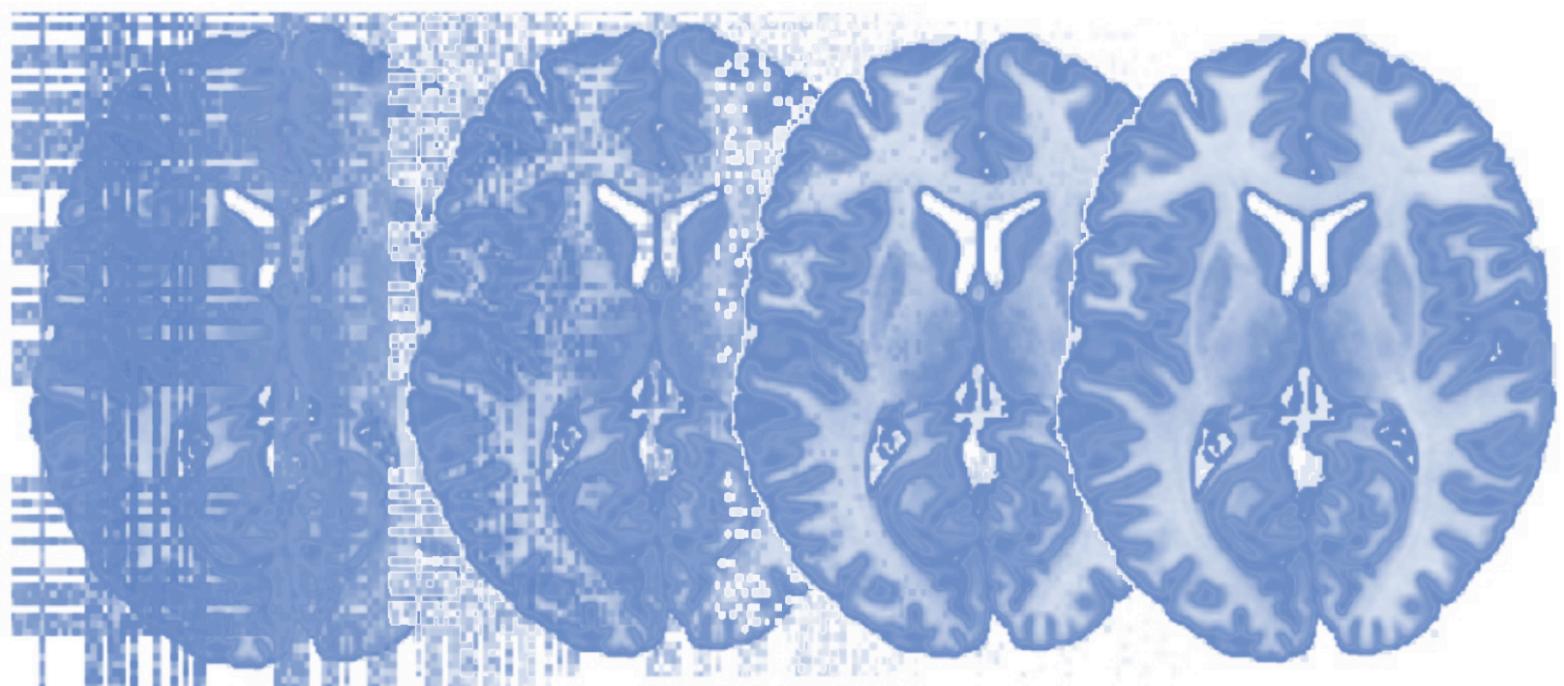


# Principles of fMRI



Martin A. Lindquist, Ph.D.

Tor D. Wager, Ph.D.

# Principles of fMRI

Tor D. Wager and Martin A. Lindquist

This book is for sale at <http://leanpub.com/principlesoffmri>

This version was published on 2015-10-12



This is a [Leanpub](#) book. Leanpub empowers authors and publishers with the Lean Publishing process. [Lean Publishing](#) is the act of publishing an in-progress ebook using lightweight tools and many iterations to get reader feedback, pivot until you have the right book and build traction once you do.

©2015 Tor D. Wager and Martin A. Lindquist

# Contents

What's in this book? . . . . .	1
About the Authors . . . . .	3
Martin Lindquist . . . . .	3
Tor Wager . . . . .	3
<b>Part 1: Motivation . . . . .</b>	<b>5</b>
<b>Chapter 1 - Introduction . . . . .</b>	<b>6</b>
MRI, PET, and beyond: A quick tour . . . . .	6
Principles . . . . .	8
<b>Chapter 2 - Why fMRI? Neuroimaging and the movement toward multidisciplinary science . . . . .</b>	<b>9</b>
Neuroimaging and the ‘common language’ of the brain . . . . .	9
Multiple roles, multiple fields: An example . . . . .	12
Challenges and motivation for multidisciplinary science . . . . .	13
<b>Chapter 3 - Types of imaging: What PET and fMRI can measure . . . . .</b>	<b>15</b>
MRI: Multiple measures, multiple modalities . . . . .	15
Structural MRI imaging . . . . .	16
Functional imaging with fMRI and PET . . . . .	18
<b>Chapter 4 - Brain mapping: A conceptual overview . . . . .</b>	<b>22</b>
What is a brain map? . . . . .	22
Fundamental assumptions and principles . . . . .	25
Types of inference: What brain maps can and cannot tell us . . . . .	30
<b>Chapter 5 - Limitations in inferences from brain maps . . . . .</b>	<b>35</b>
Seven caveats in brain map inferences . . . . .	35
A non-imaging example . . . . .	38
<b>Chapter 6 - How to lie with brain imaging . . . . .</b>	<b>39</b>
How to tell a story about the “one brain region” . . . . .	39
How to make your results look really strong . . . . .	40
Overlapping processes: How to make two maps look the same . . . . .	41

## CONTENTS

How to make two maps look really different . . . . .	42
Conclusions . . . . .	43
<b>Part 2: Fundamentals: The origins of PET and fMRI signals in the brain . . . . .</b>	<b>44</b>
<b>Chapter 7 - fMRI basics: Processing stages, terminology, and data structure . . . . .</b>	<b>45</b>
fMRI basics . . . . .	45
Data structure in fMRI experiments . . . . .	54
Conclusions . . . . .	57
<b>Chapter 8 - The MRI environment and human factors . . . . .</b>	<b>58</b>
MR basics and safety . . . . .	58
Physical limitations on data collection . . . . .	59
<b>Chapter 9 - A head-to-head comparison of PET and MRI . . . . .</b>	<b>63</b>
Acquisition options and fidelity . . . . .	63
Available signal types and their interpretability . . . . .	63
Spatial and temporal resolution . . . . .	64
Accessibility to a broad community . . . . .	65
Conclusions . . . . .	65
<b>Chapter 10 - Fundamentals of MRI Physics . . . . .</b>	<b>66</b>
<b>Chapter 11 - Physiological basis of fMRI signals . . . . .</b>	<b>68</b>
<b>Chapter 12 - Constraints on fMRI spatial and temporal resolution . . . . .</b>	<b>70</b>
Spatial Limitations . . . . .	70
Temporal Limitations . . . . .	71
<b>Part 3: Basics of fMRI signal processing and analysis . . . . .</b>	<b>73</b>
<b>Chapter 14 - Experimental Design . . . . .</b>	<b>74</b>
Block designs . . . . .	74
Event-related fMRI . . . . .	75
Optimized experimental designs . . . . .	76
<b>Chapter 15 - Resting state, natural viewing, and non-experimental designs . . . . .</b>	<b>77</b>
Resting state . . . . .	77
Natural viewing and non-experimental designs . . . . .	79
<b>Chapter 16 - Essentials of fMRI signal processing . . . . .</b>	<b>80</b>

## CONTENTS

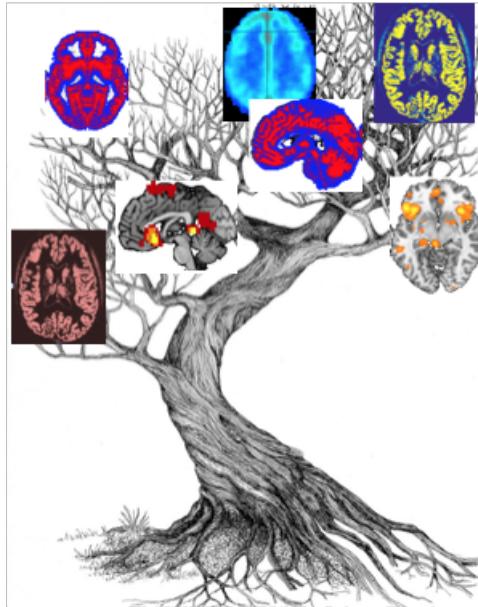
BOLD signal . . . . .	80
Noise and nuisance signal . . . . .	82
<b>Chapter 17 - Preprocessing . . . . .</b>	<b>85</b>
Reconstruction. . . . .	85
Slice-timing correction . . . . .	86
Motion correction . . . . .	86
Co-registration . . . . .	87
Normalization . . . . .	88
Spatial smoothing . . . . .	89
<b>Chapter 18 - The General Linear Model and Foundations of Analysis . . . . .</b>	<b>90</b>
Setting up the GLM . . . . .	91
GLM Estimation . . . . .	91
<b>Chapter 19 - Conditions and Contrasts . . . . .</b>	<b>93</b>
<b>Chapter 20 - Design Specification: Flexible Hemodynamics and Mis-modeling . . . . .</b>	<b>94</b>
<b>Chapter 21 - Design Specification: Dealing with Artifacts and Noise . . . . .</b>	<b>97</b>
<b>Chapter 22 - Group Analysis . . . . .</b>	<b>100</b>
<b>Chapter 23 - Multiple Comparisons . . . . .</b>	<b>103</b>
FWE correction . . . . .	104
FDR control . . . . .	105
<b>Chapter 24 - Assessing Brain Connectivity . . . . .</b>	<b>106</b>
Functional connectivity . . . . .	106
Effective connectivity . . . . .	107
Network analysis . . . . .	109
<b>Part 4: Predictive Mapping . . . . .</b>	<b>110</b>
<b>Chapter 25 - Multivariate brain analysis: From maps to models . . . . .</b>	<b>111</b>
From univariate mapping to multivariate brain models . . . . .	112
Basic criteria for a good model . . . . .	120
<b>Chapter 26 - Advantages of MVPA from a neuroscientific perspective . . . . .</b>	<b>121</b>
MPVA analysis choices: Spatial scope and flexibility . . . . .	121
Sensitivity to neural topography . . . . .	125
Sensitivity to distributed representations . . . . .	128
Benefits in testing generalizability across individuals and studies . . . . .	130
<b>Resources and further reading . . . . .</b>	<b>133</b>

CONTENTS

Other books about fMRI . . . . .	133
----------------------------------	-----

# What's in this book?

The field of neuroimaging is spreading its tendrils far and wide, and is reaching into various fallow fields and corners of the public consciousness. Neuroimaging results are bread and butter for psychologists and neuroscientists, and are becoming more and more relevant for physicians, economists, lawyers, engineers and physicists curious about biology, biologists curious about computation and the human mind, and anyone who wants to understand science news at a deeper level and is curious about what we're discovering, or not, about the human mind and brain.



The increasing popularity of neuroimaging, particularly Functional Magnetic Resonance Imaging (fMRI), poses a particular challenge. It is inherently a technical enterprise, requiring a mish-mash of biological, computational, statistical, and psychological expertise that spans fields and is contained in no single training program.

That is why we created this book. We have been teaching fMRI experimental design and analysis for 15 years, to various groups of people—from doctors to engineers to statisticians to businesspeople—and we have gained an appreciation for how challenging it can be to put all the pieces together and be a smart consumer, and a smart creator, of neuroimaging research. This book is designed to convey essential knowledge about how fMRI works, and the principles that underlie it, for both practitioners and readers who may not sit down at a computer and analyze fMRI data themselves, but want to take their understanding of how it works to the next level.

The book is divided into five sections. In the first section, we provide some basic context on fMRI in relation to other brain-focused techniques, and discuss why fMRI is being intensively developed and what some of those developments are. In the second section, we cover the origins of PET and fMRI signals in the brain, their spatial and temporal resolution, and what we can and cannot measure with MRI and PET. In the third section, we review key concepts underlying the process of fMRI data analysis, from signal processing fundamentals to statistical inference. In the fourth section, we focus on inference, and particularly on the kinds of claims that are currently being made based on brain imaging data. The data analysis strategy is often not matched to the types of inferences that researchers want to make, and the result is a number of 'crises' in neuroscience that are largely avoidable. A community of researchers and public consumers of research that is better educated on how valid inferences can be made, and the limitations involved, will be better positioned to make discoveries that will shape our world for the better. Finally, in the fifth section, we discuss some emerging approaches that are changing the way we analyze fMRI data and the kinds of valid claims that we can make about the human brain and mind.

# About the Authors

## Martin Lindquist

Professor, Department of Biostatistics Johns Hopkins Bloomberg School of Public Health



Dr. Lindquist is a Professor of Biostatistics at Johns Hopkins University. He received his Ph.D. in Statistics from Rutgers University in 2001, and served as an Assistant and Associate Professor at Columbia University from 2003-2012. His research focuses on mathematical and statistical problems relating to functional Magnetic Resonance Imaging (fMRI). Dr. Lindquist is actively involved in developing new analysis methods to enhance our ability to understand brain function using human neuroimaging.

Twitter: @fMRIstats

[Martin's web page at JHU<sup>1</sup>](#)

## Tor Wager

Professor, Department of Psychology and Neuroscience and the Institute of Cognitive Science  
University of Colorado, Boulder

---

<sup>1</sup><http://www.biostat.jhsph.edu/~mlindqui/>



Dr. Wager is a Professor of Psychology and Neuroscience and a faculty member in the Institute for Cognitive Science at the University of Colorado, Boulder. He received his Ph.D. from the University of Michigan in cognitive psychology in 2003, and served as an Assistant and Associate Professor at Columbia University from 2004-2009. Since 2010, he has directed Boulder's Cognitive and Affective Neuroscience laboratory. He has a deep interest in how thinking influences affective experiences, affective learning, and brain-body communication. His laboratory also focuses on the development and deployment of analytic methods, and has developed several publicly available software toolboxes for fMRI analysis. He has been teaching fMRI analysis methods since 2003.

Twitter: @torwager, @canlab

[Cognitive and Affective Neuroscience Lab web page<sup>2</sup>](#)

---

<sup>2</sup><http://wagerlab.colorado.edu>

# **Part 1: Motivation**

# Chapter 1 - Introduction

Welcome to Principles of fMRI. This book, its companion volume, and their accompanying courses on Coursera are part of a series on Functional Magnetic Resonance Imaging designed to benefit both those who are practitioners of fMRI and those not actually engaged in fMRI research but who want to be “smart consumers” of contemporary brain science.

*Neuroimaging* refers to any technique used to obtain and integrate multiple measures of brain structure or function into a picture - or a series of pictures - of the brain. One of the most exciting advances in the psychological, behavioral, and brain sciences over recent decades is the ability to non-invasively image the living human brain. This human imaging reveals both brain structure (i.e., anatomy) and function: it demonstrates images of brain electrical, neurochemical, and metabolic processes which occur as living humans engage in any number of various mental states including solving problems, experiencing emotion, feeling one’s bodily state, thinking about other people, or dreaming.

Many types of human neuroimaging are available today. Some are invasive and require opening the skull to make brain measurements. These include *optical imaging* and intracranial recording or *electrocorticography* (known as ECoG), which are electrophysiology techniques that can be used to create maps or, in some cases, images. The most widely used methods, however, are minimally or non-invasive and can support large-scale studies of healthy humans. These methods include *electroencephalography* (EEG), *magnetoencephalography* (MEG), *single-photon emission tomography* (SPECT), *near-infrared spectroscopy* (NIRS), *positron emission tomography* (PET), and *magnetic resonance imaging* (MRI). MRI, in turn, encompasses multiple techniques employed to image brain structure and function. Each of these methods has strengths and weaknesses, and provides a complementary window into the function of the brain and mind.

Among the diverse neuroimaging methods, MRI is now the most widely used. The family of techniques used to assess brain function is called “functional magnetic resonance imaging”, or fMRI, and is the primary focus of this book. Though we concentrate on fMRI, many of the principles and techniques we cover apply broadly to other types of imaging and beyond to further scientific fields. Many of the principles underlying experimental design and statistical modeling and inference are not unique to neuroimaging at all and accordingly pertain across various areas of scientific inquiry.

## MRI, PET, and beyond: A quick tour

In later chapters, we discuss more fundamentals of how the techniques above relate to one another and what their relative strengths and weaknesses are. To orient you, however, here are a few key ideas that also provide a rationale for limiting the scope of this book to MRI.

Among the various human neuroimaging techniques, MRI and PET are unique in their capacity to create images of the entire living human brain with fairly veridical localization of where in the brain the signals originate. EEG and MEG record, respectively, electrical and magnetic information from the brain surface. These measurements directly reflect neuronal electrical activity at high temporal resolution (e.g., every millisecond), which is an enormous advantage. NIRS also makes measurements at the skull, but does so using light. However, EEG and (to a lesser degree) MEG rely on complex mathematical models to make inferences about where the signals in the brain originate, which require one to make strong statistical assumptions. These postulations are difficult to validate and are often violated, thus leaving quite a bit of uncertainty about the origination of the signals within the brain. In addition, these three techniques - EEG, MEG, and NIRS - are all quite limited in their sensitivity to brain areas. Each is largely restricted to detecting signals from the cortical surface, though one can find a number of MEG papers that also make claims about signal sources in deeper brain structures.

By contrast, both MRI and PET provide the capacity to reconstruct three-dimensional brain volumes with fairly high spatial resolution (in some cases less than 1 mm) across the entire brain. For this reason, the term "neuroimaging" is often used to describe primarily MRI and PET.

MRI can provide measures related to many physiological features of interest, including:

- \* Gray-matter density and cortical thickness
- \* White-matter tract density and location
- \* Brain elasticity and shearing forces
- \* The sizes, location, and course of blood vessels
- \* The flow of cerebrospinal fluid
- \* Measures of a few selected neurotransmitter levels, like GABA, and specific proteins related to metabolism
- \* Moment-by-moment measures of cerebral blood flow
- \* Moment-by-moment measures related to blood oxygenation, flow, and oxygen metabolism.

This last measure refers to Blood Oxygen Level-Dependent (BOLD) signal, which is the most widely used measure of functional activity in fMRI. With modern BOLD imaging, we can sample signals throughout the whole brain with spatial resolution around 3 mm and temporal resolution of one brain volume per second. This entails sampling about 100,000 brain locations, or  $\text{Ovoxels,}^{\circ}$  every second, which provides a rich picture of activity in the cortex and dynamics across brain networks.

PET provides unique measures which complement the strengths of fMRI. PET can image cerebral blood flow and glucose metabolism, which, along with fMRI measures, are often referred to as "activity" or "activation". PET is an invasive technique that involves injecting radioactively labeled compounds into the bloodstream. It can therefore be used to image a wide variety of molecular and cellular processes, and is limited mainly by the types of compounds which can be radiolabeled and their actions on the brain's receptors. There are now hundreds of radiolabeled compounds which have been used to examine brain function, including those processes related to dopamine, serotonin, acetylcholine, opioids, microglial function, and neuroinflammation. Though in theory many compounds are available, their radioactive labeling means each compound must be developed and manufactured on-site next to the scanner with precise control over the time from manufacture to injection into the research participant. These constraints place practical limitations on the scope of molecular imaging research.

## Principles

Using each of these techniques effectively requires a great deal of specialized knowledge. However, it also requires understanding the basic principles of experimental design, statistical analysis, inference, and brain localization and function that cut across all of these techniques. We focus on MRI and fMRI in part because of their advantages and widespread availability, and in part because they, it's what we do. However, many of the principles contained in this book apply to other types of neuroimaging and beyond.

# **Chapter 2 - Why fMRI? Neuroimaging and the movement toward multidisciplinary science**

## **Neuroimaging and the 'common language' of the brain**

Human neuroimaging, especially fMRI and PET, is a growing new field now with thousands of publications per year. Why all the excitement? One of the goals of neuroimaging is a movement towards multidisciplinary science. This is one thing we're particularly excited about. For many years, people in different fields have been studying diverse aspects of the mind, the brain, and the body. Psychologists study the mind and behavior while neuroscientists study the brain. Medical and clinical researchers study the treatment and prevention of illness including those of the mind and the brain, which we increasingly understand to be interconnected with other body systems. Clinical trials study health related interventions and biologists study living systems. The fields of statistics, engineering, and computer science have each emerged as leading disciplines in the study of complex computational and biological processes with different traditions of techniques and approaches.

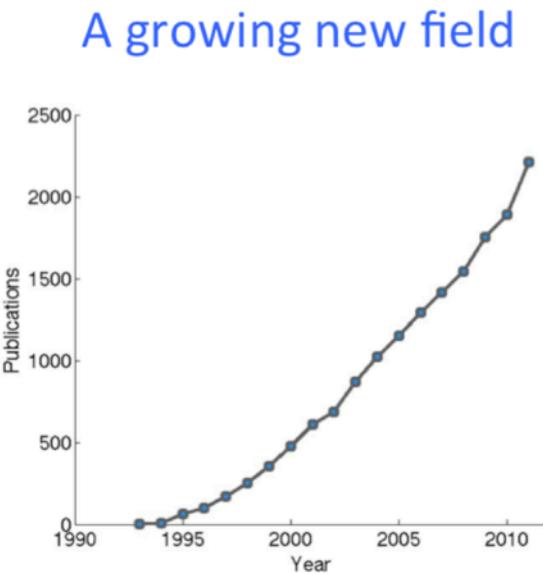
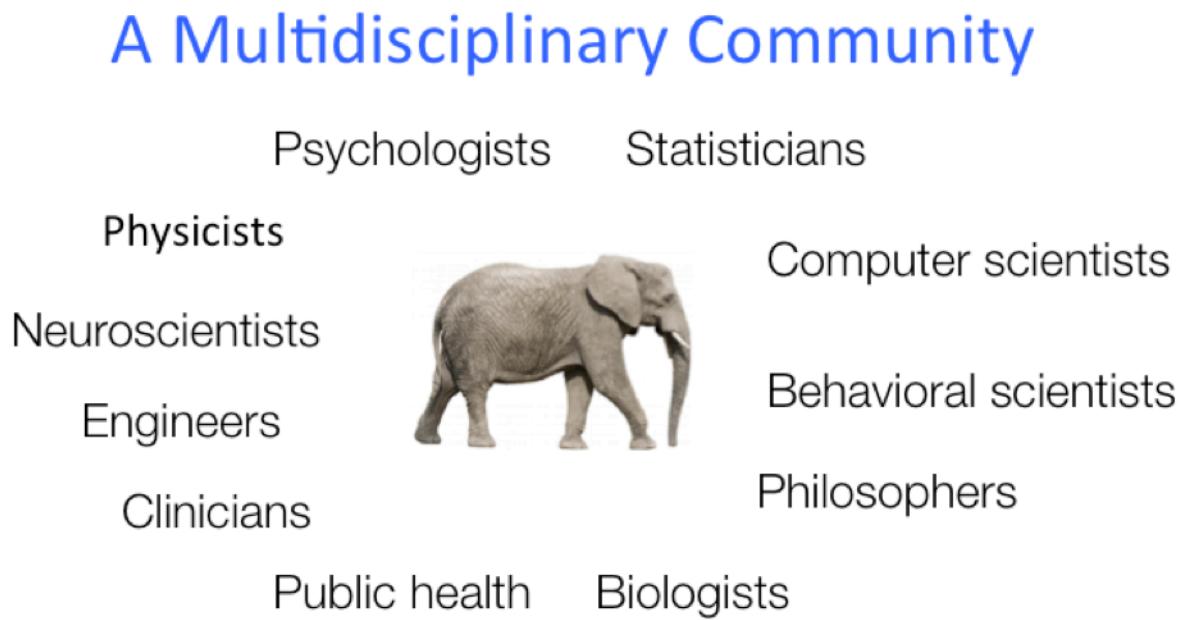


Figure 2.1. A plot of the number of publications per year in PubMed with the term *fMRI* in either its title or abstract.

These fields form rich but largely separate traditions. This is in some sense inevitable as a field grows and matures with a strongly shared history of knowledge and increasingly specialized techniques among its practitioners. This canalization and deepening of roots is complemented by new growth of fields that evolve at the intersections among established fields before developing their own research traditions. Psychophysicists study the mind as related to peripheral physiology. Neuroimmunologists study the brain as related to the immune system. Psychoneuroimmunologists study intersections of the mind, brain, and immune functions. Each of these disciplines provides a crucial but incomplete window into the most exciting frontier in contemporary science: the study of the mind and the brain - the study of us.

There is an old story about a group of blind people who each feel an elephant and try to understand together what they are observing. One person feels something long, rubbery, and flexible. Another perceives a smooth, firm surface and a third identifies a flat, delicate membrane. The study of the mind and the brain is a really, really big elephant. Its study spans several dimensions of analysis. One is a dimension of scale ranging from molecules to cells to systems. Another is a dimension of time from the opening and closing of ion channels in nanoseconds to the long-term relationships between brain and mind over a human lifetime or perhaps over the lifespan of a culture or a species. A third is a dimension of abstractness from concrete physiology to our capacities for abstract thought and emotion: for love, hope, cruelty, and empathy. Each discipline brings something unique to the table, but each specializes in a different “piece of the elephantee. To understand the whole image, we need to study these pieces deeply and rigorously, and then put them together into a picture of the integrated function of the human brain, mind, body, and environment.



- Need experts in each discipline working together
- Need individuals with multiple types of expertise

Figure 2.2. An illustration of the diverse disciplines working in neuroimaging.

The potential for such integration is one of the most exciting things about fMRI as a technique. Not only does the technology for collecting fMRI data draw on knowledge and techniques from at least a half dozen disciplines but fMRI can also be used to study just about anything related to the brain and the mind. This includes everything from abstract thought to cognitive performance, to mental illness and psychopathology, to brain regulation of inflammation in the body. For a practitioner to integrate the information and techniques required to do these studies well draws on knowledge from dozens of other disciplines.

fMRI and other types of neuroimaging also provide a way for practitioners of different disciplines to come together and speak in the “common language” of the brain. For example, consider a neuroscientist studying the molecular basis of learning, a pharmacologist interested in antipsychotic drugs, a psychiatrist examining depression, and a social psychologist investigating the nature of altruistic behavior. What do all these researchers have in common and what could they possibly converse about relating to each of their core scientific interests? Why, the dopamine system, of course! It is very likely each of these researchers has been studying brain processes related to the mesolimbic dopamine system, which connects the midbrain, ventral striatum, and prefrontal cortex. The researchers each might have results related to brain activity in the ventral striatum that could help inform the others’ ideas about what the system is doing in relation to their outcomes of interest.

Neuroimaging research can even help establish bridges between researchers in the same field who didn’t realize their ideas were grounded in similar neurophysiological processes. For example, some

social psychologists study motivation and appetite, others the effects of psychological distance, still others emotion regulation, and another group stereotyping and prejudice. All these areas contain a proliferation of theories, many of which include specific names and concepts (e.g., “construal level theory”). How do the mechanisms underlying these theories relate? Do some rely on the same core processes and systems and, if so, what are they and how are they related? Once again, the ventral striatum and medial prefrontal cortex likely play prominent roles in all these areas of social psychological inquiry. Grounding theories in models of brain function can help establish premises in measurable processes. These theories can then be shared across researchers and fields to facilitate building a cumulative science of social cognition and behavior.

## Multiple roles, multiple fields: An example

Let's look at some of the unique roles different disciplines play in an fMRI study by using an example of a basic fMRI study on how antidepressants work. Yes, we still don't really know much about how antidepressants, opioids, or any of the other systemic drugs (which we have been administering for decades or longer) work. This is in large part because these drugs affect neurons and glia all over the brain and we don't know much about the effects on the various systems that support thought, emotion, and decision-making. We don't even have a good consensus on which brain systems sustain those processes and which implement basic functions like attention, learning, and emotion. We do know a lot, but - to continue with our example - if we find that an antidepressant affects the prefrontal cortex, it is difficult to say what that means regarding the course of a person's mental health or their life.

So, back to our study - we won't try to solve the whole mystery at once. Rather, this study will simply seek to establish which brain regions change with antidepressant treatment in order to test whether the drugs do indeed alter the function of the prefrontal cortex and other brain regions. The *psychologist* uses expertise in experimental design to construct a task which can isolate particular mental processes related to depression. The psychologist and *statistician* both have expertise in ascertaining that the design is efficient and well powered, and that it will produce valid causal inferences about the effects of the drug on the brain. A *pharmacologist* has information about the cellular and molecular mechanisms of the drug's action and the kinetics of its absorption into brain tissue; the pharmacologist possibly also has data about its effects on brain vasculature and blood gas levels that may produce artifacts. A *psychiatrist* knows how drug dose and time course relate to expected clinical efficacy. A *neuroscientist* may have unique knowledge about how the drug penetrates into the brain and about the effects on neurons, glia, and/or various neural systems. The right training uniquely positions an *MR physicist* or *biomedical engineer* to ensure that we can obtain high-quality functional and structural images, and ideally minimize artifacts in the brain areas about which we care the most. The physicist or biomedical engineer may also have crucial information about how vascular and physiological drug effects might impact the fMRI signal independent of neural function. A *computer scientist* can manage and process the potentially huge volume of data acquired during the experiment, likely by borrowing signal processing techniques from *mathematics* and *electrical engineering*. During data analysis, the *statistician* again plays a

critical role in examining the data and the assumptions underlying the statistical tests, ultimately giving us a (hopefully valid!) picture of which brain areas the drug affects. A *neuroanatomist* can help localize the effects that emerge. The neuroscientist's purview, together with the psychologist and psychiatrist, is interpreting the results and their meaning.

That provides an overview of the different roles and contributions of various fields in an fMRI study. This description does not imply we need a team of 12 experts to do the study 6 in fact, that would be highly impractical. For the best science, we need collaboration of experts in multiple disciplines and individuals with proficiency in diverse aspects of design, analysis, and interpretation. A scientist using fMRI might come from any one of these disciplines, but likely has some capability in nearly all of them. While it's probably impossible to truly be an expert in each of these areas, a good scientist will know something about all of them, have some idea about what she or he doesn't know, and recognize when and how to ask for advice from colleagues.

A confluence is the running together of rivers into a greater river. This is what the collaboration of disciplines is like: many great rivers running together with their ideas and techniques intermingling and combining. This process is very good for both science and society far beyond the immediate applications of fMRI. This confluence can help those who learn and practice collaboration become educated in a rich set of scientifically grounded ideas. It can lead to new ways of thinking about the mind, health, and disease.

## Challenges and motivation for multidisciplinary science

All this sounds great, right? The catch is that it's actually not easy for people from different disciplines to work together because they must learn and talk about unfamiliar concepts and be willing to not be the expert. Collaboration requires scientists from different disciplines to care about ideas and problems outside the scope of their defined interests and perhaps to publish in journals unfamiliar to or not prestigious in their particular field (very few journals are prestigious across all fields). It also requires time spent educating other team members about basic concepts which are not groundbreaking within one's own discipline but which may be crucial and perhaps innovative in the context of interdisciplinary science.

For example, many MRI physicists are rewarded for innovating new methods to acquire data, not for explaining the basics of tried-and-true clinical study methods like our example above or for spending time tweaking those methods to minimize the artifacts in the brain structures which impact neuroscientists. Those who are willing to talk to the rest of us should be treated like gold, as should statisticians and others with specialized knowledge to contribute.

So how do we get people to talk to one another and work together? One answer lies in individual scientists developing multiple types of expertise, so that the gulf between the psychologist and the physicist, or the pharmacologist and the statistician, is not so great that they have nothing to say to one another. "Bridge" scientists are the glue that holds the team together. A little knowledge goes a long way in that respect, just like knowing a few words of someone else's language can produce a

dramatically different social interaction than sharing no words. Offering a route to develop expertise is one of the reasons we wanted to write this book.

Another answer is the movement towards multidisciplinary science, which is a challenging but laudable goal. Multidisciplinary refers to the idea that the study makes novel contributions to multiple disciplines. Take our example of the antidepressant fMRI study. If it is the study of a relatively novel drug with still unexplored mechanisms of action in the brain, it will be of interest to pharmacology. If it links two strong changes in thought and emotion, it may be of interest to psychologists and clinicians. If it involves novel innovations in data acquisition, it may be of interest in the field of MR physics. And if it involves novel computational methods to analyze brain networks, it might be of interest to the fields of computer science, informatics, and related disciplines. Not only is this difficult to pull off but also most studies should probably not try to be novel in so many different ways. However, the potential for innovation in multiple disciplines is one of the things that draw scientists from different areas together to contribute their expertise, creativity, and ideas.

# **Chapter 3 - Types of imaging: What PET and fMRI can measure**

## **MRI: Multiple measures, multiple modalities**

MRI is one of the primary tools in the modern neuroscientist's toolbox. Along with other, complementary techniques, it helps us develop integrated models of the human brain and discover how the brain relates to performance and health. In addition to being the seat of cognition, emotion, personality, motor control, perception, and social behavior, the brain is a central node in many body systems. It communicates with - among others - the muscular, circulatory, digestive, neuroendocrine, and immune systems. Understanding these relationships offers great prospects for improving our lives and helping us achieve our fullest potential.

One of the great things about fMRI and MRI is that we can use them to look at the complicated brain system in multiple ways. In a single MRI session lasting approximately two hours, we can obtain multiple types of images related to diverse aspects of brain physiology. Some of these are shown in Figure 3.1 and are described in more detail below. We can associate these measures with many aspects of cognitive function, emotional function, and peripheral physiology (e.g. the skeletomotor, digestive, autonomic, endocrine, circulatory, and immune systems) to provide a comprehensive set of relationships between the brain, the mind, and our health.

The two main types of brain measures that MRI and PET can collect are *structural* and *functional* images; see Figure 3.2. Structural brain imaging deals with the study of the brain's gray and white matter through static pictures of the distribution of neurochemical receptors. A close link exists between structural images and the diagnosis of disease and injury. For example, if your physician checks you for brain trauma after an accident, suspects that you had a stroke, or believes that you have Alzheimer's or Parkinson's disease, s/he might obtain structural images to help diagnose the problem. Functional brain imaging, on the other hand, includes measures of "activation" related to oxygen use, glucose use, and/or blood flow. It also incorporates molecular imaging techniques, which study the dynamics of neurotransmitters and other brain chemicals.

