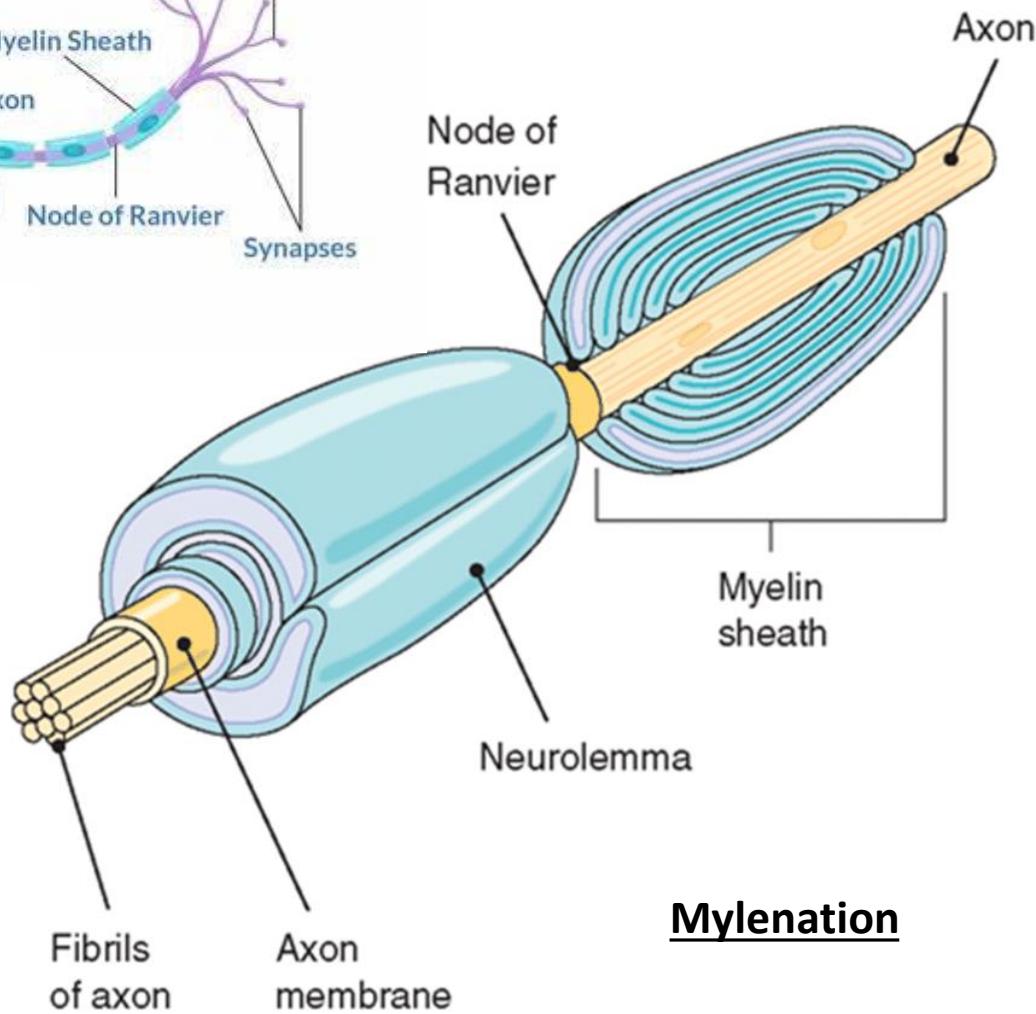
Section 5

This last part of the equation includes more ghosts. These ones are called Faddeev-Popov ghosts, and they cancel out redundancies that occur in interactions through the weak force.

Axons



Neuron



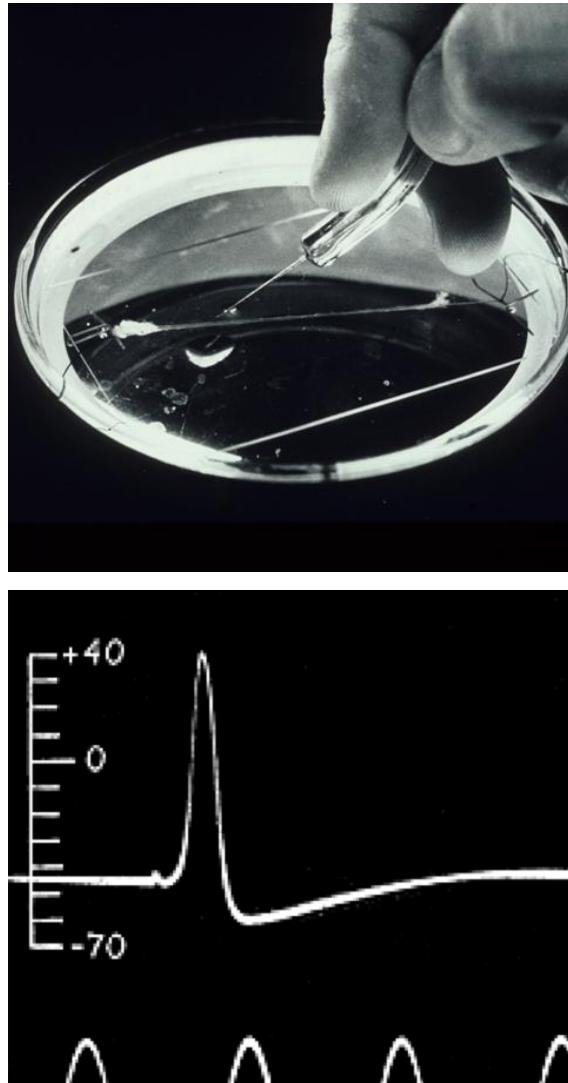
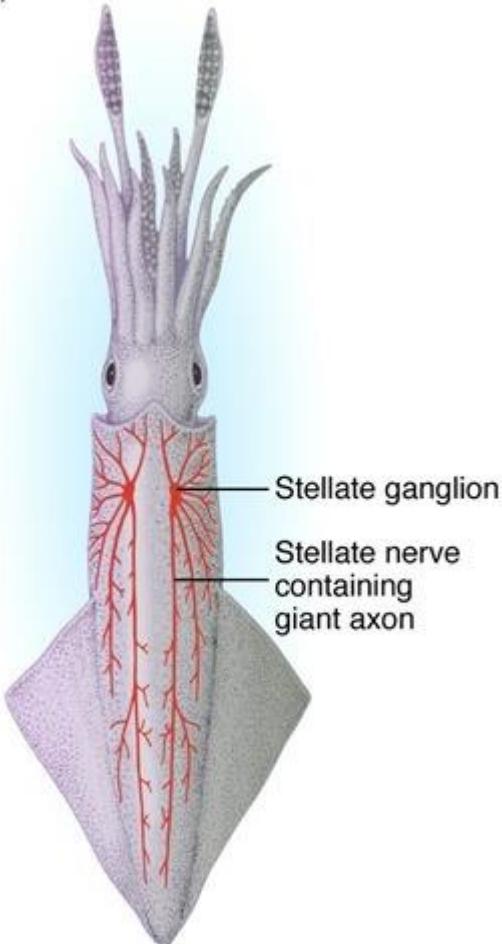
Mylenation



White Mater Fiber Bundles

Giant Squid Axon and Conduction

(a)



The Cable Equation

$$\pi d C_M \frac{dV(x,t)}{dt} + \pi d \frac{V(x,t) - E_r}{R_M} = \pi \frac{d^2}{4 R_A} \frac{\partial^2 V(x,t)}{\partial x^2} + \pi d i_{ext}(x,t)$$

C_M : Specific Capacitance (F/cm^2)

R_M : Specific Resistance ($\text{Ohm}\cdot\text{cm}^2$)

R_A : Specific Axial Resistance ($\text{Ohm}\cdot\text{cm}$)



Photo from the Nobel Foundation archive.
Sir John Carew Eccles
Prize share: 1/3

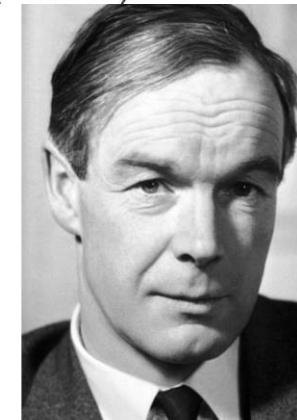


Photo from the Nobel Foundation archive.
Alan Lloyd Hodgkin
Prize share: 1/3

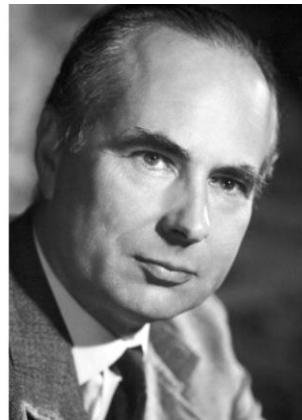


Photo from the Nobel Foundation archive.
Andrew Fielding Huxley
Prize share: 1/3

Telegrapher's Equations

$$\frac{\partial^2}{\partial x^2} V = LC \frac{\partial^2}{\partial t^2} V + (RC + GL) \frac{\partial}{\partial t} V + GRV,$$

$$\frac{\partial^2}{\partial x^2} I = LC \frac{\partial^2}{\partial t^2} I + (RC + GL) \frac{\partial}{\partial t} I + GRI.$$

Terms

R (Ohms) = resistance = mass·length² ·time⁻³·current⁻²

G (siemens) = conductance = mass⁻¹·length⁻³·time³·current²

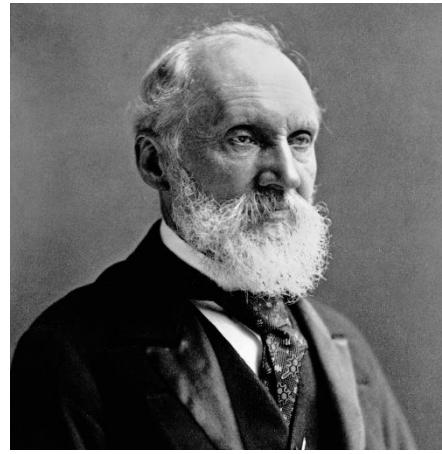
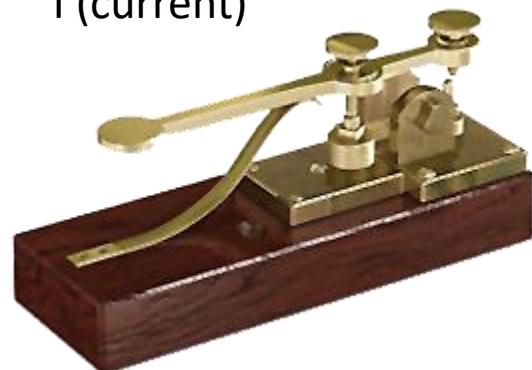
L (henry) = inductance = mass·length²·time⁻²·current⁻²

C (farads) = capacitance = mass⁻¹·length⁻²·time⁴·current²

A (Amperes)

V (volts)

I (current)



James
Clerk
Maxwell
(1831-1879)



Lord Kelvin (1824–1907)

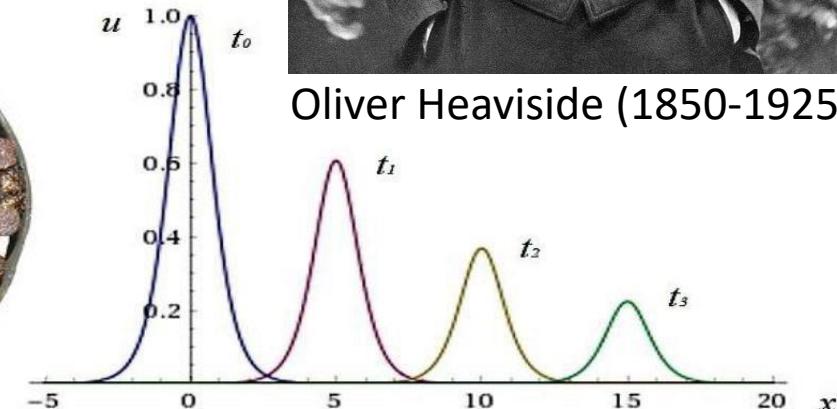
Products

LC = time²

RC = time

GL = time/length

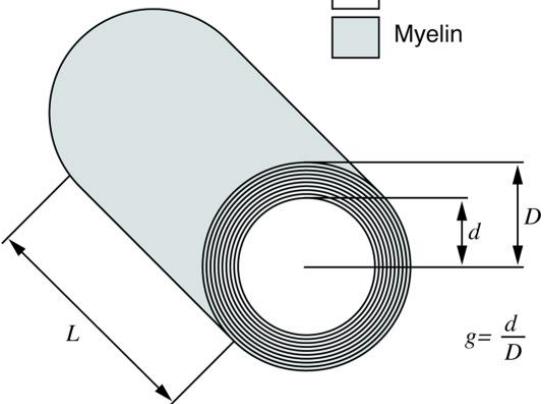
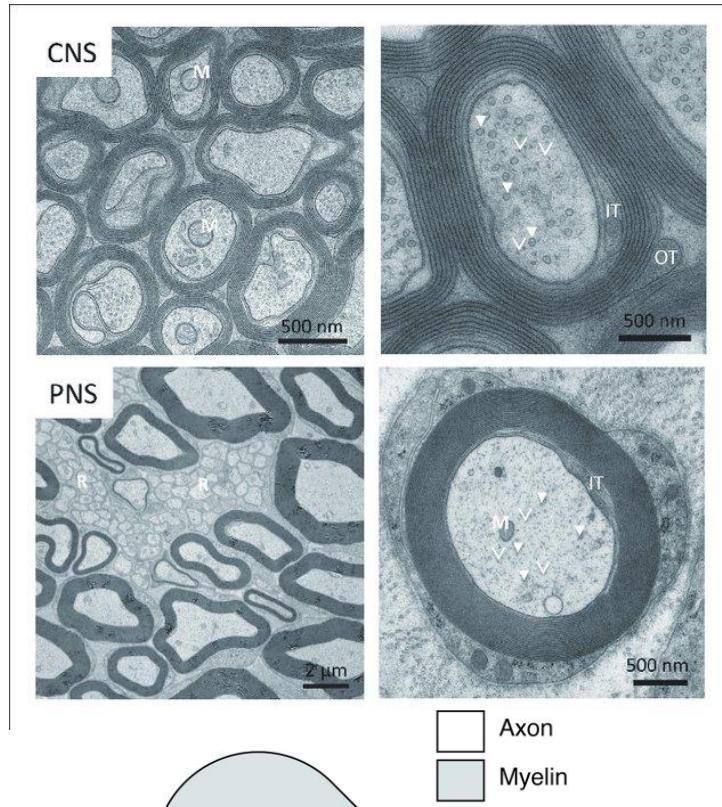
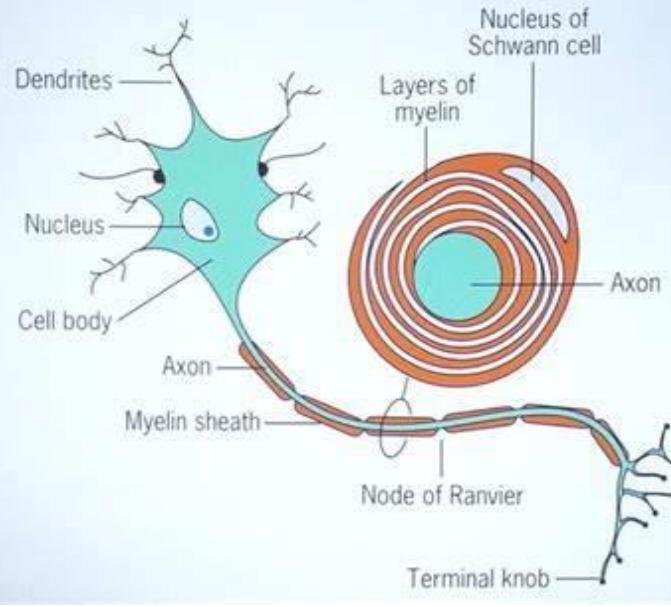
GR = 1/length