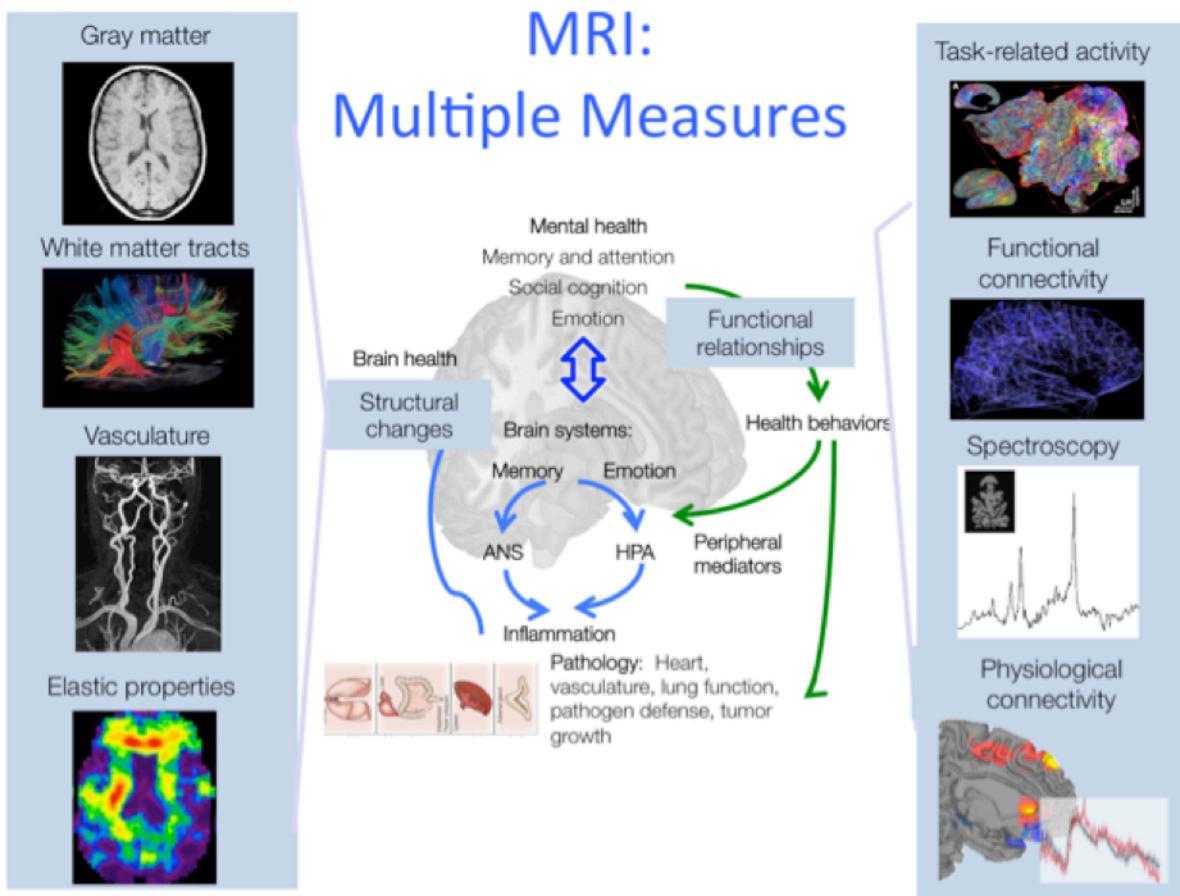


Figure 3.1. The multiple measures of MRI. In a single MRI session, we generally obtain multiple types of images related to diverse aspects of brain physiology. These measures can be associated with many aspects of cognitive function, emotional function, and peripheral physiology to provide a comprehensive set of relationships between the brain, the mind, and our health.

## Structural MRI imaging

The most commonly collected structural images are so-called *T1-weighted* and *T2-weighted* images, which provide basic anatomical pictures of the brain. Researchers most frequently collect T1 images, which they use to register functional images as well as for multiple types of anatomical analyses related to outcomes. This type of image is sensitive to the water content of tissue, so it produces different image intensities in the major in-brain tissue classes of gray matter, white matter, and cerebro-spinal fluid (CSF). T2 images provide a different type of contrast between tissue types; they are particularly useful for identifying the boundaries of certain iron-rich nuclei, such as the subthalamic nucleus and the substantia nigra.

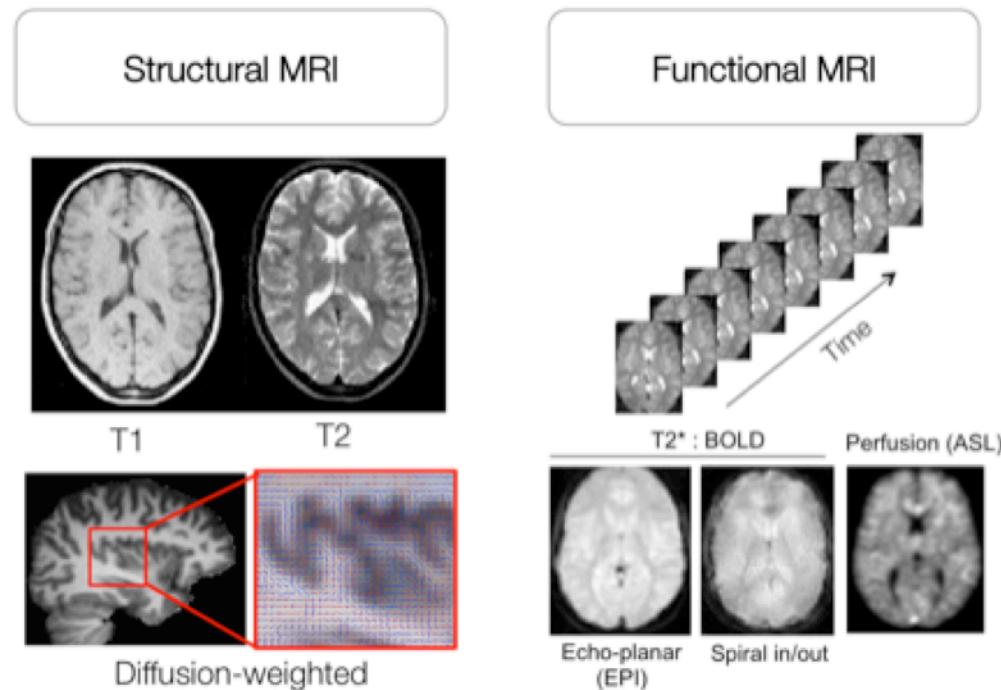


Figure 3.2. An illustration of various structural and functional images.

In addition to demonstrating basic anatomy based on tissue types, MRI can identify the major fiber tracts (fiber tracts are white-matter pathways composed of axon bundles connecting the disparate brain regions). The family of techniques researchers use for this identification is called diffusion weighted imaging (DWI) because of its sensitivity to directional water molecule diffusion in three-dimensional space. The first type of DWI was diffusion tensor imaging (DTI), which uses information about anisotropic (directional) water diffusion to estimate tensors (mathematical constructs that, for our purposes, we can think of as three-dimensional ellipsoids oriented along the direction of fiber bundles). There are now many varieties of DWI, which differ in both signal acquisition and in analysis techniques used to reconstruct the locations and the directions of white-matter tracts.

A third type of structural imaging is *vascular imaging*, which includes **MR angiography** and **veinography**<sup>3</sup>. These types of images are sensitive to flow in the large blood vessels, and thus produce an image of the vessels' locations.

These various image types demonstrate the versatility of MRI. The pulse sequences that control the parameters of the magnetic fields and radiofrequency applied to the brain can be configured to be sensitive to multiple properties of tissue. Creative uses continue to emerge. For example, **MR**

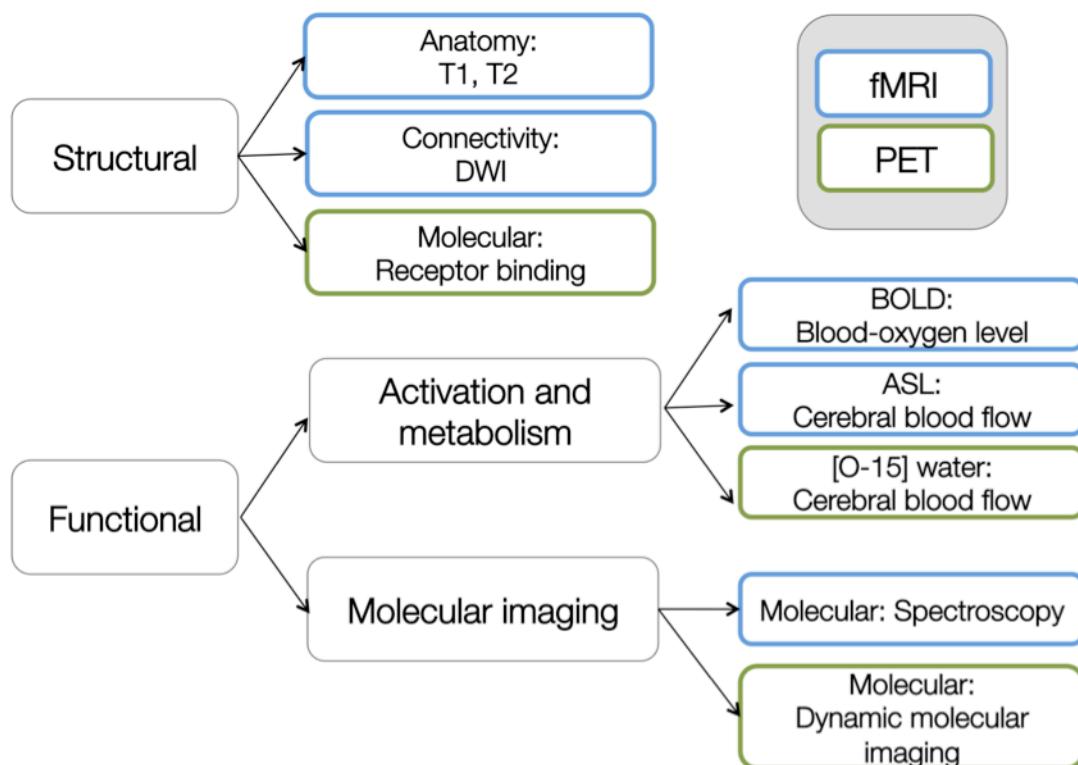
<sup>3</sup><http://www.ajnr.org/content/21/1/74.full>

*elastography*<sup>4</sup> measures the mechanical properties of brain tissue by ‘palpating’ the brain with sound waves. This technique may be useful for detecting subtle types of damage due to shearing forces of closed-head injuries and perhaps also for identifying normal variants.

## Functional imaging with fMRI and PET

Functional images measure parameters related to variations in metabolic activity, oxygen use, and release and reception of neurotransmitters in local brain regions. Both PET and fMRI can provide useful, complementary windows onto brain function, as shown in Figure 3.2. Meaningful effects can occur on time scales from seconds to years (or longer) and on spatial scales from neural column-level resolution to large-scale brain systems. As we discuss later on, different MRI and PET techniques are sensitive to different subsets of these effects depending on the imaging type and the analysis.

An overview of imaging measures



Both PET and fMRI can provide useful, complementary windows onto brain function

<sup>4</sup><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3066083/>

## ***Imaging brain “activity”: blood flow and metabolism***

Most functional imaging papers describe their results in terms of brain “activity” or “activation”. This is admittedly a vague term that can mean different things depending on which imaging modality the researchers used.

PET can measure several types of activation related to (among other possibilities) local cerebral blood flow (CBF), oxygen metabolism (CMRO<sub>2</sub>, or the “cerebral metabolic rate of oxygen”), and glucose utilization. PET involves synthesizing a radioactive isotope or “label” on-site and attaching it biochemically to a molecule of interest. Researchers then inject the radiolabeled compound into the blood. When the radioactive isotope decays, it emits two positrons - subatomic particles with the same mass but the opposite charge as an electron -that travel in opposite directions through the brain; an array of detectors positioned around the head then detects the positrons. Researchers mathematically reconstruct the frequencies of these emissions at different positions into three-dimensional volumes. The most common radioactive tracers for assessing “activation” are <sup>15</sup>O (“oxygen-15”), which assesses CBF, and <sup>18</sup>F (fluorine), which aids in deoxyglucose mapping. Researchers attach other radiolabels - for example, <sup>11</sup>C (carbon) or <sup>123</sup>I (iodine) - to other molecules and use them in a wide variety of molecular imaging applications.

In fMRI, “activation” typically refers to Blood Oxygen Level Dependent (BOLD) signal, which is a complex set of changes usually coupled to the oxygen demand and the blood flow in local blood vessels. For details of the relationships between these parameters and how they relate to BOLD signal, we refer the reader to a number of [more detailed<sup>5</sup> descriptions<sup>6</sup>](#) published elsewhere. The physiology of BOLD is also complex in its relationships to neural and glial processes, a subject we return to later in the book. Relative to other “activation” measures obtained with fMRI, BOLD signal is large in magnitude, robust, and obtainable via standard commercial pulse sequences (programs that run the scanner hardware). It is also safe to repeat, according to current wisdom, and does not require radioactive compound injection. For these reasons, about 95-99% of current fMRI studies use BOLD signal.

An increasingly popular alternative to BOLD is *perfusion imaging*. Perfusion imaging is a family of techniques that can obtain more direct measurements of local cerebral blood flow (CBF). Most notable among this family is [arterial spin labeling \(ASL\)<sup>7</sup>](#), a method that allows for quantitative measurement of regional CBF, in many cases, across long time scales (e.g. before and after cognitive training, mood induction, or clinical intervention). ASL uses radiofrequency pulses to magnetically label water molecules entering the brain through the carotid arteries, and then researchers compare the labeled MR images to the unlabeled MR images. With [appropriate models<sup>8</sup>](#), one can estimate local blood flow throughout the brain. There are many variants of ASL, but in recent years a technique called pseudo-continuous ASL (“PCASL”) has emerged as a stable and advantageous technique; it is now commercially available from scanner vendors. ASL can test the same types of functional

<sup>5</sup><http://www.ncbi.nlm.nih.gov/pubmed/9621908>

<sup>6</sup><http://www.ncbi.nlm.nih.gov/pubmed/11449264>

<sup>7</sup><http://www.ncbi.nlm.nih.gov/pubmed/8068529>

<sup>8</sup><http://www.ncbi.nlm.nih.gov/pubmed/9621908>

effects as BOLD, which include task-induced activation and connectivity, resting-state connectivity, and relationships between brain activity and performance (or other outcomes).

## ***Beyond activation: Molecular imaging***

Both MRI and PET have other ways of acquiring signal that go beyond “activation” and blood flow. These allow researchers to investigate regional functional brain changes in specific neurochemical systems. These include MR spectroscopy and PET molecular imaging. We turn to these next.

Another branch of MR techniques is *spectroscopy*. MR spectroscopy provides a way of testing a brain volume of interest for the presence of biochemicals and some kinds of gene expression. Nuclear magnetic resonance (NMR) spectroscopy takes advantage of the fact that molecules’ resonant frequency depends on their atomic characteristics, including the electron quantity and the proximity and composition of the tissue’s atomic nuclei. The resonant frequency determines the frequencies of the radiofrequency energy absorbed by local tissue. Thus certain molecules produce a relatively unique molecular “signature” in the power spectrum, with peaks at specific frequencies. NMR spectroscopy is a field in its own right, which has a growing number of detectable biochemical properties including GABA - a major, generally inhibitory neurotransmitter - and proteins related to specific aspects of the Krebs cycle - a fundamental series of molecular interactions that governs energy production within cells. Though promising, spectroscopy has shortcomings in limits in the detectable number of molecules and the relatively long time it takes to image each local brain region. Researchers do not yet widely apply spectroscopy in the cognitive neurosciences, although this may change as the field of neuroimaging matures.

PET provides the most comprehensive and versatile way of assessing specific neurotransmitters, neuropeptides, and glial cell function markers related to cognitive and health outcomes. Radioactive labels can attach to hundreds of different compounds - though for each compound, a great deal of work to understand the pharmacology and develop each research site’s procedures must be done before it can be used. Researchers have developed imaging techniques for many of the major neurotransmitters and neuropeptides, which include (among others) dopamine, serotonin, acetylcholine, norepinephrine, and opioids. The compounds are usually sub-pharmacological doses of receptor agonists or antagonists, which, like all drugs, bind to particular classes of receptors. In dopamine imaging, for example, a common technique is [<sup>11</sup>C]raclopride, which has high affinity for D2 receptors concentrated mainly in the striatum. An increasingly popular alternative is [<sup>18</sup>F]fallypride, which binds more powerfully to other dopamine receptor classes that are strongly cortically concentrated. Some other examples include muscarinic cholinergic receptors using [<sup>11</sup>C]scopolamine, mu-opioids using [<sup>11</sup>C]carfentanil, and benzodiazepines using [<sup>11</sup>C]flumazenil. Recent years have seen development of ligands for many other substances and cell markers as well, like those related to [neuroinflammation](<http://www.ncbi.nlm.nih.gov/pubmed/18006619> and **glial-cell activity**<sup>9</sup>.

These types of molecular imaging may be very useful for both basic research and to understand and diagnose clinical disorders. For example, PET imaging with a compound called “Pittsburgh

---

<sup>9</sup><http://www.ncbi.nlm.nih.gov/pubmed/25582579>

Compound B” or “PIB”<sup>10</sup> is sensitive to molecules found in neurofibrillary tangles characteristic of Alzheimer’s disease, so it is now used clinically as a marker for early-onset Alzheimer’s.

---

<sup>10</sup><http://www.ncbi.nlm.nih.gov/pubmed/14991808>

# **Chapter 4 - Brain mapping: A conceptual overview**

## **What is a brain map?**

Understanding the basics of brain mapping is increasingly important for a broad segment of society as brain images make their way into media, medical practices, courtrooms, advertisements, and other sectors of public life. However, without an explanation of the process and some of the ground rules, our understanding of how we construct brain images and what they can and cannot tell us about the brain and the mind is not obvious.

Both functional and structural imaging rely on construction of brain maps, which are maps of localized signals. There are many types of brain signals that we can map which relate to many external (outside the brain) conditions and outcomes. However, different types of brain maps rely on many of the same principles and underlying assumptions. We will devote this chapter to a conceptual overview of how researchers construct brain maps, what we can learn from them, and what some of those assumptions and limitations are.

The brain maps like those shown in Figure 4.1, generally speaking, are statistical constructions. In some cases, brain images display actual data values; this is typical in neuroradiology, in which experts ‘read’ an image and come up with an opinion or diagnosis. However, in most scientific areas, researchers want to make quantitative *inferences*, which means statistically comparing image data across conditions or individuals and then showing maps of the statistical results. We often call this practice *statistical parametric mapping*. Such maps show brain areas where researchers have deemed some effect of interest statistically significant.

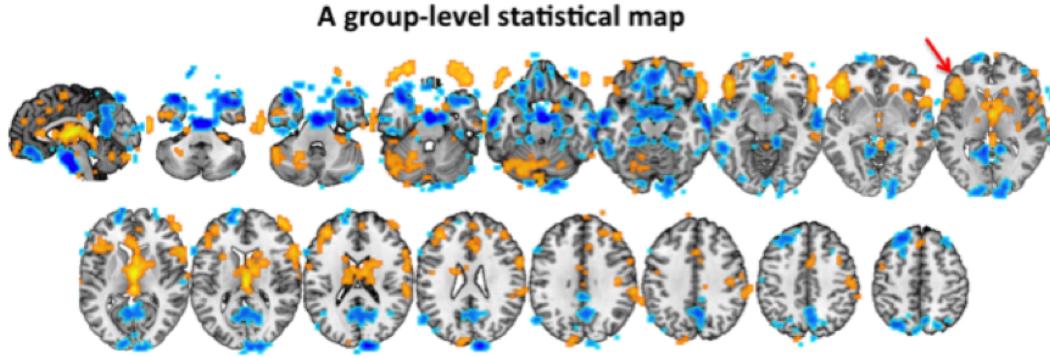


Figure 4.1. An example of a statistical map. These types of maps show brain areas where researchers have deemed some effect of interest statistically significant.

## ***Types of maps***

The types of processes that researchers map to local brain regions or networks are numerous. They include:

- Effects of experimental manipulations
- Correlations with behavior, clinical status, or other person-level outcomes
- Correlations with performance or other within-person variables
- Brain areas' correlation with other specific areas
- Brain areas' that are part of a group of areas (e.g. a *cluster* or *network*)

Accordingly, a first question to ask about any brain map is what effect it actually maps.

## ***Types of inference***

A second question to ask is to whom does the map apply - which individual or population of individuals? Data from only a single individual, scanned repeatedly, can be used to construct some maps as shown in Figure 4.2's top panel. We refer to these as *single-subject* maps. These maps are common in some sub-fields, such as vision science or primate neuroimaging, and are increasingly present in clinical and legal applications. Researchers can construct *single-subject* maps by comparing data from one condition (e.g. one experimental task) with another across repeated measurements, which thus test statistical significance in each brain region or 'voxel' (a three-dimensional cube of brain). Another method to construct these maps is by comparing an individual with a population of other individuals. If the statistics are valid (a big if!), such maps can say something useful about how an individual's brain differs from others' brains.

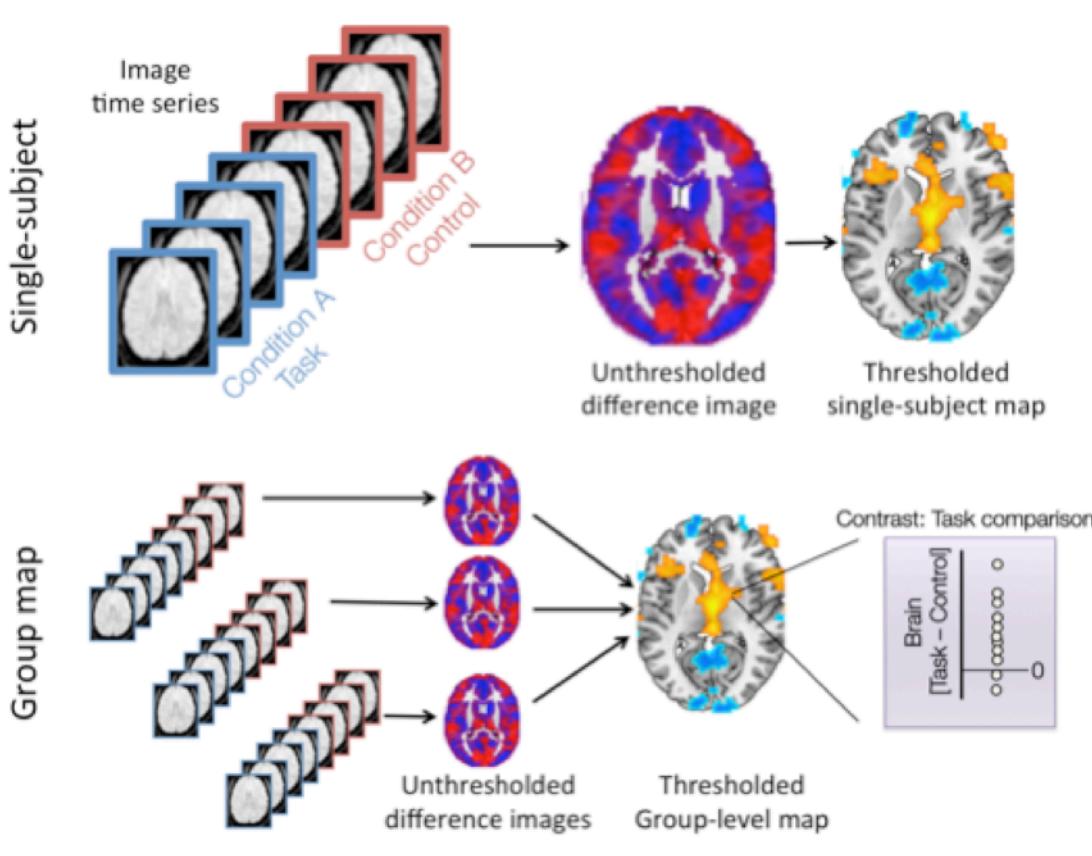


Figure 4.2. Examples of single-subject maps (top row) and group-level maps (bottom row).

However, single-subject maps cannot tell us much about the brain's general organization: researchers cannot use them to make *population inferences*, which are claims about how the brain functions in general. To make such claims, it is necessary to scan a group of participants and conduct statistical tests which explicitly evaluate how well the findings likely will generalize to new individuals. We refer to these as *group-level maps*; they identify brain areas that show consistent effects across individuals. The bottom panel of Figure 4.2 shows a schematic view of a group-level map's construction. Technically speaking, such maps require a statistical procedure which we often call *random effects analysis* because the statistical model treats each participant as a random effect. We will return to this in more detail in later chapters. The map shown in Figure 1 is a group-level brain activity map.

Maps can widely vary in what they reflect, but they all share the same underlying basic distinction between single-subject and population inference. For example, Figure 4.3 shows maps of three different kinds of brain connectivity. In this case, the colored regions do not show the significant effects of interest; here the lines connecting regions show the effects: they indicate significant functional associations across regions. The left map comes from a dynamic causal model (DCM), which analyzes dynamic regional changes from one second to the next, while controlling for other regions and experimental task variables, to examine relationships among regions. Lines show significant associations at a population level. The center map comes from a method which identifies

the most likely connections among regions and their variations across time. The connections the lines identify are not necessarily individually significant. This is common practice with many multivariate map types: one must be careful to make the correct inference because regions associated with a ‘network’ are not necessarily all significantly associated. Finally, the map at the right shows a large scale network in which each colored circle represents a brain region or system and each line shows significant associations across *studies*. Clearly, knowledge of a map’s construction process and its level of analysis are crucial for understanding what it means.

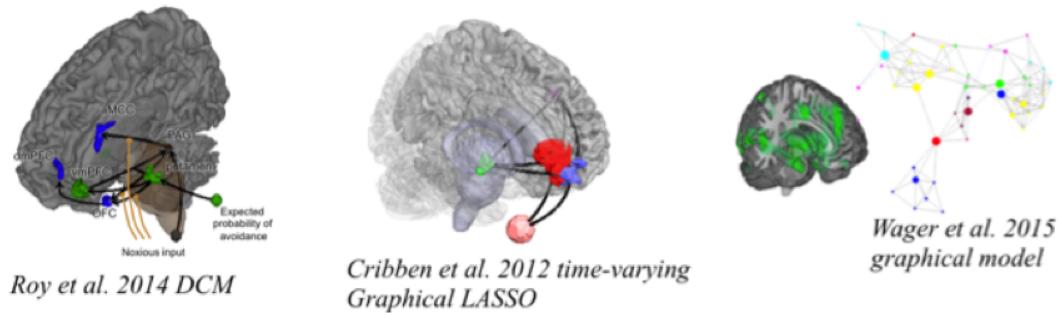


Figure 4.3. Maps illustrating three different kinds of brain connectivity.

## Fundamental assumptions and principles

In order to make statistical maps of all kinds, we rely on the assumption that the brain signals we measure reflect both effects of interest and noise. Researchers further assume that the noise is independent from the effects of interest (e.g. “random”). Repeated measurements in which the noise varies independently and stochastically allow us to obtain an average map that contains the true effect and reduces noise to a minimum. As the noise randomly varies around the true effect, it ‘averages out’, so the more data we collect, the closer the average noise will get to zero - as long as the noise is independent of the interest effect.

Consider the example in Figure 4.4. The brain - we show one representative horizontal slice here - contains some areas with a true effect, shown in blue. Perhaps this is a working memory task that requires people to maintain more versus less information in their minds; the map reflects concentration of the blue areas in frontal and parietal cortico-striatal networks. We observe a mixture of the true effects (signal) plus random noise, in red here.

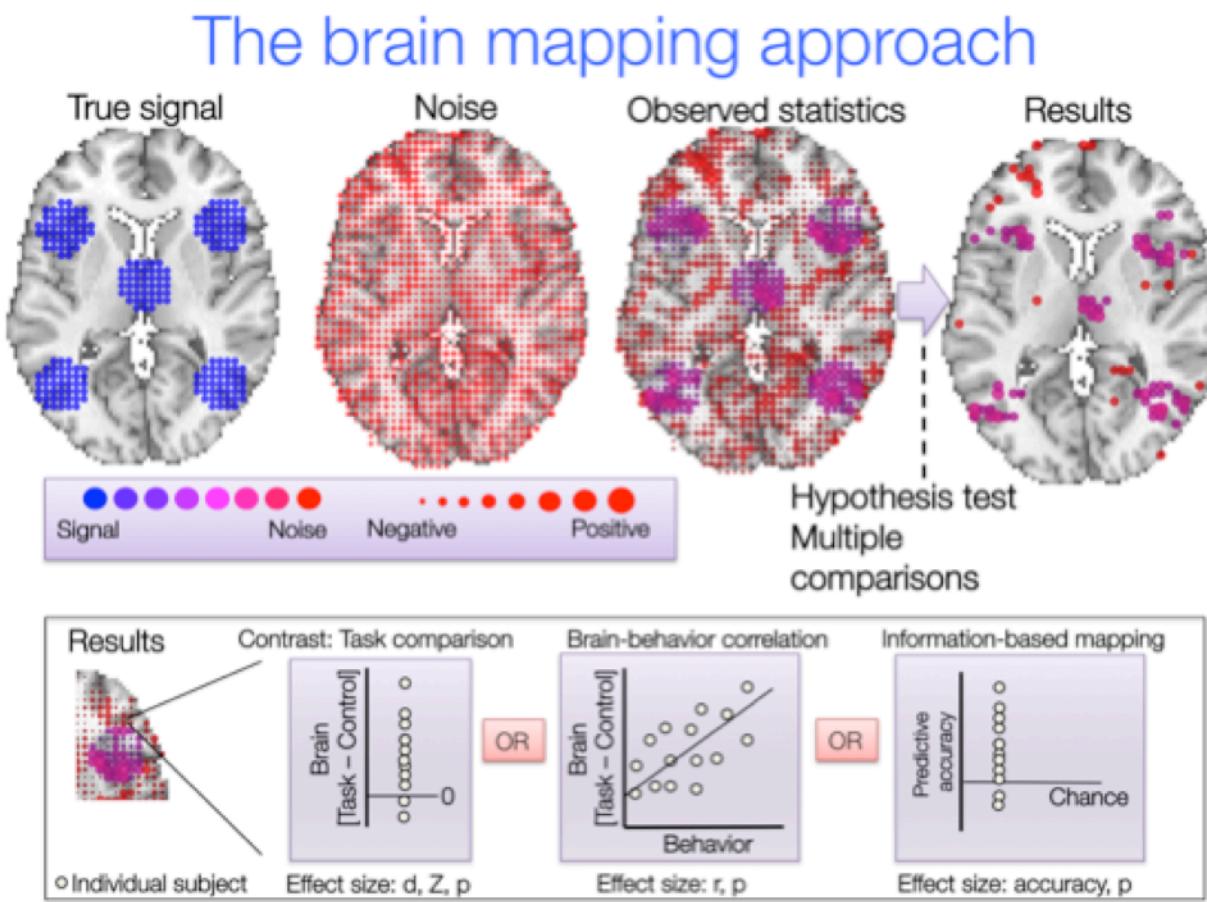


Figure 4.4. (Top) A single slice of the brain contains some areas with a true effect, shown in blue. We observe a mixture of the true effects (signal) plus random noise, in red here. Statistical tests are used to infer which voxels show true effects. (Bottom) Three common data types that go into such maps: task-related group analyses that compare a task of interest to a control task; brain-behavior correlations; and the average accuracy in predicting a stimulus category or behavior from each voxel's local multivariate patterns of brain activity.

Importantly, this noise is non-zero even averaged across the observed data, so we need to first separate it from the signal and then decide which areas really show the effect. We do this with a statistical test which compares each voxel's observed effect with its noise level (i.e. signal/noise). Common statistics, which include T-scores, F-values, and Z-scores, are all examples of such signal-to-noise ratios. We then compare the resulting statistic value with an assumed distribution to obtain each voxel's *p-value*. The p-value reflects the probability of observing a statistic value (e.g. a T-score) as or more extreme than that actually observed under the *null hypothesis* - that is, if there is no true effect. The lower the P-value is, the less likely that we believe the null hypothesis is true. We compare p-values with a fixed value to *threshold* the map and to infer which voxels show true effects. Because of the many possible tests, researchers often set a very high bar for significance (i.e. low p-values) by *correcting for multiple comparisons*.

When we use standard statistic values like T-scores and compare them with their canonical, assumed distributions, we are using *parametric* statistics. When we use the data itself to estimate the null

hypothesis' — which often involves fewer assumptions — we are using *nonparametric* statistics.

In most cases, we test each voxel in the brain separately, ignoring other voxels' potential influence, to construct brain maps. This is the case whether one maps activations which respond to a task, structural differences between groups, or functional correlation of areas with a 'seed' region of interest. It is a big assumption that the rest of the brain doesn't matter, so many multivariate analyses relax this assumption in certain ways (depending on the specifics of the multivariate model). However, the assumption is in some ways quite useful as we can interpret one brain area's effects independently of other area's responses. For example, a brain map which correlates activity levels in an anger-induction task with self-reported anger levels can provide a simple picture of which areas are associated with anger and so can be a starting point for more sophisticated models.

This basic brain mapping procedure applies to the vast majority of published neuroimaging findings, including both structural and functional imaging using MRI and PET. Figure 4.4's bottom panels show three common data types that go into such maps. On the left, the statistical brain map's voxels reflect a task-related group analysis that compares a task of interest to a control task. Each data point that goes into the test at that voxel (the circles) is the [task - control] contrast magnitude from one participant; the null hypothesis here is that the population's [task - control] differences are zero. The center map shows a brain-behavior correlation in which the test statistic is the correlation between the activity levels (often in a [task - control] contrast) and an external outcome, as in the anger example above. The right map shows an "information-based mapping" test in which the test statistic is the average accuracy in predicting a stimulus category or behavior from each voxel's local multivariate patterns of brain activity. In all of these cases, the above principles and assumptions apply.

### ***Bringing prior information to bear: Anatomical hypotheses***

Regardless of the type of map constructed and the variables involved, researchers' basic question is, "is there some effect at this location?" As Figure 4.5 shows, researchers can apply hypothesis tests to each brain voxel or to a set of voxels in pre-defined regions of interest (ROIs). They can also apply hypothesis tests to voxels in a single ROI or to signals averaged over voxels in one or more ROIs. These examples illustrate a progression from conducting many tests across the brain to performing few tests, a movement that depends on the prior information brought to bear to constrain hypotheses.

## Using prior information to constrain hypotheses

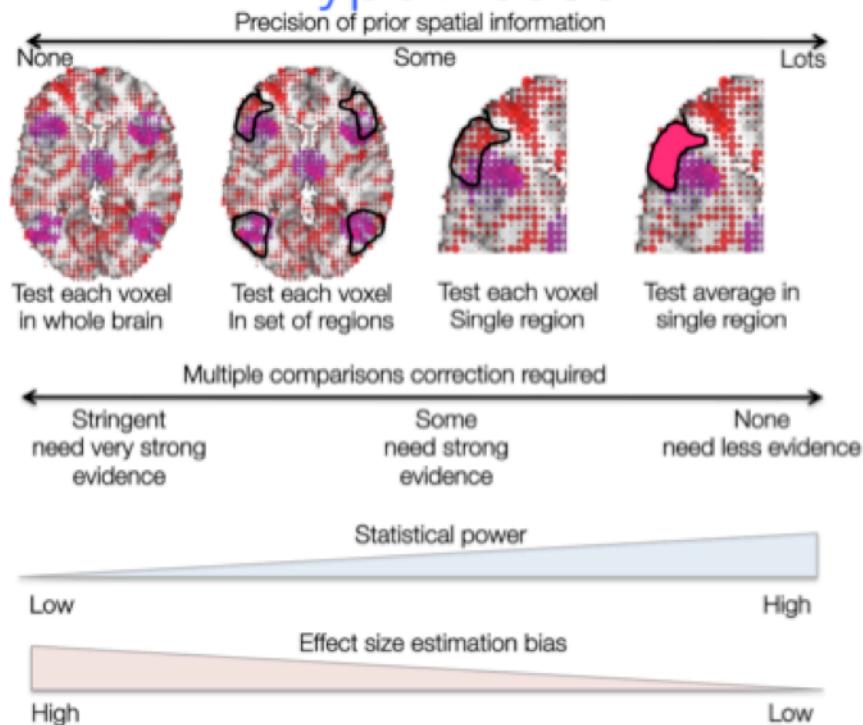


Figure 4.5. Researchers can apply hypothesis tests to each brain voxel, to a set of voxels in pre-defined regions of interest (ROIs), to voxels in a single ROI, or to signals averaged over voxels in one or more ROIs, depending on the prior information brought to bear to constrain hypotheses.

The more tests researchers perform then the more stringent the correction for multiple comparisons must be if they are to interpret all significant results as ‘real’ findings. As the threshold becomes more stringent, statistical power - the chance of finding a true effect if it exists - drops, often dramatically, which entails increasingly missed activations. In the extreme case in which there is only one ROI and the signal in its voxels is averaged, researchers perform only one test and do not need multiple comparisons correction.

Researchers need not limit *a priori* hypotheses to single regions; it is also possible to specify a pattern of interest, in which an average or a weighted average is taken across a set of brain region, and a single test is performed. Figure 4.6 shows an example from a [working memory study<sup>11</sup>](#). We first defined a pattern of interest based on previous working memory studies from [neurosynth.org<sup>12</sup>](#), which is an online repository of over 10,000 studies’ activation results. Then we applied the pattern to working memory-related maps from two participant groups - a group exposed to a social evaluative

<sup>11</sup><http://cercor.oxfordjournals.org/content/early/2014/09/22/cercor.bhu206.full>

<sup>12</sup><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3146590>

threat (SET) stressor and a control group - by calculating a weighted voxel activity average in the pattern of interest. Applying the pattern allowed us to (a) establish that, in our study, working memory produced robust activation in the pattern expected from previous studies and (b) test for SET effects on working memory-related activation without needing multiple comparison correction.

### Defining and using a pattern of interest based on previous studies

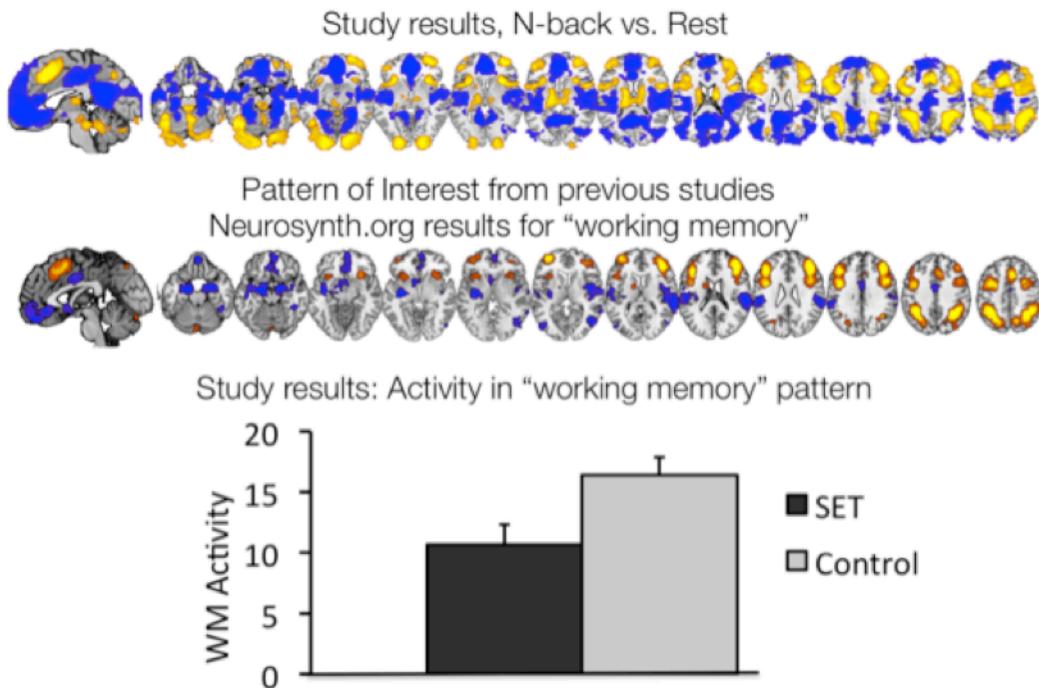


Figure 4.6. An example from a working memory study. A pattern of interest based on previous working memory studies was created. The pattern was applied to data from two groups (one exposed to a stressor and a control group) by calculating a weighted voxel activity average in the pattern of interest. This allowed for a test of stressor effects on working memory-related activation without requiring multiple comparison correction.

There are thus many benefits to specifying anatomical hypotheses *a priori*. However, when we specify *a priori* hypotheses, we must truly specify the region or pattern in advance based on data whose errors are *independent* from the dataset testing the effect, otherwise the p-values and the inferences will not be valid. We suspect there are many unreported cases of post-hoc “*a priori*” selection of ROIs.

## Types of inference: What brain maps can and cannot tell us

What can we infer from thresholded brain maps of all types, regardless whether they concern anatomy, neurochemistry, or functional activation? What we can make inferences about is rather specific and may not be exactly what you expect. Below, we discuss inferences about brain effects, a term which applies to many types of images that span beyond task-based activation, anatomical relationships with behavior, or maps of molecular imaging. First we discuss inferences about brain effects' presence, size, and location. Then we discuss forward and reverse inference, which, respectively, relate to making inferences about the brain and our psychological states (or other outcomes).

### ***Inferences about the presence, size, and location of effects***

The basic brain mapping procedure involves a test of significance at each voxel; this is a hypothesis test. This allows us to reject the *null hypothesis* that a subset of voxels has *no effect* in favor of an alternative hypothesis. That alternative hypothesis, however, is not very precise: it is merely that there is *some non-zero effect*.

As we will see in the next chapter, this does not let us conclude anything about *how big* or *how meaningful* the effects are; attempts to do so using standard hypothesis testing procedures can be highly misleading. At best, then, brain maps can allow inference that a set of significant voxels has *some effect*, but not how much effect.

Standard brain maps are also not very good for determining which voxels *do* versus *do not* show effects. Thus they are not useful to show us the complete *pattern* of activity (or structural effects, etc.) across the brain. This is primarily because of the stringent thresholds that usually limit the false positive findings. Current thresholding procedures do not optimally balance the number of false positives and false negatives (missed findings).

Another thing in which standard brain maps are not particularly good for is precise determination *where* the effects are in the brain. This may seem very surprising as researchers nearly always interpret thresholded brain maps in terms of where the most statistically significant results lie in the brain. However, the trouble lies in brain maps providing confidence intervals (which researchers use as a guide for how strongly to believe in the effect) on *whether* each voxel is significant but not on the significant voxels' *locations*. They provide a 'yes/no' value for whether a significant effect appears at each voxel. Inferences about result locations, then, are heuristic rather than quantitative.

This limitation becomes intuitive if we consider the brain map in Figure 4.1. The map contains significant activation (yellow) in the ventrolateral prefrontal cortex (vlPFC), marked with a red arrow. Imagine repeating this experiment again. What are the chances that the exact same voxels in vlPFC would be active? Or that the most active voxel would fall in the exact same location? We do not know. Standard mapping procedures do not provide p-values or confidence intervals on the activation's location or shape. However, we know from meta-analyses like the one in Figure 4.7

that the location of the peak voxel will likely be quite variable, possibly around plus or minus 1–1.5 cm. Incidentally, Figure 4.7 does show spatial 95% confidence intervals for the across study mean location for positive (green) and negative (red) emotions, drawn as 3-D ellipsoids. In addition to noise related uncertainty about local effects' locations and shapes, we also must keep in mind that artifacts and imprecision in anatomical alignment can also cause mis-localized effects. All brain images have an intrinsic point-spread function, or a blurring of localized true effects at one local brain point into a broader 'blob' of observable signal. BOLD images in particular are susceptible to arterial inflow and draining vein artifacts; they are also typically overlaid on an anatomical reference image which may not perfectly align with the functional map.

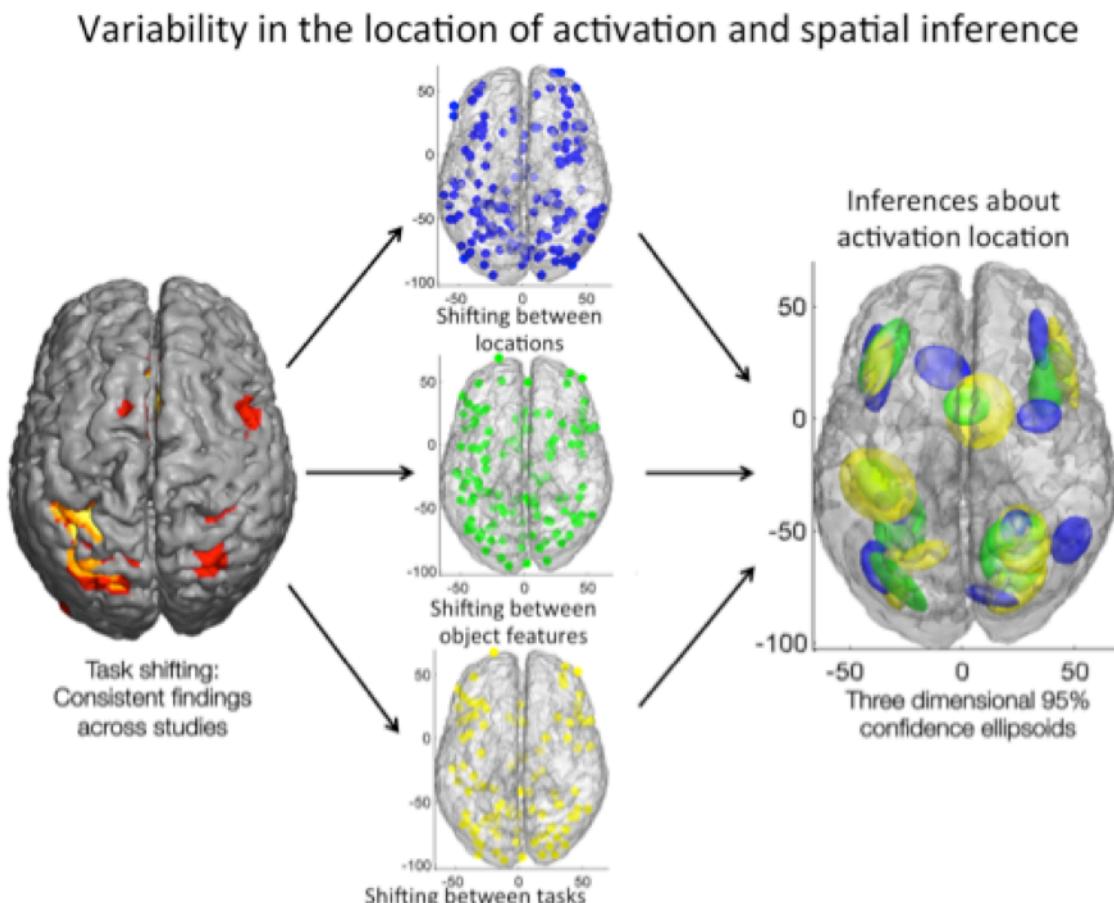


Figure 4.7. An illustration of the variability in the location and shape of activation.

The upshot of all this is that though we can make inferences about certain areas' activity, we must be cautious about over-interpreting size and location of significant findings and about the completeness of the picture thresholded maps provide.

If these types of inferences sound limited, we agree! Standard brain maps are very limited - we devote much of the next chapter to further unpacking their limitations. Fortunately emerging alternative methods avoid some of standard brain maps' problems. These include (a) specific multivariate

pattern analyses types that build predictive models and (b) spatial models that we can use to make inferences about the location of effects.

## **Forward and reverse inference**

Inferences drawn from brain maps have another limit. Typically, researchers either (a) induce a psychological state by manipulating experimental variables or (b) observe a behavior of interest or other outcome. Then researchers assume that the state or behavior is known and make inferences about the statistical reliability of brain activity given (or *conditional on*) the state or behavior. In Bayesian terms, we infer the probability of brain activity given a psychological state or behavior, or  $P(\text{Brain} | \text{Psy})$ . This is *forward inference*, which can tell us about how the brain functions under different psychological or behavioral conditions but not much about the psychological state or behavior itself (see Figure 4.8).

Standard brain maps provide information on forward inferences. Though above we expressed them in terms of probability, the same concept applies to effect size measurements. The stronger a brain map's statistical effects then the more likely we are to observe a significant result in probabilistic terms.

# Forward and reverse inference

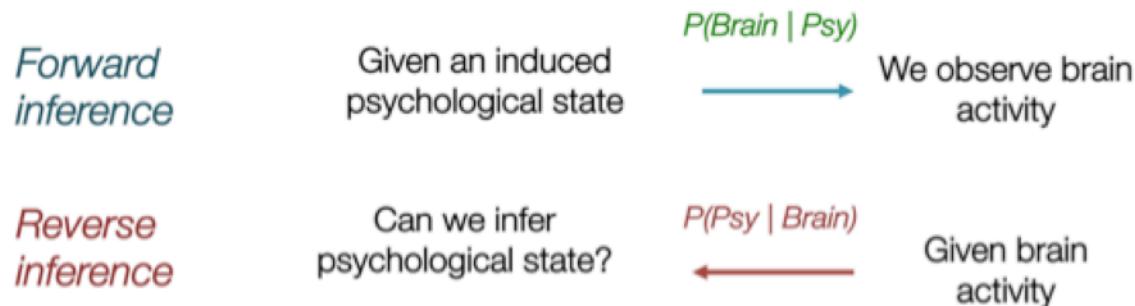


Figure 4.8. An illustration of forward and reverse inference.

Why can't standard brain maps teach us much about psychological states? Forward inferences take psychological states as given. They do not tell us how brain measures constrain our theories of which psychological processes are engaged. For that, the inference we want concerns  $P(\text{Psy} | \text{Brain})$ , the probability (or, heuristically, the strength) of a psychological process' engagement given activity in a particular brain region or pattern. Neuroimaging literature has termed this reverse inference. Though related through Bayes' Rule, forward and reverse inference are not the same thing qualitatively or quantitatively.

The field of logic calls fallacious reverse inference ‘affirming the consequent’. For example, assume this statement is true: ‘If one is a dog, then one loves ice cream’, or  $P(\text{Ice Cream} | \text{Dog}) = 1$  for short. Then given that Mary loves ice cream, i.e.  $P(\text{Ice cream}) = 1$ , one might erroneously infer that Mary is a dog. The problem is that all dogs love ice cream, but not all ice-cream lovers are dogs.  $P(\text{Ice Cream} | \text{Dog}) = 1$  does not imply that  $P(\text{Dog} | \text{Ice cream}) = 1$ .

Standard brain maps’ limitations in constraining psychological theory have led many researchers to be critical of neuroimaging, often<sup>13</sup> rightly so<sup>14</sup>. Examples of papers that make fallacious reverse inferences - like, for example, inferences that long-term memory processes were engaged (Psy) because the hippocampus was activated (Brain) - litter neuroimaging literature. In fact, some psychologists have argued that neuroimaging has not taught us *anything*<sup>15</sup> about the mind yet.

Reverse inference is actually possible; it is a major piece of the puzzle in constraining psychological (and behavioral and clinical) theory with brain measures. To understand how, let’s revisit forward and reverse inference from a diagnostic testing perspective.  $P(\text{Brain} | \text{Psy})$  is the ‘hit rate’ of significant activity given a psychological state; testing theory calls it sensitivity. In a standard test, e.g. a diagnostic test for a disease, Brain is analogous to having a positive diagnostic test, and Psy to having the disease.  $P(\text{Psy} | \text{Brain})$  is the test’s positive predictive value - how likely one is to have the disease given a positive test. High positive predictive value requires both high sensitivity and high specificity, which entails a low probability of a positive test if one does not have the disease - or, in brain imaging terms, low  $P(\text{Brain} | \sim\text{Psy})$ , where  $\sim$  means ‘not’. To use a brain example, before we can infer that hippocampal activity implies memory involvement, we must first show that hippocampal activity is specific to memory and that other processes do not activate it.

Thus to make reverse inferences about psychological states we must estimate the relative probabilities of a defined psychological hypotheses set given the data, typically by using Bayes Rule. This requires analysts to construct brain maps of multiple - ideally many - psychological conditions and assess the brain findings’ positive predictive value formally.

In addition to assessing positive predictive value, analysts can optimize maps and models of brain function to maximize function - that is, to strongly and specifically respond to particular classes of psychological events, behaviors, or other prompts. This is the goal of an increasing number of studies which use multivariate pattern analysis with machine learning or statistical learning algorithms. This is a promising direction; we devote a great deal of space to these techniques later in the book.

Ability to infer a psychological process’ presence or strength is important in its own right. It opens up various possibilities for testing and constraining psychological theories - or at least, their biological bases. Valid reverse inferences could allow, in some cases, researchers to infer a number of processes otherwise problematic or impossible to confidently measure. Among others, these states include being in pain, experiencing an emotion, lying or hiding information, and engaging in cognitive work. Researchers can use reverse inferences to probe the unconscious and to help study mental processes in cognitively impaired, very young, or otherwise unresponsive individuals. And, finally, comparing brain markers for different psychological processes could allow us to develop new mental

<sup>13</sup><http://www.ncbi.nlm.nih.gov/pubmed/8585670>

<sup>14</sup><http://www.ncbi.nlm.nih.gov/pubmed/16406760>

<sup>15</sup><http://www.ncbi.nlm.nih.gov/pubmed/16771037>

process typologies - including emotion, memory, and other processes - which, regardless whether they match our heuristic psychological categories, may have their own diagnostic value.

In conclusion, standard brain maps provide specific types of inference about brain activity. Though there are a number of fundamental limits to these inferences, new techniques are circumventing many of those limitations and providing a more complete range of inferences about the brain and mind.

In the next chapter, we further explore those limitations, some ways that researchers exploit brain maps to support erroneous conclusions, and how you can become a savvy consumer of neuroimaging results.

# Chapter 5 - Limitations in inferences from brain maps

In these next chapters, we explore the dark side of neuroimaging results. In this chapter, we elaborate on seven limitations in what one can infer from significant results in standard neuroimaging maps. In the next chapter, we discuss several fallacious arguments for which to watch as inspired by the classic book, ‘How to lie with statistics’.

We discuss statistical inference and multiple comparisons extensively in later chapters, but for now it is worth remembering a few relevant points which fundamentally constrain what we can say about the brain based on maps like those discussed above. First we start with seven caveats in what brain maps allow us to infer about local regional activation. Then we provide a non-imaging example to drive home some of the statistical points.

## Seven caveats in brain map inferences

### ***Not significant does not mean not important***

Statistical significance in only some brain areas does not mean that effects in other brain areas are negligible. Most studies have extremely low statistical power, which is the chance of finding a true effect if it exists. Additionally, most tasks and outcomes of interest likely involve many more brain areas than are visible on thresholded maps. Multiple comparison correction reduces the false positive rate but also reduces power, often dramatically.

### ***Maps are noisy***

Not only is every statistical map noisy but also the smaller a sample size is, the greater noise-related variability becomes. This fact means one should not over-interpret voxel significance without evidence that the significant voxels’ pattern is reproducible. In addition, maps across multiple tasks and outcomes, even if they reflect the same underlying processes, are likely to yield different activation patterns directly proportionate to the maps’ noisiness. In addition, more stringent multiple comparisons correction increases the dependence of the choice which voxels are or are not significant on the tails of the applied statistical distribution, which means increased influence of noise. More studies are starting to examine the reproducibility of the maps themselves.

## We cannot infer the size of the effects in significant regions

Thresholded maps cannot effectively tell us how large (i.e. how important and meaningful) effects in ‘significant’ regions really are for two reasons. First, the statistical tests we described above are hypothesis-testing procedures which test whether an effect is zero or not, but don’t tell us how large it is. Second, thresholding the maps creates a *voxel selection bias* which inflates apparent effect sizes in significant regions and makes it impossible to determine how large they are. Every significant region contains some true signal (if they are true positive findings), but also contains noise that may favor the hypothesis by chance. This makes effects appear very large in some studies, particularly those with small sample sizes, with higher variability. Multiple comparison correction does not help; it actually makes this problem *worse* by increasing ‘selection pressure’. This may sound surprising, but it’s true; below, we illustrate why this problem occurs through a non-imaging example.

## Maps are relative comparisons

Brain maps nearly always compare one condition with another: [Task - Control], individuals high in a trait versus those who are low, patients versus controls, etc. That means that the baseline condition against which researchers compare an active condition matters too. Though many maps do not seem to have a comparison condition, they have an implicit baseline, which is often resting activity levels. This is true for all image types, even ‘quantitative’ cerebral blood flow (CBF) measures made with PET or ASL imaging. For example, we might measure motor cortex blood flow as  $50\text{ml}/\text{min}$  during motor task performance, but this value is only meaningful when compared to a no-movement or related baseline, or to another group of individuals. The same is also true for molecular imaging, anatomical imaging, and spectroscopy. With BOLD fMRI, the comparisons we can make are even more limited. Because BOLD’s absolute units are not very stable from person to person and scanner to scanner, BOLD maps nearly universally involve comparisons between conditions rapidly fluctuating over time (i.e. every few seconds to every few minutes).

Functional connectivity measures, including correlations in BOLD time series among brain areas, can potentially reveal changes happening across longer time scales. However, like other brain measures, their usefulness in this respect depends on how reproducible the levels of those connectivity measures are across time, participants, and, ultimately, scanners. Random variations in the absolute levels of the response or in the signal scaling and noise levels can affect the utility of brain measure.

## Significant does not mean causal

All brain imaging is essentially correlational, so it is difficult to confidently conclude that brain effects cause behavioral effects and other outcomes. For example, researchers robustly associate Alzheimer’s disease with reduced medial prefrontal cortex gray matter density. But those associations alone do not warrant inferences that those changes cause memory impairment or other symptoms. Researchers often base such inferences on converging evidence from other neuroimaging studies and methods, particularly on animal studies which can manipulate the brain and provide for invasive measurements. Showing that such structural changes correlate with memory symptoms

is helpful, but such effects are subject to *indirect effects* of other brain regions' changes and other unmeasured *common causes* of brain changes and memory impairment. Experimental manipulations, which can be exogenously randomized, provide the strongest evidence for causality. But these also do not strongly imply that an experimentally manipulated task independently activates each significant brain area shown in the thresholded brain map. Connections with areas that are more central for task processing may indirectly activate some areas. Others may be driven by alternative extraneous processes or even by artifacts like head movement. With any brain map, it is useful to consider all the processes that might drive brain activation. For example, seeing faces evokes some relatively face-specific perceptual processes. But a map comparing face viewing versus rest does not necessarily isolate those processes. Activity could reflect any process which differs between face viewing and rest, including attention direction, associative memory retrieval, eye movements, emotional and motivational responses, arousal, or autonomic physiological responses. Just because a brain map is nominally 'about' some process doesn't mean all observed activity reflects that process.

### ***Statistically stronger does not mean more important***

We often use an effect's statistical strength as a guide to its importance, but one should make such inferences with caution. Significance in a group-level map implies an effect which is consistent across individual participants. However, both large effects and low inter-individual variability can drive strong effects. Thus the most statistically reliable effects may be the least important for determining individual differences in performance or other outcomes. In addition, noise and artifact levels vary widely across the brain. The most important regions for a task or an outcome may often be those with high noise or poor signal quality which result in weak effects. Lastly, large effects do not guarantee importance. Many of the players critical in virtually every system are often not the most numerous. This applies to the brain even at neural or neurochemical levels: take, for example, motivating events which induce dopamine, opioid, serotonin, and other neurochemical release. Finding that more dopamine D1 than D2 receptors are bound during reward states does not imply that D2 receptors are more important for any particular behavior.

### ***Anatomical localization is imprecise***

The functional images used to create statistical brain maps are usually inherently blurry and are often subject to artifacts including spatial distortions. In addition, these maps usually overlay an anatomical image that, while higher resolution, does not perfectly align with the functional images and is not subject to the same distortions. Often researchers register and overlay group results on a standard brain atlas constructed from averaged brains, which necessarily reduces spatial precision. All this means that we should interpret locations of functional results cautiously. Higher-resolution imaging at higher field strengths and individualized single-person analysis have potential to circumvent, to some degree, many such problems. However, these techniques have their own costs and so researchers do not yet widely use them.

## A non-imaging example

Let's consider a non-imaging example to illustrate some issues with inference in the brain mapping framework. Imagine observing a casino's roulette games over a year's time: maybe 100,000 such games would occur, each which involves a gambler's attempt to win money by spinning the roulette wheel exactly 8 times. Each game is like one voxel's test of significance with the null hypothesis that the table's odds are 'fair', or as stated. The set of 100,000 games together is like a brain map composed of 100,000 voxels, which is a typical number.

In roulette, a player can bet on black or on red with just under a 50% chance of each occurring - 47.37%, to be precise, because the house constructs the odds in its favor. What are the chances the roulette ball will land on black all 8 times? The answer is  $0.4737^8$ , or  $p = 0.0025$ . If we just observed one game which came up black 8 times in a row, we might provisionally reject the null hypothesis and conclude the table may be biased. However, now imagine a fair table (there are no true biases) on which we observe 100,000 games, just like we might observe 100,000 voxels. What is the chance we will observe 8 consecutive black outcomes (at  $p = 0.0025$  per outcome) in at least one game? It is essentially 100%. In fact, there is a 92% chance that we'll get at least one 'significant' test - a game with 8 black outcomes, or equivalently a result somewhere in the brain - with only 1,000 tests. Brain image noise is not independent across space, so though it's tricky to determine the effective number of tests in our 100,000 voxel map, it is safe to say that we'll find a significant result at  $p < 0.0025$  in a brain map every time. This is why we need to correct for multiple comparisons. Now comes the rub in terms of estimating how large effects are. Let's say we threshold our 100,000 test results at  $p = 0.0025$  uncorrected, then consider only games (voxels) with significant results. We'd expect 250 of 100,000 games to be significant on average even if the table is fair. The true value for the average number of blacks is just under 4 per game. However, if we examine only the winners - that is, we *condition on* a significant result - the average number of blacks is 8, double the true effect size. Similarly, picking out brain voxels with significant effects always inflates the apparent effect size, often dramatically.

# Chapter 6 - How to lie with brain imaging

In this chapter, we explore the dark side of neuroimaging results. We discuss several fallacious arguments for which to watch out. In writing this, we are inspired by two classic books. One is called ‘How to lie with statistics’, which, of course, really tells you how you should not lie with statistics or at least how to avoid being fooled by those who do. The other book is Bob Cialdini’s terrific ‘Influence’, in which he claims that his own gullibility inspired him to study persuasive power and resistance. Accordingly, this is not really a chapter about how you can lie with brain imaging, in case you were wondering. It’s really a chapter about what not to believe.

Below, we describe five tricks to make your results look specific, strong, and compelling, and also to make them come out like your theory predicted. For example, if you have a theory that requires two psychological tasks to produce highly overlapping brain activity, we can help you make that happen. Or if your theory specifies that patients and controls engage very different brain systems, we can help with that too.

Of course, these are not the only ways to lie with brain images. There are the obvious ways - plain old making stuff up or engaging in a little self-deception like defining ‘*a priori*’ ROIs after peeking at the statistical maps (because you would have expected activation in the precuneus, right?). There are also techniques like ‘P-Hacking’, which include sleights of hand such as continuing data collection, adding and removing covariates, or transforming outcome measures until you have a significant result. We’ll discuss those more later. Here, we’re interested in techniques that are, at least in some cases, a little bit more subtle and that apply even to brain maps generated through otherwise valid means.

## How to tell a story about the “one brain region”

### *The high-threshold*

Most clinical disorders and many processes which psychology studies are likely distributed across multiple brain systems. How can we make such a bold claim? To be encapsulated in one brain region, a process must be relatively pure, which implies that localized lesions produce complete and specific deficits. This is true in a few cases: V1 lesions produce cortical blindness and specific inferior temporal lesions produce prosopagnosia, a face recognition deficit. But most processes, even evolutionarily conserved and sensory-driven ones like pain, are highly distributed. The trouble with this is that a neuropsychological tradition which focused on selected cases of specific deficits after focal lesions created a past ‘culture of modularity’. Prestigious journals like Nature and Science have

historically vastly preferred simple results with one-point headliner messages like ‘this brain region implements this complex psychological process’ (we won’t pick on any specifics). So how do you get your results to tell that simple story?

The answer is very simple: the *high-threshold*. Simply raise the bar for statistical significance until you have one region (or very few) left in your map. Not only is this useful for writing a paper around a single brain region which enables emotion, goal setting, attention shifting, hypothesis testing, or whatever you’re studying but it is also really useful if you see significant activation in the white matter or the ventricles - places you shouldn’t see activation in artifact-free statistical maps. The antidote is to (a) choose the threshold a priori and (b) require researchers to show the entire map, including the ventricles (or at least to check it).

## How to make your results look really strong

Strong results mean large effect sizes which include high correlations between healthy-sized, meaty-looking blobs with bright colors and brain measures and outcomes. There are two techniques to ensure your brain map looks as it should no matter how weak the effects actually are.

### ***Circular selection (this technique is also known as the voodoo correlation)***

Let’s face it: most complex personality traits and clinical symptoms are unlikely to strongly correlate with any one brain voxel. The reliability of both brain and outcome measures limit such correlations’ true values. The heterogeneity of outcome measures also limits them: there is no single reason why people feel depressed, experience neuropathic pain, or are schizophrenic, courageous, or optimistic. Additional limits include person-level factors which affect brain response magnitude unconnected to outcomes of interest: among these factors are individual differences in hemodynamic responses and vascular compliance, blood iron levels, alertness, and caffeine intake. However, isn’t it more convincing if your brain findings correlate with optimism or anxiety above  $r = 0.8$ ?

Yes, virtually any study can achieve this. The procedure is simple: first run a correlation map across the whole brain, then select the peak region and test that region’s correlation. If your sample contains 16 participants, then any voxel with a p-value less than 0.005 will show a correlation of at least  $r = 0.8$  or so. Now, maybe you’re worried about not finding any voxels with such a low P-value.... but don’t be. If you test only 1,000 independent comparisons, you have a 99% chance to get at least one significant result, even with *no* true signal anywhere in the brain. Add to this that brain maps can easily contain 100,000 voxels, though they are not independent. And, of course, if you have some voxels with more modest true correlations - say, in the  $r = 0.1$  range - then the chances are even greater that you will select a voxel with an apparent  $r = 0.8$  correlation, or higher. Small sample sizes will increase your success, too, because they are more variable across the brain. With only 8 participants, the average significant voxel at  $p < 0.005$  will correlate above  $r = 0.93$ .

There is more good news as well: this technique will work for any effect size measure whether it is a correlation, a difference between experimental conditions, or a multivariate pattern analysis classification accuracy.

If you do not want others' circular selection to fool you, you will need to know that (a) there was a priori selection of all tested regions and (b) the report includes all tested effects. And keep in mind that (c) if there are many tests, some will show large effects by chance.

### ***The low-threshold extent correction***

Circular selection will make your effects look really strong, but won't create those large, fruit-colored blobs on your brain map. Such blobs are important because human minds naturally confuse 'lots of reported areas' with 'strong effects', even if the two are unrelated. The solution is to lower the statistical threshold until you get large blobs - and possibly to mask out the pesky white matter and ventricle activations that tend to appear at low thresholds. The problem is that reviewers are savvy and will ask you to report results with multiple comparisons correction.

There is a method to lower your statistical threshold and still claim rigorous multiple comparisons correction. How is this possible? Fortunately the technique called cluster extent-based correction lets you set as liberal a 'primary threshold' as you want (say,  $p < 0.05$  uncorrected) and then correct for multiple comparisons based on the extent of the blob. Among other problems, correction methods are too liberal with such low thresholds (<http://www.ncbi.nlm.nih.gov/pubmed/24412399>). The bonus is that your figures' maps will show all the voxels significant at the liberal, uncorrected threshold even though you can at best actually only claim that the activated area has some true signal somewhere in the activated area.

This antidote to this trick is to use more stringent primary thresholds, to clearly indicate each significant region's identity in figures, and to make it evident that most voxels which appear in the figure may not actually be activated. Or, of course, to avoid extent-based thresholds altogether.

## **Overlapping processes: How to make two maps look the same**

### ***The overlap zoom-in***

Let's say that your theory focuses on overlap across two or more processes such as two types of emotion, pain, or cognitive control. You scan two tasks and compare each one's activation maps with its respective control condition. To support your theory, simply focus on the overlapping voxels and assume non-overlapping ones are due to noise. Now even if the maps are 95% different across the brain, you can still claim support for your theory. You might also do a multivariate 'searchlight' analysis that looks explicitly for similar brain regions across the two processes. Anything significant in the map is positive evidence and the remaining brain areas in which the tasks are dissimilar are just inconclusive null results attributable to low power.

If you are not getting enough overlap, the **low-threshold extent correction** can greatly amplify the extent of your activation patterns and thus increase the apparent overlap. Hopefully most reviewers will not realize that this is not a valid test as your comparison is between two maps with ‘some true signal somewhere’ at each individual voxel, as though each voxel were significant. And, finally, to enhance any of these techniques, you can make a figure that focuses selectively on the overlap locations.

The antidote of this technique is to provide unbiased similarity measures across the whole brain including regions that might be shared or unique. Such approaches are not common in neuroimaging literature yet, which makes this technique particularly hard to counteract.

### ***The low-level control***

If the overlap zoom-in does not provide enough ‘evidence’ for overlapping activation, try this additional technique. Similarity is relative: an apple and a banana are dissimilar when compared to an orange but are quite similar when compared to roast beef. Likewise the technique for making the activity maps of two tasks very similar is to compare them to a very dissimilar control condition. Of course, reviewers might object if you compare your two tasks to a third which is very dissimilar. Fortunately, however, there is a perfect comparison condition that will not raise eyebrows, namely rest.

Imagine you have a theory that altruism is an automatic human response (which it actually may be). You posit that punishing others produces internal decision conflict even if they deserve it. Thus you would like to demonstrate that brain responses are similar when unfairly punishing others and within a cognitive ‘conflict’ task. No problem. Simply compare each to rest then look at the overlap of the resulting activation maps. Many low-level processes will activate in each map: processes involved in most cognitive tasks such as orienting attention, making basic motor decisions, and executing them. If your study is sufficiently powered, you will observe beautiful overlapping activation in areas including the anterior cingulate, the anterior insula, and the supplementary motor cortices.