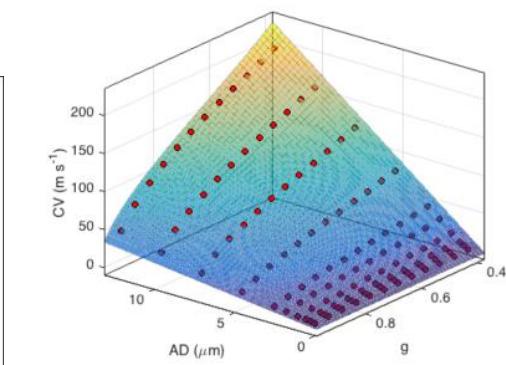
Oliver Heaviside (1850-1925)

G-ratio

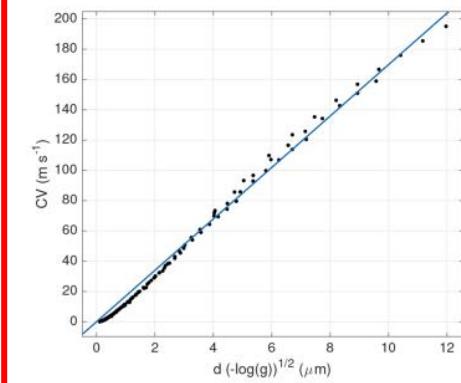
Neurons = Nerve cells



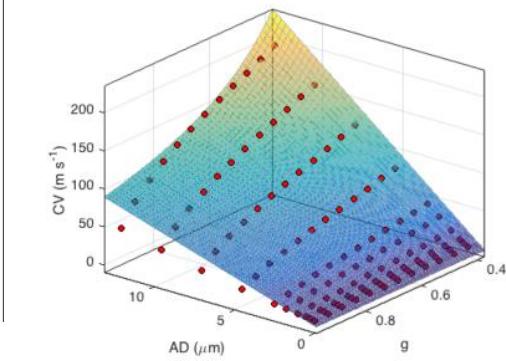
(a)



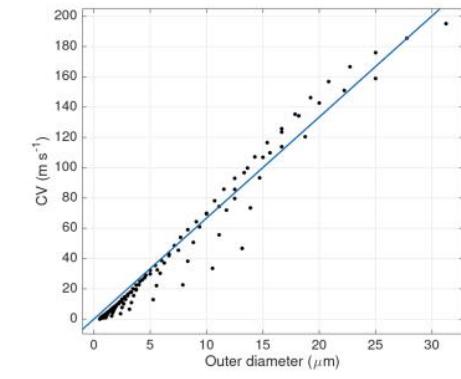
(b)



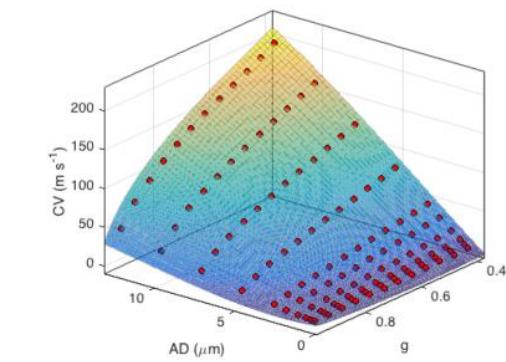
(c)



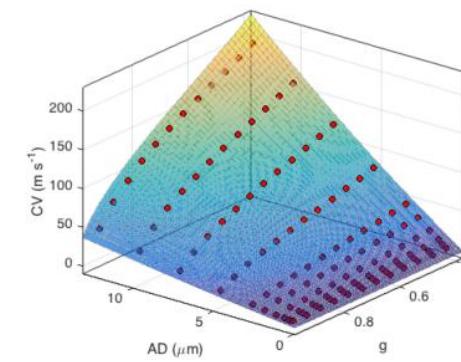
(d)



(e)



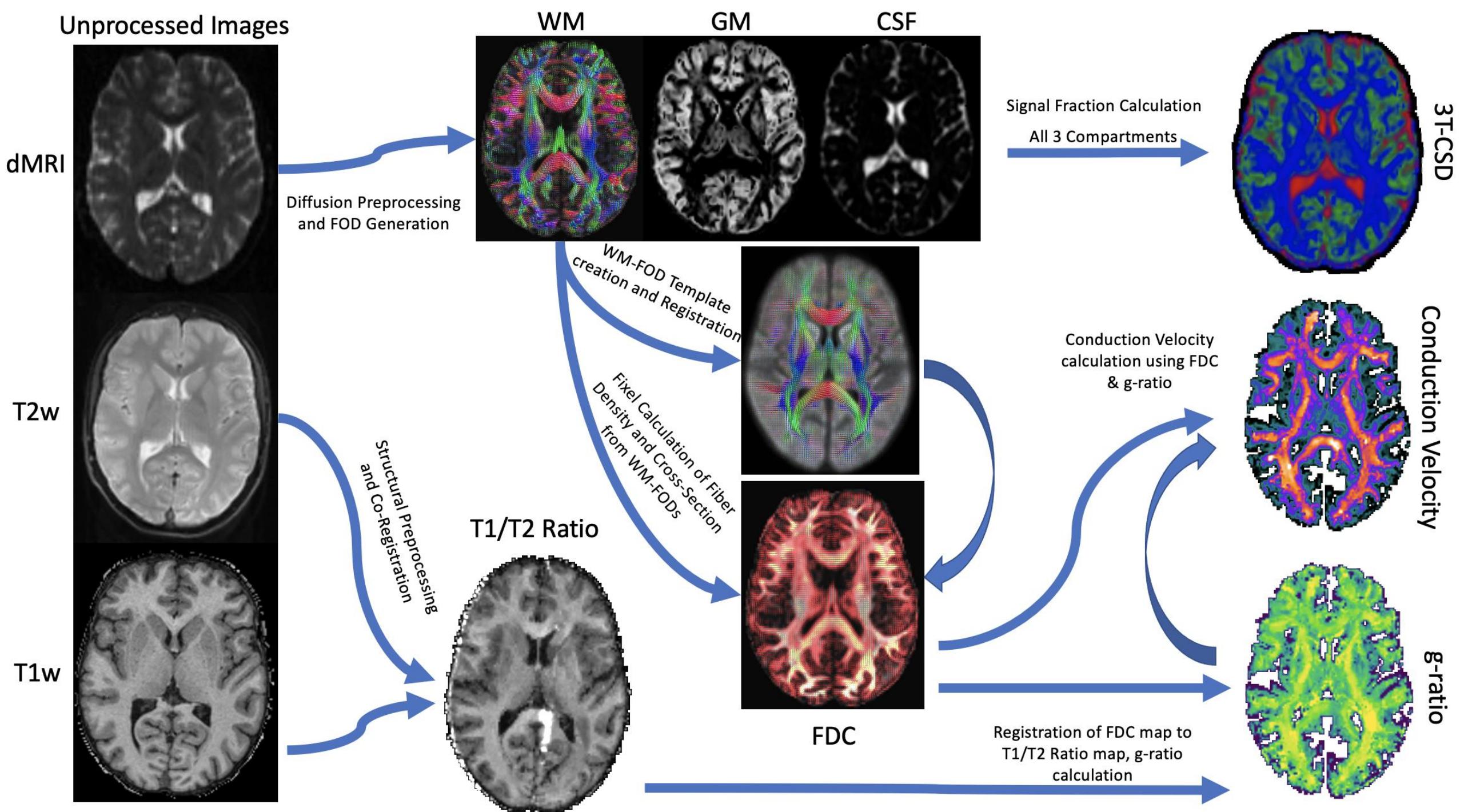
(f)



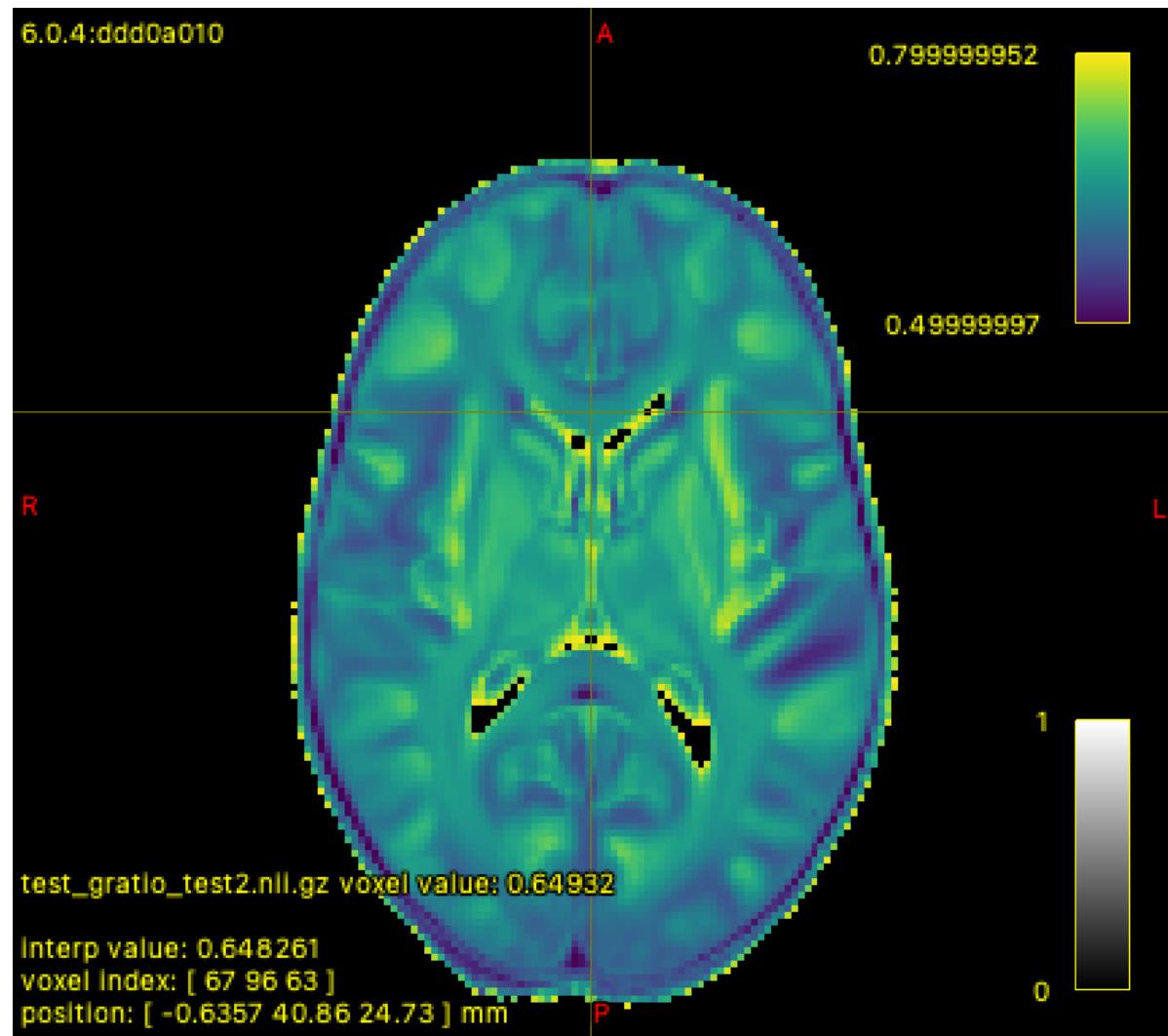
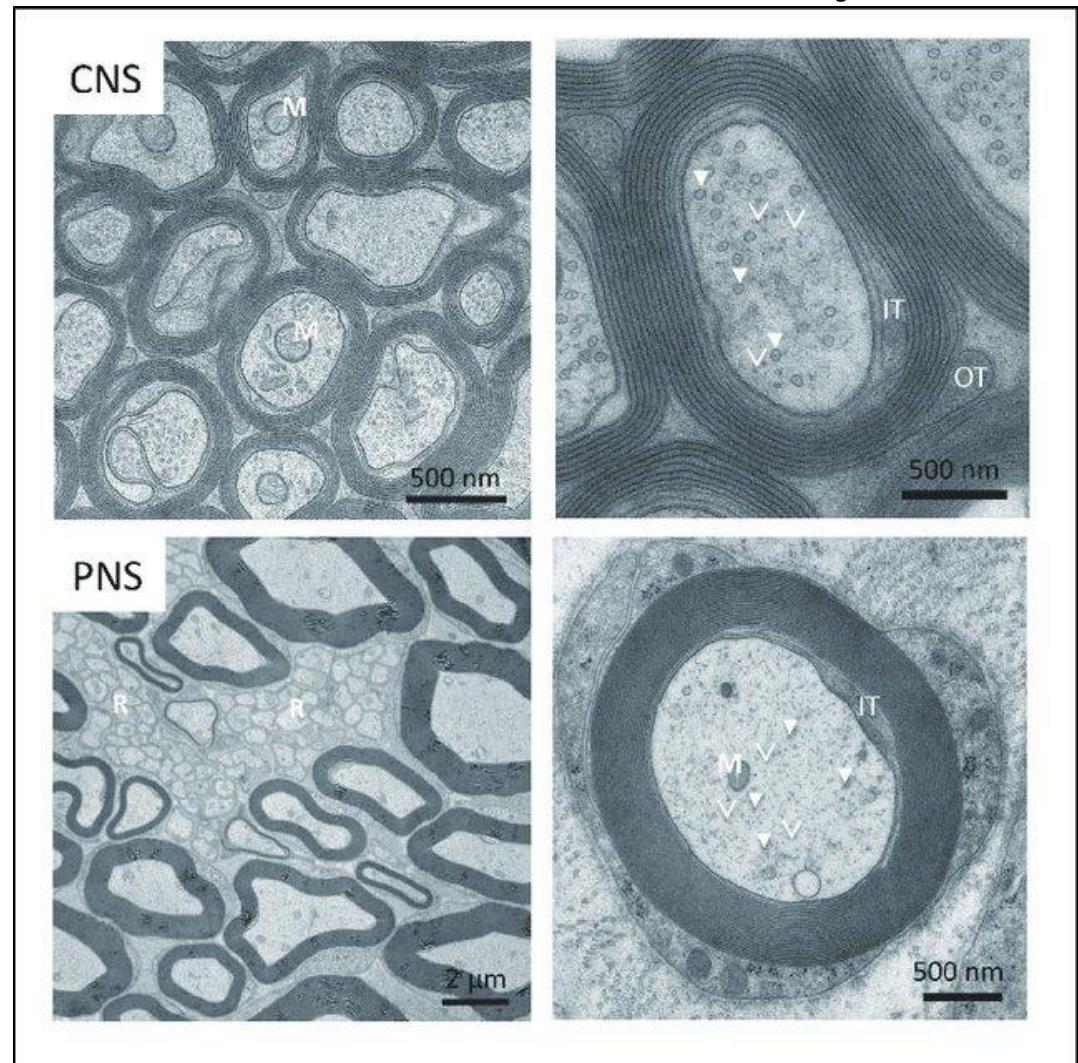
<http://neuroscientist.weebly.com/blog/lesson-2-the-materialistic-mind-your-brains-ingredients>