The antidote to this technique is to require tight control of the tasks or, even better, to track parametric strength increases of each process in which you are trying to assess the overlap. Then the maps you compare will be more tightly constrained to reflect the cognitive processes of interest.

## **How to make two maps look really different**

Now let’s assume that you have the opposite problem. Your theory dictates that different processes should be involved in two or more maps. Perhaps you suppose that children with attention deficit disorder process cognitive stimuli differently than those without the disorder. No matter how similar the underlying brain processes are, you can always conclude that the activation patterns are distinct if you so desire.

The **high-threshold** can come to your aid again here. Because every brain map is variable and the locations of significant voxels vary, the chances that any two maps will produce overlapping voxels

decreases as the threshold increases. Alternatively the **low-threshold extent correction** can also be helpful as it will produce large blobs in mostly non-significant constituent voxels. If you focus on the differences in maps rather than similarities and zoom in on areas with apparent differences, then you will be able to convince most readers that the activation maps are quite distinct. If you analyze the spatial patterns, e.g. by correlating the maps across voxels, then the noisier your maps are, the more likely your maps will be uncorrelated.

The antidotes here involve spatial tests, analyses in which one generates a p-value for whether two tasks activate distinct brain locations, and treating participants as a random effect (we discuss this in more detail later). Without spatial tests, no principled null hypothesis exists for how many voxels should or should not overlap in two truly similar underlying processes, so you can essentially say whatever you want. However, if a reviewer should require you to do a spatial test, you would need to demonstrate statistically significant differences in the activated areas' location or shape. This is a much higher bar to pass, especially considering that interpreting the voxel overlap heuristically is really no bar at all.

There are also antidotes which relate to spatial pattern tests. Reviewers may require you to demonstrate that each of the two patterns you correlate (a) are reliable, with high correlations with themselves or related within-task measures at re-test and (b) strongly correlate with a task state or outcome. If so, then the bar is again raised: a null correlation across tasks is contextually meaningful with positive correlations within-tasks.

## Conclusions

With this chapter, we hope we have been able to show you that you can take valid, albeit noisy, statistical brain maps and shape their presentation to fit your theories in multiple ways, independently of the truth. Of course, we do *not* want you to actually do this (in case anyone missed that point). We want you to be aware of these deceptions and self-deceptions to ensure that your analyses are unbiased so data shapes theory rather than the opposite. The best way to make sure this happens is to care more about discovering something *true* than about finding supporting evidence for a particular view or theory. This is sometimes difficult when the truth does not line up with our publication goals and cherished beliefs, but it is the austere path of science.

## **Part 2: Fundamentals: The origins of PET and fMRI signals in the brain**

# **Chapter 7 - fMRI basics: Processing stages, terminology, and data structure**

## **fMRI basics**

In this chapter, we'll talk about analysis of functional magnetic resonance imaging, or fMRI, data. We'll start with some nomenclature, talk about data types and structures, and end with a little bit about fMRI data analysis goals. fMRI is a noninvasive technique for studying brain activity. By noninvasive, we mean that it generates pictures of the inside of your head without using any implants or injections. There are also no known side effects of being scanned frequently with fMRI. For example, one of our colleagues scanned himself nearly 100 times with no apparent adverse effects. Scans are now also routinely performed on both infants and children.

A single session in the scanner allows researchers to collect many image types, both anatomical (or 'structural' MRI) and functional (related to dynamic brain activity changes). Here, we are concerned particularly with functional imaging.

During the course of an fMRI experiment a series of brain images are acquired, often while the subject performs a task set. Those 'tasks' can include cognitive paradigms, viewing, hearing, feeling, tasting, or smelling various stimuli. An increasingly popular 'task' is simply lying in the scanner doing nothing; there is now a whole subfield devoted to 'resting-state' fMRI.

### ***Processing and analysis stages***

Figure 7.1 shows a basic flowchart for a typical fMRI experiment. Throughout this book, we're going to keep coming back to this flowchart to unpack each of the basic data processing and analysis steps in more detail. All studies begin with experimental design, which is perhaps the single most crucial factor in determining how well the experiment will go. There are a number of design principles rooted in statistics which apply to all design types, neuroimaging or otherwise. Other principles are specific to fMRI experiments and relate to the properties of fMRI data and their specific analyses.

# Data Processing Pipeline

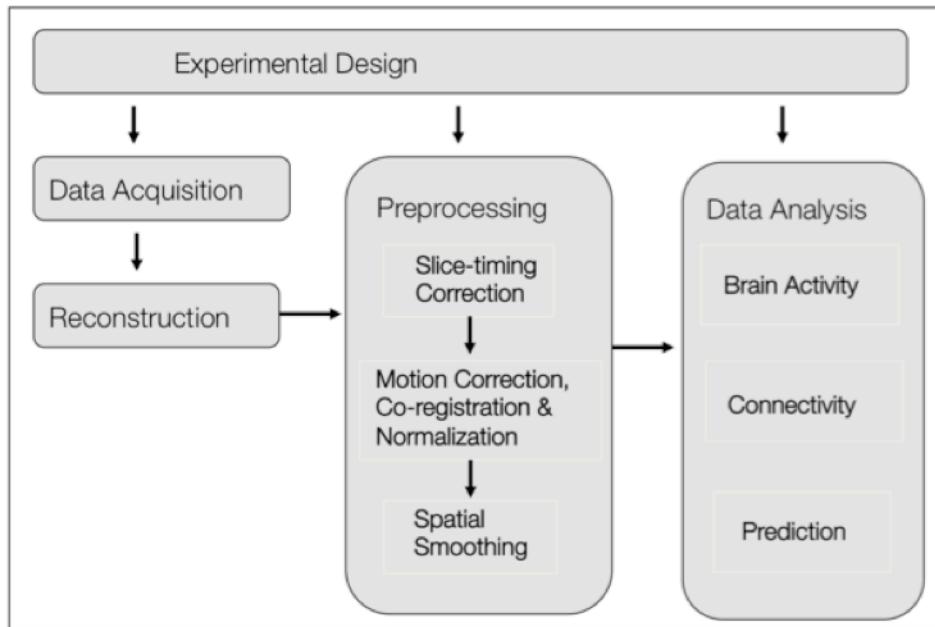


Figure 7.1. An illustration of processing pipeline.

Data acquisition is the next step after experimental design, followed by reconstruction of the data into images. Researchers perform a series of *preprocessing* steps before statistical analysis. These deal with anatomical alignment of the various image types (*coregistration*), timing issues (*slice-timing correction*), head movement (*motion correction*), and image transformation onto a standard anatomical reference space (*spatial normalization* or *warping*). Researchers also commonly perform artifact-mitigation procedures and physiological noise correction. After preprocessing, the images are ready for statistical analysis. Analyses can test task- or outcome-related brain activity, assess *functional connectivity*, and develop multivariate *predictive models* designed to correlate optimally with experimental variables or outcomes.

## Acquisition

We acquire MRI and fMRI data by applying radiofrequency (RF) pulses to the brain. These pulses perturb the magnetic spins of the protons of hydrogen atoms (mostly in water molecules) so they give off energy with particular spatiotemporal characteristics. The RF antenna reads off this signal, which is then used to reconstruct images. During data acquisition, magnetic gradient coils are applied

in particular patterns so signals from different spatial locations in the image are given particular characteristics, which enables accurate spatial reconstruction. The *pulse sequence*, the software that runs the RF antenna and gradient coils which acquire the signal, determines what type of data the process acquires - including whether the image is structural or functional. This will all be covered in greater detail in a later chapter.

It's useful to know some basic terminology related to MR image acquisition. Figure 7.2 shows some of the basics. The bounding box that defines the image acquisition volume depends on the *field of view* (the slice dimensions), the number of slices, and the *slice thickness*. Data are sampled within small cubic volumes called 'voxels' or *volumetric pixels*. Voxel size depends on slice thickness and in-plane *matrix size*, which is the number of grid elements on which each slice's data are sampled. The field of view divided by the matrix size is the in-plane resolution, measured in *mm*. Thus in-plane resolution and slice thickness determine voxel size. Researchers typically desire *isotropic* voxels, which are the same dimension on all sides, though unequal sizes work as well. A typical size is  $3 \times 3 \times 3\text{ mm}$  voxels; this is close to optimal for many purposes when using a 3-Tesla scanner.

## Images: Basic terminology

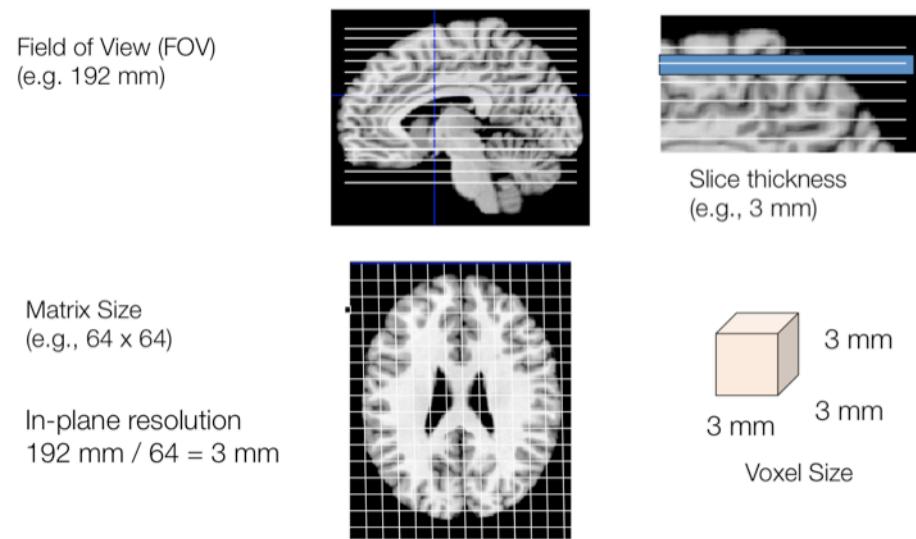


Figure 7.2. Basic terminology related to MR image acquisition.

Designing an fMRI study requires a series of tradeoffs given the study's particular goals. One fundamental tradeoff is between spatial and temporal resolution. You can either collect data with high spatial resolution or collect data fast, but you can't do both. Spatial resolution defines our ability to distinguish how an image changes across different spatial locations and thus also our ability to extract location-coded information about brain states and behaviors. Both the voxel size and the image's underlying smoothness (blurriness), which depends on the main magnetic field's strength

and gradients and on underlying physiological limitations (because most of our signal is blood flow-related), determine spatial resolution. Temporal resolution determines our ability to separate brain events in time. Both the TR and the hemodynamic response to neural and/or glial events' time course (more on this is below) determine this.

## ***Image orientation and dimensions***

Understanding and interpreting which part of the brain one is viewing requires some practice. The brain is a complex three-dimensional structure with many curved ‘C’ shape sub-structures that wrap around the brain’s center, the thalamus. It is typical to show neuroimaging results on anatomical brain slices. Figure 7.3 provides a basic orientation to those slices and their spatial relation to the overall head and brain surface. Each of the three dimensions of brain space has a special name. The left-to-right dimension is conventionally the *X* direction in standard brain coordinate space. The back-to-front dimension is the *Y* dimension which ranges from *posterior* at the brain’s back to *anterior* at the front. Sometimes anterior is also called *rostral*, which means ‘toward the head’, and posterior is called *caudal*, ‘toward the tail’. The bottom-to-top dimension is the *Z* dimension which ranges from *inferior* to *superior* locations. These locations are sometimes also called *ventral* ('towards the belly') and *dorsal* ('towards the back').

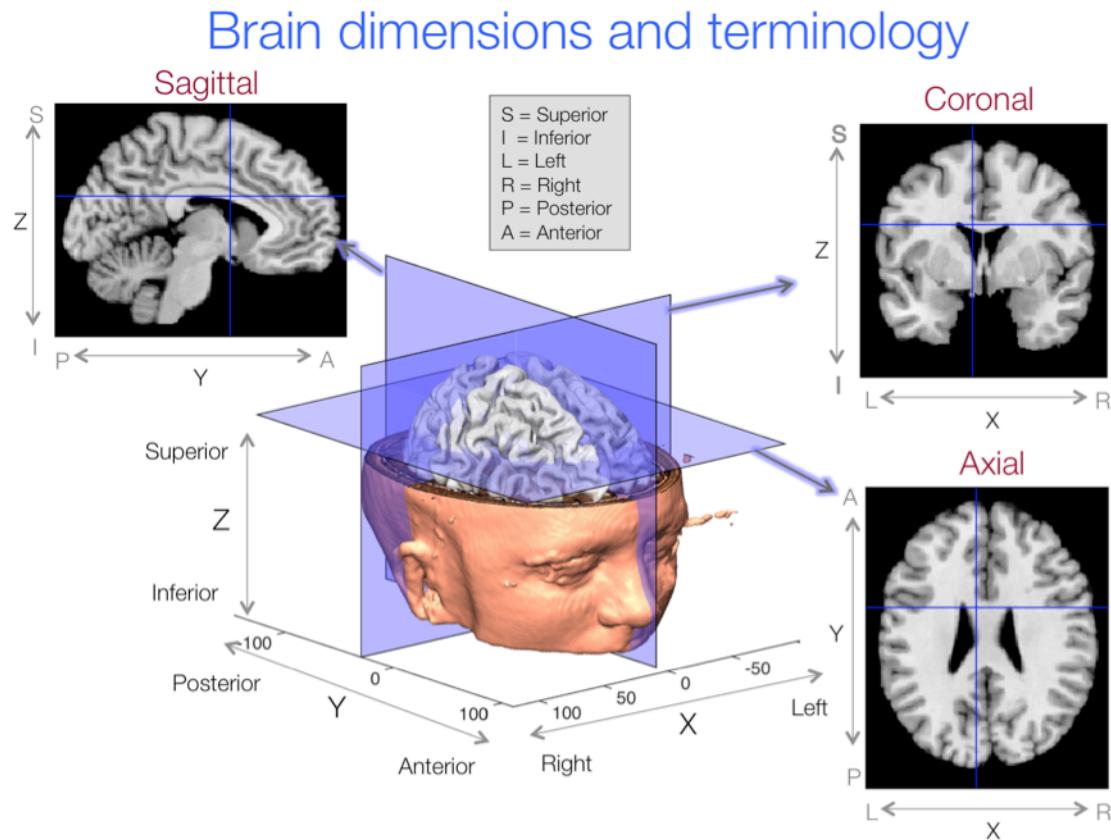


Figure 7.3. A basic orientation to anatomical brain slices and their spatial relation to the overall head and brain surface

Researchers typically report locations along these dimensions in  $[x, y, z]$  coordinate triplets with  $x$ ,  $y$ , and  $z$  values indicating distances in millimeter units relative to a zero point. The  $[0, 0, 0]$  point is, by convention, the anterior commissure, a small white-matter bundle which connects the brain's two hemispheres. Figure 7.4 shows this point.

## The anterior commissure

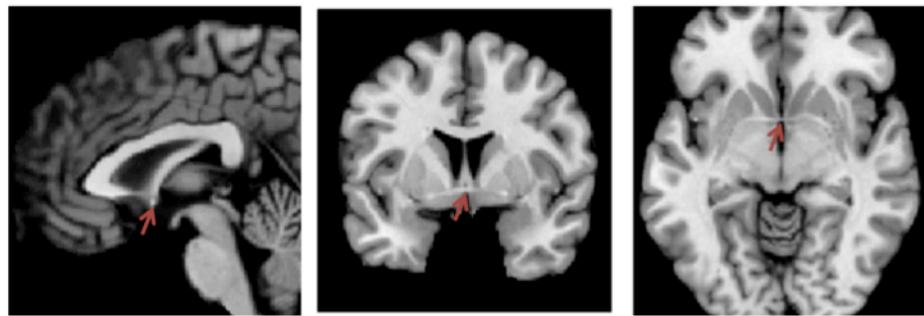


Figure 7.4. The location of the anterior commissure.

In the brainstem, some of the dimension names are not very intuitive because they describe the dimensions as in an animal which walks on four legs with the spinal cord toward the rear (caudal) and the midbrain, which lies just below the thalamus, at the rostral end. Thus, the part of the brainstem toward the back of the head is the *dorsal* brainstem and the part toward the front is the *ventral* brainstem. Figure 7.5 shows these directions and some of the most important structures' locations.

## Brain dimensions in the subcortex

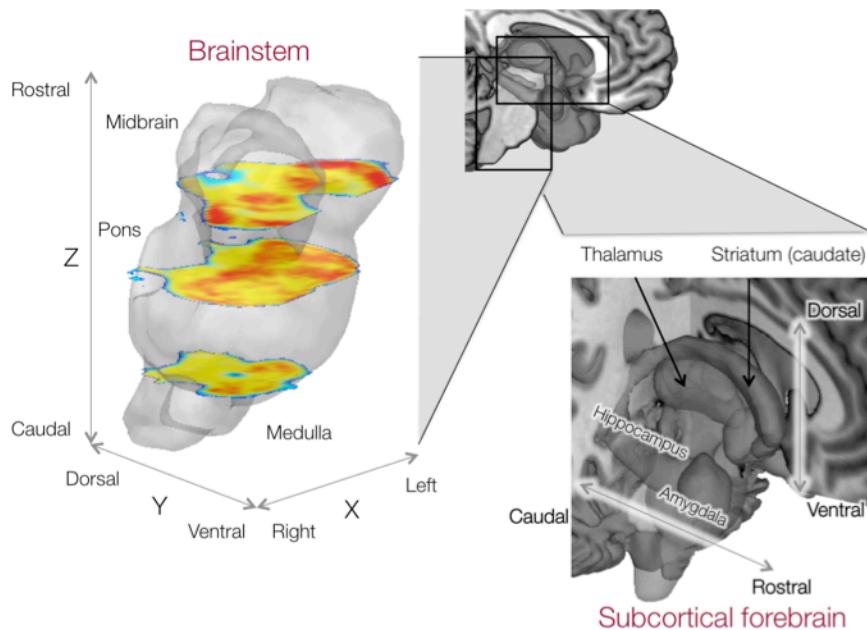


Figure 7.5. Directions and some of the most important structures' locations in the sub cortex.

The sections of the brain that are typically used to display neuroimaging results have particular names too. Figure 7.3 shows these. *Coronal* slices are sections that span the left-right and inferior-superior dimensions at one location from front to back. *Sagittal* slices span the front-to-back and inferior-superior dimensions at one location from left to right. And *axial* or \*horizontal\* slices span the left-right and front-to-back dimensions at one location from inferior to superior.

## **fMRI time series**

Functional images (also called  $T2^*$ -weighted images) have lower spatial resolution than structural images. That is, they're much blurrier than their structural counterparts. However, we can measure many of them, so they have higher temporal resolution and we can use them to relate signal changes to experimental manipulations or other outcomes that vary from second to second.

One participant's fMRI dataset contains a time series of 3-D images, or 'volumes', shown in Figure 7.6. The volumes often cover the entire brain but can also cover just one brain tissue section or slab at a higher spatial resolution. The data for each volume are usually acquired slice-by-slice; after completing one volume, the scanner moves on to the next image. As they are collected, the data are sampled onto a rigid voxel grid.

## fMRI data time series

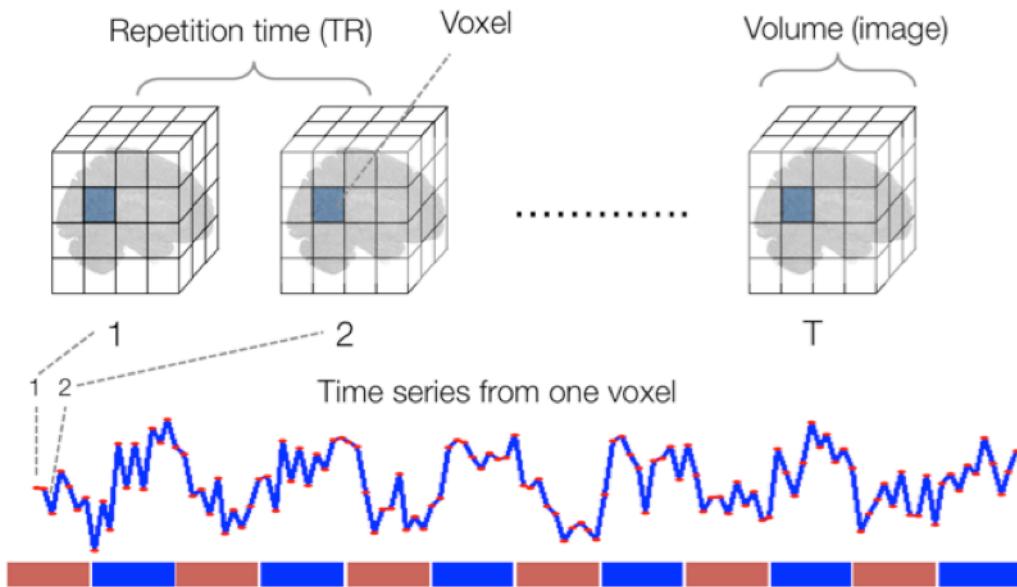


Figure 7.6. fMRI dataset consists of a time series of 3-D images, or ‘volumes’, measured at every TR.

It is not uncommon for each volume to contain 100,000 or more voxels, though the number varies depending on the acquisition choices. The *pulse sequence*, or the software that runs the radiofrequency antenna and magnetic coils which acquire the signal, determines how researchers acquire data. The repetition time between volumes, or *TR*, varies quite a bit across studies but typical values for a whole-brain acquisition have historically been about 2-3 seconds. However, recent imaging advances now make it possible to collect a whole brain volume in < 500 msec. Experiments can be brief at 6 minutes of functional time (e.g. 180 volumes), but experiments that include 40 or more minutes of functional time, with over 1,000 volumes measured at a typical TR, are not uncommon in practice.

Thus fMRI data comprise hundreds to thousands of images in a time series. As local regions’ oxygen metabolism and blood flow change, researchers use fluctuations in the measured signal to make inferences about brain activity and connectivity. The usual approach towards assessing brain activity is based on examining average fluctuations locked to particular experimental conditions or events. We refer to this as task-based fMRI. Researchers assess brain connectivity by examining associations in the fluctuations among voxels with or without task condition influence analyses.

A simple, canonical example of a task-based fMRI experiment is a motor task. Let's say we want to examine activity increases in the motor cortex when participants execute simple finger movements. Researchers often use such tasks as quality control assessments in order to check signal and analysis quality. Participants might alternate between 20-second long blocks of finger tapping and 20 seconds of rest. This is a 'block design', illustrated in Figure 6.6's bottom panel. Not all designs are equally efficient or powerful, but 20-second long blocks have good properties in particular; we will return to this concept in later chapters.

## **Statistical analysis**

Once we have run an experiment and say obtained motor task data for a group of participants, we are ready to analyze the data. Recall that the fMRI dataset contains a time series of each voxel's signal values. A basic analysis will study each participant's data one person and one voxel at a time. fMRI data are quite noisy, so we use statistical analysis to determine whether a signal change is consistently associated with the finger-tapping task.

The first step of statistical analysis is to fit a model to each voxel's time series. In this case, the model simply states that activity levels are different between finger-tapping and control periods. We can use a t-test to examine how large the difference (or 'contrast') is between finger tapping and rest divided by the noise measure (i.e. error variability). Then we make a map of the resulting voxel t-values and their associated p-values, which provides evidence to evaluate the null hypothesis of no task effect. Again, each voxel corresponds to a spatial location and has an associated statistic that represents the evidential strength for task-related effects. Researchers usually *threshold* these maps by applying a statistical cutoff related to the p-value, so scientific papers only plot and discuss voxels with sufficient evidence of an effect.

The description above covers the basics of a simple statistical analysis. However, it leaves out one key detail. fMRI activity, either BOLD or ASL, does not rise instantaneously when the task begins. Rather it increases over several seconds as blood flow function increases, peaks at about 5-6 seconds after the increase in local brain metabolic demand and decreases after 10-15 seconds. This function over time is the *hemodynamic response function* or the *HRF* which researchers must measure, or else assume a canonical function, to be able to perform a reasonably accurate analysis. Figure 7.7 shows a canonical HRF widely used as a model for fMRI responses. One piece of good news is that even very brief neural events (e.g. a 17 msec stimulus presentation) can reliably elicit measurable hemodynamic responses, so fMRI can be sensitive to short events. Another is that even with a complex neural event series or sustained blocks as in our finger-tapping experiment, we can still account for the HRF in our analysis.

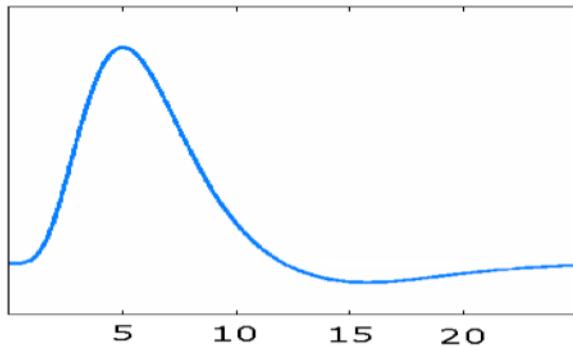


Figure 7.7. An illustration of the canonical hemodynamic response function.

The description above highlights the fact that fMRI data analysis is fundamentally a time series problem. However, it's a time series problem on steroids because every voxel has its own time series and there are about 100,000 voxels. The concept of analyzing voxels individually is sometimes called *mass univariate* analysis; in this, we treat all of the voxels separately then construct a map of the statistical results at each voxel. Other techniques that do not separate the voxels are becoming more widely used. We typically refer to these techniques as *multivariate* analyses because they are multivariate in brain space and model multiple voxels simultaneously.

Clearly fMRI data analysis is a massive data problem. Each brain volume consists of roughly 100,000 different voxel measurements. Each experiment might contain 1,000 brain volumes or more. And we might repeat each experiment for multiple subjects, maybe 20, 30, or 40, but sometimes hundreds or thousands, to facilitate *population inference*, i.e. making generalizable conclusions about human brain function. Because of both the amount of data and its complexity, fMRI data statistical analysis is challenging. The signal of interest is relatively weak and the data exhibits a complicated temporal and spatial noise structure. Thus there are ample opportunities to develop new increasingly sophisticated and powerful statistical techniques.

## Data structure in fMRI experiments

### ***Hierarchical data structure***

fMRI data has a hierarchical structure, as Figure 7.8 shows. Understanding this structure and dealing with it appropriately is important when undertaking fMRI data preprocessing and statistical analysis.

## Hierarchical Structure of fMRI Data

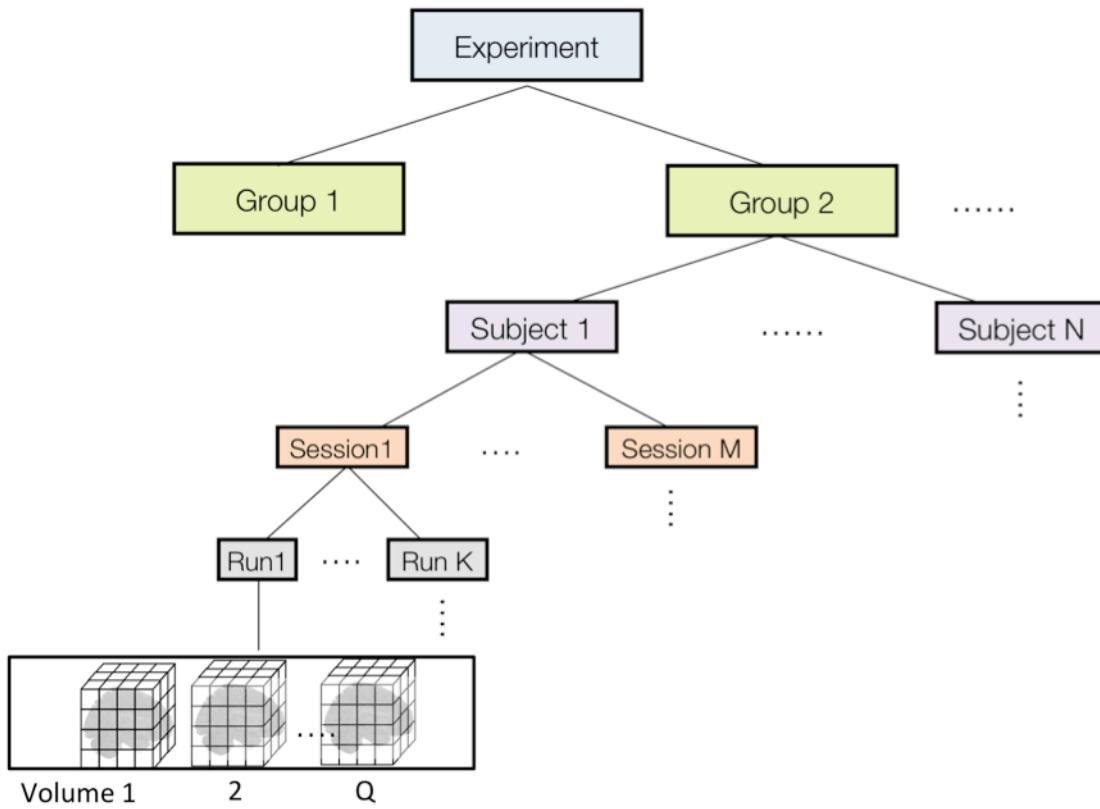


Figure 7.8. An illustration of the hierarchical structure of fMRI data.

The vast majority of experiments include many different participants' data. This is critical in order to obtain population generalizable results - i.e. results that are not just idiosyncratic features of the individuals we happened to study but rather that constitute general conclusions that apply to new individuals. Each participant (sometimes called 'subject') performs the same task or tasks. Sometimes we nest participants within groups, such as patient group versus controls or elderly individuals versus young. In other cases, there is just one group and researchers' interests are in studying experimental manipulations, behaviors, or other outcomes measured within-person. Even if we do not organize participants into groups, it is still possible relate brain activity differences to individual person-level variables (e.g. age, performance, or other variables).

Experiments entail collecting many repeated measurements on each participant over time. We may scan each participant longitudinally in multiple *sessions*. During a session, it is typical to start and stop the scanner multiple times, collecting data for brief periods - usually 4-10 minutes. We refer to these as *runs*. Because head movement during the scans is particularly problematic, short runs are advisable to give participants a break and allow them to communicate with the experimenters if necessary (though speaking can induce additional head movement!). Each run, in turn, entails a series of brain volumes, one per TR, nested within task conditions (e.g. finger tapping and rest). Each volume consists of multiple slices acquired sequentially, and each slice contains many voxels.

Often we analyze each participant separately with a mass univariate analysis of each voxel's experimental effects. This is a *first level* analysis. The resulting maps of experimental effect magnitudes, called contrast maps, become the data for a *second level* analysis of effect reliability across participants, which includes differences between groups and effects of individual differences.

## **Image file formats**

One barrier to entry in fMRI analysis is that the image data are not simple text files. Rather they are stored in specific customized formats along with associated 'meta-data' or information about the imaging parameters.

Historically, data formats differed widely across scanner manufacturers and software packages. For example, raw data on General Electric scanners are in a proprietary format called 'P-files' and on Siemens scanners as DICOM files (which stands for Digital Imaging and Communications in Medicine, <http://www.dicomlibrary.com/dicom/>).

DICOM files contain a single slice's data at a single time point, with extensive 'header' information, though different scanners use and store this differently. A study can thus include millions of files, which presents logistical challenges with many file management systems. Therefore preprocessing and statistical analysis packages usually require that we convert these images into other, standard formats.

Though these standard formats also differed across packages, most now have the facility to read and write NIfTI (which stands for [Neuroimaging Informatics Technology Initiative<sup>16</sup>](#)) images, a standard 3-D or 4-D file format. A 3-D NIfTi file contains one image per volume, while a 4-D NIfTi file often contains a single person's time series of image volumes. Both these files have a .nii extension. A related older file format less standardly used across software packages is the Analyze image format which has .img extensions and associated separate header files containing meta-data with .hdr extensions.

It is important to exercise caution when reading and writing files across various software packages because these packages use the meta-data differently. Some researchers thus feel uncomfortable about mixing and matching algorithms from different software packages, though it can be done if one is cautious and meticulous.

One of the biggest issues of which to be aware relates to flipping images in the *X* direction, from left to right. The brain is largely symmetrical, which makes it difficult to tell if an image display's left side, such as in Figure 7.9, is on the left or right side of the brain. The fact that there are two orientations typically used to view images further complicates this point of potential confusion. If the image display is in *radiological* format, the brain's left side is on the displayed image's right side. This display is as though one looks up at a person's brain from their feet. If images are in *neurological* format, the brain's right side is on the displayed image's right side. This is the format most cognitive neuroscience research uses. Though imaging software should keep track of format, different packages use header information related to flipping differently and custom reconstruction

---

<sup>16</sup><http://nifti.nimh.nih.gov/nifti-1/>

and stacking code at different research centers can also treat the image orientation information differently. As a result many, many errors have undoubtedly occurred.

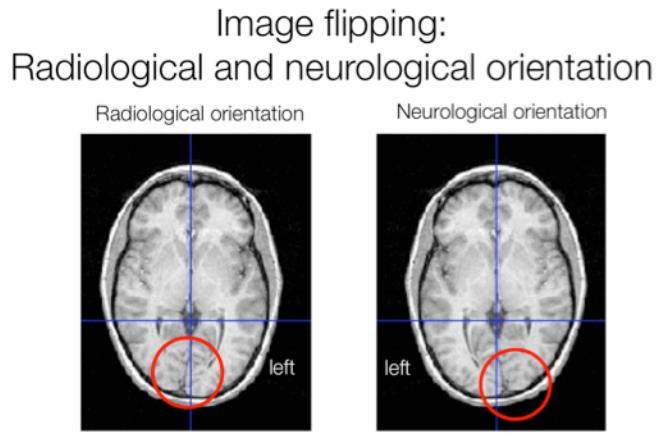


Figure 7.9. The same brain slice shown in radiological and neurological format.

A number of strategies can help avoid this error. Using NIfTi format helps, as does consistent use of a single software package. Researchers often tape a Vitamin E capsule to the same side of each participant's head. This produces a bright spot on the image in one hemisphere. Finally, we can heuristically check the flipping by viewing the images, because the left occipital lobe's larger size in most people causes the calcarine fissure's deviation to the right as it courses from front to back. This asymmetry is prominent in the structural image shown in Figure 6.9.

## Conclusions

In this chapter, we briefly covered the major steps in fMRI data processing and analysis and some of the most commonly used terminology. In addition, we reviewed the hierarchical structure of fMRI data and common image file formats. In later chapters, we will cover the design, acquisition, preprocessing, and statistical analysis of fMRI data in more detail.

# Chapter 8 - The MRI environment and human factors

In this chapter, we review some important basic facts and safety issues when working in the MR environment, and some considerations on what types of studies are amenable to MRI scanning. These include *MR basics and safety* and *Physical limitations on data collection*.

## MR basics and safety

### ***Ferromagnetic objects and particles***

The MR environment is highly magnetic as the magnetic field in the scanner is *always on* (barring unusual events that may require a shutdown or ‘quench’).

Among the many types of materials which have the potential for some interaction with the magnetic field, there are multiple levels of magnetic susceptibility. The most magnetically interactive substances are *ferromagnetic*, which is defined roughly as materials that can interact with magnets, have noticeable effects, and can remain magnetized after the external magnetic field is gone. Common ferromagnetic materials include iron, nickel, and cobalt, as well as their various alloys including stainless steel. In contrast, *paramagnetic* materials interact with magnetic fields and become magnetized, but do not stay magnetized after the external field is removed. Many compounds, including hemoglobin which has lost its oxygen, exhibit paramagnetic properties; this is one of the reasons that MR works! Molecules of paramagnetic materials have dipoles, i.e. unequal charges across atoms at either end of the molecule, but these are oriented randomly and have no net magnetization until an external magnetic field is applied. Most other materials are strongly diamagnetic, meaning their molecules are very weakly repelled by magnetic fields. Such materials include water, wood, and plastic, as well as most other organic compounds and non-magnetic metals which include copper, mercury, and gold.

Ferromagnetic materials can be disruptive or dangerous in the MR environment since they interact strongly with the scanner’s magnetic field. Ferromagnetic objects will be pulled toward the bore of the magnet, so they should never be brought into the scanner room. It may not feel like there is a strong pull on such objects when one is at the edge of the room, but this is deceptive! The pull increases by the square of closeness to the magnet. Metal objects can fly out of hands or pockets as one moves closer to the bore, which means larger objects pose a serious health hazard because they become missiles in the MR environment. For a helpful review of safety concerns, see Ester et al<sup>17</sup>.

---

<sup>17</sup><http://www.ncbi.nlm.nih.gov/pubmed/8273665>

Another concern is metallic or electronic objects in and around the participant's body. These objects include medical devices such as artificial joints, pacemakers, and fragments of metal lodged in a person's body from an old injury. The concern even extends to tattoos and makeup or hair products with ferromagnetic particles (e.g. eyeliner). Ferromagnetic objects can become dislodged in the intense magnetic field and cause serious injury or death. Because of the intensity of the magnetic field, participants with implants that have any chance of being magnetic (e.g. pacemakers or any electronic implants) should be kept out of the MR environment. In addition, the presence of other metal in the body (as in metal workers who may have tiny metallic shreds in their eyes) is a strict contraindication for scanning.

## ***Radiofrequency energy exposure***

Electromagnetic fields can cause heating of tissue if applied with sufficient intensity over enough time. This is typically not an issue since most MRI scanners have built-in safeguards to prevent too much RF power deposition into the subject. However, should any metal conductors be inside the RF coil, induced currents can make them quite hot to the point that they cause burns even at RF power levels otherwise harmless to the participant. This is the same principle which underlies the production of sparks from metal objects in the microwave oven. Small metallic objects far from the participant's head (such as buttons on jeans) haven't presented a problem in our experience. However, many MRI centers prefer that participants wear scrubs or hospital gowns to reduce issues of unexpected magnetic objects (underwire bras are common objects in this category) that in some cases cause burns.

Additionally, in some rare instances the gradient coils produce changes in the magnetic field which can induce electric currents in long nerves and cause them to depolarize and produce mild twitching. This is referred to as peripheral nerve stimulation (PNS). It occurs only rarely during fast imaging sequences but is more likely at higher field strength.

A safety concern particular to the PET environment is radiation exposure, which is limited by FDA regulations of 5 rem per session or 15 rem annually. NIH guidelines<sup>18</sup> are 3 rem within 13 weeks or 5 rem annually. Radiation exposure is substantially lower with MRI, which currently has no limitations on how frequently a person may be scanned.

## **Physical limitations on data collection**

In addition to safety precautions, there are other considerations one must address to protect the integrity of the acquired data.

### ***Head movement***

For both PET and fMRI, the participant must remain motionless for the duration of the session, especially while imaging data is being collected. Though PET tolerances for head movement are

---

<sup>18</sup>[http://www.cc.nih.gov/ccc/protomechanics/chap\\_6.html](http://www.cc.nih.gov/ccc/protomechanics/chap_6.html)

generally higher, task-correlated head movement can be a serious confound in either modality. In fMRI, head motion is particularly problematic since it induces changes in the local magnetic field. While linear motion-correction algorithms generally do a reasonable job adjusting for gross displacement of the head, they do not correct for the more complex artifacts movement creates. Participants' heads are usually restrained with a vacuum bag (a soft pillow that becomes hard when air is pumped out of it), a forehead strap, some foam pads, a bite bar, or some combination of these restraints. Participants who move too much are excluded from analysis.

The restrictiveness of the scanning environment means it is not generally advisable to use neuroimaging for tasks in which head movement is unavoidable (e.g. studies involving overt speech or pain studies involving sudden-onset electric shocks). Researchers have partially circumvented the problem: they collect vocal responses during tasks by pausing data collection during verbalization or including a set of regression predictors to account for motion artifacts. However, the latter in particular may be only a partial solution.

The enclosed MR environment can also be a problem for individuals with claustrophobia. It is a good idea for groups working with special populations - children or individuals with psychiatric disorders - to familiarize participants in a mock-scanning environment (including an enclosed bore and simulated scanner noise) before they enter the magnet proper.

## ***Auditory noise***

Functional MRI scanning creates repeated loud tapping and buzzing noises (approximately 100dB at the patient's location), which makes presentation of stimuli through auditory means more difficult than other sensory modalities. Because the noise can be reduced with earplugs or shielded earphones, some kinds of auditory studies are possible. However, earphones, like other electrical and electronic devices, may cause magnetic susceptibility artifacts in the images if present in the scanner room.

For visual stimulation, participants in PET and MRI scanners typically view a display projected by LCD onto a screen in the magnet room, shown on a shielded in-scanner LCD display, or projected onto each eye with fiber optics or LCD screens mounted in goggles. Small mirrors mounted in the head coil (in MRI) often project the screen onto participants' retinas. The visual angle of presentation is often limited to about 15 degrees. Contrast and display image quality should be assessed before imaging, particularly for tasks that require viewing photographs or other fine-grained visual discriminations.

## ***Participant or patient response options***

Response devices vary from scanner to scanner, but often the need for adequate RF-shielding limits options for in-scanner devices. Responses are often limited to pressing buttons, making eye movements (several manufacturers provide scanner-compatible eye trackers), and moving joysticks or track-balls. MR compatible keyboards and cameras also exist. Most MR compatible devices are free of ferromagnetic parts. In some cases, they contain wires with shielded connections.

## ***Sources of artifacts***

In general, any device with wires or ferromagnetic parts distorts the magnetic field and thus can induce artifacts in the images. If severe, these artifacts may be visible as stripes or distortions in the structural (T1 or T2) images. However, T2\* contrast is much more sensitive to artifacts, so distortions may not be visible in the structural scans but will present in the functional data. Such artifacts have been observed (in our experience) from electrodermal response (EDR) leads, earphones, joysticks, mice, or keyboards in the scanning room. Some of these were not designed for use in the scanning environment, so there is no surprise that they result in artifacts. However, other commercial products intended to be RF-shielded for fMRI use may also lead to artifacts in images. It is generally advisable to look carefully at structural and functional scans (and do a complete analysis of a simple paradigm such as visual stimulation) with and without any new piece of equipment in the room.

RF artifacts can also be caused by improperly shielded electrical cables running through the wall between the control room and the scanner room, or even by hair products and makeup containing ferrous material (not uncommon!). The closer a metallic object is to the patient's head, the greater the potential for artifacts. Jewelry, watches, and credit cards should never be taken into the MR environment, not only because of the possibility for artifacts but also because electronics are likely to be destroyed.

## ***Effects of the MR environment on task performance and mental status***

The MR environment involves exposure to an incredibly strong magnetic field which is always on, as we stated above. A common question is whether this environment itself may change performance of the task or alter other physiological measures. After all, people have been using much weaker magnets to heal others, cast off evil spirits, and realign energy balance for over 100 years. In spite of the continued persistence of these practices, however, we're not aware of any evidence suggesting that the magnetic field changes one's mental state in any appreciable way. Thousands of MR experiments have been done, many of which test facts related to behaviors and emotions, each of which has been established outside narrow environments and tested extensively at various field strengths without any noticeable changes.

However, that is not to say that there are no effects of being in an MR environment and that it is safe to ignore possible impacts when designing experiments. Alterations may occur because of anxiety about the scanning environment, changes in temperature (many scanner rooms are chilly), changes in posture inducing physiological changes (lying down reduces orthostatic load), practice, responses to the medical context imaging suites present, or other variables. Being in an MR scanner for the first time has been shown to induce cortisol release, cortisol being commonly measured as a marker for stress which has diverse effects on the central nervous system and body. In addition, prolonged scanning requires participants to move very little; accordingly, after an hour or so of scanning it is common to have bodily aches and pains, including a sore back of the head or a stiff back depending on cushion positions. And finally, the scanning environment is quite noisy, which may be distracting.

That does not mean that the scanning environment is always stressful, however. One of the most common side effects of scanning is participants falling asleep. A recent study found that a substantial portion of people studied in resting state scans fall asleep during the scan. The scanning environment may be noisy, enclosed, and uncomfortable, but it is also quite hypnotic once one relaxes into the environment. While some researchers do not allow participants to use caffeine, it is more common to allow participants to drink their normal levels of daily caffeine, which keeps participants awake and has the added benefit of enhancing the BOLD response.

All of these are potential considerations to think about when designing, running, and evaluating MRI studies. Though there are many potential factors at play, the challenges are not insurmountable; many thousands of studies have now collected high-quality data. It is advisable to test paradigms outside the scanner and to make sure that the subjective and behavioral effects that you are studying are comparable outside and inside the MR environment. Many research groups also familiarize patients with a mock scanner which can acclimatize participants to some of the novelty and the discomfort (e.g. noise) in the MR environment. This can also provide some training to reduce head movement during scanning, which is particularly useful with children and patient populations who may be more prone to anxiety in the scanner and therefore tend to move their heads more than others.

# **Chapter 9 - A head-to-head comparison of PET and MRI**

Both PET and fMRI can measure brain activity, though each has unique advantages over each other and alternative techniques. Below we list some strengths and weaknesses of PET and fMRI in five key areas. These are: *Acquisition options and fidelity*; *Available signal types and their interpretability*; *Spatial and temporal resolution*; *Accessibility to a broad community*; and *Multimodal potential: potential for combination with other techniques*.

## **Acquisition options and fidelity**

An advantage of MRI is that it allows repeated scanning of the same participants. There are currently no known intrinsic risks associated with scanning, other than those associated with the MR environment and the strong magnetic field. Thus MRI is ideal for longitudinal designs and for designs that require large amounts of data on individual participants. Because PET requires radioactive material to be injected into the bloodstream, researchers cannot frequently repeat PET scans on a participant.

MRI is also superior in the number of different types of images that can be collected in a single session . These include structural anatomical images, white-matter density images, vascular images, functional images of experimentally induced activation and functional connectivity across brain regions, and molecular imaging with MR spectroscopy.

However, some unique features of PET balance out these advantages. Susceptibility artifacts, including signal loss in particular brain regions and spatial image distortion, are an intrinsic part of BOLD fMRI signal. PET does not have this problem; it provides coverage with minimal artifacts and geometric distortions throughout the brain. Thus PET may be more sensitive to signal in high-susceptibility areas such as the amygdala, the inferior temporal lobes, and the orbitofrontal cortex, though meta-analyses have not identified systematic activation differences in these regions.

## **Available signal types and their interpretability**

PET signal is more straightforwardly interpretable in terms of cerebral blood flow (CBF) or glucose metabolism because it can more directly image these processes. BOLD fMRI, the most widely used fMRI technique, is much more complex; it reflects a mix of blood oxygenation, volume, and flow. ASL fMRI techniques, as an alternative, can quantitatively measure CBF and researchers have demonstrated their high reproducibility across long time periods (e.g. 1 month). However, the signal

magnitude in these techniques is much smaller than BOLD signal. ASL is a technique poised for expanded use if stable sequences become more widely available.

In terms of dynamic imaging, fMRI is superior. fMRI can link brain activity to specific psychological and physiological events which vary across seconds. This enables event-related analyses and within-person functional connectivity analyses, both of which have exploded in use and become dominant techniques over the past years. fMRI also permits inferences on the relative timing of psychological event-related responses, which, though researchers seldom make now, may become much more popular as rapid imaging sequences - which can collect whole-brain volumes in hundreds of milliseconds - become widely used. PET, by contrast, requires a minimum of 20-30 seconds to collect enough data to form an image. The time can be much longer for molecular imaging, around 10-40 minutes. Thus, PET can only image blocks of events and is insensitive to temporal dynamics. MRI spectroscopy is similarly insensitive.

Conversely, however, PET's stability across long time periods means researchers prefer it for statistical comparisons over significant periods, with the proviso that they cannot repeat it frequently. ASL has similar advantages but is repeatable, so ASL-fMRI may prove more advantageous in the long run as its use becomes more widespread.

PET really shines in molecular imaging. We can, in principle, label hundreds of compounds with radioactive isotopes and thus provide brain images related to specific aspects of particular classes of neurotransmitter, neuropeptide, and neuroinflammatory molecules related to glial function. This has already proven critical in translational research on pharmaceutical development and on early Alzheimer's diagnoses using neuroinflammatory signaling. Thus researchers use PET much more widely in preclinical and clinical drug development than they use fMRI.

## Spatial and temporal resolution

fMRI is superior to PET in both spatial and temporal resolution. Its spatial resolution varies considerably depending on inference level (within-person or between-person) and on particular acquisition and analysis techniques. However, its upper bound is clearly [sub-millimeter resolution<sup>19</sup>](#), a level where much information about specific psychological processes is coded in the brain. Standard group analysis may be comparable to high-resolution PET, around 1 - 1.5 cm. Thus both techniques are viable for many purposes; they are also often combined and compared in neuroimaging data [meta-analyses<sup>20</sup>](#).

The temporal resolution of fMRI is also superior, as discussed above. The relatively sluggish BOLD hemodynamic response, with a usual peak at 5-6 seconds following brief events and more than 30 seconds before it returns to baseline, limits it. However, its event-related averaging capability combined with fast imaging makes it possible to resolve average temporal differences between conditions in approximately 100-200 milliseconds, depending on data amount and quality. PET's

<sup>19</sup><http://www.ncbi.nlm.nih.gov/pubmed/24082116>

<sup>20</sup><http://www.ncbi.nlm.nih.gov/pubmed/22617651>

resolution capability is 20-30 seconds at best and it cannot feasibly collect enough data in a session to meaningfully attempt event-related averages.

## Accessibility to a broad community

fMRI is more accessible than PET, which has been a major factor driving fMRI's increasing popularity. fMRI costs about 1/3 as much as PET, on average, in part because of PET's requirement for a co-localized cyclotron and a radiochemistry laboratory. These requirements also increase exertion of regulatory control over PET access, which makes it more difficult for a broad set of researchers to conduct studies. There are also simply more MRI facilities available. Many non-clinical centers and academic departments have an MRI scanner, as do virtually all hospitals: these can in principle, with the appropriate software, conduct fMRI as well. It is not uncommon, for example, for psychology departments to have a departmental scanner in the building.

Because the fMRI community has grown and expanded across multiple disciplines, fMRI has a broader research community which can provide varied perspectives of discussion, input, and support. Some early leaders in fMRI research also opted to make their analysis tools freely available under an 'open source' model. The major software packages researchers used to analyze data, including Statistical Parametric Mapping (SPM), FMRIB Software Library (FSL), and Analysis of Functional Neuroimaging (AFNI), all operate under this model. The choice to share freely has been a tremendous advantage. In other fields, including PET, much software is proprietary or is only available to collaborators, so it is not uncommon that laboratories pay tens of thousands of dollars for relatively simple software.

In addition, this spirit of open sharing in fMRI has also recently extended to dataset sharing, many which involve 1,000 or more individuals. Some are downloadable full dataset repositories; this includes the 1,000 Functional Connectomes project, OpenfMRI, and the new OpenPAIN repository. NIH has increasingly joined this effort, so data from the first 500 Human Connectome Project participants is available for download or is actively shared on hard drives shipped to investigators at cost. Other enterprises are 'grass-roots' repositories for meta-analytic data and published study maps, including NeuroSynth, NeuroVault, and BrainMap. These sharing initiatives make neuroimaging data available to yet wider scientific and analytical communities interested in developing and testing methods from statistics, computer science, or other fields.

## Conclusions

In sum, fMRI and PET both have unique advantages and are quite complementary techniques. fMRI has emerged as a dominant and increasingly popular technology in large part due to its accessibility and its development of a large community which is willing to share tools and data.

# Chapter 10 - Fundamentals of MRI Physics

The same scanner can obtain both structural and functional MRI images, though it needs different programming depending on the image type one wants to obtain. For each type, a subject is placed in a strong magnetic field (usually 1.5 to 7 Tesla) generated by an electromagnet. The main magnetic field aligns the protons in the brain's hydrogen atoms (spins) along its axis. During imaging, the scanner exposes the subject to a radiofrequency (RF) electromagnetic field pulse. The protons absorb the RF pulse's energy at a very specific frequency band dependent upon the field strength, which causes them to become 'excited' (i.e. they change their quantum energy state). The nuclei then emit the energy at the same frequency as they 'relax'. The same antenna that produced the RF field detects the returned energy as a one-dimensional series of fluctuations over time. It is possible to use these signals in reconstructing a three-dimensional image by applying gradients, which are magnets that change the magnetic field strength systematically across space so the signals' frequencies and phases will encode their location within the brain.

Researchers use pulse sequences, or software programs that implement particular RF patterns and gradient magnetic field manipulations, to acquire data which they can then reconstruct into an underlying MR signal source map: this provides a brain image. There are various pulse sequence types, including echo-planar imaging (EPI) and spiral imaging. With a standard EPI sequence, researchers perform 3-D localization by selectively exciting tissue in a 2-D slice series one at a time, then by applying gradients to encode the signal's location within the 2-D slice. Note that this technique does not acquire MR signal voxel-by-voxel, but rather collects the acquired data in the frequency domain, or k-space. Applying a Fourier transform converts the data from k-space to obtain the image.

We can describe the relaxation of the nuclei using three values:  $T_1$ ,  $T_2$ , and  $T_2^*$ . The  $T_1$  and  $T_2$  values are constants determined by the spin frequency, field strength, and tissue type. These values are based on hydrogen content, which in turn depends on the amount of water in the tissue.  $T_1$  refers to the rate spins relax back to alignment with the main magnetic field after removal of the RF pulse and  $T_2$  refers to the attenuation rate of the magnetic field applied by the RF pulse. Finally,  $T_2^*$  relates to  $T_2$  but depends additionally on local inhomogeneities in magnetic susceptibility caused by changes in blood flow and oxygenation, among other factors.

Different pulse sequences produce images sensitive primarily to  $T_1$ ,  $T_2$ , or  $T_2^*$ . Because  $T_1$  and  $T_2$  vary with tissue type but are insensitive to functional changes and local magnetic field homogeneity,  $T_1$ - and  $T_2$ -weighted images produce high-resolution images depicting the boundary between gray matter (mostly cell bodies), white matter (mostly axons), and cerebrospinal fluid (CSF, mostly water). Regions with high iron content (which is magnetic), such as the substantia nigra, have very fast  $T_2$  relaxation and appear as dark spots in  $T_2$ -weighted images, providing anatomical definition in

these regions. Changing the contrast mechanism can help differentiate brain structures or lesions, since some structures will be apparent in certain image types but not in others. For example, multiple sclerosis lesions appear very brightly in  $T_2$ -weighted images but are virtually invisible in  $T_1$ -weighted images. Because  $T_2^*$  is sensitive to flow and oxygenation, unlike  $T_1$ - and  $T_2$ -weighted images,  $T_2^*$ -weighting can create images useful for studying brain function. Hence  $T_2^*$ -weighted images form the basis for functional MRI.

# Chapter 11 - Physiological basis of fMRI signals

The most popular approach to fMRI uses the **Blood Oxygenation Level Dependent (BOLD)<sup>21</sup>** **contrast<sup>22</sup>**, which is based on signal difference between a series of  $T_2^*$ -weighted images. Other methods for performing fMRI are available but less widely used. These include several varieties of Arterial Spin Labeling (ASL) which use pulse sequences sensitive to blood volume or cerebral perfusion. Because it is by far the most common method currently used, we focus here exclusively on reviewing BOLD physiology.

BOLD imaging takes advantage of the  $T_2^*$  difference between oxygenated and deoxygenated hemoglobin. As neural activity increases, so do affected brain regions' metabolic demands for oxygen and nutrients. Neural firing signals oxygen extraction from hemoglobin in the blood, which causes the hemoglobin to become paramagnetic as exposure of iron atoms to the surrounding water increases. This creates small magnetic field distortions that in turn cause decreased  $T_2^*$ , thus leading to faster signal decay and local BOLD signal decrease. A subsequent over-compensation in blood flow tips the balance towards oxygenated hemoglobin (and less signal loss due to dephasing), which leads to increased BOLD signal.

BOLD fMRI allows one to study hemodynamic responses to neural firing. We typically refer to the MR signal change caused by a neural event as the *hemodynamic response function* (HRF). Increased metabolic demands of neuronal activity lead to increased inflow of oxygenated blood to active brain regions. Since the body supplies more oxygen than it actually consumes, the deoxygenated hemoglobin concentration decreases, which leads to increased signal. This positive signal rise has an onset approximately 1 - 2 seconds after neural activity onset and peaks 5 - 8 seconds after peak neural activity. After reaching its peak level, the BOLD signal decreases to a below baseline level for roughly 10 sustained seconds. This effect, known as the post-stimulus undershoot, is due to blood flow's more rapid decrease than blood volume's, which allows for a greater deoxygenated hemoglobin concentration in previously active brain regions.

Many studies have shown evidence of an oxygenation level decrease immediately following neural activity which causes decreased BOLD signal in roughly the first 0.5 seconds following activation. We believe this is due to oxygen extraction taking place prior to oxygenated blood's inflow; we typically refer to this as the **Öinitial dipÖ<sup>23</sup>**. Researchers have reported the ratio of the dip's amplitude compared to the positive BOLD signal rise to be roughly 20% at 3 Tesla, though it ultimately depends on the magnet's strength. There is also evidence that the dip may be more localized to neural activity

---

<sup>21</sup><http://www.pnas.org/content/89/12/5675>

<sup>22</sup><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC55275/>

<sup>23</sup><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3389272/>

areas than the rise, which appears less spatially specific. Due in part to these reasons, the existence of the negative response remains controversial as researchers have so far not reliably observed it.

Questioning how well BOLD signal reflects neural firing increases is natural. Not only is the answer complex but understanding the BOLD response's physiological basis is a topic of intense<sup>24</sup> research<sup>25</sup>. Essentially the BOLD signal corresponds relatively closely to the local electrical field potential surrounding a cellular group, which in turn likely will reflect post-synaptic activity changes under many conditions. Demonstrations by Logothetis and colleagues<sup>26</sup> have shown that BOLD activity closely tracks neural firing positions and local field potentials in monkeys' visual cortices, even to specific cell columns' locations responding to particular line orientations. However, under other conditions, neural activity and BOLD signal may become decoupled. Thus it is likely that the BOLD signal only reflects a portion of neural activity changes in response to a task or a psychological state.

Another important question is whether BOLD signal increases reflect neural excitation or inhibition. Some researchers support the idea that glutamate metabolism, usually a major brain excitatory transmitter which 60-90% of the brain's neurons releases, drives much of the blood's glucose and oxygen extraction<sup>27</sup>. This is based on the belief of glutamate's involvement in generating the signals that trigger glucose uptake from blood vessels. However, this is not the complete story as inhibitory interneuron activation may in some cases also cause BOLD increases.

Given these ambiguities, it is reasonable to ask whether BOLD signal increases linearly with increased cognitive effort, which we define for present purposes as the metabolic demand involved in mental process engagement. In addition to issues regarding which physiological processes BOLD signals sample, floor and ceiling effects could result in insensitivity to task/mental state demands, which would create null findings. The answer to this question depends on precise task, mental state, experiment, subject's expertise, and brain regions which researchers are testing. A distinction between cognitive effort and work (what the cognitive effort accomplishes) illuminates two sets of findings. First, it has been shown that 'expert' subjects (with expertise related to the task of interest) achieve the same outcome with less cognitive effort than 'novices'. Second, stimulus repetition can lead to lower BOLD sensory cortex responses while subjects still perceive the stimulus - the same percept accompanies reduced fMRI signal. These examples illustrate instances in which BOLD signal may not increase linearly with task demands. Fortunately researchers have shown that BOLD signal increases approximately linearly in appropriate brain regions with increasing demand on conditions including visual processing, reaction time, subjective value, and pain. These demonstrations of BOLD signal sensitivity to particular mental processes in a specific psychological intensity range are important because they help ensure that brain measures will be sensitive to subsequent tests which try to augment or inhibit the mental state.

<sup>24</sup><http://www.ncbi.nlm.nih.gov/pubmed/8978388>

<sup>25</sup><http://www.ncbi.nlm.nih.gov/pubmed/9558643>

<sup>26</sup><http://www.ncbi.nlm.nih.gov/pubmed/11449264>

<sup>27</sup><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC21753/>

# Chapter 12 - Constraints on fMRI spatial and temporal resolution

## Spatial Limitations

One of fMRI's primary benefits is that it provides a relatively high spatial resolution compared to many other commonly used functional imaging modalities, such as PET, MEG, and EEG. For example, we can acquire fMRI data at a spatial resolution less than  $1mm^3$  in high-field animal imaging, though it is typically around  $27 - 36mm^3$  for human studies.

Limiting factors in the spatial resolution of fMRI data include signal strength and BOLD imaging's point-spread function, which tends to extend beyond neural activation sites into adjacent draining veins<sup>28</sup>. Estimates of BOLD's point-spread function, which limits effective resolution based on the fact that BOLD samples local vasculature oxygenation and flow, are around  $3mm$  at 3 Tesla. This is true no matter how small the acquired voxels are. For these reasons, separating out information encoded in brain features, such as cortical columns and even major sub-nuclei (there are 30 or so each in the amygdala and the thalamus), requires high-resolution techniques with customized acquisition parameters in order to achieve the necessary resolution. Despite limits to the BOLD point-spread function, it is still possible to obtain differential information encoded in brain structures with a spatial frequency around  $1 - 2mm$ . For example, careful work in individual participants has demonstrated the ability to image human ocular dominance columns<sup>29</sup>.

In addition, there are multiple analysis choices that ultimately limit many studies' effective spatial resolution. First, researchers commonly spatially smooth fMRI data prior to performing analysis, which causes a decrease in the data's effective resolution. Second, making inferences about populations requires analyzing groups of individuals who each have inherent differences in brain structure. To circumvent this issue it is therefore typical to align individual brains to one another through a registration or warping process which introduces substantial blurring and noise in the group average. One meta-analysis estimate is that the spatial variation in an activation peak's location among comparable group studies is about  $2 - 3cm$ <sup>30</sup>.

Use of high-resolution fMRI imaging to overcome these limitations is a challenging and rapidly developing area of research. By focusing on particular regions and limiting much of the brain's data collection, it is possible to acquire voxels of  $3.375mm^3$ , which will yield maps with resolution closer to functional sub-regions' physical size - sub-regions such as hippocampus cortical fields or brainstem nuclei. We can potentially considerably enhance resolution with high-field imaging and

<sup>28</sup><http://onlinelibrary.wiley.com/doi/10.1002/mrm.10252/abstract>

<sup>29</sup>[http://www.cell.com/abstract/S0896-6273\(01\)00477-9](http://www.cell.com/abstract/S0896-6273(01)00477-9)

<sup>30</sup><http://www.sciencedirect.com/science/article/pii/S105381190400223X>

analysis techniques that remove some fMRI signal spread due to the draining veins. In addition, collecting thinner slices can reduce susceptibility artifacts and improve imaging around the brain's base. There are, however, costs to this as smaller voxel volume means substantial signal loss. Ultimately high-resolution studies are very promising when a small set of subcortical nuclei or nearby cortical regions are the primary interest.

Finally, we can partially overcome limitations due to inter-individual variability in group analyses by identifying regions of interest on individual participants' anatomical images or by using advanced cortical unfolding and inter-subject warping techniques. A more recent idea, called 'hyperalignment', is to match voxel response profiles between subjects so their alignment is in a 'representational' space rather than an [anatomical space<sup>31</sup>](#). These techniques are making it increasingly possible to do group studies at higher effective spatial resolution, obtaining more precise population inferences about performance, clinical status, and other outcomes.

Similarly, in the area of preprocessing, the introduction of enhanced spatial inter-subject normalization techniques and improved smoothing techniques could potentially help researchers avoid data blurring's most dramatic effects. Adaptive smoothing techniques likewise can more efficiently retain boundaries between different tissue types.

## Temporal Limitations

A given fMRI study's temporal resolution depends on the TR, which in most fMRI studies ranges from 0.5–4.0s. These values illustrate the fundamental disconnect between the underlying neuronal activity, which takes place in tens of milliseconds, and the study's temporal resolution. However, it is important to note that most fMRI data statistical analysis focuses on using the BOLD response's positive rise to study the underlying neural activity. Hence the limiting factor in determining the appropriate temporal resolution is generally not data acquisition speed, but rather is response speed of the underlying evoked hemodynamic to a neuronal event. Since the basis of inference is oxygenation patterns which take place roughly 5-8 seconds after neuronal activation, researchers have generally considered a time resolution in the range of 2 seconds adequate.

That said, currently resolutions are not conducive to modeling the physiological artifacts present in the fMRI signal. For example, a subject's heart-rate and respiration give rise to periodic fluctuations particularly difficult to model without explicitly monitoring this signal. This is true because at standard resolutions these fluctuations violate the Nyquist criteria, which states that it is necessary to have a sampling rate at least twice as high as the frequency of the periodic function which one seeks to model; therefore heart rate and respiration are often left un-modeled. Because of aliasing, signal arising from physiological effects tends to be distributed throughout the time course, which gives rise to temporal autocorrelation in the signal. As fMRI signal generally suffers from low signal-to-noise ratio (SNR) and physiological artifacts can potentially comprise a large portion of the noise component, this can seriously impede accurate inference. There has recently been active research on increasing fMRI studies' temporal resolution, which would make TRs around hundreds

---

<sup>31</sup><http://www.ncbi.nlm.nih.gov/pubmed/22017997>

of milliseconds possible. These advances may ultimately allow us to circumvent many of the issues involved with physiological artifacts.

Because of relatively low temporal resolution and the hemodynamic responses sluggish nature, the ability to perform inference about when and where activation takes place finds its basis in oxygenation patterns outside the immediate vicinity of the underlying neural activity on which we want to base our conclusions. Since the positive rise in BOLD response occurs on a larger time scale than brain operation speed, there is a risk of unknown confounding factors influencing time-to-peak ordering relative to brain activation ordering in different ROIs. For these reasons, it is difficult to determine brain activity's absolute timing with fMRI data. However, studies have shown<sup>32</sup> that well-designed experiments can accurately capture a voxel's relative timing in response to different stimuli.

In addition, it is possible to measure the HRF at a finer temporal resolution than the actual  $TR$  as long as we jitter repeated stimuli's onsets in time relative to data collection time. For example, if we shift a repeated stimulus' onset by half a  $TR$  in a fraction of the stimuli, it may possible to estimate the HRF at a temporal resolution of  $TR/2$ , compared to a  $TR$  resolution without jittering.

A series of recent technological developments called 'multiband' or 'simultaneous multi-slice' MRI] (<http://www.ncbi.nlm.nih.gov/pubmed/20432285>) have sped up the temporal resolution of fMRI by up to an order of magnitude (i.e. from 2s to 0.2s for whole-brain imaging); they appear likely to offer the possibility for even further acceleration. In contrast to standard acquisition techniques, multiband MRI excites multiple slices at the same time, then separates MR signals arising from these slices by using multiple receiver coils and special encoding techniques. The introduction of multiband MRI promises to change the manner in which we acquire and analyze fMRI data, by providing access to data with both high spatial and temporal resolution.

---

<sup>32</sup><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC27993/>

# **Part 3: Basics of fMRI signal processing and analysis**

# Chapter 14 - Experimental Design

Designing a neuroimaging study involves a series of tradeoffs between experimental power and ability to make strong inferences based on the results. Researchers use two major design classes in most fMRI experiments: block designs and event-related designs (though we can intermix or hybridize these). Block designs typically yield high experimental power but provide imprecise information about the particular psychological processes which activate a brain region. In contrast, event-related designs allow us to more precisely relate brain activation to the particular cognitive processes certain mental events engage. However, they often suffer from decreased power, depending on the studied process.

Researchers may choose to focus intensively on testing one comparison of interest and will maximize the power to detect this particular effect or they may test multiple conditions to draw inferences about a brain region's general involvement in a class of similar psychological processes. Below we describe several experimental design types and discuss the types of applications which they best suit.

## Block designs

A block design separates experimental conditions into extended time intervals, or blocks. For example, to image a briefly occurring psychological process (e.g. activations due to attention switching) with a block design, one might repeat the process of interest during an experimental block (A) and have the subject rest during a control block (B). The A - B comparison is the most basic contrast type for this design type, but it imposes limitations on result interpretability. While block designs capture activations related to slowly changing factors well, they are not well suited for determining neural responses to individual stimuli. In addition, the A - B contrast does not provide information about whether A alone activates a region or B alone deactivates it, or if it is some combination of the two. Including multiple controls and comparison conditions can limit this problem to some degree.

Block designs' main advantage is that they generally offer increased statistical power to detect a change. In fact, researchers have shown<sup>33</sup> that under ideal conditions, block designs can be over six times as efficient as randomized event-related designs. Research assessing fMRI design experimental power indicates that the optimal alternating-block design with respect to statistical power is a 16-18 s task / 16-18 s control<sup>34</sup>. However, a block design's relative power ultimately depends on whether the target mental process continuously engages in A but does not engage at all in B and whether imposing a block structure changes the task's nature.

---

<sup>33</sup><http://www.ncbi.nlm.nih.gov/pubmed/12595184>

<sup>34</sup>[http://fmriserver.ucsd.edu/ttliu/papers/liu04\\_mfmri2.pdf](http://fmriserver.ucsd.edu/ttliu/papers/liu04_mfmri2.pdf)

## Event-related fMRI

In an event-related design, the stimulus consists of short discrete events, such as brief light flashes, with randomized timing. These design types are flexible and allow estimation of key HRF features (e.g. onset and width) which we can use to make inference about relative activation timing across conditions and about sustained activity. Event-related designs allow one to discriminate different conditions' effects as long as one either inter-mixes different event types or varies the inter-stimulus interval between trials. Another advantage to event-related designs is that we can avoid the effects of fatigue, boredom, and systematic thought patterns that are not task related during long inter-trial intervals. A drawback is that the power to detect activation is typically lower than in block designs, though the ability to obtain data over more trials per unit time can counter this power loss.

Event-related designs rely strongly on assumptions about the time course of both the evoked neural activity and the HRF. For example, it is common to assume a near-instantaneous neural brief event response and a canonical HRF shape to generate linear models for statistical analysis. In practice, however, we know the HRF's timing and shape vary across the brain within an individual and **across individuals**<sup>35</sup>. Because block designs depend on the total activation caused by a *train* of stimulus events, this makes the overall predicted response less sensitive to variations in the individual events' response shape. However, predicted responses in block designs can still be quite inaccurate if the HRF model is extremely inaccurate or if the neural activity's density and time-course model is not appropriate.