<https://pubmed.ncbi.nlm.nih.gov/31542512/#&gid=article-figures&pid=fig-4-uid-3>

https://www.researchgate.net/publication/336804352_Crowd_Control_Effects_of_Physical_Crowding_on_Cargo_Movement_in_Healthy_and_Diseased_Neurons/figures?lo=1

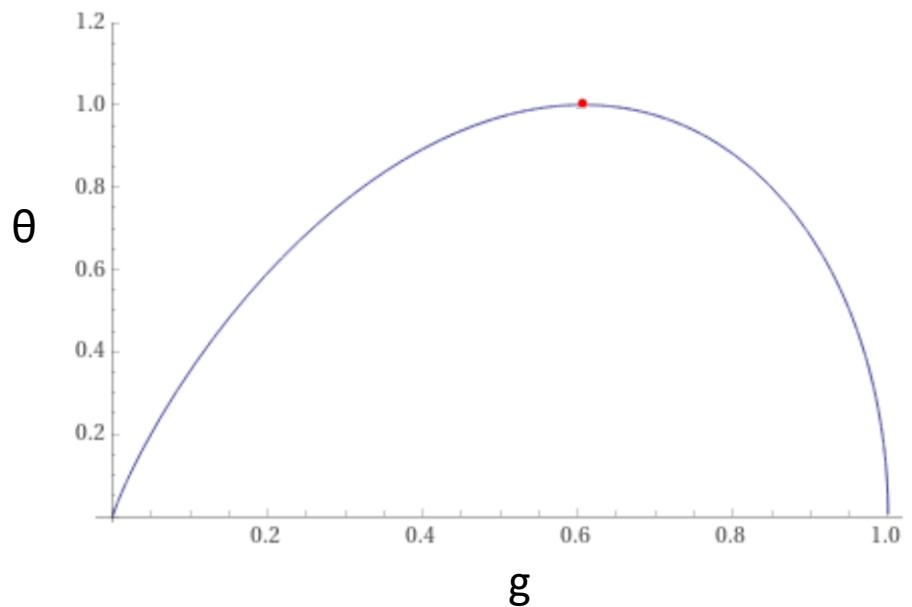


Axonal Geometry

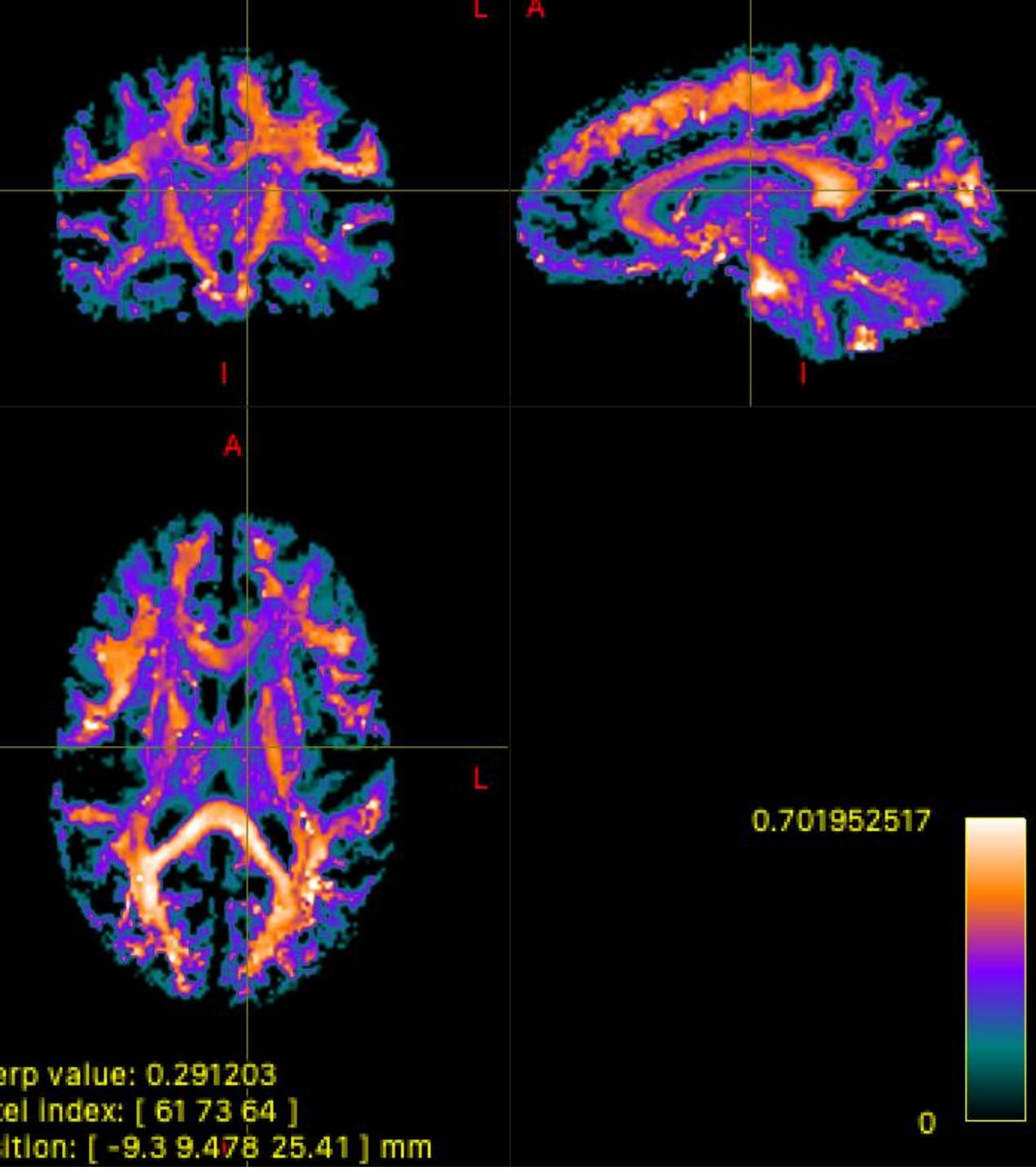


Conduction Velocity

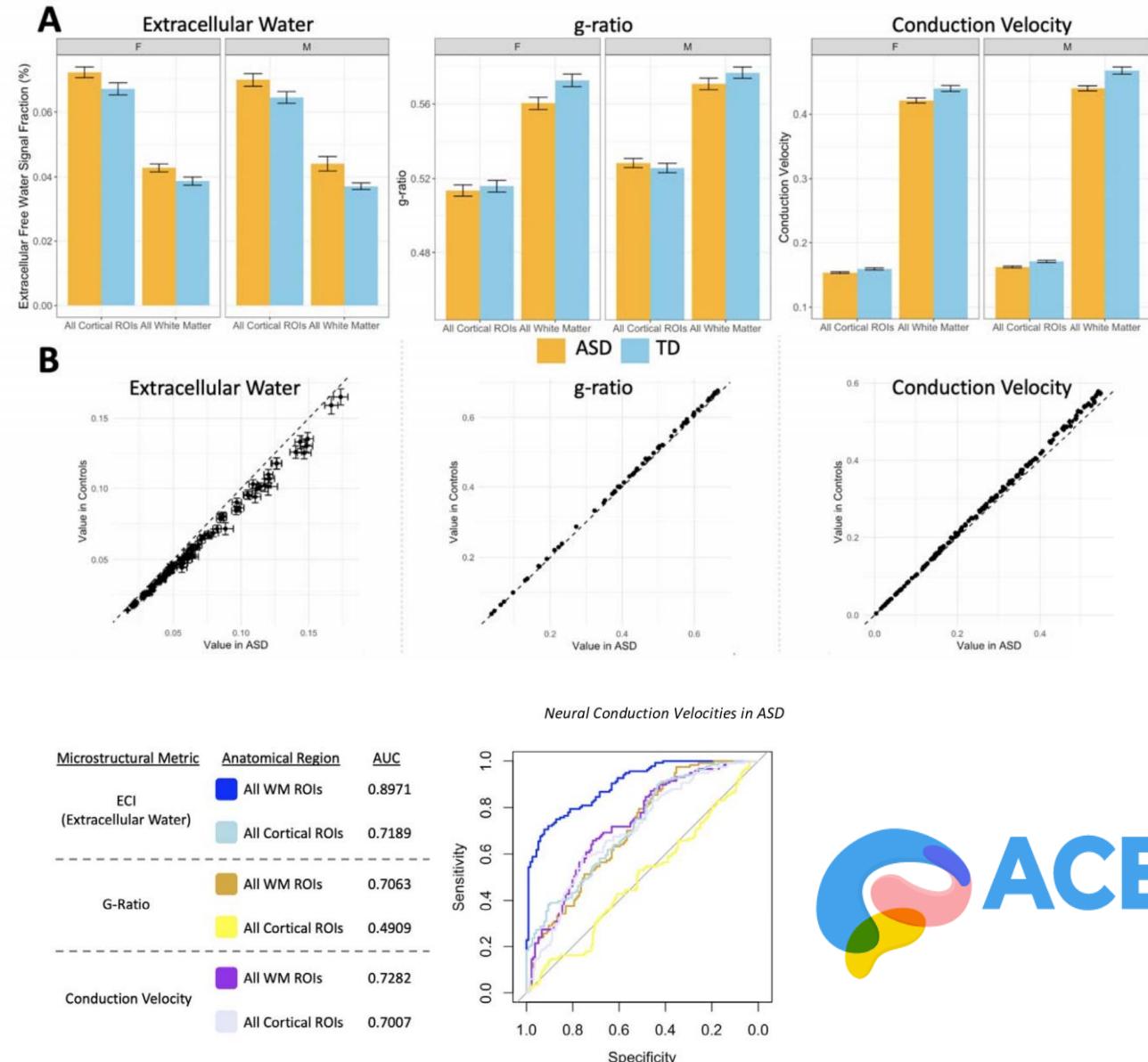
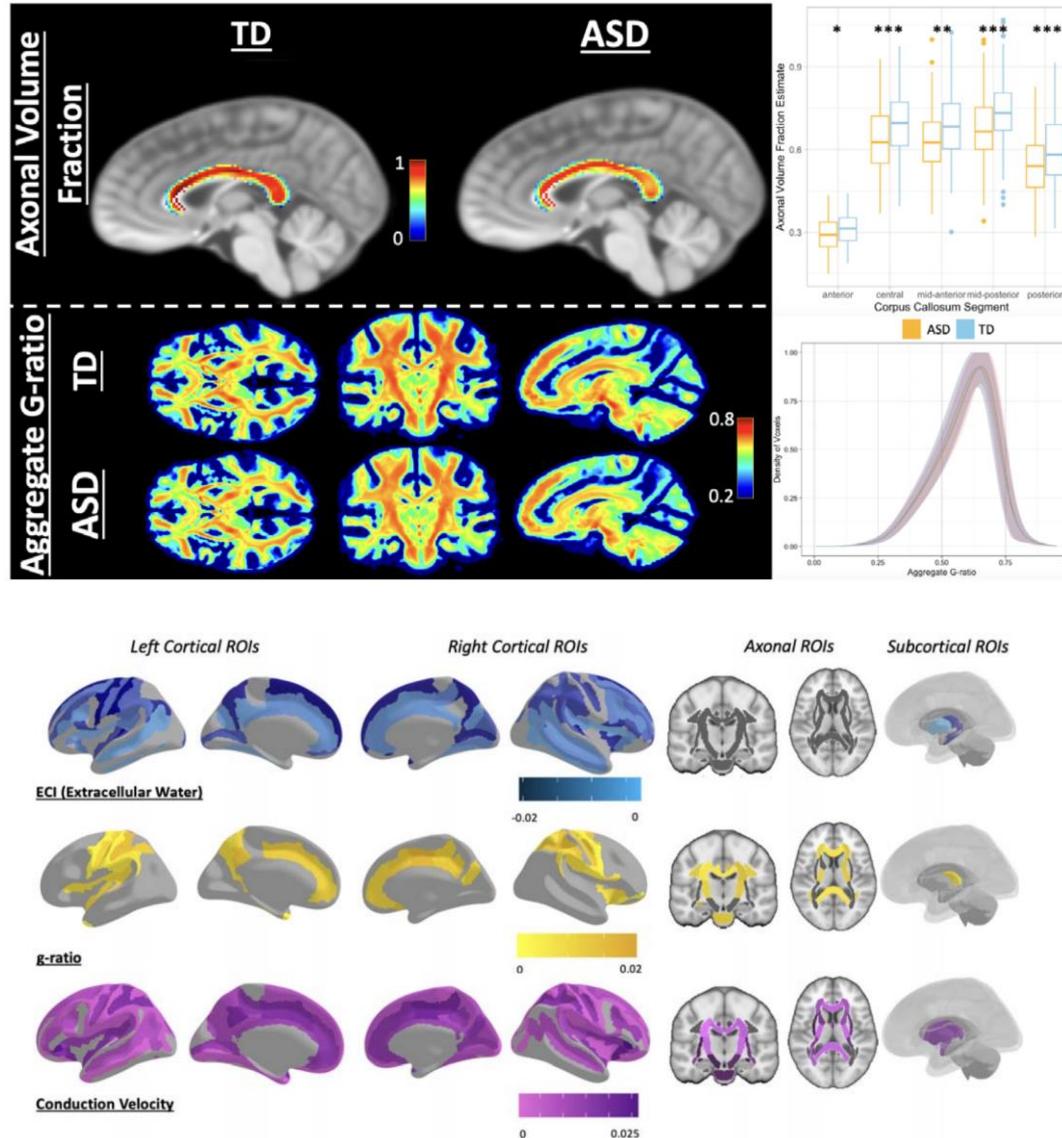
$$\theta \propto l \propto Dg \sqrt{\ln\left(\frac{1}{g}\right)} \propto d\sqrt{-\ln(g)}$$



$$\theta \propto l \propto de^{\frac{1}{2}} \sqrt{-2\ln(g)}$$



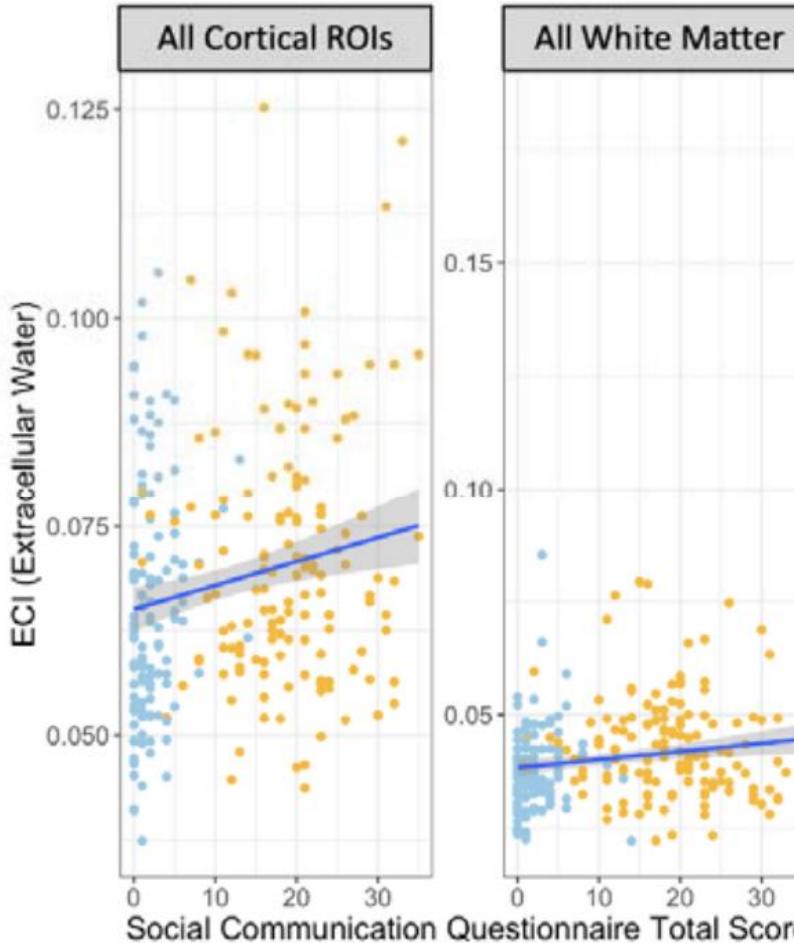
White Matter Microstructure in ASD vs. TD