Event-related designs rely on voxels' response to single trials or brief events. The underlying assumption is that the BOLD response's magnitude and shape do not change depending on the preceding stimuli. Studies have found that nonlinear effects in rapid sequences (e.g. 1 – 2s) can be **quite large**<sup>36</sup> but that responses are roughly linear if events occur at least **5s apart**<sup>37</sup>. If properly designed, rapid designs still allow one to discriminate the effects of different conditions. A key idea is incorporating 'jitter', or variable inter-stimulus intervals (ISIs), between events, which allows comparison of event-related responses to an implicit resting baseline - i.e. determination whether the events are 'activations' or 'deactivations' relative to rest.

With a randomized and jittered design, sometimes several trials of a single type will occur in a row and so, because hemodynamic response to closely spaced events sums roughly linearly, that trial-type's expected response will build to a high peak. Jitter's introduction allows activation peaks and valleys to develop specific to a particular experimental condition. If the goal is to compare event types (e.g. A  $\neq$  B), randomizing event order creates optimal rise and fall without additionally jittering the ISI. However, **jittering**<sup>38</sup> the ISIs is critical if the goal is to compare events to baseline activity and thus to determine whether events *activate* or *deactivate* a voxel relative to that baseline.

Finally, we must weigh the advantages of rapid pacing, including faster trials and possible increased statistical efficiency, against potential problems with nonlinearity, regressor multicollinearity, and

<sup>35</sup><http://www.ncbi.nlm.nih.gov/pubmed/9811554>

<sup>36</sup><http://www.ncbi.nlm.nih.gov/pubmed/9558643>

<sup>37</sup><http://gablab.mit.edu/downloads/Miezin.NeuroImage.2000.pdf>

<sup>38</sup><http://www.ncbi.nlm.nih.gov/pubmed/12595184>

model mis-fit. A popular choice is to use ‘jittered’ designs with inter-stimulus intervals of at least 4s and with exponentially decreasing delay frequencies up to 16s.

## Optimized experimental designs

What constitutes an optimal experimental design ultimately depends on the task, as well as on the fMRI signal’s ability to track changes the task introduces over time. It also depends on the study’s specific comparisons (i.e. contrasts) of interest. In addition, factors that include the delay and shape of the fMRI signal, scanner drift, and nuisance factors, such as physiological noise, conspire to make fMRI experimental design more complicated than experiments that solely measure behavior. Not all designs with the same trial quantity of a given condition set are equal, so determining events’ appropriate spacing and ordering is critical.

Some intuitions and tests of design optimality follow from deeper understanding of the statistical analysis of fMRI data. We also note that several computer algorithms are available for constructing statistically optimized designs, including an approach based on [m-sequences<sup>39</sup>](#), which are mathematical sequences that are near-optimal for certain design types, and one based on a [genetic algorithm<sup>40</sup>](#), which incorporates m-sequence designs as a starting point and considers various contrasts’ relative importance to the study goals in determining optimality.

---

<sup>39</sup><http://www.ncbi.nlm.nih.gov/pubmed/12169264>

<sup>40</sup><http://www.ncbi.nlm.nih.gov/pubmed/12595184>

# Chapter 15 - Resting state, natural viewing, and non-experimental designs

## Resting state

The majority of fMRI experiments study cognitive tasks, perception, and action related brain activations. However, a number of years ago Biswal et al. (1995)<sup>41</sup> observed that the BOLD time-courses in left and right sensorimotor cortices were highly correlated at rest, which suggested that much of these regions' 'noise', and possibly that within the rest of the brain, was due to coherent spontaneous activity. Further studies<sup>42</sup> identified a set of large-scale networks that showed correlated activity in a task's absence. We typically identify these networks with clustering approaches based on pairwise correlations or data-decomposition algorithms, such as Independent Components Analysis (ICA) or Principal Components Analysis (PCA), in which we consider voxels loading highly on the same component to comprise a 'network'.

Different samples have reliably identified these networks, which we often label using psychological terms and use in many studies as units of analysis. The ventromedial and dorsomedial prefrontal cortices (vmPFC/dmPFC), posterior cingulate, medial temporal lobe, superior temporal cortices, and a number of other areas comprise the 'default mode network' (DMN)<sup>43</sup>. This name comes from many of these regions showing high metabolic activity in a person at rest and decreased activity during the performance of several cognitive tasks. However, it is critical to note that a number of tasks which focus on internal reflection in fact activate DMN regions above resting levels; these tasks include retrieving semantic memories, imagining the future, experiencing psychological stress, experiencing emotion, reflecting on one's self, reflecting on others' minds, and 'mind-wandering', which is a mix of often self-focused thoughts and memories<sup>44</sup>.

Researchers have identified and labeled many other networks with terms which imply that they implement specific functions. For example, the 'salience network'<sup>45</sup> includes regions which activate many cognitive and affective states, such as the dorsal anterior cingulate, anterior insula, and amygdala. While these regions respond to many salient events, conclusion that a task activates the network because it is 'salient' would be a mistake.

<sup>41</sup><http://www.ncbi.nlm.nih.gov/pubmed/8524021>

<sup>42</sup><http://www.pnas.org/content/98/2/676.abstract>

<sup>43</sup><http://www.pnas.org/content/98/2/676.abstract>

<sup>44</sup><http://www.ncbi.nlm.nih.gov/pubmed/20188659>

<sup>45</sup><http://www.ncbi.nlm.nih.gov/pubmed/17329432>

Resting-state studies have increased in popularity with the hope that they will provide characteristic markers related to aging, psychopathology, performance, and clinical symptoms. These studies do not employ a specific task or an experimental manipulation but instead acquire data while the subject rests in the scanner. Most studies ask subjects to focus on a fixation cross during the experiment or, alternatively, to close their eyes. Typical scan durations are 5 - 12 minutes per subject, which makes it easy and cost-effective to acquire many subjects' data. In addition, one primary resting-state fMRI benefit is the ability to compile data across research labs, as researchers need not synchronize experiments. This has led to many large-scale data sharing initiatives (e.g. 1000 Functional Connectomes Project).

Resting-state fMRI primarily studies low-frequency BOLD fluctuations. Researchers have found functionally relevant, spontaneous BOLD oscillations in the lower frequency ranges ( $0.01 - 0.08\text{Hz}$ ), which is separable from frequencies that correspond to respiratory ( $0.1 - 0.5\text{Hz}$ ) and cardiovascular ( $0.6 - 1.2\text{Hz}$ ) signal. We often band-pass filter data at  $0.01 - 0.08\text{Hz}$  for these reasons. In addition, research has shown that non-neuronal physiological signals may interfere with resting state BOLD data, and that removing confounding signals such as respiratory or cardiovascular noise considerably improves data quality attributable to neural activity.

Resting state data analysis differs from experimental fMRI studies. Since no experimental manipulation occurs, conventional general linear model (GLM) analyses are not feasible. Instead most analyses study correlational structure among voxels or regions. Resting state analysis involves first estimating brain connectivity measures using 'seed' regions, ICA, or voxel-by-voxel pairwise correlation matrices across the brain. Researchers then correlate these connectivity metrics with outcomes of interest (e.g. clinical symptom scores).

Though increasingly popular, resting state analyses are not without problems. Some pitfalls are ambiguity and variability in which person-to-person mental states and physiological processes researchers actually image. In addition, we can show at least some coherent brain activity in rest depends on physiological noise, including artifacts from head movement, respiration, pulsatile motion, and vascular oxygenation due to heartbeat. Though we implicitly assume that participants comply with task instructions and are all equally awake and alert, this is clearly not the case as a recent study found that 50% of resting state study participants are asleep after 10 minutes (<http://www.ncbi.nlm.nih.gov/pubmed/24811386>). Since the transition from wakefulness to sleep changes activity patterns and neuronal oscillations drastically, it is important to control for wakefulness during the scan and carefully check for potential group differences.

Finally, different resting state connectivity patterns relate to different spontaneous thought types<sup>46</sup>. While the goal of experimental paradigms is to explicitly control types of mental processes in which a participant engages and to study brain activity in relation to those processes, there is no such control in resting state studies.

While some researchers view resting state scans as a window into the brain's intrinsic architecture, others view them as a window into mental states or mental status, or simply as physiological artifacts they will discard. It is difficult to determine how related the resting brain connectivity patterns are

---

<sup>46</sup><http://www.ncbi.nlm.nih.gov/pubmed/20188659>

to each of these three alternatives. Even if an outcome's association with resting state networks is reliable, it may not be clear why or whether the association has interesting neuroscientific implications. Instead they may merely be from physiological or image artifacts. Resting-state fMRI's utility, like all scientific inquiry areas, is ultimately an empirical question currently being explored in various ways.

## Natural viewing and non-experimental designs

The fast growth in computing power together with the introduction of multivariate techniques into the fMRI literature have paved the way for large-scale decoding studies. Many of these studies aim to study the brain processes of natural vision. To achieve higher external validity as in natural conditions, researchers reduce experimental control. Early approaches searched for brain regions whose activity held information about the current conscious percept with quasi-experimental designs. These studies used multi-stable visual stimuli (e.g. Necker cube) that led to fairly regular conscious percept switches. Researchers asked subjects to report the perceptual switches via button-presses and then analyzed responses following perceptual switches. An early univariate fMRI study reported phasic positive fusiform gyrus responses and negative thalamus responses (Kleinschmidt et al., 1998). A later study<sup>47</sup> using multivariate analyses estimated the current percept from lateral geniculate nucleus activity, an early visual area. To achieve even more natural viewing conditions, some researchers present their subjects with movies or podcasts while measuring fMRI data. These scans can last up to 10 hours over multiple sessions. Researchers can then use the enormous amount of acquired data to predict current perceptions from brain activity by exploiting the unique covariation patterns between brain activity and the current stimulus' composition features<sup>48</sup>.

---

<sup>47</sup><http://www.sciencedirect.com/science/article/pii/S0896627305004083>

<sup>48</sup>[http://www.cell.com/neuron/abstract/S0896-6273\(12\)00934-8](http://www.cell.com/neuron/abstract/S0896-6273(12)00934-8)

# Chapter 16 - Essentials of fMRI signal processing

In order to appropriately model fMRI data, it is important to first gain a better understanding of the various components present in the time series data. In general, the data consists of the BOLD signal, which is the component of interest, and a number of nuisance parameters and noise. Here we discuss each component in detail and various modeling strategies.

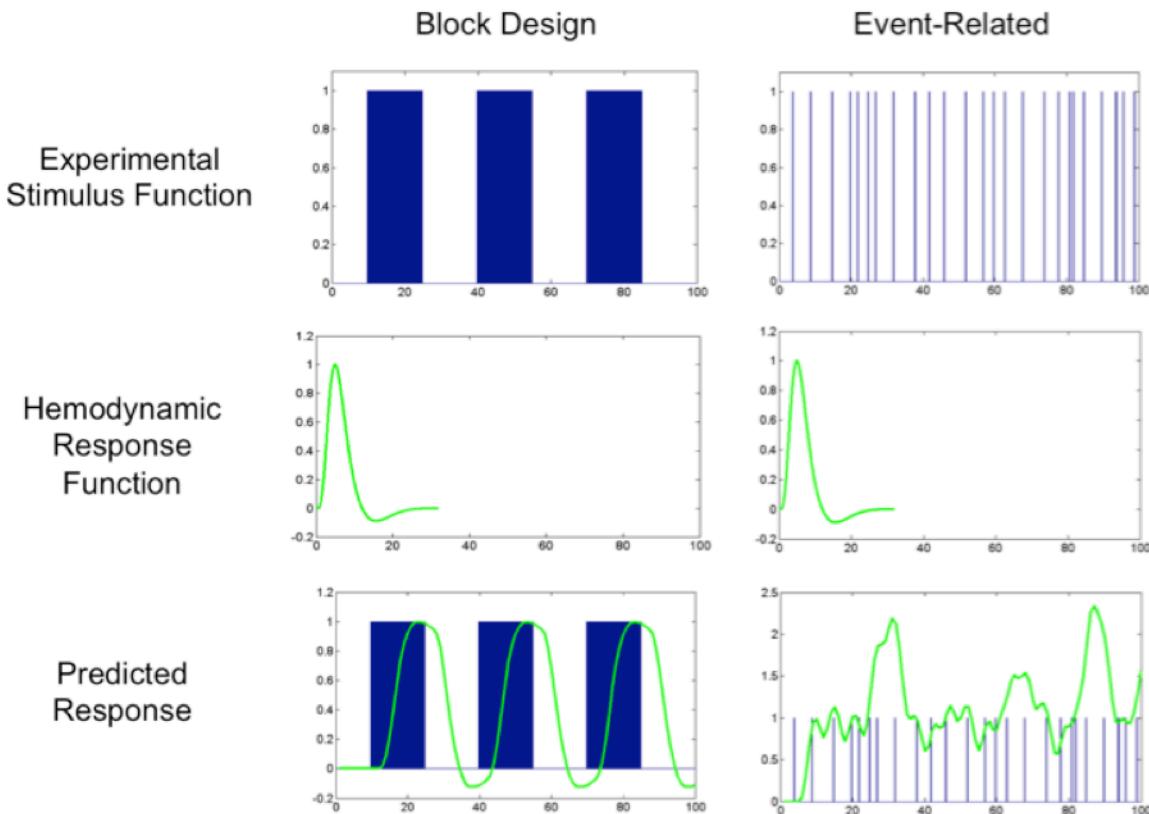
## BOLD signal

The evoked BOLD response is a complex nonlinear function of neuronal and vascular changes. The shape of the response depends both upon the applied stimulus (i.e. task) and the hemodynamic response to neuronal events. A number of methods for modeling the BOLD response as well as the underlying hemodynamic response function (HRF) exist in research literature. A major difference between the methods lies in how each method models the relationship between the stimulus and BOLD response. In particular, we differentiate between non-linear physiologically based models, such as the Balloon model, and models that assume a linear time invariant (LTI) system.

The Balloon model describes the dynamics of cerebral blood volume and de-oxygenation, and their effects on the resulting BOLD signal. It consists of a set of ordinary differential equations that model changes in blood volume, blood inflow, deoxyhemoglobin, and flow-inducing signal, and then describes how these changes impact the observed BOLD response. While models of this type tend to be more biophysically plausible than linear models, they have several shortcomings. First, they require estimation of a large number of model parameters, second, estimates are often noisy, and third, they do not provide a direct framework for performing inference. In general, researchers do not yet consider these as feasible alternatives for performing whole-brain multi-subject fMRI data analysis data in cognitive neuroscience studies. However, they are a critical component in brain connectivity studies using Dynamic Causal Modeling (DCM), which we will discuss in a later chapter.

While nonlinear models' flexibility makes them attractive, linear models allow for robust and interpretable characterizations in noisy systems. Standard practice assumes a linear relationship between neuronal activity and BOLD response in which linearity implies that the magnitude and shape of the evoked HRF does not depend on the preceding stimuli. Studies have shown that under certain conditions, we can consider the BOLD response approximately linear with respect to the stimulus, particularly if events occur at least 5 seconds apart. However, studies have shown there are still some nonlinearities (roughly 10%) even at 5 second spacing. Other studies have found that nonlinear effects in rapid sequences (e.g. stimuli spaced fewer than 2 seconds apart) can be quite large.

The ability to assume linearity is critical, as it allows modeling of the relationship between stimuli and the BOLD response with a *linear time invariant system*. Here the assumed neuronal activity (based on task manipulations) constitutes the input, or impulse, and the HRF is the impulse response function. In a linear system framework, the signal at time  $t$ ,  $x(t)$ , is modeled as the convolution of the stimulus function  $v(t)$  and the hemodynamic response  $h(t)$ , i.e.  $x(t) = (v * h)(t)$ . In this setting, we assume  $h(t)$  either takes a canonical form or, alternatively, we model it using a set of linear basis functions. Figure 16.1 illustrates this concept for block and event-related designs.



**Figure 16.1.** An illustration of the linear system framework used in fMRI. The experimental stimulus function is convolved with a canonical HRF to obtain the predicted response. Here there are two examples: a block and event-related design.

LTI systems' scaling, superposition, and time-invariance properties characterize them. Scaling implies that if the input scales by factor  $b$ , then the BOLD response will scale similarly. This is important as it implies that the signal amplitude provides a measure of the neuronal activity amplitude. Therefore researchers can use the relative difference in amplitude between two conditions to infer that the neuronal activity was similarly different. Because of this assumption, many of the activation analyses performed on fMRI data focus on studying contrasts between the responses to stimuli at different levels. Superposition implies the response to two different stimuli applied together equals the sum of the individual responses. Finally, time-invariance implies that a stimulus' shift by a time  $t$  will mean the response similarly shifts by  $t$ . These three properties together allow us to differentiate

between responses to multiple closely spaced stimuli in various brain regions.

As with any model, one makes a number of assumptions with an LTI system. First, the system assumes the BOLD response is linear. Studies have shown this is reasonable, though some departures from linearity have been observed. There is some evidence of refractory effects, or reductions in response amplitude as a function of inter-stimulus intervals, which may cause overestimation of closely spaced stimuli's amplitude. Second, LTI systems assume the correct modeling of the neural activity function. As researchers typically take this to be equal to the experimental paradigm, one must assume this provides a reasonable proxy for the underlying neuronal activity. Third, the system assumes the correct modeling of the HRF. Researchers often assume a canonical shape for the HRF. A popular choice is the linear combination of two gamma functions, i.e.

$$h(t) = \frac{t^{\alpha_1-1} \beta_1^{\alpha_1} e^{-\beta_1 t}}{\Gamma(\alpha_1)} - c \frac{t^{\alpha_2-1} \beta_2^{\alpha_2} e^{-\beta_2 t}}{\Gamma(\alpha_2)}$$

where  $\alpha_1 = 6$ ,  $\alpha_2 = 16$ ,  $\beta_1 = \beta_2 = 1$ ,  $c = 1/6$  and  $\Gamma$  represents the gamma function. See Fig. 7.7 for an illustration.

It is critical to note that the timing and shape of the HRF are known to vary across the brain within and across individuals. Part of the variability is due to the vascular bed's underlying configuration, which may cause HRF differences for purely physiological reasons across brain regions within the same task. Differences in evoked neural activity patterns in regions performing different functions related to the same task are another source of variability.

In order to deal with this issue, modeling of the HRF often uses multiple basis functions since a linear combination provides a better fit for the evoked BOLD response. To illustrate, suppose we model the HRF as a linear combination of temporal basis functions:  $f_i(t)$ , such that  $h(t) = \sum \beta_i f_i(t)$ . Then we can rewrite the BOLD response:  $x(t) = \sum \beta_i (s * f_i)(t)$  where each corresponding  $\beta_i$  describes the  $i^{th}$  component's weight. We will discuss this issue in greater detail in a later chapter.

## Noise and nuisance signal

In addition to the signal of interest, random noise and various nuisance components arising from hardware limitations and the subjects themselves also corrupt the measured fMRI signal. Fluctuations in the MR signal intensity caused by thermal electron motion within the subject and the scanner are one source of variability. These give rise to highly random noise independent of the experimental task. The amount of thermal noise increases linearly as a function of field strength, with higher field strengths giving rise to more noise. However, these fluctuations do not tend to exhibit spatial structure. Averaging the signal over multiple voxels can help minimize the effects.

Another source of variability in the signal is due to scanner drift, as scanner instabilities result in slow changes in voxel intensity over time (low-frequency noise). The effects of drift vary across space; it is important to either remove this source of variation before fitting your model or, alternatively, include

it in your models. Because of drift, most of the power in the time course lies in the signal's low-frequency portion. To remove drift effects, it is common to eliminate fluctuations below a specified frequency cutoff with a high-pass filter. Researchers can do this either by applying a temporal filter as a preprocessing step or by including covariates of no interest into the model.

When subjects move their heads in the MRI scanner, the sequence of measurements corresponding to a given voxel may actually comprise values originating from different locations. This necessitates motion correction: prior to analysis, researchers estimate the between-scan movement using a rigid body transformation, then realign the images. However, this procedure does not correct for so-called 'spin history artifacts', which are changes head motion causes in the magnetic field that lead to nonlinear, time-varying distortion of the resulting brain images. There has been some debate on how to deal with these residual artifacts.

Some researchers suggest including motion regressors as 'nuisance covariates' in the BOLD response model to adjust for this error; they argue that this yields estimates which seem more reasonable than those obtained without these covariates. But as head motion tends to be task related, concern exists that these covariates' inclusion can lead to underestimating the signal component from 'true' activation. However, we recommend that motion regressors be included in models of the BOLD response.

Physiological noise due to patient respiration and heartbeat can cause spatial and temporal fluctuations in signal. In certain situations, we can directly estimate this physiological noise from the data. A properly designed band-pass filter can aid in removing some of it. However, in most studies with TR values ranging from 2-4s, one cannot hope to estimate and remove the effects of heart rate and respiration based solely on the observed fMRI time series. According to the Nyquist theorem, a sampling rate at least twice as high as the periodic function's frequency which one seeks to model is necessary. If the TR is too slow, as in most fMRI studies, problems with aliasing will occur and periodic fluctuations will be distributed throughout the time course, giving rise to temporal autocorrelation. Some groups have therefore begun measuring heartbeat and respiration directly during scanning and using this information to remove physiological fluctuation signal from the data. These groups do this either as a pre-processing step or by including these terms as model covariates.

In standard time series analysis, researchers use model identification techniques to determine the appropriate type and order of a noise process. In fMRI data analysis, the large number of time series that researchers analyze means this approach is not feasible, so noise models are typically specified *a priori*. An AR(p), with  $p$  set to either 1 or 2, or an ARMA(1,1) process is typically used to model noise in fMRI. Researchers generally attribute autocorrelation to unmodeled nuisance signal; if these terms are properly removed, there is [evidence<sup>49</sup>](#) that the resulting error term corresponds to white noise.

In our own work, we typically use an auto-regressive process of order 2. We use an AR model instead of an ARMA model because it allows us to use method of moments rather than maximum likelihood procedures to estimate the noise parameters. This speeds up computation time when

---

<sup>49</sup><http://www.ncbi.nlm.nih.gov/pubmed/16099175>

repeatedly fitting the model to tens of thousands of time series. The order 2 of the AR process has been empirically determined to provide the most parsimonious model which accounts for signal autocorrelation present due to aliased physiological artifacts.

# Chapter 17 - Preprocessing

Prior to statistical analysis, fMRI data undergoes a series of preprocessing steps which aim to (i) minimize the influence of data acquisition and physiological artifacts, (ii) check statistical assumptions and transform the data to meet these assumptions, and (iii) standardize the location of brain regions across different subjects to achieve validity and sensitivity in group analysis.

The major pre-processing steps performed on fMRI data (shown in Figure 17.1) include reconstruction, slice-timing correction, motion correction, co-registration of structural and functional images, normalization to standard space, and spatial smoothing. Normalization, which introduces spatial uncertainty in terms of anatomical locations, is not required for single-subject data (which therefore can provide higher anatomical resolution). Group studies, however, permit population inference and help guard against false positives due to fMRI time series artifacts. Some group studies do not employ smoothing in order to increase spatial resolution.

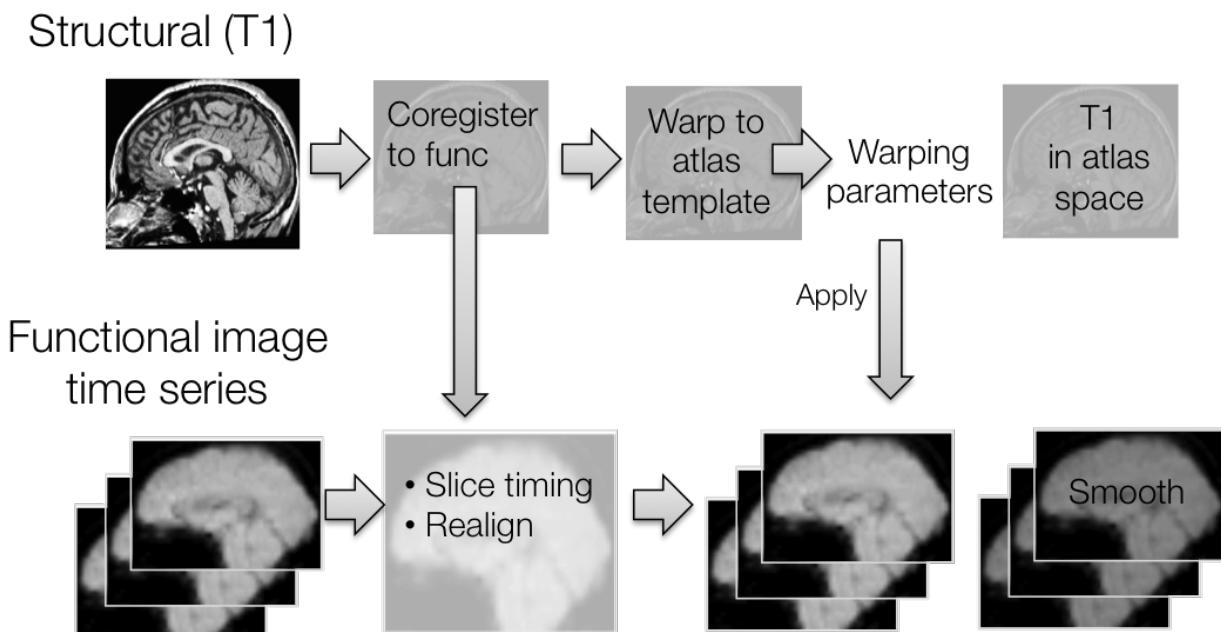


Figure 17.1. An illustration of the pre-processing steps performed on fMRI data.

## Reconstruction.

Prior to any analysis the images must first be reconstructed from the raw MR signal acquired in k-space. This step typically automatically occurs at the scanner site before most researchers actually

have access to the data. Both the raw and reconstructed data can be stored in a variety of formats, but reconstructed images are generally stored in a 3-D array of data containing the signal intensity at each voxel with a header containing information about the important image parameters including dimensionality and voxel size. A popular format is the so-called nifti-format (extension .nii), which can contain either single or multiple 3-D volumes in each file. This particular format therefore allows multiple images to be stored in a 4-D array in which the fourth dimension is time.

## Slice-timing correction

Most statistical analysis assumes that every voxel in an image was acquired simultaneously. In reality, voxels from different slices are (relative to each other) shifted in time. fMRI typically acquires data in a slice-wise manner, with some collected later in the acquisition process than others. Slice-time correction shifts each voxel's time series so that the voxels appear to have been sampled simultaneously. To perform this correction, we need to estimate the signal intensity in all voxels at the same moment in the acquisition period. We can do this by interpolating the signal intensity at a chosen time point using data from the previous and subsequent acquisitions at the same voxel. A number of different interpolation techniques exists, each of which have varying degrees of accuracy and speed. For example, sinc interpolation is the slowest but generally the most accurate approach. A number of researchers elect not to use slice-timing correction as it adds interpolation error to the data. These researchers instead use more flexible hemodynamic models in later analysis to account for variations in acquisition time.

## Motion correction

Head motion is a serious problem in most neuroimaging experiments. Even minor movements of the head during an experiment can be a major source of error if not handled properly. It is assumed that a time series associated with a particular voxel depicts the same region of the brain at every time point. However, if the head moves between acquisitions, the voxel's signal intensity gets "contaminated" by the signal from its neighbors.

To correct for this, one must rotate and translate each individual image to compensate for the subject's movements. Realignment is typically performed by choosing a reference image (e.g. the first or the mean image) then using a rigid body transformation to match all the other images in the time series to it. This allows the translation (shifted in the x, y, and z directions) and rotation (altered roll, pitch, and yaw) of the images, to find the best possible alignment with the reference image. The transformation can be expressed as a pre-multiplication of the image spatial coordinates altered by a  $3 \times 3$  affine matrix: the elements of this matrix are the parameters to be estimated. An iterative algorithm searches for the estimates that provide the best match between the image and the reference image. Usually, the matching process is done by minimizing sums of squared differences between the two images. Often the translations and rotations are plotted as a function of time as shown in Figure 17.2, and saved for subsequent statistical analysis.

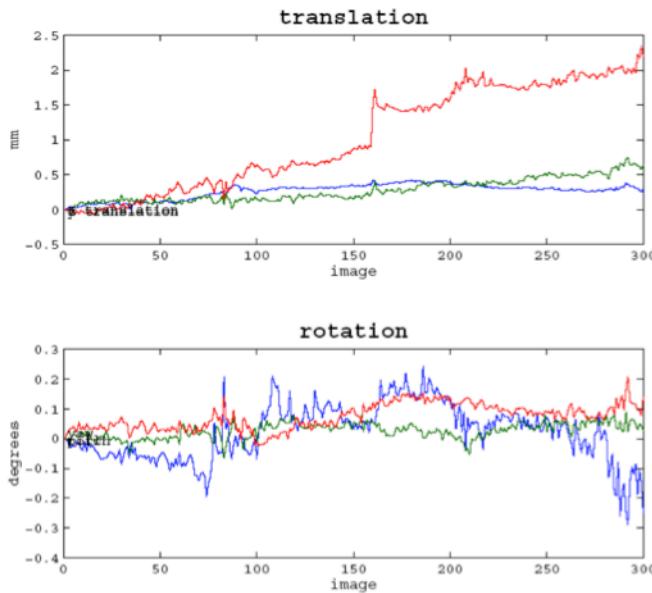


Figure 17.2. Plots of the 3 translations and 3 rotations as a function of time estimated using a rigid body transformation.

While this procedure effectively corrects for small head movements, it does not correct for more complex spin-history artifacts created by the motion, so the parameters at each time point are often included in subsequent statistical analysis as covariates of no interest. However, even this additional step does not completely remove the effects of head motion as residual artifacts remain in the data and contribute to noise. Sometimes this noise can be correlated with the task, particularly if the task somehow causes the motion, and can create false results in single-subject analyses. Because these artifacts typically differ in sign and in magnitude across subjects, group analysis remains valid. Large movement artifacts, if present, can severely compromise statistical power, therefore it is common to exclude subjects who move their heads substantially during the scan.

## Co-registration

The scanning session typically includes the acquisition of high-resolution structural images (T1 and/or T2), which the process of co-registration aligns with the functional images. This procedure typically uses an affine transformation (seven degrees of freedom: scaling and the same six used in rigid body transformations); maximizing the mutual information between the two images allows estimation of the transformation parameters. Typically a single structural image is co-registered to the first, or mean, functional image. Co-registration allows one to visualize single-subject task activations overlaid on the individual's own anatomical information. It also simplifies later transformation of the fMRI images to a standard coordinate system (see below).

## Normalization

To perform group analysis, each voxel must lie within the same brain structure for all subjects. Even though individual brains have different shapes and features, every non-pathological brain shares regularities; normalization attempts to register each subject's anatomy with a standardized atlas space defined by a *template* brain. The two main template brains used in practice are the Talairach atlas, based on a single subject, and the Montreal Neurological Institute (MNI) atlas, based on the combination of many MRI scans on normal controls. The process is illustrated in Figure 17.3. where a high-resolution structural T1 image is aligned to a template to obtain a normalized T1 image in a standard space.

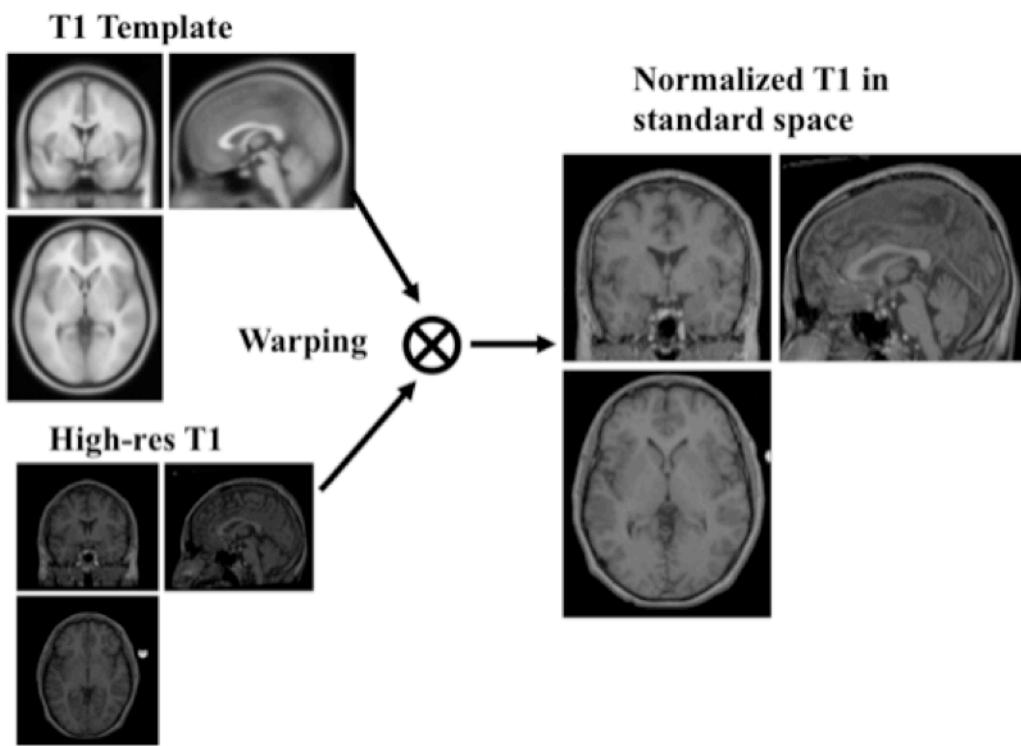


Figure 17.3. High-resolution structural T1 image is aligned to a template image to obtain a normalized T1 image in a standard space.

Normalization can be linear, simple registration of the brain's gross shape, or non-linear, warping to match local features. In intensity-based normalization, matching is performed using image intensities corresponding to gray/white matter/fluid tissue classes. Surface-based normalization, however, uses extracted features such as gyral and sulcal boundaries explicitly. In contrast to rigid body and affine transformations, normalization can locally stretch and shrink different image regions to achieve the closest match, through shifting voxel locations by differing amounts depending on their original location. A set of cosine basis functions can express the function that describes how to shift the voxels. The goal is to find a set of coefficients which corresponds to each basis function and minimizes the least square's difference between the transformed image and

the template. The ability of the algorithm to match the template's local features depends on the number and spatial frequency of the basis functions. Surface-based warping uses similar principles, but matches features on extracted cortical surface representations instead of image intensities.

Inter-subject registration is one of the largest sources of error in group-level analysis. Therefore it is important to inspect each normalized brain and, if necessary, take remedial measures. These include improving the alignment manually, using a mask to exclude problematic regions of atrophy or abnormality (e.g. a lesion), altering the number of basis functions and other fitting parameters, and, in some cases, developing specialized template brains (e.g. for children or elderly populations).

## Spatial smoothing

It is common practice to spatially smooth fMRI data prior to statistical analysis. One reason for this is to overcome limitations in the normalization step by blurring any residual anatomical differences. Another reason is that *Gaussian random field theory* a popular multiple-comparisons correction procedure, requires a certain degree of smoothness.