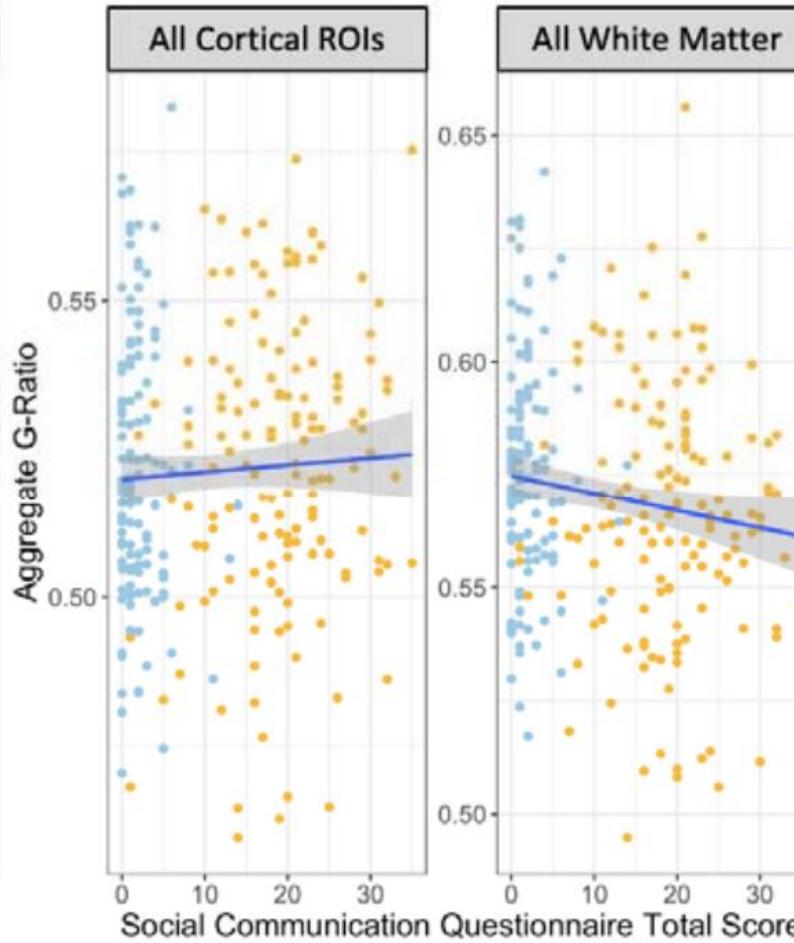
Microstructural Changes in ASD vs. TD



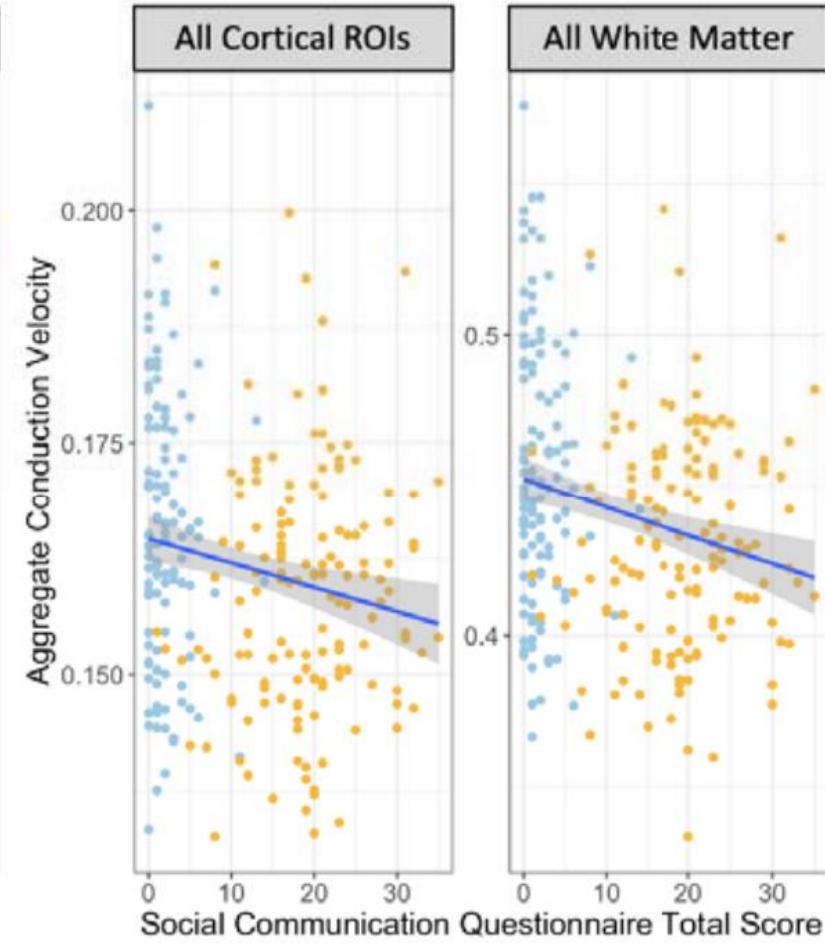
Extracellular Water



Aggregate G-Ratio

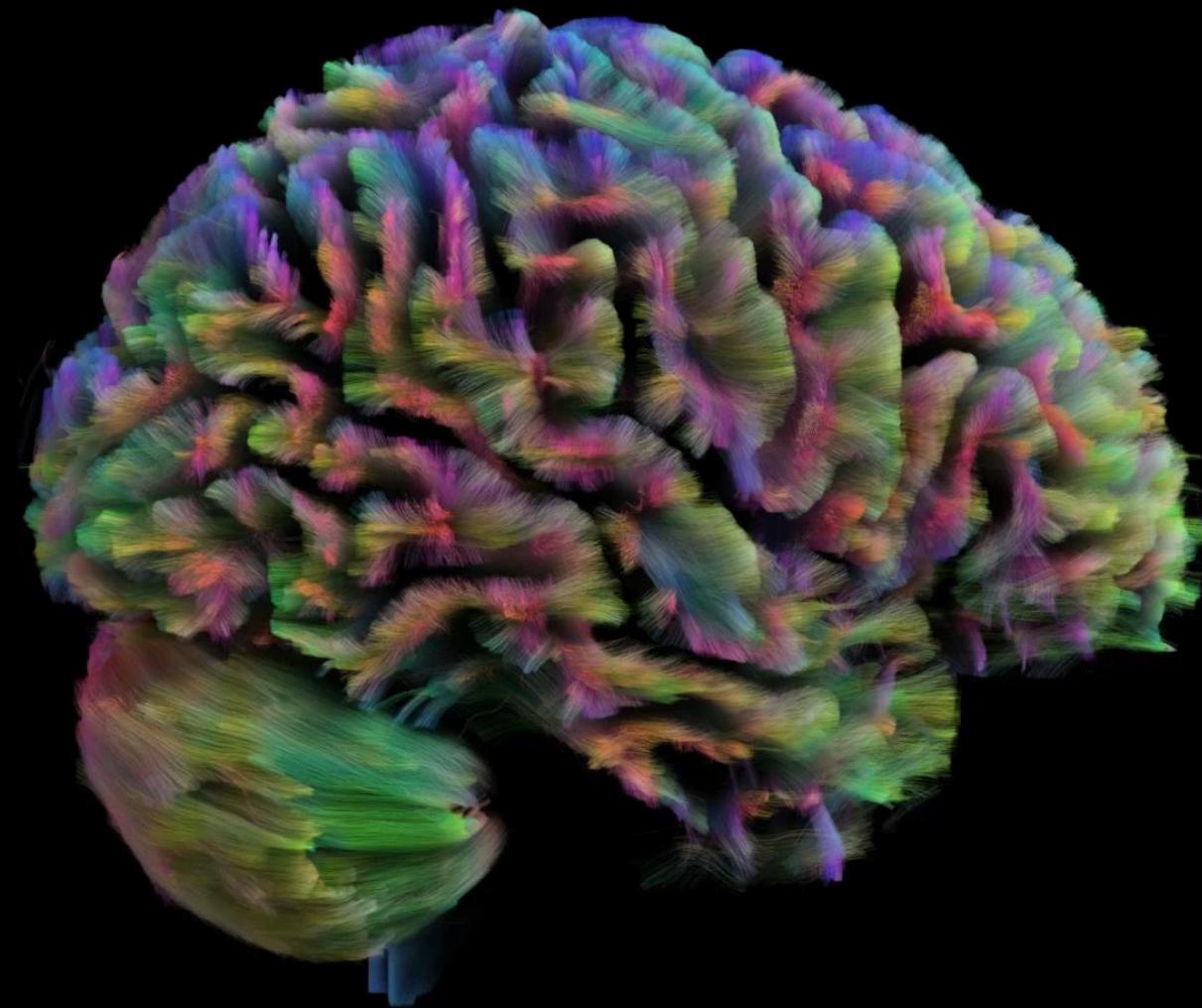


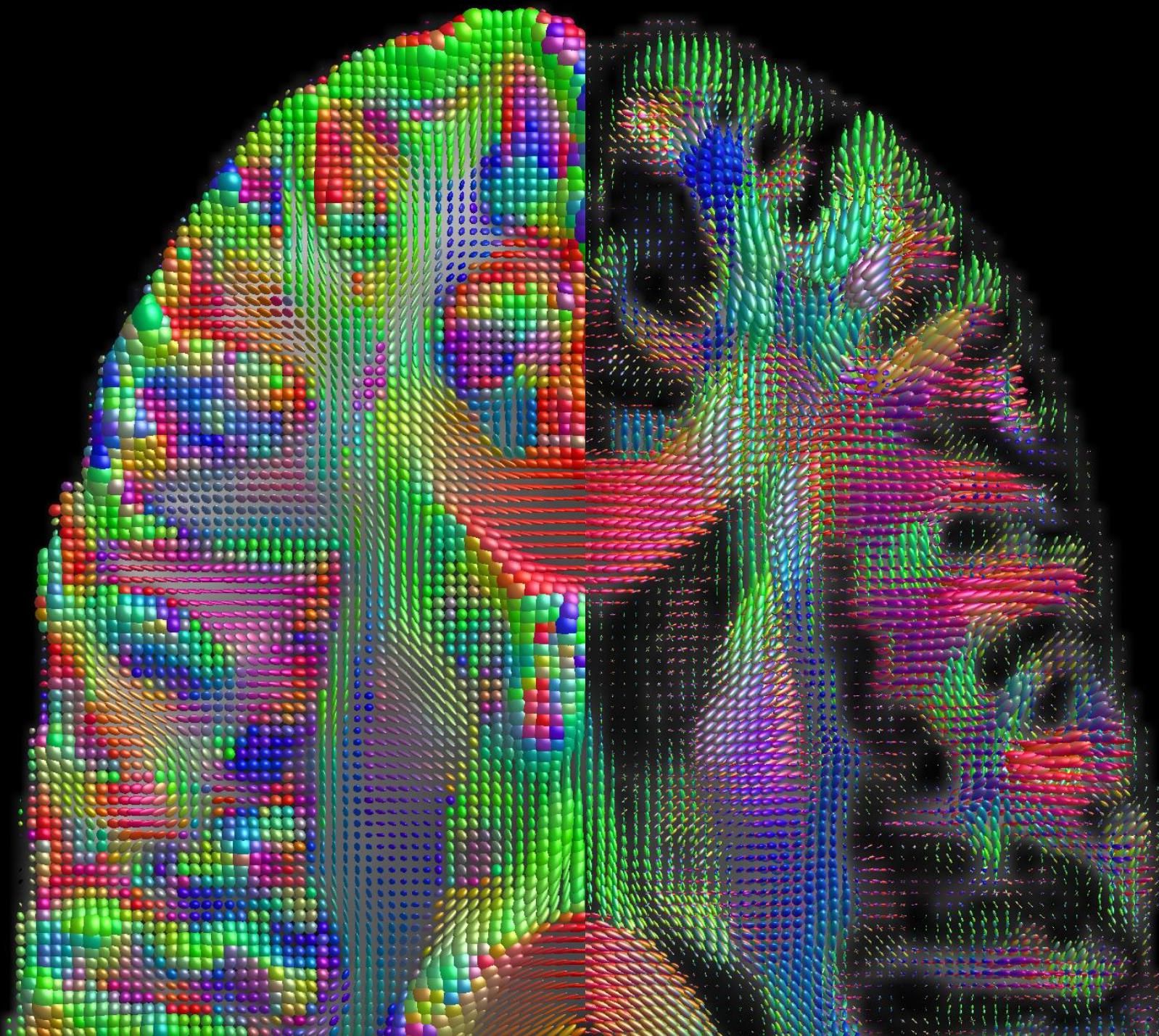
Conduction Velocity

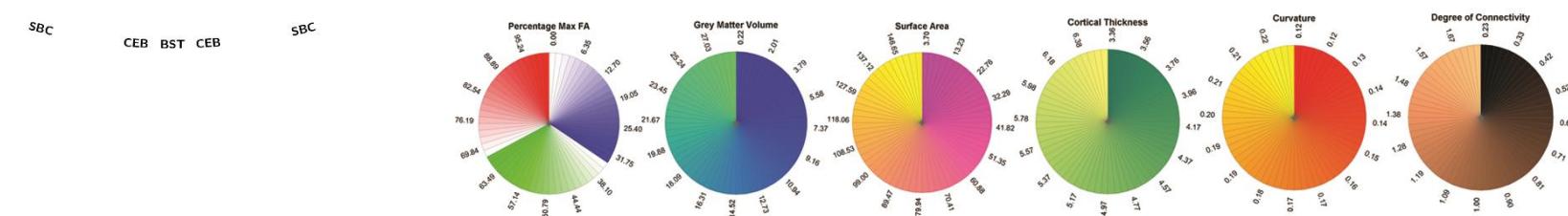
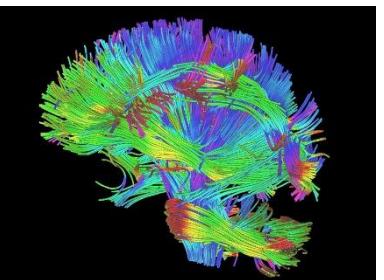
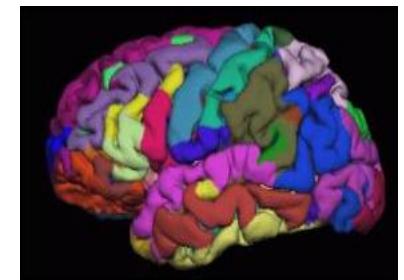
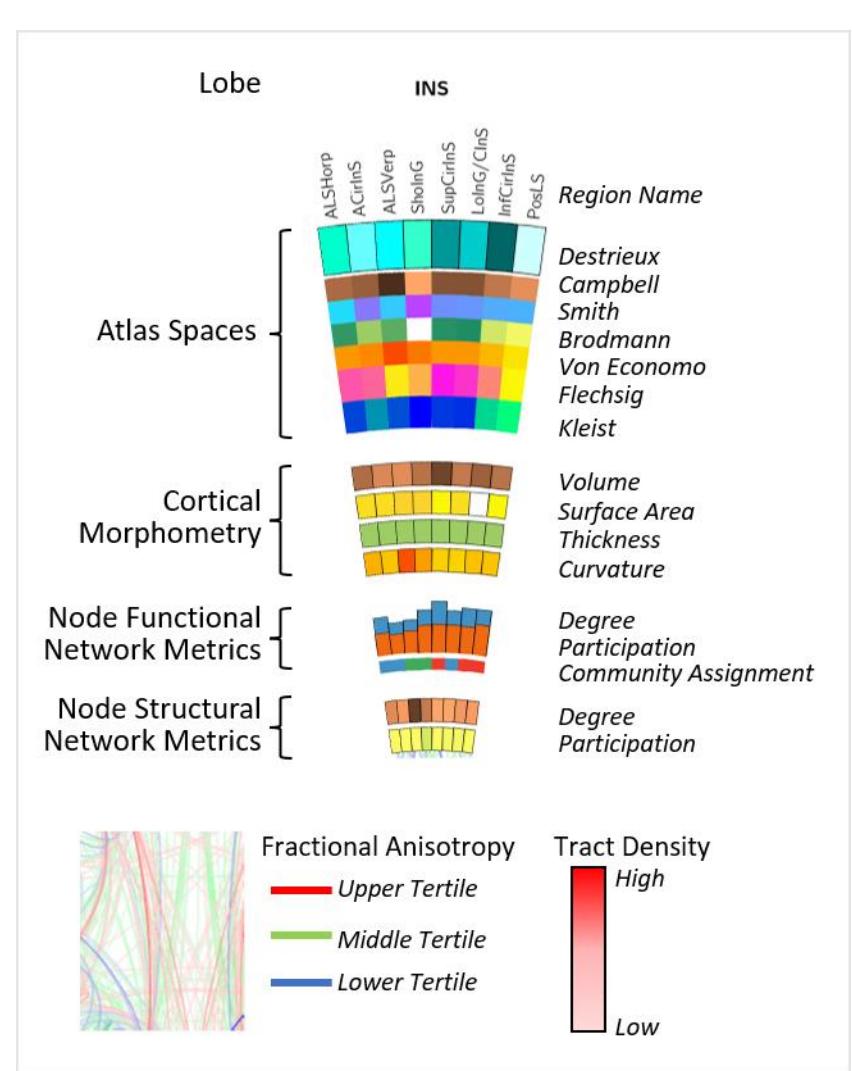
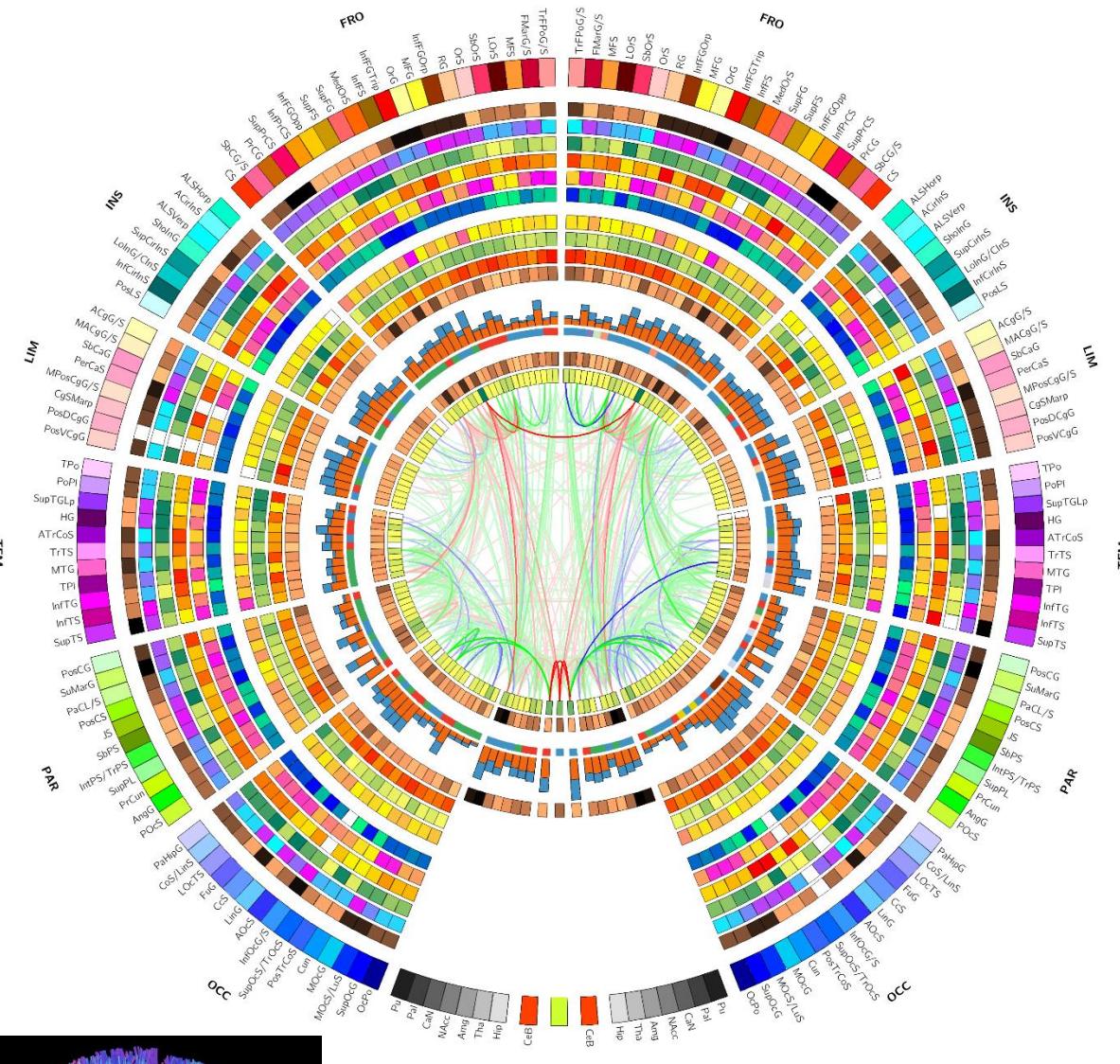
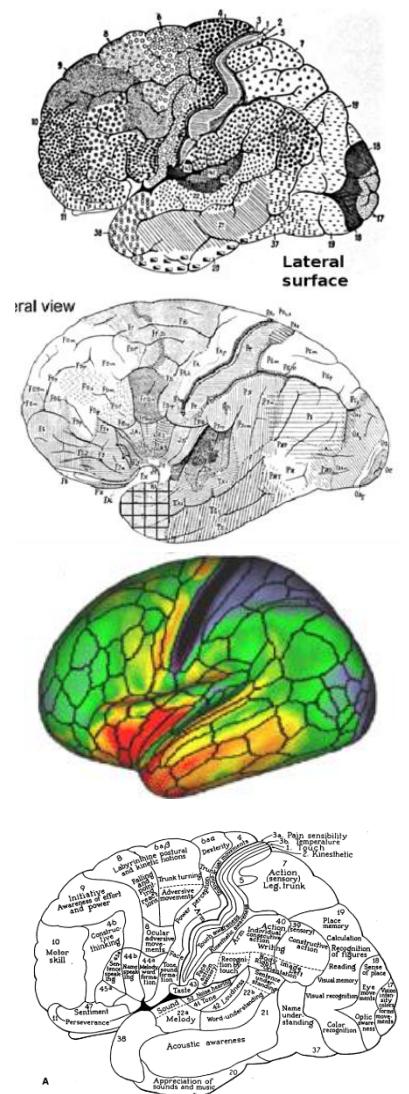


ASD TD

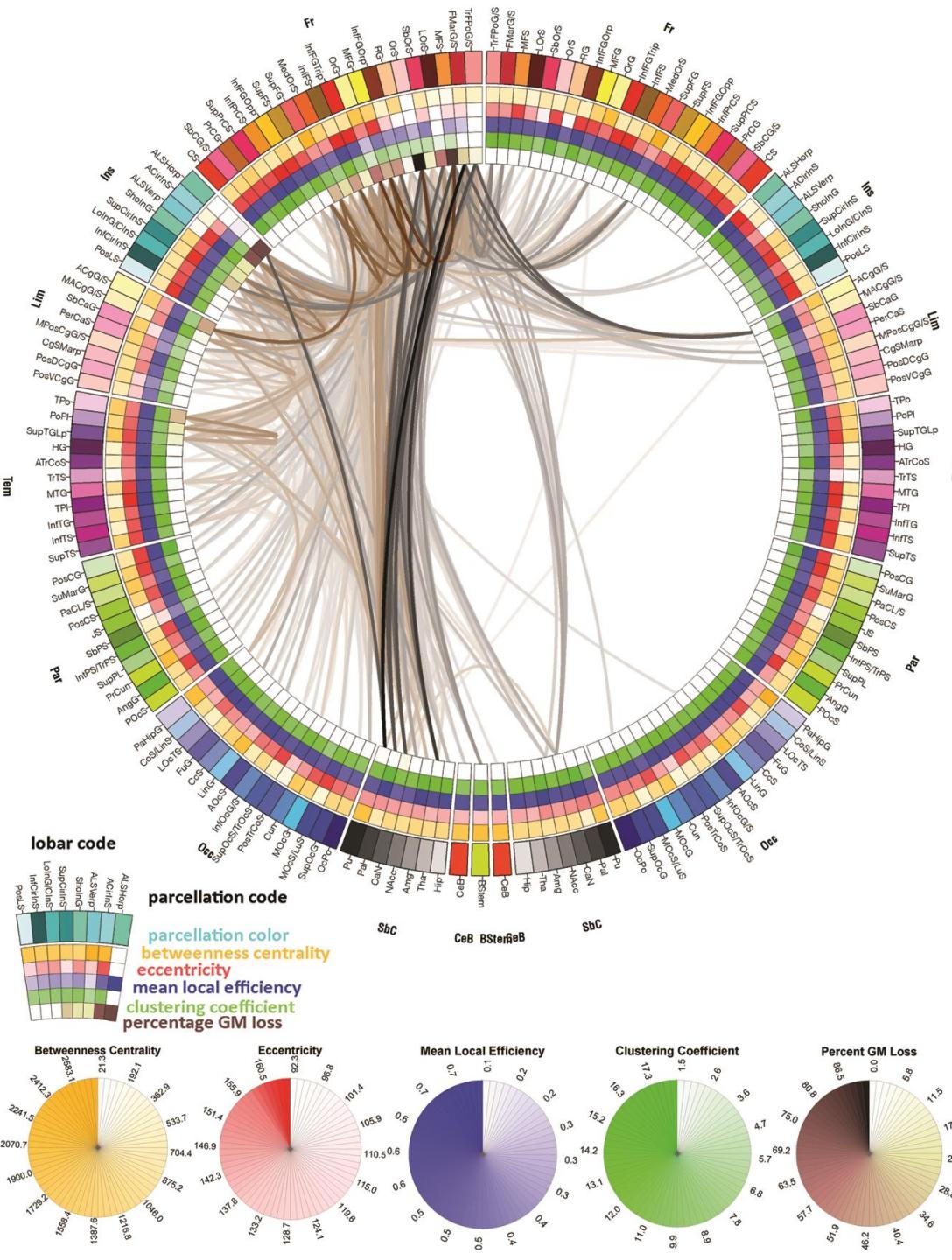
But what about $N = 1$?



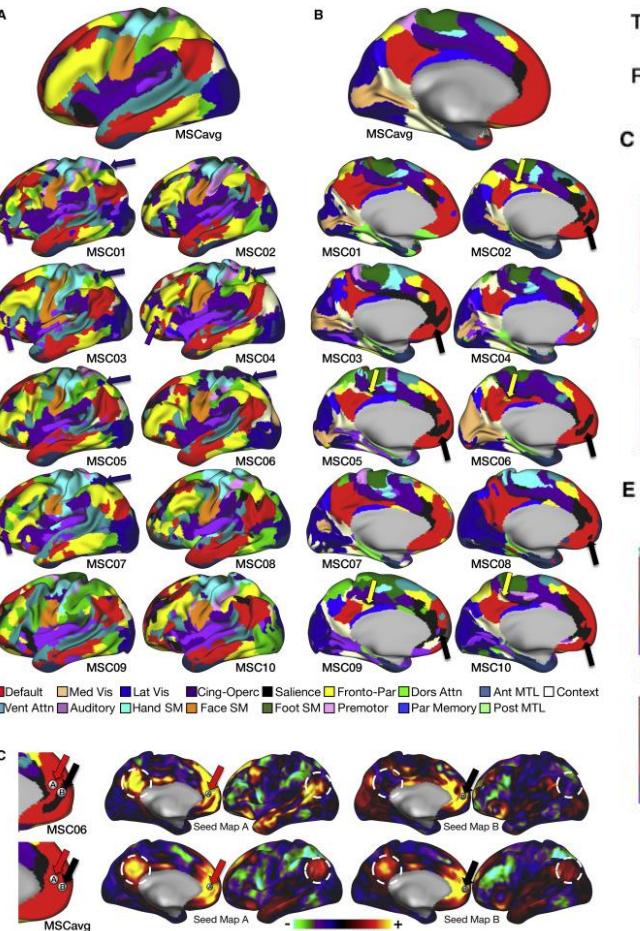
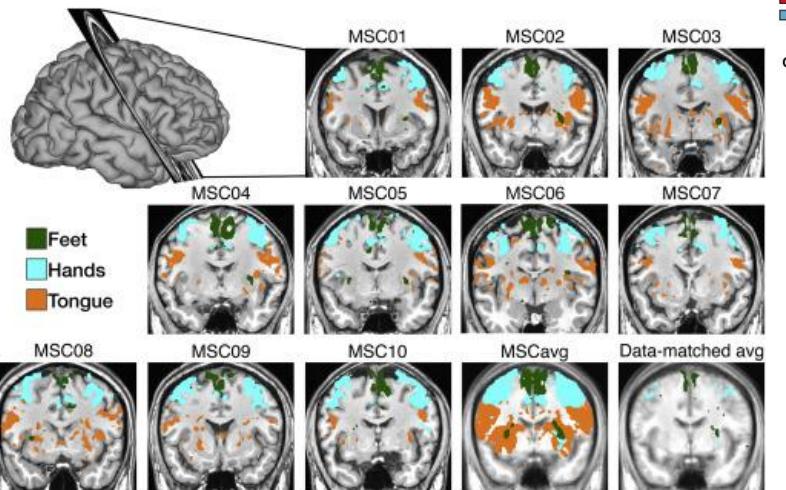




Connectomic Damage in the Case of Phineas Gage



fMRI of Individual Subjects



<https://doi.org/10.1016/j.neuron.2017.07.011>

Consensus recommendations for clinical functional MRI applied to language mapping

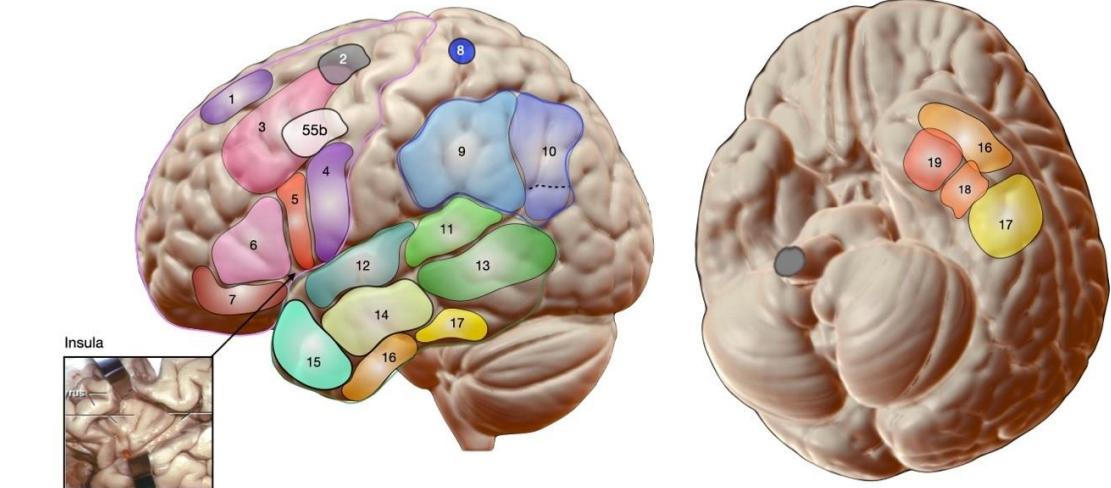
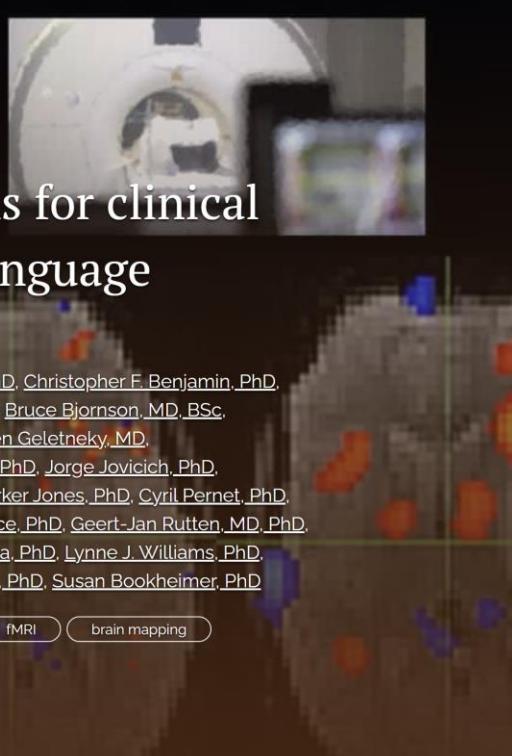
Natalie L. Voets, PhD, Manzar Ashtari, PhD, Christian F. Beckmann, PhD, Christopher F. Benjamin, PhD, Tammie Benzinger, MD, PhD, Jeffrey R. Binder, MD, Alberto Bizzzi, MD, Bruce Bjornson, MD, BSc, Edward F. Chang, MD, Linda Douw, PhD, Jodie Gawryluk, PhD, Karsten Geletreky, MD, Matthew F. Glasser, MD, PhD, Sven Haller, MD, MSc, Mark Jenkinson, PhD, Jorge Jovicich, PhD, Eric Leuthardt, MD, Asim Mian, MD, Thomas E. Nichols, PhD, Qiwi Parker Jones, PhD, Cyril Pernet, PhD, Puneet Plaha, MD, MS, Monika Potczyńska-Bletsos, PhD, Cathy J. Price, PhD, Geert-Jan Rutten, MD, PhD, Michael Scheel, MD, Joshua S. Shimony, MD, PhD, Joanna Sierpowska, PhD, Lynne J. Williams, PhD, Ghoufran Talib, MSc, Michael Zeineh, MD, PhD, Andreas Bartsch, MD, PhD, Susan Bookheimer, PhD

functional MRI
language
neurosurgery
surgical planning
fMRI
brain mapping
brain / surgery
functional laterality

CC BY-4.0 • <https://doi.org/10.52294/001c.128149>

“Presurgical planning also includes deciding on a strategy to reach a surgical target while minimizing damage to surrounding functionally important brain tissue. fMRI - if properly performed - allows the non-invasive visualization of gray matter functions at an individual patient level. But what constitutes ‘high-quality’ clinical fMRI?”

-- Voets, et al. (2025) *Aperture Neuro*,
<https://doi.org/10.52294/001c.128149>



Frontal Lobe: Top-down control		Parietal Lobe: Attention, Prediction, Multimodal integration	Temporal Lobe: Categories & Meaning																																																										
Superior	1 pre-SMA	Speech initiation & sequencing	Superior																																																										
Middle	2 Exner's 3 pMFG/IFS Area 55b	Writing (and reading) Cognitive control Articulatory coordination & planning	Inferior	4 vPreM 5 Pop 6 Ptr 7 Por	Phonological processing & articulation Phonological programming / control Lexico-semantic integration & control Associative semantic processing	Inferior		Insula	Complex articulatory planning			SMA: supplementary motor area. pMFG: posterior middle frontal gyrus. IFS: inferior frontal sulcus. vPREM: ventral premotor. Pop: pars opercularis. Ptr: pars triangularis. Por: pars orbitalis. aIPS: anterior inferior parietal sulcus. SMG: supra-marginal gyrus. PT: planum temporale. pSTG: posterior superior temporal gyrus. mSTG: middle superior temporal gyrus. pMTG: posterior middle temporal gyrus. mMTG: mid-partition of middle temporal gyrus. TP: temporal pole. ant/midITG: anterior-to-mid inferior temporal gyrus. pITG: posterior inferior temporal gyrus. VOT: ventral occipito-temporal area. VWA: visual word form area. BTLA: basal temporal language area.	Superior				11 PT/pSTG 12 mSTG				Middle				13 pMTG 14 mMTG				Anterior				15 TP				Inferior				16 ant/midITG 17 pITG/VOTC				Basal				18 mid fusiform 19 ant fusiform				WVFA-1: Lexical access & categorization (e.g., orthography)				WVFA-2: integrating orthography with sound & meaning				BTLA: Lexical-semantic processing
Inferior	4 vPreM 5 Pop 6 Ptr 7 Por	Phonological processing & articulation Phonological programming / control Lexico-semantic integration & control Associative semantic processing	Inferior																																																										
	Insula	Complex articulatory planning			SMA: supplementary motor area. pMFG: posterior middle frontal gyrus. IFS: inferior frontal sulcus. vPREM: ventral premotor. Pop: pars opercularis. Ptr: pars triangularis. Por: pars orbitalis. aIPS: anterior inferior parietal sulcus. SMG: supra-marginal gyrus. PT: planum temporale. pSTG: posterior superior temporal gyrus. mSTG: middle superior temporal gyrus. pMTG: posterior middle temporal gyrus. mMTG: mid-partition of middle temporal gyrus. TP: temporal pole. ant/midITG: anterior-to-mid inferior temporal gyrus. pITG: posterior inferior temporal gyrus. VOT: ventral occipito-temporal area. VWA: visual word form area. BTLA: basal temporal language area.	Superior				11 PT/pSTG 12 mSTG				Middle				13 pMTG 14 mMTG				Anterior				15 TP				Inferior				16 ant/midITG 17 pITG/VOTC				Basal				18 mid fusiform 19 ant fusiform				WVFA-1: Lexical access & categorization (e.g., orthography)				WVFA-2: integrating orthography with sound & meaning				BTLA: Lexical-semantic processing							
		SMA: supplementary motor area. pMFG: posterior middle frontal gyrus. IFS: inferior frontal sulcus. vPREM: ventral premotor. Pop: pars opercularis. Ptr: pars triangularis. Por: pars orbitalis. aIPS: anterior inferior parietal sulcus. SMG: supra-marginal gyrus. PT: planum temporale. pSTG: posterior superior temporal gyrus. mSTG: middle superior temporal gyrus. pMTG: posterior middle temporal gyrus. mMTG: mid-partition of middle temporal gyrus. TP: temporal pole. ant/midITG: anterior-to-mid inferior temporal gyrus. pITG: posterior inferior temporal gyrus. VOT: ventral occipito-temporal area. VWA: visual word form area. BTLA: basal temporal language area.	Superior																																																										
			11 PT/pSTG 12 mSTG																																																										
			Middle																																																										
			13 pMTG 14 mMTG																																																										
			Anterior																																																										
			15 TP																																																										
			Inferior																																																										
			16 ant/midITG 17 pITG/VOTC																																																										
			Basal																																																										
			18 mid fusiform 19 ant fusiform																																																										
			WVFA-1: Lexical access & categorization (e.g., orthography)																																																										
			WVFA-2: integrating orthography with sound & meaning																																																										
			BTLA: Lexical-semantic processing																																																										

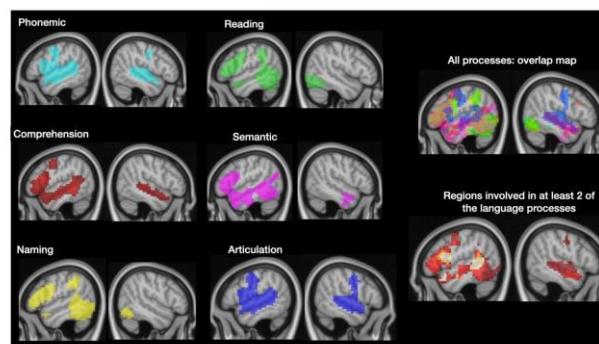


Fig 2. Relative lateralization of language processes based on fMRI

Different language processes are lateralized to different extents, depending on the language process engaged during a given fMRI task, and – importantly – what control condition the target language process is compared against (see Fig 5). Activation results are shown from predictions based on meta-analysis of ~13,500 neuroimaging studies in Neuroquery (<http://neuroquery.org>), separately querying the search terms “phonemic”, “comprehension”, “naming”, “reading”, “semantic” and “articulation”. The Neuroquery-derived z-score maps are shown thresholded at $z=3.1$ (corresponding to $p < 0.01$). Overlap maps show all processes overlaid and a heat map of voxels engaged by at least 2 of the language processes (brighter voxels indicate more overlapping processes, up to a maximum of 5 shared processes in the brightest voxels).