Smoothing typically convolves the data with a Gaussian kernel, which is a 3-D normal probability density function often described by the full width of the kernel at half its maximum height ("FWHM") in mm. The appropriate amount of smoothing is an open question. One estimate of the smoothing required to meet the assumptions of random field theory is a FWHM of 3 times the voxel size (e.g., 9 mm for 3 mm voxels). In theory a smoothing kernel matched to the activation signal extent will give optimal results in terms of signal-to-noise ratio; hence if we have *a priori* knowledge of the spatial extent, we should smooth by that amount. However, the amount of smoothing is typically chosen independently of the data. It is therefore unlikely that the smoothing kernel will actually match the signal extent. Furthermore, the whole image receives the same amount of smoothing though the spatial extent of activation is likely to differ across the brain.

Finally, it is also important to note that acquiring an image with large voxels is not the same thing as acquiring one with small voxels then smoothing it. The signal-to-noise ratio during acquisition increases as the square of the voxel volume, so acquiring small voxels means that much signal is irrecoverably lost. Researchers using multivariate analyses methods often do not smooth their images in order to retain the information contained in individual fine-grained activation patterns. This is more useful when the evaluation of the multivariate model is within one subject. When the study aims to accurately predict variables across subjects (e.g. from new fMRI data sets), some smoothing can help increase inter-subject alignment and predictive performance.

# Chapter 18 - The General Linear Model and Foundations of Analysis

The general linear model (GLM)<sup>50</sup> is arguably the most common statistical method for assessing task-brain activity relationships in neuroimaging. It is a linear statistical analysis method that subsumes many basic analysis techniques, including t-tests, ANOVA, and multiple regression. The GLM can assist with numerous tests including whether the brain responds to a single type of event, comparison of different types of events, and assessment of correlations between brain activity and behavioral performance or other psychological variables.

The GLM is appropriate when multiple predictor variables that together constitute a simplified *model* of the data's variability sources are used to explain variability in a single, continuously distributed outcome variable. In a typical neuroimaging experiment, psychological events associated with the predictors result in the outcome of a time series from a brain voxel or region of interest. Analysis is typically “massively univariate”, which implies that the analyst performs a separate GLM analysis at every single voxel in the brain and saves summary data in statistic value maps across the brain.

It is usually advantageous to design studies and statistical analyses to permit inferences about a population of participants, which is important in all kinds of studies. For instance when testing a new drug, researchers perform statistical tests that allow inference about whether the drug is likely to produce a benefit on average for a certain population. They do so by invoking the additional assumption that all participants will behave similarly to those actually observed in the study. In almost all domains of human neuro-psychology, this is not a safe assumption, so researchers should perform statistics that permit a standardized population inference. In neuroimaging, considering the multi-level nature of the data (typically described using a two-level hierarchical model) can allow generation of these statistics. In the first level, we analyze within-subject effects for each individual in the study. In the second level, we perform analyses across subjects or across groups in a group analysis. Researchers can do this either in stages or combined into one integrated model.

A key to population inference is to treat the variation across participants as an error term in a group statistical analysis, which allows the results to be generalized to new participants drawn from the same population. The most popular group analysis is the one sample t-test on differences between two conditions (e.g. Task A - Task B) at each voxel. This analysis tests whether the difference is non-zero on average for the sampled population and provides a starting point for our discussion on population inference. The principle, however, applies to any kind of statistical model, including more complex ANOVA, regression models, and multivariate analyses such as group independent component analysis (ICA).

---

<sup>50</sup><http://www.ncbi.nlm.nih.gov/pubmed/9343600>

## Setting up the GLM

The GLM approach treats the data as a linear combination of model functions: predictors plus noise or error. We can envision this as breaking the data up into the part a model can explain and the unexplainable part. Though we assume model functions have known shapes (the simplest being a straight line), we need to estimate their amplitudes or the slopes as they are unknown.

For a single subject, the fMRI time series from a single voxel is the outcome variable ( $y$ ). That voxel's activity is modeled as the linear combination of an independent predictor series ( $x_1, x_2$ , etc.) related to task conditions and other nuisance covariates of no interest (e.g. head movement estimates). fMRI analysis includes, for each task condition or event type of interest, construction of a time series of the signal response's predicted shape, which uses prior information about the shape of the vascular response to a brief impulse of neural activity. Most often, analysis implements a canonical hemodynamic response function (HRF). The vectors of predicted time series values for each task condition are collated into the columns of the design matrix,  $X$ , which contain a row for each of  $n$  observations collected over time and a column for each of  $k$  predictors. The structural model for the GLM can be expressed as:

$$y = X\beta + \epsilon$$

In this calculation,  $\beta$  is a  $k \times 1$  vector of regression slopes,  $X$  is an  $n \times k$  design matrix,  $y$  is an  $n \times 1$  vector containing the observed data, and  $\epsilon$  is an  $n \times 1$  vector of unexplained error values. We assume error values are independent and follow a normal distribution with mean 0 and standard deviation  $\sigma$ .

## GLM Estimation

The GLM fitting procedure estimates the best-fitting amplitude for each column of  $X$ , so that the sums of fitted values across predictors best fit the data. An advantage of the GLM approach is that an algebraic solution exists for the values  $\hat{\beta}$ , which minimizes the squared error. This is the *ordinary* least-squares solution:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Here  $T$  indicates the transpose operator. We generally conduct inference by calculating a t-statistic by dividing the  $\hat{\beta}$  by its standard errors and obtaining p-values using classical inference. This procedure can then be repeated at each voxel.

In practice, fMRI data are not independent. They exhibit autocorrelation which must be removed for valid single-subject inference. To do so, one estimates the autocorrelation in the residuals after model fitting then removes it by applying a filtering matrix  $W$  that counteracts the autocorrelation. What is known as the *generalized* least-squares solution incorporates this process, so that:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}$$

Note that the standard errors and degrees of freedom change as well due to the whitening process. Because the estimation of  $\mathbf{W}$  depends on  $\hat{\beta}$ , and vice versa, a one-step algebraic solution is not available. The parameters, then, are estimated using an iterative algorithm.

There are many ways of designing  $\mathbf{W}$ , ranging from estimates that make strong simplifying assumptions about the data's form, such as the one-parameter autoregressive AR(1) model, to empirical estimates which use many parameters. As with any model fitting procedure, a tradeoff exists between using few and many parameters. Many-parameter models generally produce close fits to the observed data. However, models with few parameters - if they are chosen carefully - can produce more accurate estimates of the underlying true function because they are less susceptible to fit random noise patterns in the data.

# Chapter 19 - Conditions and Contrasts

Once we have estimated the parameters of the GLM, we can turn our attention toward addressing scientific questions of interest. This can involve estimating the signal magnitudes in response to a single condition, as an average over multiple conditions, or as a difference in magnitude across conditions. Within the GLM framework using *contrasts*, or linear combinations of the  $\beta$  values, we can easily handle all these questions. For example, consider that we have two conditions, A and B, which reflect periods of auditory and visual stimulation, respectively. The contrast [A - B] estimates the activation difference for auditory - visual stimulation, which we can test by applying a linear contrast vector  $c$  (i.e. calculating  $c^T \beta$  where  $c^T = [1 \ -1]$ ).

In general, contrasts can estimate signal magnitudes in response to a single condition (i.e.  $c^T = [1 \ 0]$ ), as an average over multiple conditions ( $c^T = [1 \ 1]$ ), or, as mentioned above, as a difference in magnitude across conditions ( $c^T = [1 \ -1]$ ). We can estimate the true contrast values  $c^T \beta$  using  $c^T \hat{\beta}$ . We can then perform hypothesis testing in the usual manner, by testing the significance of the contrasts with a t-test.

There are a few general rules of thumb for working with contrasts. First, if you test a difference between conditions, then contrast weights should always sum to zero (i.e.  $(c^T = [1 \ -1])$ ). In contrast, if you test the average of one or more conditions against the implicit baseline, this need not hold (i.e.  $c^T = [1 \ 0]$  or  $c^T = [1 \ 1]$ ). Second, while scaling of weights affects magnitude (e.g.  $c^T = [1 \ -1]$  compared to  $c^T = [0.5 \ -0.5]$ ), it does not influence inference (i.e. t-values and p-values).

Most imaging statistics packages write not only a series of images to disk containing the betas for each condition throughout the brain but also another set of *contrast images* containing the values of  $c^T \hat{\beta}$  throughout the brain, with one image per contrast. A group analysis typically later uses these contrast images. In this setting one must be careful that the applied contrast weights are the same for all participants.

Finally, another kind of contrast, the F-contrast, can test the significance of multiple parameters simultaneously. For example, we could test a subset of parameters (e.g. do a set of nuisance covariates explain a significant amount of variation in the data?) or test two means at the same time. Now,  $c$  is a contrast matrix with a separate row for each parameter (or linear combination of parameters).

For example, suppose that simultaneously testing whether the two conditions (with parameters  $\beta_1$  and  $\beta_2$ , respectively) in the example above both equal zero interests us. We could then use the contrast matrix  $c^T$  which takes the form of a  $2 \times 2$  identity matrix. This gives us  $c^T \beta = [\beta_1 \ \beta_2]^T$ . We perform hypothesis testing in the usual manner by testing the significance of the contrasts using a F-test.

# Chapter 20 - Design Specification: Flexible Hemodynamics and Mis-modeling

The most challenging aspect of using the GLM is creating realistic task-related signal predictions to include as the columns of the design matrix  $\mathbf{X}$ . To build the model, researchers start with an *indicator* vector representing the neuronal activity for each condition sampled at the resolution of the fMRI experiment. This vector takes the value zero, except during the hypothesized neural activation periods in which it has a value of 1. Researchers then convolve each indicator vector with a canonical hemodynamic response function (HRF) to yield a predicted time course related to that event. These time courses form the columns of  $\mathbf{X}$ . See Figure 20.1 for an example with four conditions (labeled A-D).

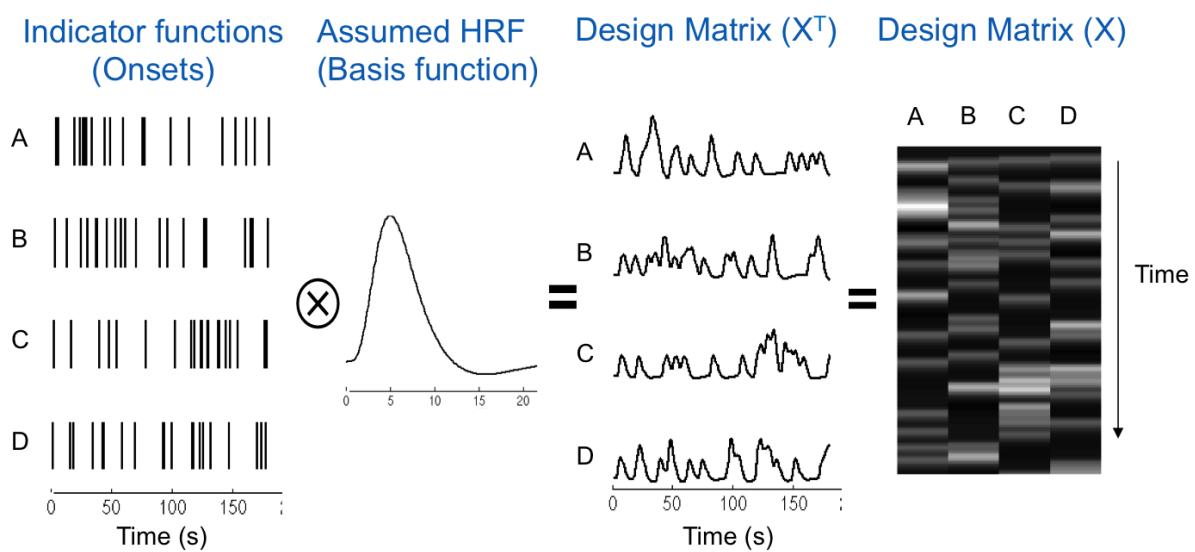


Figure 20.1. Indicator functions corresponding to the timing of four conditions are convolved with a canonical HRF to obtain 4 predicted responses, one for each condition, which is placed into the design matrix

If the canonical HRF fits the shape of the BOLD response to psychological events, this simple approach is greatly sensitive to detect differences between conditions. However, if the canonical HRF does not fit, at best there is a drop in power and, at worst, **false positives<sup>51</sup>** and **mis-interpretation of results<sup>52</sup>** The canonical HRF used in most standard software packages consists of two gamma

<sup>51</sup><http://www.ncbi.nlm.nih.gov/pubmed/17094118>

<sup>52</sup><http://www3.stat.sinica.edu.tw/sstest/j18n4/j18n410/j18n410.html>

functions linearly combined. However, empirical evidence has shown that assuming a fixed HRF is usually not appropriate as its shape is known to vary, both across the brain and across subjects.

As an alternative, researchers can convolve the same ‘neural’ indicator vector described above with multiple canonical waveforms then enter this data into multiple columns of  $\mathbf{X}$  for a single event type. We call these reference waveforms *basis functions*. The predictors for a specific event type, constructed using different basis functions, linearly combine to provide a better fit for the evoked BOLD responses. A basis set’s ability to capture hemodynamic variations depends on both the number and the shape of the reference waveforms that the researchers use. There is a fundamental tradeoff between power and flexibility to model variations. This stems from the fact that each parameter is estimated with error, so models that are too flexible risk modeling noise and thus tend to produce noisier parameter estimates.

One of the most flexible models, a finite impulse response (FIR) basis set, contains, for each modeled cognitive event-type, one free parameter for every time point following stimulation (Glover, 1999). This model makes minimal assumptions about the HRF’s shape because the  $\hat{\beta}$  values estimate the average response at each time point following an event’s onset. The FIR model, which is implemented in all the major software packages including AFNI, SPM, and FSL, is a preferred way to estimate and visualize the shape of BOLD responses. To guard against the possibility of over-fitting, one can use a smooth (or regularized) FIR model constrained to produce smooth response functions. Other choices of basis sets include those composed of principal components, cosine functions, radial basis functions, spectral basis sets, inverse logit functions, or other functions. For a critical evaluation of various basis sets, see [Lindquist & Wager<sup>53</sup>](#) and [Lindquist et al.<sup>54</sup>](#).

Figure 20.2 shows an example where three different models are used to fit the actual experimental data. The first model is a single canonical HRF. It clearly does a poor job of fitting the signal and will give rise to biased estimates of the amplitude. The second model is the canonical HRF plus its temporal and dispersion derivatives. This model does a nice job of fitting the rise in signal, though it doesn’t capture the undershoot. However, for most analysis this is less important. The third model uses the FIR approach. Clearly, the fitted response very accurately follows the actual signal. However, here the results may be ‘too good’ as it appears we may be modeling the noise as well as signal, and the results will not be generalizable.

<sup>53</sup><http://www.ncbi.nlm.nih.gov/pubmed/17094118>

<sup>54</sup><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3318970/>

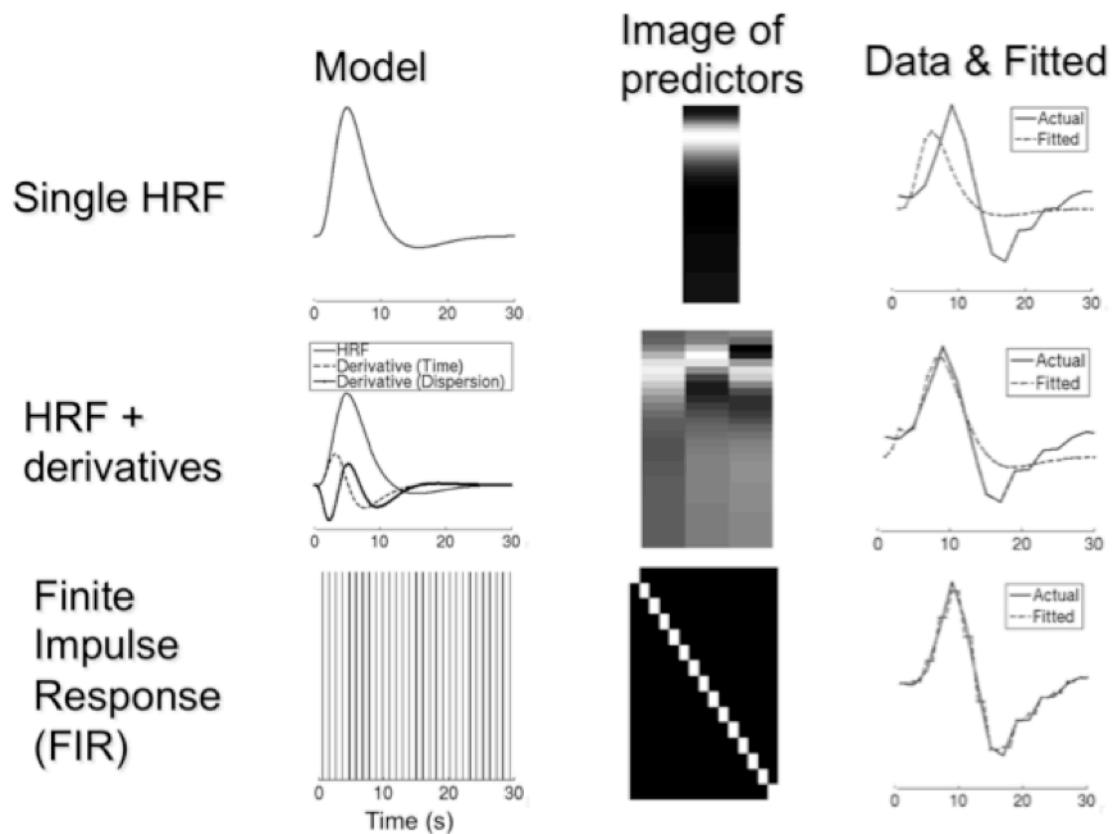


Figure 20.2. An example of the use of basis functions to fit actual data. Here we show models using a single canonical HRF, the canonical HRF and its derivatives, and FIR basis sets.

Though basis sets offer the major advantage of more accurate HRF modeling across subjects and across the brain, they also introduce additional technical difficulties which hinder more widespread usage. Calculating contrasts across conditions is not straightforward when there are multiple parameter estimates per condition, but leaving out some basis functions when calculating contrasts, as is typically done, is **not** generally advisable. An alternative is to calculate one contrast per basis function for each contrast of interest. Researchers can then do group analysis using repeated measures analyses at the second level rather than the usual one sample t-test. However, adding basis functions results in a cost in power, as does, in general, comparing more parameter estimates. As an alternative, some have suggested to instead use the norm of the coefficients or to simply re-create the HRF after estimation and use the resulting amplitude as the contrast of interest<sup>55</sup>.

<sup>55</sup><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3318970/>

# Chapter 21 - Design Specification: Dealing with Artifacts and Noise

After adding realistic predictions of task-related signals to the design matrix  $X$ , researchers typically include additional predictors to account for other sources of variation in the data. One particular class of predictors is *nuisance covariates*, which reduce noise and prevent signal changes related to drift, head movement, and physiological (e.g. respiration and cardiac pulsation) artifacts from influencing the contrast estimates. These are referred to as nuisance covariates because they are model factors associated with known sources of variability not directly related to the task or to the experimental hypothesis.

In fMRI, the signal typically drifts slowly over time, even in the absence of an explicit task, so that the most power lies in the lowest temporal frequencies. Therefore researchers often include covariates at this stage which implement high-pass filtering, or removal of signal frequencies below a specified cutoff. One often performs high-pass filtering in GLM analysis by adding covariates of no interest (e.g. either low-frequency cosines or polynomial functions). However, a researcher must take special care to ensure that fluctuations which the task design induces are not in the frequency range removed by the filter, as then one risks removing important signal of interest.

Figure 21.1 shows an example of a design matrix with both a task-related regressor and nuisance covariates in the form of low-frequency cosines. The figure illustrates the relative contribution of each portion of the design matrix to modeling the actual data (shown in blue). The red curve shows the predicted response with the low-frequency drift explained away, which corresponds to the signal of interest.

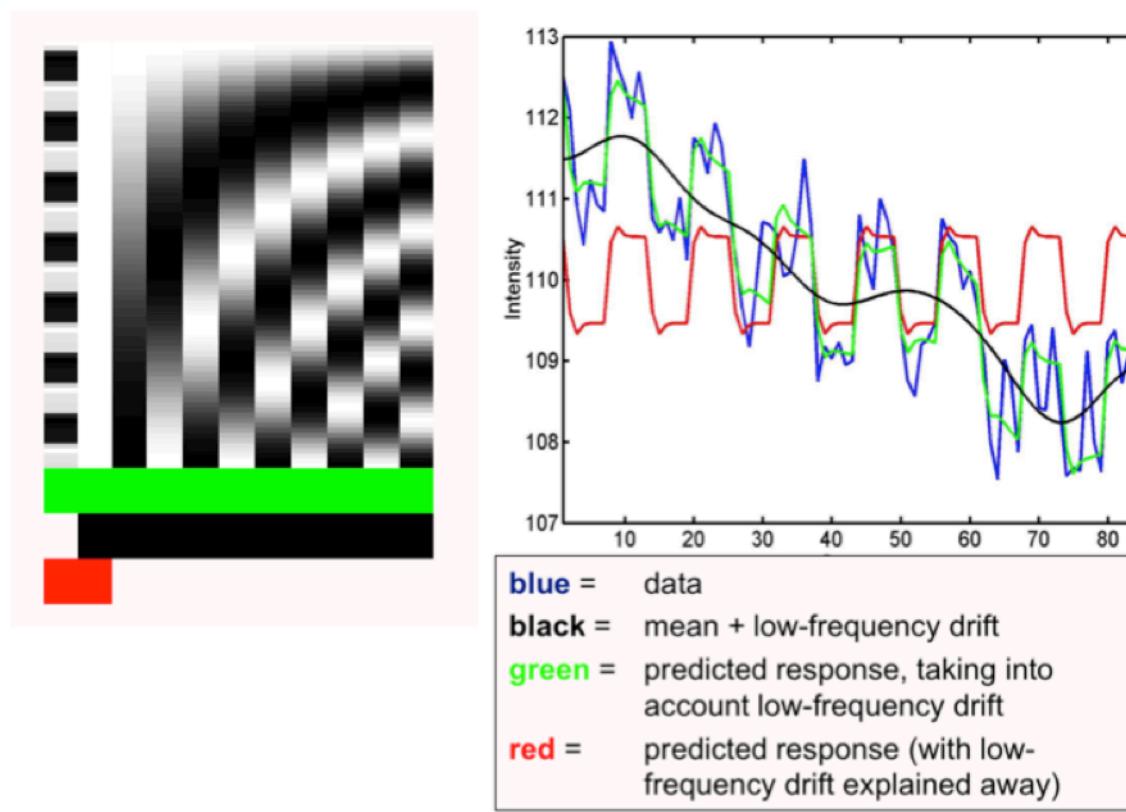


Figure 21.1. An example of a design matrix with both a task-related regressor and nuisance covariates in the form of low-frequency cosines. The panel to the right shows the relative contribution of the different columns of the design matrix.

Head movement presents one of the biggest challenges in fMRI data analysis and artifact correction. The pre-processing stage of the analysis includes basic motion correction, or image realignment. This procedure typically deals well with gross differences in position across the images. However, motion also induces complex artifacts due to the spin-history and due to changes introduced to the magnetic field by motion, effects which standard motion correction algorithms cannot remove.

There are two basic approaches to deal with the residual effects of head motion in a GLM analysis. The first approach is to include nuisance regressors in your design matrix in order to model the movement. This entails including the six regressors, comprised of three translations and three rotations, which were estimated during the pre-processing stage. Researchers often also include transformations of the six regressors: it is, for example, increasingly common to include the squares of the six regressors, their successive differences (related to the derivative), and the squared successive differences. Hence researchers, many times, include up to 24 additional movement-related covariates in the GLM design matrix. The second approach, known as scrubbing, refers to the practice of dropping images with high estimated movement. Essentially, this entails removing images which exhibit high degrees of motion from the time series and treating them as missing data.

Physiological noise, such as respiration and heart rate, can induce signal modulations in fMRI signal

that can lead to increased noise. Physiological signals give rise to periodic noise, which most fMRI experiments' current sampling rates alias into the task frequencies. For this reason, it is often difficult for researchers to remove these types of noise so they are typically left in the data, which gives rise to temporal autocorrelations. However, a number of ways of monitoring physiological artifacts (e.g. RETROICORR and RVHRCOR) and including them in your model exist. These techniques work in slightly different ways with regard to how they take factors, such as neuronal activation, respiration cycle, cardiac cycle, respiration volume, and heart rate, into consideration.

Finally, another type of artifact for which it is important to control is transient gradient artifacts. Once either PCA or an outlier detection approach identifies these, they can be modeled using one regressor per bad image. This is a type of scrubbing, which uses one degree of freedom to account for the variation due to that spike.

# Chapter 22 - Group Analysis

Though making inferences about single subjects is common in some areas (i.e. fMRI of early visual processes), most researchers' primary interest lies in generalizing their results to a population of unobserved individuals. Science is, after all, about generalizable knowledge. To achieve this goal requires group analysis with multiple subjects performing the same type of experiment.

Group analyses are useful as they can increase the experiment's overall sensitivity (as more data is available), allow you to determine whether observed effects are common and stable across, or between, groups, and allow your conclusions to be generalized to the entire population from which the subject was drawn.

Multi-subject fMRI data is hierarchical in nature, with lower-level observations nested within higher levels (i.e. subjects nested within groups). In order to effectively compare data for specific voxels across subjects, it is important that all subjects are normalized to a stereotaxic space prior to comparison; see the pre-processing chapter for more information.

A standard group analysis consists of specifying a second-level GLM model, this time using a set of contrast images corresponding to a single contrast (one per subject) as the response variable ( $y$ ). The simplest design matrix is an intercept term (all values constant) which performs a one-sample t-test to determine whether each voxel's contrast values are non-zero; see Figure 22.1. However, it is possible to specify designs that implement a two-sample t-test (e.g. regressors with 1 or -1), regression relating a continuous predictor's individual differences to contrast values, and ANOVA or other designs. This type of analysis has come to be known as a *random effects* analysis in neuroimaging literature. It is critical to avoid the common error associated with including more than one image per subject into the group analysis. If there is more than one, then the group analysis has repeated measures, so researchers must appropriately deal with correlated errors across repeated measures.

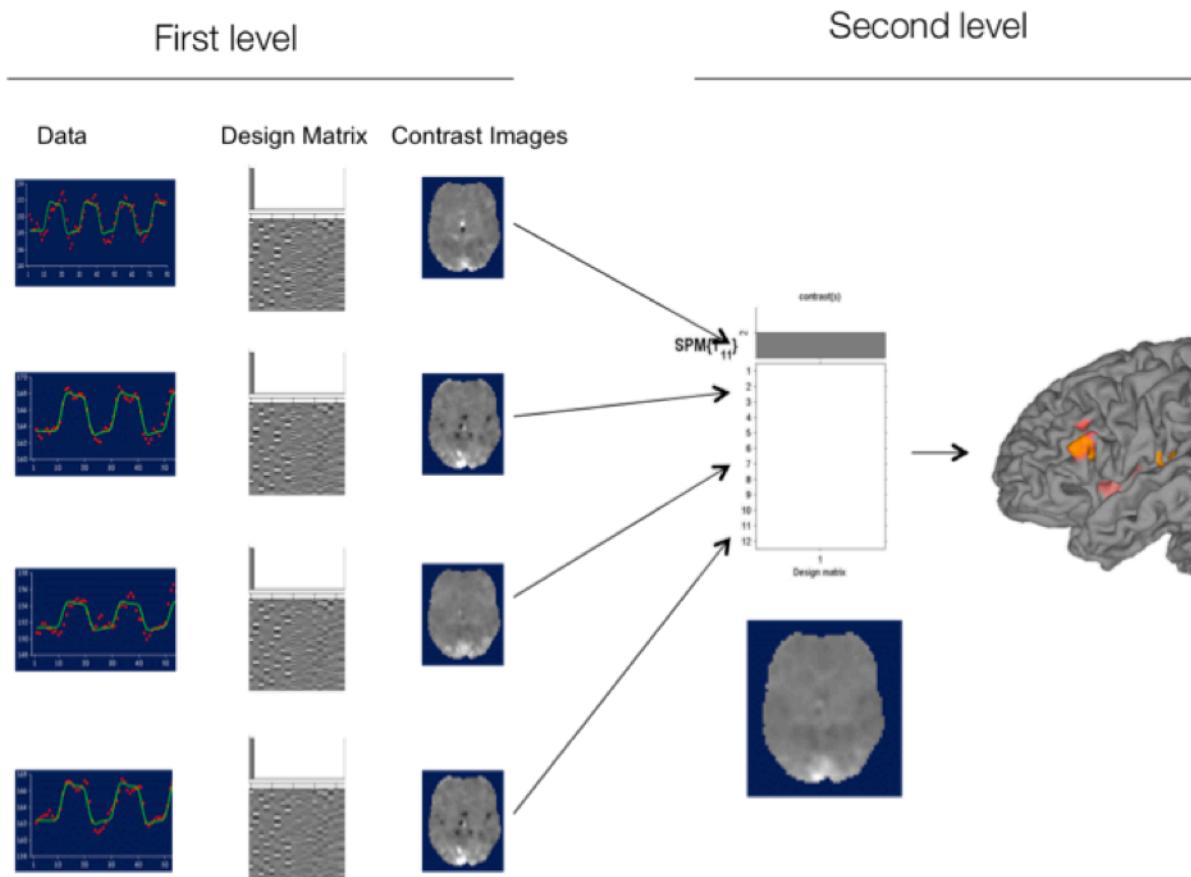


Figure 22.1. An illustration of the so-called ‘summary statistic’ approach towards group analysis. In a first level a GLM is fit separately to each subject and contrast images are created. In the second level, these contrasts are used to perform statistical tests (e.g., a one-sample t-test to determine whether the contrast is significantly different from 0).

Many early studies concatenated data from different participants together into a single “super subject” then analyzed the data using a single statistical model. We typically refer to this as a *fixed effects* analysis because it treats participants as fixed effects and assumes the only noise present in the signal is measurement error within subjects. This analysis is not appropriate for population inference because it does not account for individual differences. Some researchers have argued that fixed effects analysis allows researchers to make inferences about specific participants’ brains, but do not extend these inferences to a broader population. While this is technically true, inferences about particular individuals are seldom useful. In fact, such a lack of generalizability would be unacceptable in virtually any field, including neuroimaging studies.

*Mixed effects* analysis, so termed because it estimates multiple error sources, including measurement error within subjects and inter-individual differences between subjects, is a more valid analysis. The one-sample t-test on contrast estimates, described above, is actually a simplified version of a mixed-effects analysis which is valid, if the standard errors of contrast estimates are the same for all subjects. This is rarely true in practice, which has led to the development and increased popularity of full mixed-effects models. Such models incorporate both first-level (within-subjects) and second-

level (between-subjects) effects into a single model. This relaxes the assumption of homogenous standard errors by appropriately weighting, in proportion to their precision, the contribution of each subject's estimates to the group estimates. The idea is that the larger a subject's standard error, the less reliable the estimate is, so the less that subject should contribute to the group results. Fitting this type of model requires variance component estimation at each level. Typically, researchers estimate separate error variances for (1) the within-subject measurement error (and model misfitting) and (2) the inter-individual differences among subjects. Restricted maximum likelihood (ReML) is a popular type of variance component estimate based on the residuals. Since variance estimates and model fits ( $s$ ) are inter-dependent, researchers use iterative algorithms such as the expectation-maximization (EM) algorithm to estimate ReML variance components.

# Chapter 23 - Multiple Comparisons

We typically summarize the results of neuroimaging studies with a set of statistical maps in which we color-code voxels whose t-values or comparable statistics (e.g. z or F) exceed a specific statistical threshold for significance to describe brain activation. The implication is that the experimental task activated these voxels. Choosing the appropriate threshold is obviously crucial for determining whether or not voxels are ‘active’. Many fields consider the p-value of test statistics taking values below 0.05 as sufficient evidence to reject the null hypothesis with an acceptable false positive rate ( $\alpha$ ) of 0.05. However, brain imaging often simultaneously performs 100,000 hypothesis tests (one for each voxel). Here a voxel-wise  $\alpha$  of 0.05 implies that 5% of the voxels *on average* will show false positive results - we actually *expect* around 5,000 false positive results. Thus even if an experiment produces *no true activations*, there is still a good chance that the subsequent activation map will show a number of activated regions which will lead to erroneous conclusions.

Researchers’ standard method to deal with multiple comparisons is to simultaneously control the probability of obtaining a false positive for every statistical test (i.e. voxel) in the brain by adjusting the threshold. In neuroimaging, researchers have suggested a variety of different approaches for controlling the false positive rate. The fundamental difference between methods is whether they control for the family-wise error rate (FWER), which is the probability of obtaining any false positives in the brain, or for the false discovery rate (FDR), which is the proportion of false positives among all rejected tests.

To briefly illustrate the difference between FWER and FDR correction, assume that when we perform tests on 100,000 brain voxels using  $\alpha = .001$  uncorrected, we find 300 ‘significant’ voxels. Theoretically we would expect 100 (or 33%) of the significant ‘discoveries’ to be false positives, though it would be impossible to tell which. Because of this, we may have low confidence that the activated regions are actually true results, so it could be advantageous to set the threshold to limit the expected false positives to 5%, which is to control the FDR at the  $q = 0.05$  level. We could then argue that most of the results are likely true activations. However, we will still not be able to distinguish truly activated voxels from false positives. FWER provides a stronger method for controlling false positives. If we control the FWER at 5%, the threshold is set in such a way that only 5 out of 100 repeated experiments would result in one or more false positive voxels. Hence when controlling the FWER at 5%, we can be fairly certain that all the voxels we deem active are truly active. That said, the thresholds would typically be quite conservative, which leads to problems with false negatives or with truly active voxels incorrectly deemed inactive. In our example, perhaps only 50 out of the 200 truly active voxels will give rise to significant results. While we can be fairly confident that all 50 are true activations, we have still ‘lost’ 150 active voxels.

It is unfortunate that many published PET and fMRI studies do not use either of these corrections. Instead, they use arbitrary uncorrected thresholds (typically  $p < .001$ ). One reason the corrected thresholds are so high is because the typically small sample sizes mean power is extremely low.

This is extremely problematic when interpreting individual studies' conclusions since many of the activated regions are possibly false positives. Sometimes researchers impose an arbitrary 'extent threshold' for reporting, which they base on the number of contiguous activated voxels. However, this does not necessarily correct the problem, as imaging data are spatially smooth and corrected thresholds should be reported when possible.

Because achieving sufficient power is not possible, it makes sense to report results at an uncorrected threshold and use meta-analysis, or a comparable replication strategy, to identify consistent results. However, it is important to note that we should not strongly interpret uncorrected results from individual studies. Ideally, studies should report both results which have been corrected and results at a reasonable uncorrected threshold (e.g.  $p < .001$  and 10 contiguous voxels) for archival purposes.