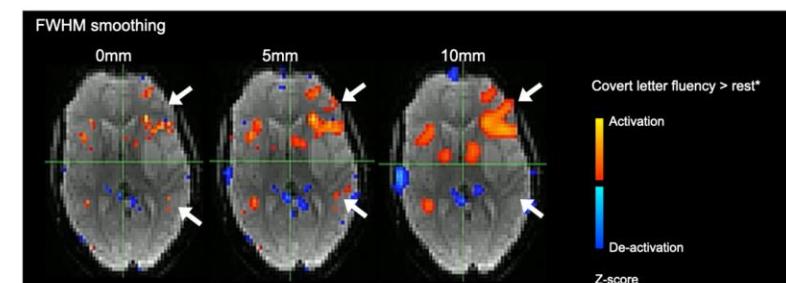
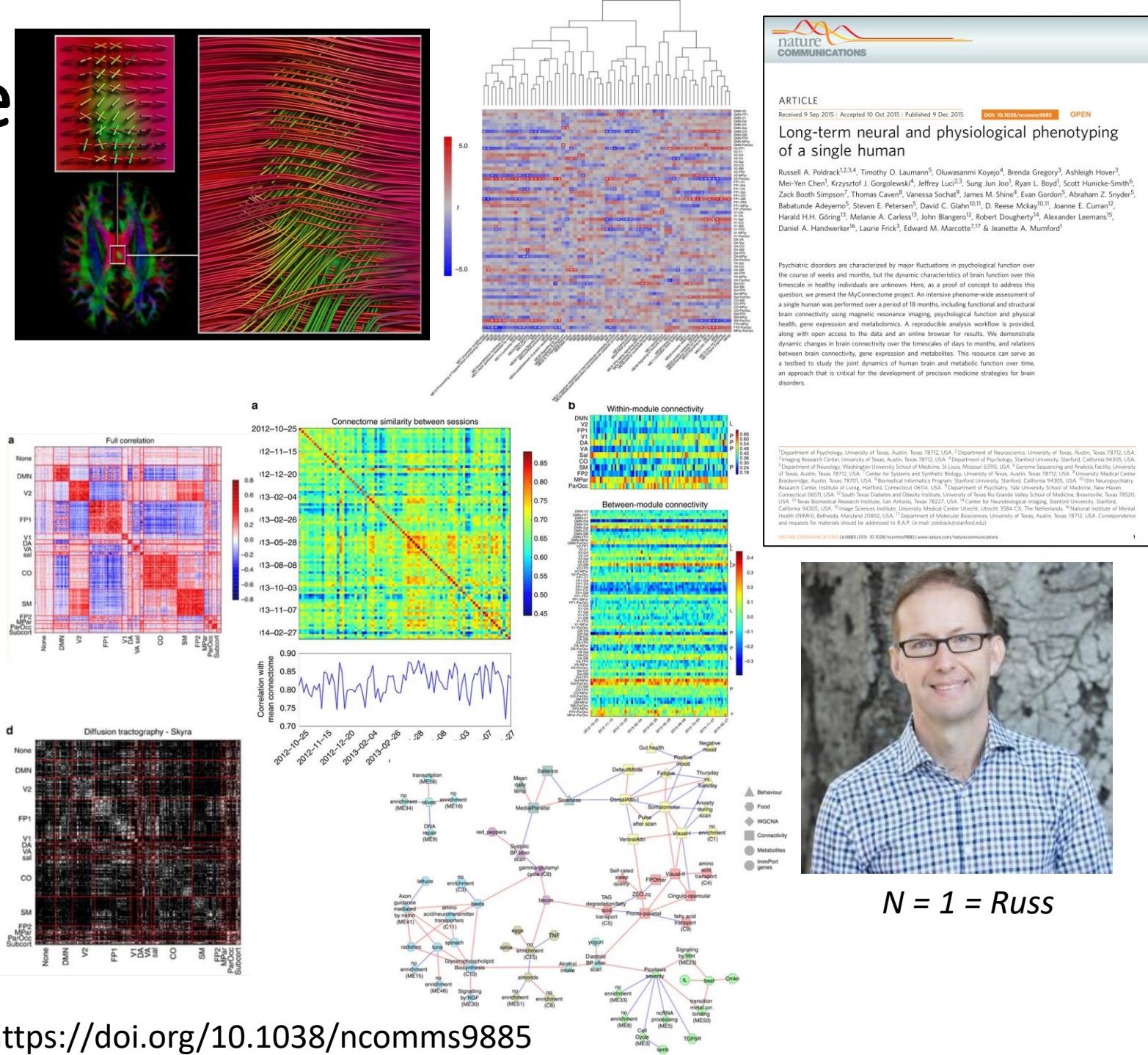
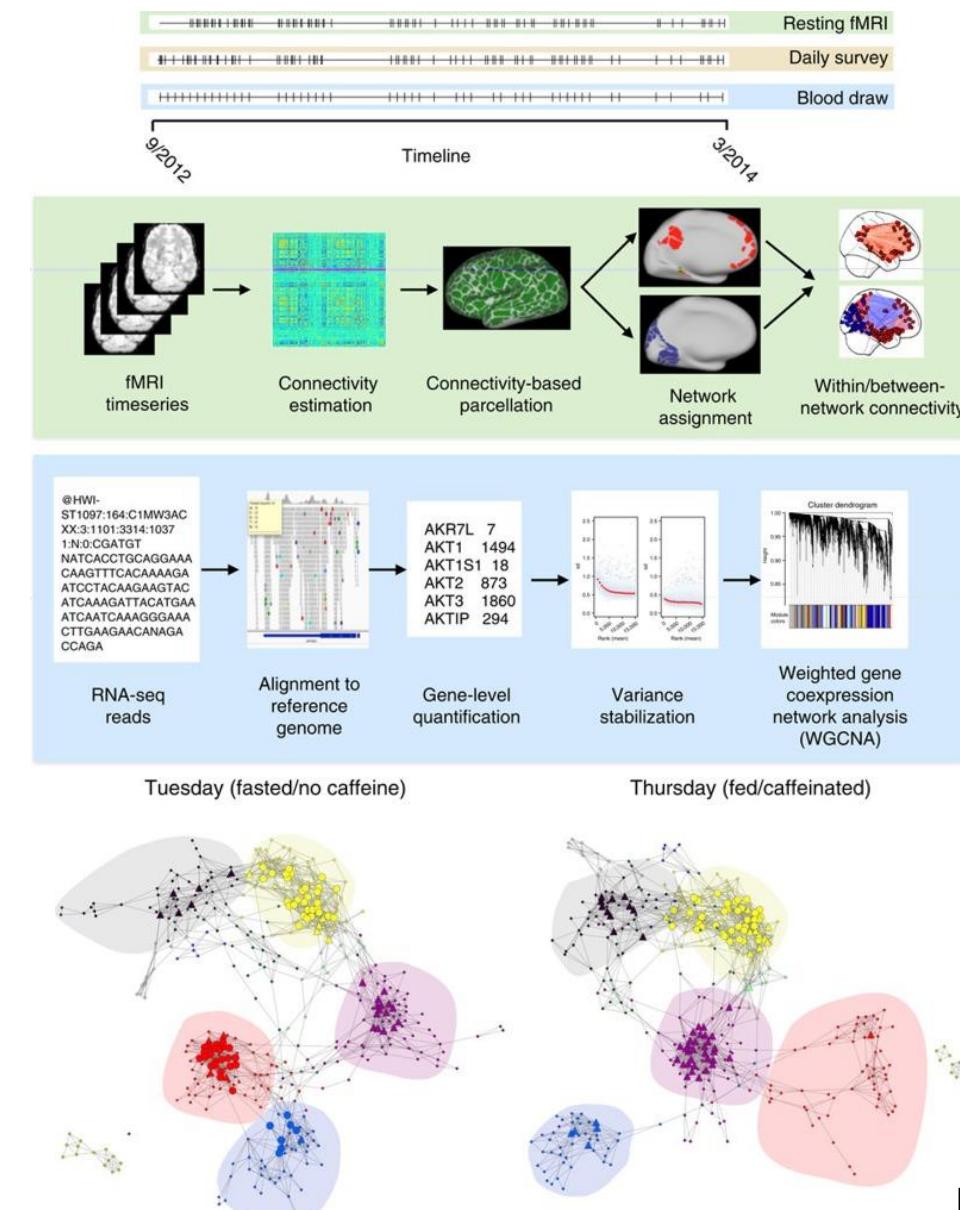


Fig 8. Effect of spatial smoothing on statistical activation maps

Effect of varying spatial smoothing on an example word generation language activation map, acquired at a voxel resolution of $2 \times 2 \times 2$ mm, illustrated in a patient with a left fronto-insular glioma. The amount of smoothing was varied from 0 to 10 mm while keeping all other analysis steps constant. The resulting spatial maps (each presented at the same statistical threshold) show substantial influence of the choice of smoothing on the spatial extent and foci of activation (e.g., white arrows). * Resting fixation is not generally recommended as a task contrast; it is used here merely to illustrate the effects of spatial smoothing on activation maps in general (irrespective of task/contrast). FWHM: Full width at half maximum.

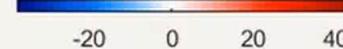
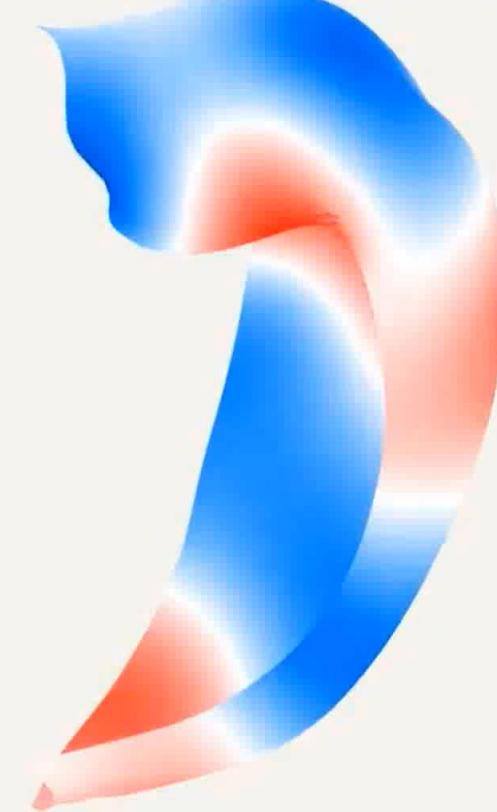
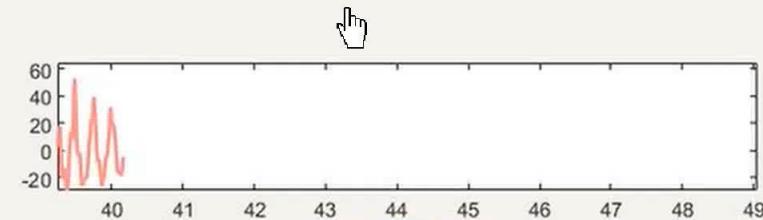
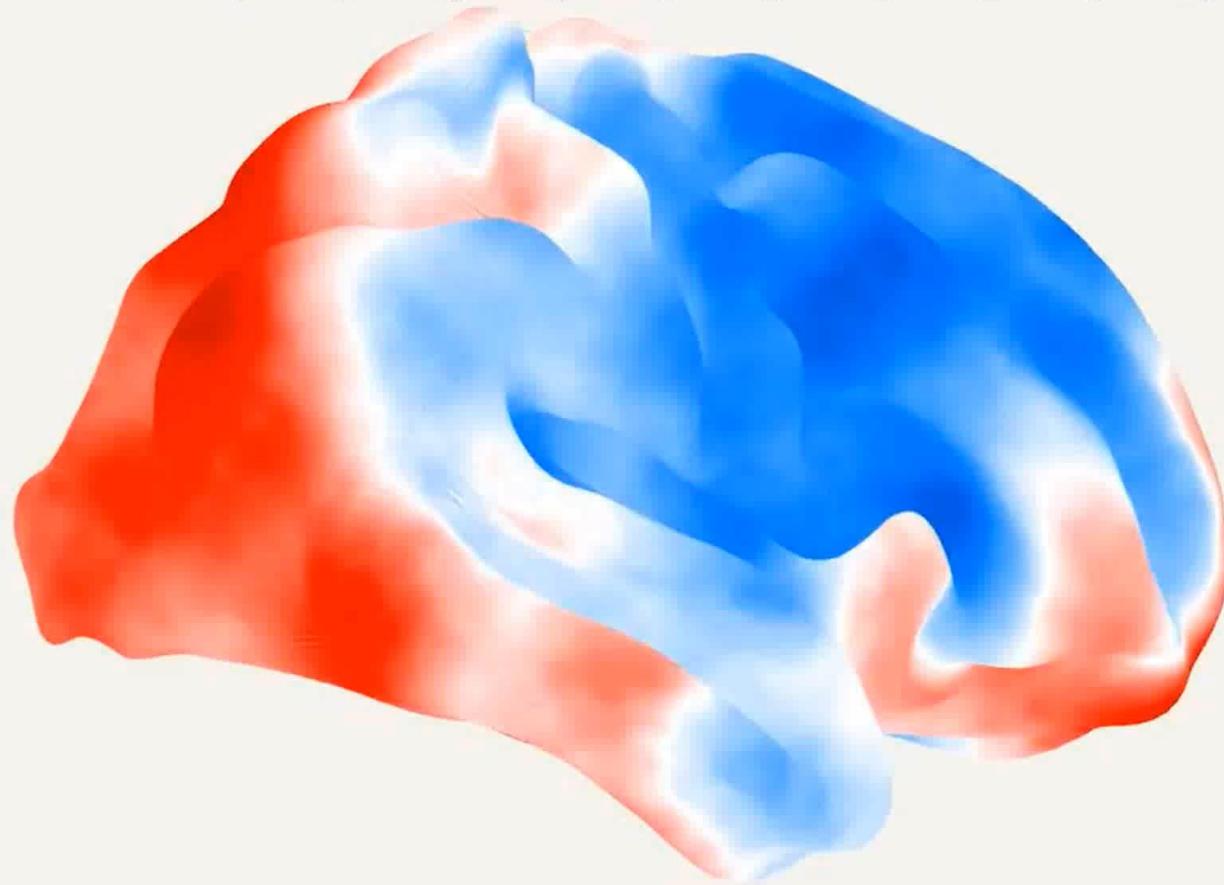
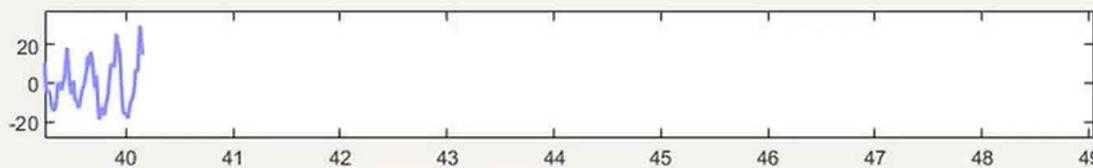
All Russ, All the Time



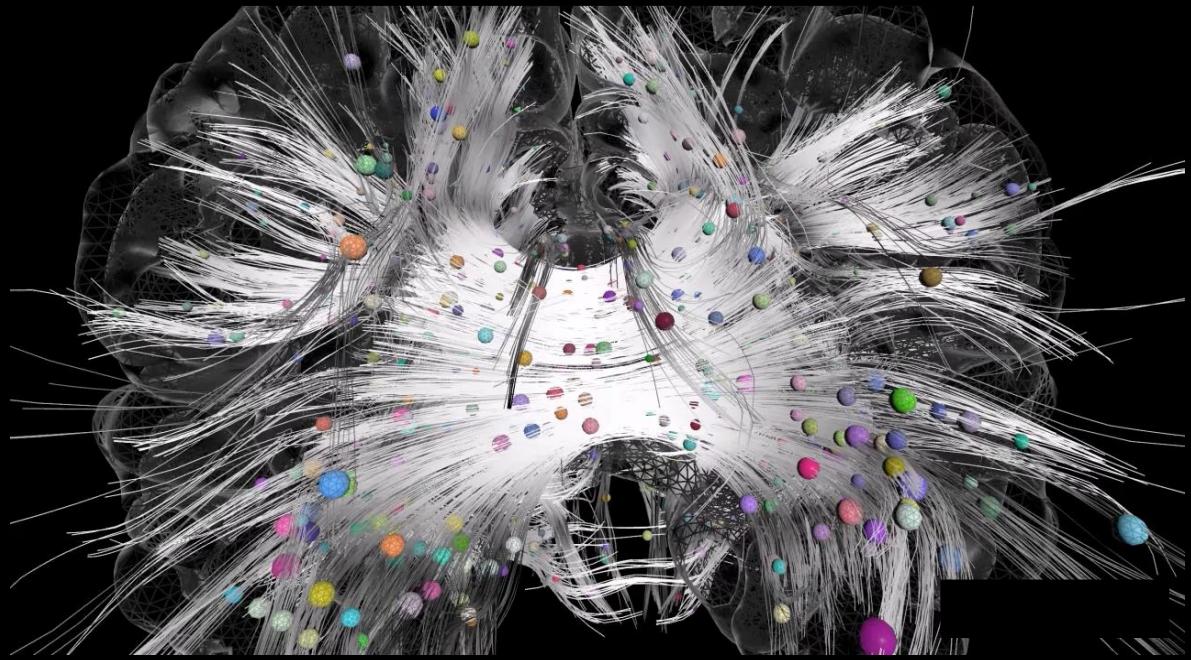
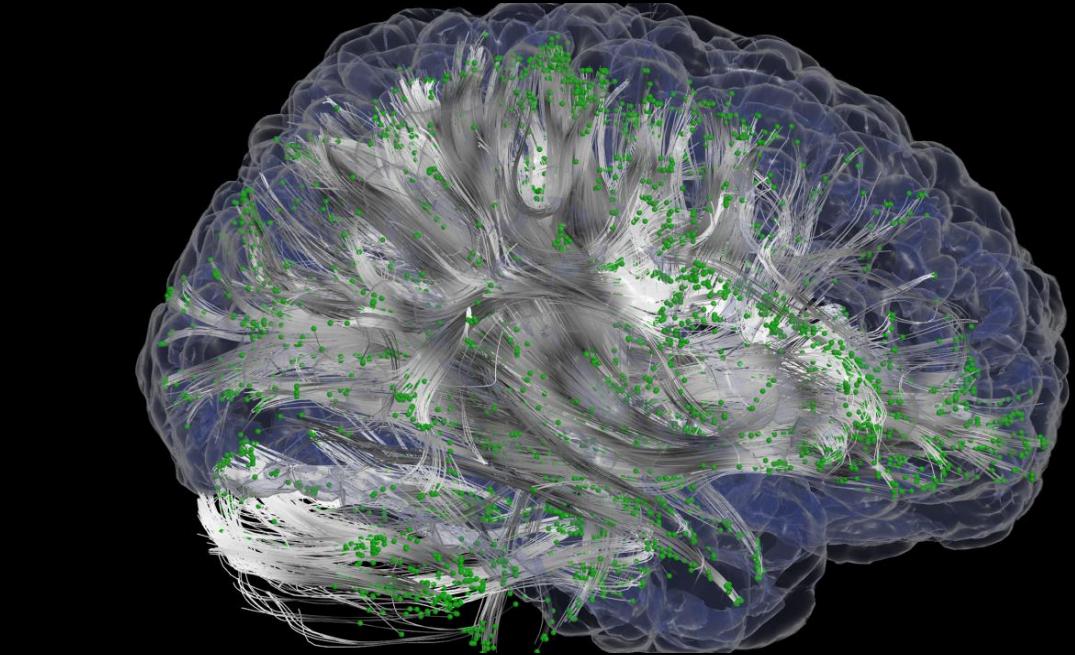
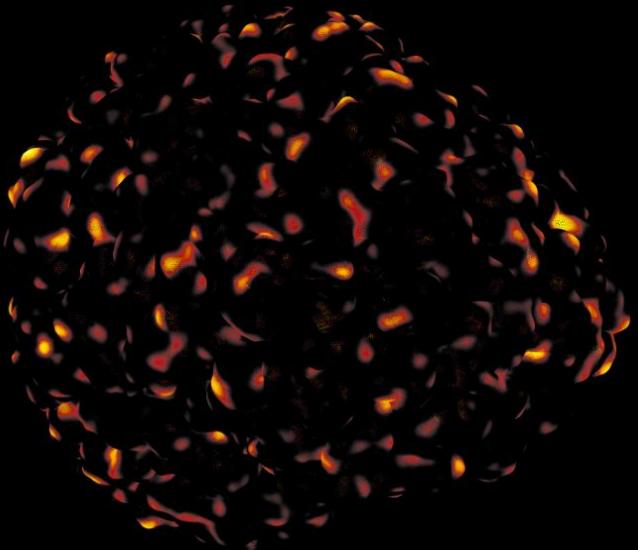
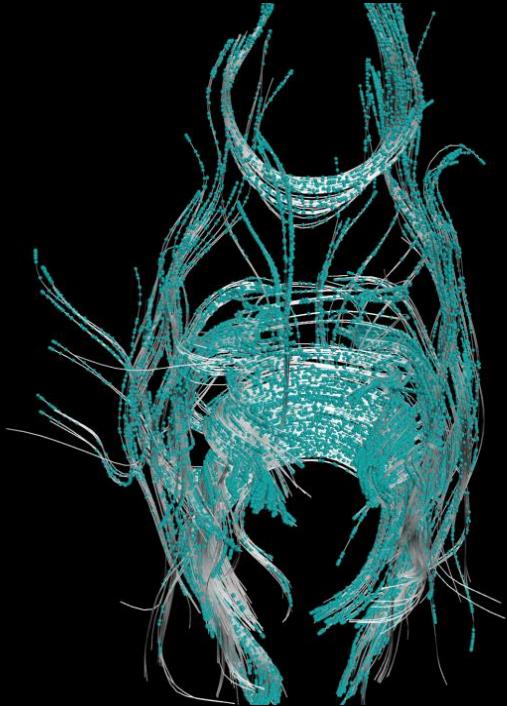
<https://doi.org/10.1038/ncomms9885>

Integrating Data Toward Digital Twins?

Time = 40.16



A Digital Twin?



Conclusions



YES, THE HUMAN BRAIN IS THE MOST COMPLEX THING IN THE UNIVERSE

<https://mindmatters.ai/2022/03/yes-the-human-brain-is-the-most-complex-thing-in-the-universe/>

Conclusions

- Data set size has tended to increase with each MR technological innovation
- MRI physicists earn their paychecks by developing imaging technologies providing more and richer data per *unit of space* and per *unit of time*
- Gathering more data to ensure sufficient statistical power is necessary....to a point(?)
- Claims that studies needing 10,000's of subjects may be exaggerated or overblown(?)
- Modest sized studies permit innovation but some new statistical considerations may be needed (e.g., SGPs, but understand the Null Distribution better)
- AI is likely the future and we should get used to it
- Improved models of brain functional dynamics governed by form and physical connectivity are needed to provide neuroscience with its own “*standard model*”
- Is now the time to integrate function with morphology and with structural connectivity – in turn, making individual and population-level data sets still larger?
- Data will need to be big as it needs to be for us to model brain functional activity and dynamics with high fidelity to make accurate statements about how the brain works.

Maybe we are the fish?

Constraints

- Efforts to “chase” smaller effects and p-values via ever larger sample sizes
- Historical dependence on the GLM and correlational methods
- Still relatively poor understanding of the relationship between the neural and BOLD responses
- Bigger data, *per se*, may not help us
- ML/DL/AI approaches are very exciting but may not tell us “how” the brain does what it does
- Sometimes, less is more – small can be good.
- Maybe, we also need bigger thinking not just bigger data!



EDITORIAL



Editorial: Is Now the Time for Foundational Theory of Brain Connectivity?

John Darrell Van Horn^{1,2}  · Zachary Jacokes²  · Benjamin Newman¹  · Teague Henry^{1,2} 

Published online: 14 August 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

For more than a century, the neuron doctrine has provided the bedrock of neuroscience, proclaiming that the neuron is the fundamental unit of the nervous system, both in structure and function (Yuste, 2015). This viewpoint emerged during an era when single-neuron techniques prevailed, emphasizing the significance of individual cells. However, the advent of advanced brain imaging methods capable of capturing the simultaneous activity from multiple neurons at the macro-scale has revealed a broader understanding. It is now apparent that ensembles of neurons, connected over long distances, rather than individual, isolated cells, form neurophysiological units and exhibit emergent functional properties and states (Hutchison et al., 2013; Martin et al., 2021). But rather than the former replacing the latter, we see the further continuation of a narrative where the architecture of individual neurons governs electrical signal characteristics which, in turn, effects broad levels of network connectivity, and is evident in patterns of cortical activity. Perhaps now is the time to consider forming a more general quantitative model of brain networks not governed by statistical measures of association between spatial signals but upon the underlying physical properties of neural tissues from which those signals emerge.

An impressive article by Pang et al. (2023) has reiterated, utilizing brain imaging, that “the close link between geometry and function is explained by a dominant role for wave-like

activity" in the brain. The connectivity of the mesocortical and physical components of the fundamental equation of Heaviside's - derived from Maxwellian signal transmission in graph cables. Hodges found that the giant squid axon as an equation in the family observed that the wavelength is nominally predicted by the calculated diameters, which of Ranvier along the length of their axonal and dendritic signal magnitudes, can be parameterized by membrane, relative characteristics, coupling. Important spatially restricted sensory and cognitive loci via a dense network of connections optimized toward the signal transduction process.

tributing to behavioral characteristics and, potentially, clinical symptoms. As tempting as it is to suggest that structural and functional connectivity measures naturally overlap, they often do not. As the principle of functional divergence illustrates, the brain is a highly adaptable and functionally versatile structure, preventing a one-to-one correspondence between its functional subunits and its emergent phenomena. This failure to overlap should not serve to divide neuroscientists into functionalism or structuralism camps but should encourage them to devise more mathematically comprehensive ways to define connectivity. Neither structural connectivity nor functional connectivity alone can fully describe the brain. Rather, structural connectivity should be viewed as the physical antecedent of functional networks and incorporated in connectivity models accordingly. In theory, this would serve to generate biologically-informed hypotheses and shift the field towards biological realism.

Structural connectivity, referring to the total set of nodes

Neuroinformatics (2024) 22:225–227
<https://doi.org/10.1007/s12021-024-09676-4>

EDITORIA



Towards Comprehensive Connectivity Modeling

Campbell Coleman¹: John Darrell Van Horn¹

Published online: 27 June 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Neuroinformatics is well positioned to help create a hyper-realistic, functionally-predictive model of the brain rooted in its underlying physiological structure. In theory, such an *in-silico* brain — often considered a “digital twin” — would serve to bridge the divide between psychology and biology, connecting underlying structural metrics such as fiber density and myelination to functional neuroanatomy. As the field develops stronger knowledge-bases of the functional and structural differences of neurological pathologies separately, it should, at the same time, be able to connect the two in order to isolate the biological mechanisms con-

How individual neural potentials over distance also an increasing amount of processing depends on the activity of nearby neurons through epiphapses. Ephapses are sites where anatomical or electrical modulation of the excitability of their close proximity enabling the transmission of nerve cell to affect neighboring neurons. This is due to the fact that the brain is a highly adaptable and functionally versatile structure, preventing a one-to-one correspondence between its functional subunits and its emergent phenomena. This failure to overlap should not serve to divide neuroscientists into functionalism or structuralism camps but should encourage them to devise more mathematically comprehensive ways to define connectivity. Neither structural connectivity nor functional connectivity alone can fully describe the brain. Rather, structural connectivity should be viewed as the physical antecedent of functional networks and incorporated in connectivity models accordingly. In theory, this would serve to generate biologically-informed hypotheses and shift the field towards biological realism.

Neuroinformatics (2024) 22:1–4

EDITORIAL



Editorial: On the Economics of Neuroscientific Data Sharing

John Darrell Van Horn¹

Published online: 15 November 2023
© The Author(s), under exclusive licence to Springer Science + Business Media, LLC, part of Springer Nature 2023

Today, the concept of sharing the data from one's neuroscientific research experiments seems second nature to many (Birgiolas et al., 2023; de Vries et al., 2023; Reer et al., 2023). Openly accessible archives readily exist for contributing raw and processed data which may then be downloaded by others (Markiewicz et al., 2021). Community-based file structures ease the assessment of data quality (Esteban et al., 2017), the application of processing workflows (Esteban et al., 2019), and the emergence of standards organizations supporting the exchange of data (Abrams et al., 2021). Research funding agencies now require statements about how data will be shared in grant applications and expect that primary data will be shared as a condition of funding. It might seem that data sharing has always been something the

The economics of neuroscientific data sharing encompasses the costs, benefits, and incentives associated with the exchange and dissemination of research data among scientists, institutions, and the broader scientific community. It involves considerations related to data acquisition, its storage, accessibility, usage, and the potential impact on scientific progress and innovation. The various aspects of the economics of neuroscientific data sharing include:

- Cost of data generation and storage:* Neuroscientific data generation can be a resource-intensive process, requiring substantial investments in equipment, materials, personnel, and infrastructure. Researchers and institutions bear the costs associated with collecting

✉ John Darrell Van Horn
jdv7g@virginia.edu

Zachary Jacokes
zj6nw@virginia.edu

Benjamin Newman
bnewm@virginia.edu

Those @ virginia.edu

¹ Department of Psychology, University of Virginia, Charlottesville, VA, USA.

² School of Data Science, University of Virginia, Charlottesville, VA, USA.

The fMRI Data Center



DARTMOUTH COLLEGE

Data Center • 6162 Moore Hall • Hanover, NH 03755-3569

2-2002-112E9
Discs 5-8 of 9

<http://www.fmridc.org>

Tel: (603) 646-0170 • Fax: (603) 646-1181

<http://www.fmridc.org>
Tel: (603) 646-0170 • Fax: (603) 646-1181

2-2003-113JE

Disks 1-2 of 2

<http://www.fmridc.org>

Tel: (603) 646-0170 • Fax: (603) 646-1181

PHINEAS GAGE SKULL
CT DATA
ORIG
COPY #2

5-118XA

1-4 of 5

<http://www.fmridc.org>
Fax: (603) 646-1181



Thank You