## FWE correction

The simplest way of controlling the FWER is the classic Bonferroni correction, which divides the applied  $\alpha$  value by the total number of statistical tests performed (i.e. voxels). If spatial dependence is present in the data (which is almost always the case because of natural resolution and applied smoothing), this unnecessarily conservative correction leads to decreased power to detect truly active voxels. Gaussian Random Field Theory (RFT) is another, significantly more theoretically complicated, approach toward controlling the FWER. Though RFT is beyond the scope of this book, if the image is smooth and the number of subjects is high enough (around 20), it does provide control closer to the true false positive rate than Bonferroni correction.

RFT can also assess the probability that  $k$  contiguous voxels exceed the threshold under the null hypothesis and thus lead to 'cluster-level' correction. The probability that researchers will find a cluster of size  $k$  under the null hypothesis is specific to an initial uncorrected significance threshold. For example, it is more likely to obtain a cluster of  $k = 300$  at an initial threshold of  $p < 0.05$  than it is with the initial threshold  $p < 0.001$  because more voxels survive the more liberal threshold. A recent analyses by Woo et al.<sup>56</sup> has shown that a liberal initial threshold (i.e. higher than  $p < 0.001$ ) will inflate the number of false positives above the nominal level of 5%. Nichols and Hayasaka<sup>57</sup> conclude that while RFT is overly conservative at the voxel level, in small sample sizes it is liberal at the cluster level. Also keep in mind with cluster-level correction that inference is only valid for the whole cluster. It is therefore not possible to make inferences about single voxels within that cluster. Instead, the appropriate interpretation is that '**there is signal somewhere in the cluster**'<sup>58</sup>. For large clusters spanning multiple regions, it is impossible to precisely state which region presents activation. This is particularly problematic with liberal initial thresholds as we consider more voxels active and they form larger clusters. Thus performing cluster-level inference with a liberal initial threshold ultimately reduces the spatial resolution of fMRI.

The RFT approach assumes normal distribution of the error values and equal error variance across all predictors' values. As an alternative to RFT, nonparametric methods use the data themselves to find

<sup>56</sup><http://www.ncbi.nlm.nih.gov/pubmed/24412399>

<sup>57</sup><http://www.ncbi.nlm.nih.gov/pubmed/14599004>

<sup>58</sup><http://www.ncbi.nlm.nih.gov/pubmed/24412399>

the appropriate distribution. These methods can provide substantial improvements in power and validity, particularly with small sample sizes, so we regard them as the ‘gold standard’ in imaging analyses. Thus these tests can verify the validity of the less computationally expensive parametric approaches.

## FDR control

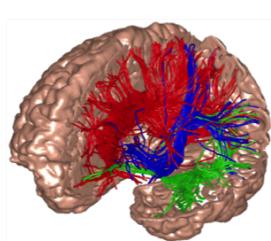
The false discovery rate (FDR) is a comparatively recent development in multiple comparison correction. While FWER is the probability of any false positives occurring in a family of tests, the FDR is the expected proportion of false positives among significant tests. In a brain map, this means that we expect approximately 95% of the voxels reported at  $q < .05$  FDR-corrected (note we use  $q$  instead of  $p$ ) to be true activations. The FDR controlling procedure is adaptive in the sense that the larger the signal, the lower the threshold. Also if all of the null hypotheses are true, then the FDR will be equivalent to the FWER: any procedure which controls the FWER will also control the FDR. For these reasons, any procedure which controls the FDR is necessarily less stringent than a FWER controlling procedure, which leads to increased power. A major advantage, since FDR controlling procedures work only on the p-values and not on the actual test statistics, is that they can be applied to any valid statistical test.

# Chapter 24 - Assessing Brain Connectivity

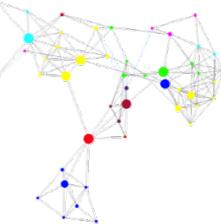
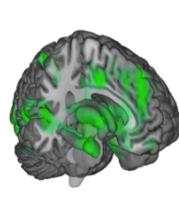
Most of the analysis techniques discussed so far focus on questions of functional specialization. Spatial resolution inherently limits the kinds of questions that fMRI can answer with regard to specialized functions. A different type of question asks about integration of cognitive functions across brain regions or about how neuronal populations work together. Here it is necessary to study multiple regions simultaneously and to investigate their relationships. There are many ways of extracting brain connectivity data measures; research literature is now replete with a growing variety of possibilities. The appropriateness of each possibility depends upon a number of factors which include the type of conclusions one wishes to make, the assumptions one will make, the level of the analysis, the modality used to obtain the data, and the number of brain regions included in the analysis.

The term ‘connectivity’ is an umbrella term which refers to a number of related aspects of brain organization. In neuroimaging literature, it is common to distinguish between anatomical, functional, and effective connectivity. For an illustration see Fig. 24.1. Anatomical connectivity deals with the description of how different brain regions physically connect; it can be approached using techniques such as diffusion tensor imaging. In contrast, functional connectivity and effective connectivity each study the functional relationships between different brain regions. These two, together with metrics for characterizing networks, are the focus of this chapter.

Structural connectivity

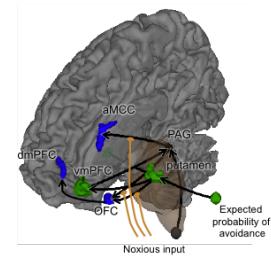


Functional connectivity



*Wager et al. 2015  
Graphical model*

Effective connectivity



*Roy et al. 2014 DCM*

## Functional connectivity

Functional Connectivity is the undirected association between two or more fMRI time series; it seeks to make statements about relationship structures among brain regions. Methods used to access

functional connectivity usually do not make any *a priori* assumptions about the underlying biology and tend to be data-driven. The simplest approaches simply compare the bivariate correlations between regions of interest or between a seed region and all other brain voxels. Researchers have also applied inverse covariance estimation methods: they use the fact that in multivariate normal data, conditional independence between regions corresponds to zero entries in the inverse covariance matrix. This allows investigation of multiple regions simultaneously and controls the effects of other regions.

Finally, it is common to use various multivariate decomposition methods to identify brain activation patterns without making *a priori* assumptions about its form. These methods include principal components analysis (PCA) and independent components analysis (ICA). Here one attempts to decompose the time-by-voxel data matrix,  $\{\$\}\{\backslash bf Y\}\{\$\}$ , into a set of spatial and temporal components according to some criteria. PCA allows one to determine the spatial patterns that account for the greatest amount of variability in a time series of images, while ICA requires the components to be independent rather than orthogonal. PCA and ICA have been especially useful for analyzing resting-state data in which subjects do not perform an explicit task. This is an area of intense research where standard methods for localizing brain activation are not applicable due to the lack of stimulus.

Methods for assessing functional connectivity can be applied at different levels of analysis, with different interpretations at each level. Connectivity across time series data can reveal networks that are dynamically co-activated over time, and is perhaps closest to the concept of communication among regions; though to be clear it does not conclusively demonstrate that. Connectivity across [single-trial response estimates<sup>59</sup>](#) can be used to identify coherent networks of task-related activations. Whereas these levels are only accessible to fMRI and EEG/MEG, which provide relatively finely sampled time series data, other levels of analysis may be examined using PET as well. Connectivity across subjects can reveal patterns of coherent individual differences, which may result from communication among regions but also from differences in strategy or other genetically determined or learned differences among individuals. Finally, connectivity across studies can reveal tendencies for studies to co-activate within sets of regions, which may be influenced by any of the factors mentioned above, and also differences among tasks or other study-level variables.

While most literature has focused on stationary correlations that are constant across time, there is increased interest in *time-varying correlations* that provide expanded measures of correlation changes across time and can estimate time-varying graph or network structures. Researchers believe changes in connectivity across time can provide insight into fundamental properties of brain networks.

## Effective connectivity

Effective Connectivity is one brain region's directed influence on other brain regions• physiological activity, and claims to make statements about causal effects among tasks and regions. Usually

<sup>59</sup><http://www.ncbi.nlm.nih.gov/pubmed/15488425>

methods which assess effective connectivity make anatomically motivated assumptions and restrict inference to networks comprised of pre-selected interest regions.

Effective connectivity analyses are inherently model-dependent. Typically, researchers specify a small set of regions and connections *a priori* then use tests to compare a small number of alternative models' fit and to assess individual connections' statistical significance. Directional specificity in connections means researchers typically consider that the models imply causal relationships. Because of the myriad possible models, anatomy and theory must motivate the choice of which regions and connections to include. Hence most effective connectivity depends upon a neuroanatomical model that describes connections between areas and a mathematical model that describes how they are connected.

Popular methods to assess effective connectivity include structural equation modeling (SEM)<sup>60</sup>, Granger causality<sup>61</sup>, and dynamic causal modeling (DCM)<sup>62</sup>.

In SEM the emphasis lies on explaining the variance-covariance structure of the data. It comprises a set of pre-defined regions and directed connections between them. Further, path coefficients are defined for each link representing the expected change in activity of one region given a unit change in the region influencing it. The path coefficient indicates the average influence across the time interval measured.

DCM is an attempt to move the analysis to the neuronal level. It includes a neuronal network model and links the observed fMRI signal to its underlying generative model via a model of neurovascular coupling. DCM explicitly specifies the nodes and the connections between nodes, and can include psychological moderator variables which affect connections or nodes. This explicit hypothesis formulation is one of DCM's strengths because it forces the researcher to clearly define hypothetical models of brain function. After the researcher has specified and estimated a set of candidate models, DCM uses Bayesian model selection to choose the model which best explains the observed data.

In contrast, Granger causality does not rely on *a priori* specification of a structural model, but rather quantifies the usefulness of past values from one brain region in predicting current values in another. Granger causality provides information about the temporal precedence of relationships between two regions, but is a bit of a misnomer because it does not actually provide information about causality in the classical sense.

While effective connectivity methods have become increasingly popular, it is important to keep in mind that the conclusions about direct influences and causality obtained using these models are only as good as the specified models. Any misspecification of the underlying model will almost certainly lead to erroneous conclusions. In particular, the exclusion of important lurking variables (brain regions involved in the network but not included in the model) can completely change the fit of the model and thereby affect both the direction and strength of the connections. Great care always needs to be taken when interpreting the results of these methods.

<sup>60</sup><http://onlinelibrary.wiley.com/doi/10.1002/hbm.460020104/abstract>

<sup>61</sup><http://www.ncbi.nlm.nih.gov/pubmed/15734358>

<sup>62</sup><http://www.ncbi.nlm.nih.gov/pubmed/12948688>

## Network analysis

In many situations, we seek to create networks which contain numerous non-overlapping brain regions. We can represent these networks with graphs, which are mathematical structures that can model pair-wise relationships between variables. Their makeup is sets of nodes (or vertices)  $V$  and corresponding links (or edges)  $E$  that connect pairs of vertices. A graph  $G = (V, E)$  can either be undirected or directed with respect to how the edges connect to one another.

As the number of regions included in the analysis grows, it can become difficult to make sense of the vast amounts of data. Network analysis attempts to characterize these networks by using a few meaningful summary measures. Researchers use large-scale connectivity matrices to estimate higher-order *graph theoretic properties* of the networks as a whole, which they can then relate to outcomes. There is currently a proliferation of such measures, including 'small worldness', path length, 'betweenness-centrality', 'rich club' indices, and degree distribution metrics. These describe, in various ways, organizational properties concerned with how the brain voxels (or regions) relate to one another. The idea is that network topology comparisons between groups of subjects may help reveal connectivity abnormalities related to neurological or psychiatric disorders.

## **Part 4: Predictive Mapping**

# Chapter 25 - Multivariate brain analysis: From maps to models

Much of the excitement about neuroimaging has been predicated on the promise that it will revolutionize our understanding of how the brain enables cognitive performance, emotions, and other aspects of human mental life. It has the capability to shed light on how mental representations are organized in the brain, and how systems-level interactions map onto cognition, emotion, and behavior. It also has the potential to help understand brain and psychological health, as well as various brain-based disorders, including mental health, substance abuse, neurological disorders, stress and bio-behavioral health, brain-body relationships, and more. The potential applications in translational settings, for understanding human health and performance, are vast. They include (a) medical screening and diagnosis of psychopathology, neurological, and cardiovascular disease; (b) optimizing performance in educational, professional, and military settings; (c) drug development; and (d) the development and targeting of brain-based interventions.

However, there is an increasing recognition that to achieve these translational goals, we need to take different approaches towards analyzing fMRI data. Most early work focused on brain mapping, or localizing various mental and behavioral events to specific regions or systems in the brain. The shortcomings of this approach are becoming increasingly recognized, and include:

1. Massive multiple testing, which either dramatically increases false positive rates or dramatically reduces power, unless very large samples are used.
2. Difficulties in estimating effect size, i.e., whether brain effects are meaningfully large.
3. The inability to model contributions from multiple brain regions in representing a mental state or other outcome.

Two main goals of translational neuroscience are to (a) understand the brain bases of mental states and behaviors, and (b) use these measures to predict outcomes of interest in the real world.

One type of prediction is diagnosing or ‘decoding’ mental states, including what a person is thinking or feeling. This is obviously useful when one cannot measure what someone is thinking or feeling in other ways. For example: Do babies feel pain? Do animals? If you are unconscious, does your brain register pain signals that can adversely affect your recovery?

Another example is ‘neuro-prosthesis’, or brain-based decoding of movement intentions in those unable to move. Measured brain signals are transformed into robot-assisted movements, so that individuals who are paralyzed can control a mouse cursor, type email, and move a wheelchair with their minds.

Brain imaging has even been used to ‘decode’ the semantic content of dreams.

However, even if one can measure a mental state and prediction is not a goal per se, there are still advantages of brain-based predictive models. The ability to use brain measures to predict an outcome, such as which type of object a person is thinking about, is a gateway to understanding how semantic categories are *represented* in brain circuits. Only if brain measures are predictive are they really informative about the representational basis of mental constructs. Otherwise, they may well be superfluous.

Likewise, one might be able to measure pain by asking a person how they feel - but even if their response is completely trustworthy, pain reports do not tell us much about the underlying mechanisms. If you report lower pain than I do in response to the same stimulus (say, eating a hot chili pepper), is it because your pain sensory systems are registering less? Or maybe it is because you are simply being tough, and unwilling to admit how much it hurts. Maybe you have convinced yourself that it doesn't matter and pay less attention to it, or you feel the pain and just don't care. Brain imaging can help identify which systems are involved, and which neural (or glial) *ingredients* differ between us. If we can establish brain activity patterns that can predict pain - i.e., track how much pain a person is in when we know their reports are reliable - we can use this to identify meaningful differences in 'pain processing'.

## From univariate mapping to multivariate brain models

The vast majority of studies thus far have been *brain mapping* studies, whose goal is to identify which brain areas encode a particular psychological condition or other outcome. They have focused on brain organization in diverse ways, from whether different regions encode different types of memory to whether cognitive and emotional control systems are similar. At their core, the statistical procedures used to develop such maps test whether there is a non-zero effect of a particular psychological manipulation or observed behavior on one or more brain voxels or regions. As shown in Figure 25.1, psychological states are treated as predictors, and activity in each brain voxel or region is treated as the outcome. Each voxel is tested individually using separate models. The psychological conditions are the regressors ( $X$ ) and the outcome is the brain data from one voxel ( $Y$ ). This is the so-called 'mass univariate' approach discussed in previous chapters, which has to date been the dominant approach towards analyzing fMRI data.

## Standard mass univariate mapping with multiple regression

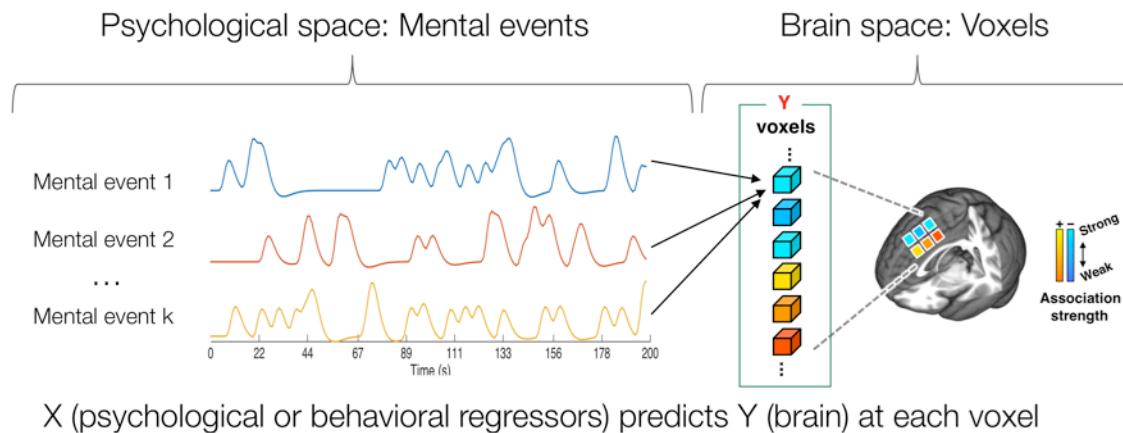


Figure 25.1. An illustration of the mass univariate approach towards neuroimaging.

### **Strengths and weaknesses of the mass univariate approach**

A strength of the *mass univariate* approach is that considering a single voxel, without taking into consideration the influences of others, provides a stable and interpretable picture of whether that voxel responds to the task or not. The same answer is obtained no matter what else is happening in the rest of the brain. Another strength is that it is possible to assess the effects of one task-related state on the voxel while controlling for other potentially correlated task-related states, often using a multiple regression framework. However, a major weakness of the mass univariate approach is that by ignoring other brain regions, it can miss regions that are only significant when controlling for other variables, and falsely identify regions whose correlations with the outcome are spurious, and actually driven by other, correlated brain regions. In addition, it is possible that many psychological states and other outcomes are related to activity in *brain networks* that span multiple regions, and cannot be captured by measuring activity in any single region in isolation.

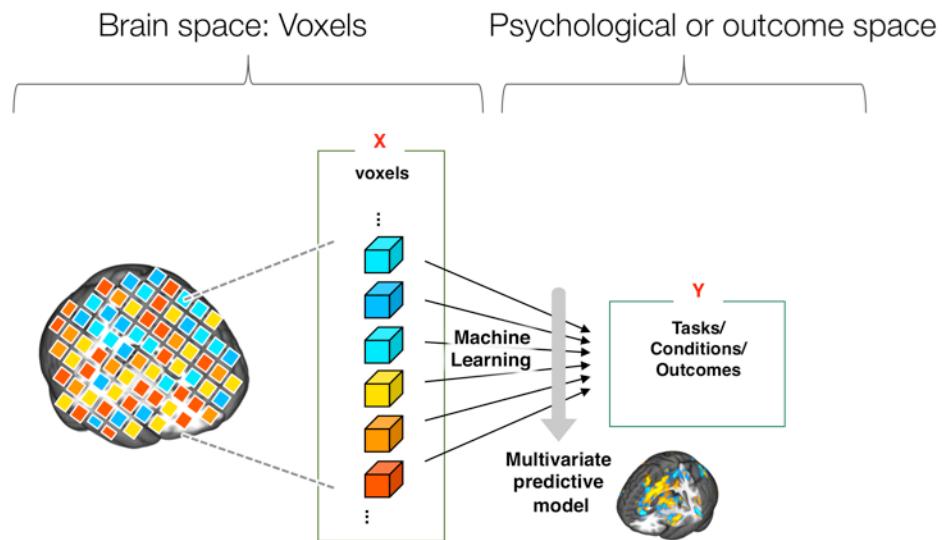
Making some analogies to other areas outside brain imaging makes some of these issues intuitive. One example is cardiovascular disease, which is not related to any single risk factor, but rather a complex web of related factors. Genetics certainly plays a role, as does diet, exercise, psychological stress, social support, body fat, drug use, and environmental risks. No epidemiologist in their right mind would analyze the correlation between any one of these factors and disease incidence without controlling for [1]others. This would represent a failure to acknowledge the effects of *indirect influences*. Appropriate controls and models are part of what makes epidemiology such a difficult science, but it is also the only way to estimate which variables are important to an outcome. The results matter, as behavioral and policy changes are often guided by inferences about which variables

are causally related to disease. Though this is extremely challenging to get right, fortunately, an easier question is to create a risk profile across a set of inter-correlated variables. This risk profile can be highly stable, predictive, and useful, even if exactly which variables are causally vs. indirectly related to the outcome is questionable. Diagnosing risk is an easier problem than knowing exactly where to intervene.

## **Predictive mapping and its place in the family of multivariate techniques**

In both translational work and understanding the representational basis of mental events, we want to use brain activity to predict an outcome or otherwise develop a ‘risk profile’ for the mental state or outcome. This suggests that we should treat the brain data as a set of predictors (for starters, activity in a set of voxels), and the outcome we would like to predict or explain as the response variable. This reverses the standard regression equation used in neuroimaging. Now, voxels are the predictors ( $X$ ) and the behavior is the dependent variable ( $Y$ ). This allows one to develop a model of how voxels jointly contribute to predicting the outcome. We refer to this as the ‘predictive mapping’ approach, shown in Figure 25.2.

Predictive mapping, modeling an outcome using multivariate fMRI data



X (a set of brain variables) predicts Y (a single outcome variable of interest)

Figure 25.2. An illustration of the predictive mapping approach towards neuroimaging.

**A place in the family.** When neuroimaging papers refer to ‘multivariate analysis’, they are broadly referring to models in which activity in multiple voxels (or other spatial units, such as regions)

are included in the same model. *Dimension reduction* approaches such as Principal Components Analysis and Independent Components Analysis, discussed earlier, are examples in which there are *no* mental state variables or other outcomes. The goal is to find a reduced set of variables that explain a large proportion of the variance in the original multivariable dataset, given specific constraints. Similarly, *cluster analysis* includes a large family of algorithms designed to identify sets of observations that behave similarly on a set of variables, and differently from other sets, without reference to other algorithms.

*Predictive mapping* is another, complementary kind of multivariate analysis that relates a set of brain variables to a single outcome variable. It can also be extended to control for other, correlated psychological variables and relate brain activity to latent psychological variables measured by patterns across psychological measures.

**Prediction and MVPA.** Because predictive maps involve identifying patterns across voxels that jointly comprise a composite ‘risk score’ for a mental state or outcome, they are often referred to in the neuroimaging literature as Multi-voxel Pattern Analysis (MVPA). So why don’t we just call it MVPA? Well, first of all, because predictive maps are not required to operate on voxels. They can operate on networks, regions, or other distributed components. In addition, MVPA is associated with particular usages that overlap only partially with the goals of predictive mapping. It can be used to predict an outcome, or to understand the local similarity structure among patterns of brain responses to different tasks. Thus, ‘MVPA’ is agnostic with respect to usage, whereas ‘predictive mapping’ implies a specific usage in which the brain basis of a psychological variable or outcome is identified and tested.

**Examples of algorithms for different types of outcomes.** *Predictive mapping* and MVPA can be used to predict categorical or continuous outcomes. If the outcome is *categorical* (e.g., is one thinking of an animal or a vegetable?), the problem is a *classification* problem, and solutions can be estimated with *classifier systems*, including support vector machines, logistic regression, discriminant analysis, multinomial regression, decision trees, and random forests, to name a few. However, predictive mapping is not limited to classification problems, and can also be used to predict continuous outcomes (e.g., how bad will your symptoms be a year from now?) In this case, the problem is a *regression* problem, and solutions can be estimated using variants of multiple regression, including ridge regression, LASSO regression, elastic nets, regression trees, support vector regression, or just plain old regression. Both types of outcomes can also be predicted using *neural networks* - a flexible framework that encompasses a large family of models - with different types of transfer functions across layers of the network.

While many predictive mapping analyses predict one outcome at a time, some types of algorithms identify patterns across multiple outcomes, and can identify multiple maps associated with different combinations of outcomes. Two common algorithms that do this are canonical correlation and PLS. Neural networks can also be used for this purpose. Traditionally, these techniques have been relatively infrequently used. This is in part because they are in some ways *too* flexible and can (like all of the algorithms above) produce estimates that reflect noise. This flexibility also reduces interpretability. Instead of developing a brain map whose activity predicts one outcome of interest, the map might often predict some combination of outcomes that is not intrinsically meaningful.

However, as we discuss below, they are an important part of the toolbox one can use to build predictive models.

## The space of approaches available for building models

### Brain space

	Univariate	Multivariate
Psychological space		
Univariate	'Mass univariate' mapping with simple regression design	Predictive mapping with MVPA
Multivariate	Brain-outcome correlations	Most local multivariate 'searchlight' analyses
Multivariate	'Mass univariate' mapping with multiple regression design	Effective connectivity: Mediation and structural models
Multivariate	Encoding-decoding models	Fully multivariate models; canonical correlation and partial least squares

Figure 25.3. An example of the space of approaches available for building models.

## ***Benefits of predictive analysis***

Predictive mapping strategies have some important intrinsic benefits over standard brain mapping techniques. These include:

1. *The ability to obtain accurate encoding models for mental constructs.* Predictive maps can average signal over many regions, reducing noise via averaging and capturing outcome-related signals that are distributed in space. This can dramatically increase power.
2. *Avoiding multiple comparisons.* Development and testing of predictive accuracy does not require correction for multiple comparisons, which can dramatically increase power.
3. *Sensitivity to brain information at multiple spatial scales.* Patterns of activity can be sensitive to predictive information at different spatial scales, including (a) relative activity levels and covariance across voxels that are not captured in univariate maps; and (b) neural topography at map-, column-, and possibly microcircuit-levels of organization.

4. *More realistic estimates of brain contributions to outcomes.* Predictive maps can provide information about which voxels contribute to an outcome controlling for other voxels. They are extendable to models of how brain regions work together to directly and indirectly influence an outcome.
5. *Replication across studies.* Replicating a univariate map usually requires showing that there exist significant effects in the same, or nearby, voxels. Such replications are unsatisfying because they are usually dramatically underpowered due to the massive multiple testing framework and they do not precisely test either the magnitude or the location of the effects. In many cases, activation of nearby (but different) voxels is interpreted as an approximate ‘replication’, but there is no rule for how close is sufficient. Predictive maps provide a precise specification of relative voxel weights that includes predictions about the magnitude of expected brain activity, which allows for direct replication.
6. *Generalization across studies.* Predictive maps are ‘research products’ that can be tested across studies for generalizability to new populations, experimental settings, and psychological or behavioral constructs.

These properties do not depend on the algorithm used to develop the predictive model or map, but they do depend on the spatial scale and scope of the data used to build the model and the level of analysis chosen. We examine some of these benefits in more detail in the next chapter.

## **A role for machine learning**

**Machine learning and statistical learning.** One of the issues with multivariate models is determining how to deal with the potentially large number of correlated brain variables. Typically, there are many more potential voxels to include - 100,000 is a typical number - than observations. In such cases, the voxel weights obtained in the regression or classification of the outcome are highly unstable at best.

One solution has been ‘searchlight analysis’, which uses brain variables as predictors but restricts the voxels in the model to a small number in a local area, thus reducing the dimensionality of the problem space. This analysis is repeated thousands of times across the brain. This can be very useful, but incurs drawbacks that undercut some of the main benefits of predictive analysis.

Another solution has come to be known as the ‘encoding/decoding’ model. It is predicated on the idea that any brain map, even a univariate map, can be used to predict outcomes by averaging the predictions made by each individual voxel. Thus, the ‘encoding’ model is a separate, univariate regression of each voxel’s activity onto psychological or behavioral regressors. Once the model has been estimated across the brain, the individual voxel models are applied to test data and the predictions averaged into one final prediction. This type of model ignores the relationships among brain voxels, but avoids the difficult job of modeling their associations in the process of predicting the outcome.

Apart from these solutions, the fields of *machine learning* and *statistical learning* have developed a wide array of techniques for (a) optimizing the prediction of outcomes given large sets of potential

variables, and (b) the stable estimation of large numbers of inter-variable relationships with limited data. These fields are sometimes referred to as *data mining*, *pattern recognition*, and *informatics*. These tools are widely used in practical applications in business, including:

1. Scheduling airline flights.
2. Organizing large searchable text databases.
3. Analysis of preferences for products (e.g., movies).
4. Targeting advertisements and information to consumers.
5. Control of energy usage, oil distribution, and other resources.
6. Expert systems in the intelligence and defense industry.
7. Processing and analysis of satellite images, cellphone usage data, and other complex datasets.
8. Novel applications such as ‘self-driving cars’.

Machine learning is really continuous with the field of statistics and model building in the behavioral and economic sciences, and there are not precise, hard-and-fast boundaries between which models should be considered ‘machine learning’ and which should not. However, there are several families of techniques that are associated with the rise of machine learning that are broadly useful.

One type of technique is the use of *penalization* or *shrinkage*, the concept of introducing small amounts of bias in model estimates in order to stabilize their estimates when there are large numbers of variables. Another technique is *dimension reduction* and *feature selection*, two families of procedures used to simplify and re-represent a set of variables so they are related to the outcome in simpler and more transparent ways. A third technique is *space expansion* and *kernel* techniques, which re-represent a set of variables beyond the original variable space - sometimes to infinite-dimensional spaces - in order to find a space in which the features are robustly and simply mapped to the outcome. Kernels can also be used to make models flexible in important ways and inflexible in ways that are implausible given the nature of the data at hand. A fourth technique is *reinforcement learning* and other types of gradual, *error-driven learning* processes that iteratively update model parameters instead of attempting to find closed-form mathematical solutions, as with many classical statistical models.

Finally, a fifth machine learning technique relates to making inferences about how accurately a model generalizes to new observations or ‘out-of-sample’ data. Testing a model prospectively on brand-new datasets (without changing the model or its parameter estimates) is the gold-standard way of assessing replicability and generalization. The more dissimilar the test dataset is in superficial respects, the more powerful the inference about generalization. However, such *data-splitting* strategies are inefficient, and can reduce the quality of models because one is always training a model (i.e., estimating parameters) on less data than one has at hand, and more data is, as a rule, better. To resolve this dilemma, *cross-validation* has become widely used as a strategy for both training (developing) a model and testing its accuracy in prediction for out-of-sample data.

Cross-validation begins by splitting the data into partitions. If one has a dataset consisting of 100 participants, for example, one might divide it into 4 partitions of 25 participants each. The model is

then trained 4 times, in *4 folds*. In each fold, one forth of the data (one partition) is left out as the test set, and the remaining three forth of the data is used as the training dataset. This is referred to as 4-fold cross validation, and in this case, each participant is left out of the training dataset and used to test the model exactly once. The final model can be either the average model parameter estimates for the 4 folds, or a model trained across all folds for the purpose of estimating the parameters. The model's accuracy is the average accuracy across all of the observations when used in the test set. There are many versions of cross-validation, and some pitfalls that can induce bias in accuracy. Overall, however, it provides a robust way of testing accuracy and generalization for arbitrarily complex models, even when there are many more variables and potential features than observations in the dataset.

## ***From maps to models***

Each of the algorithms discussed above can be used ‘out of the box’ to develop predictive brain maps, and this is the most common approach so far in the neuroimaging community. However, they also constitute a toolbox that can be used to build predictive models in flexible and creative ways.

As mentioned above, there is a continuum of models and approaches across many computational fields, and machine learning and statistics now offer a rich variety of tools for building models. The best models will both predict outcomes, and thus have demonstrable utility, and represent the processes that contribute in ways that add to our understanding of the process.

Models can be simple specifications of several brain regions that jointly and independently contribute to an outcome. They may include specific multi-voxel patterns within local regions, with parameters that govern how these regions relate to one another and how they directly and indirectly contribute to the outcome(s). And they may include intermediate features that are transformations of the original variables, which are designed to capture brain representations that map onto the outcome in transparent ways.

*Hierarchical models* include representations at multiple levels or stages of transformation, often structured in multiple *layers* of neural networks or belief networks. For example, one layer of a network may represent the original input variables - say visual input at different points on the retina. Another layer may represent more complex features composed of multiple lower-levels - e.g., bars of light or even object shapes. These intermediate representations may be mapped from earlier inputs in ways that make them *scale and transformation invariant*, which implies they respond to the object shape even if its size and position on the retina vary. Additional layers may represent semantic features associated with objects and even action tendencies and motivational properties. The output layer of such *deep networks* usually reflects an action to be performed or semantic category, depending on the goal of the model.

The brain is able to accomplish amazing feats in areas like object representation, inference and generalization, and generative recombination of elements into new emergent constructs. One of the main ways may be by using multi-layered, structured networks with rich and varied types of representations. In fact, a large class of such networks is called ‘neural networks’ because they were

inspired by ideas about how the brain works. For a number of examples and models, see the excellent book by O'Reilly and Munakata<sup>63</sup>.

Though neural networks are one class of models, there are many kinds of hierarchical models that are formally related to how neural networks work, but with variations related to the meaning of variables (or 'nodes'), the rules for transfer of information across 'nodes', and the way the models are trained. Such models can include hierarchical models in classical (frequentist) statistics, hierarchical Bayesian models, hidden Markov models and partially observable Markov decision processes, belief nets, and variations thereof.

## Basic criteria for a good model

The number of options and variations are bewildering, but the principles involved in building a useful model are simple. Here are some priorities that we feel should be kept in mind for many types of models that relate brain imaging to outcomes:

1. Models should make quantitative, testable predictions about outcomes and be falsifiable.
2. Models should predict and explain outcomes outside the brain itself, e.g., thoughts, behaviors, percepts, and clinical status.
3. It should be easy to test the model prospectively, by 'feeding in' new test datasets, and methods should be made available for this.
4. Models should be simple to interpret. They are simplified explanations of a complex world, and a goal is to increase understanding as well as prediction. In addition, understanding how a model works can help us understand when its predictions will hold and when they will break down.
5. Models should be as simple as they can be and only as complex as they need to be. Given models that are equally predictive, the simpler model (either in its parameter space or interpretability) is preferred.
6. Accuracy counts: models that make accurate predictions are more likely to be more 'correct' than less accurate models.
7. Models should not be over-interpreted as complete explanations of a phenomenon. There is no guarantee that either the simplest model or the most accurate model is the most 'correct' model. To be correct, a model must be better than all other possible models, and must also capture the true structure of causal relationships among modeled processes and the outcome. Models can still be useful without being over-interpreted.

---

<sup>63</sup><https://grey.colorado.edu/CompCogNeuro/index.php/CECN>

# **Chapter 26 - Advantages of MVPA from a neuroscientific perspective**

As mentioned in the previous chapter, MVPA and machine learning have a great deal of promise as a framework for predicting outcomes, tracking brain processes relevant for outcomes, and other things relevant for translational neuroscience. They are also likely to be essential in the future for building more realistic, and thus accurate, models of the neural bases of psychological process. This is important not only for prediction, but for understanding the brain and mind as well. Brain maps or models that are predictive of an outcome are gateways to understanding its representation.

In this chapter, we explore two of the major benefits of predictive analysis and MPVA in more detail. We first describe two major analysis choices that one must make, independent of the actual algorithm or model. These are determining (a) the spatial scope of the data subjected to MVPA and (b) whether the model is allowed to vary for each individual (within-participant or idiographic) or be developed to capture consistencies across individuals (between-participant or group level).

Then, we explore benefits related to (a) sensitivity to neural topography, (b) sensitivity to distributed representations, and (c) benefits in testing generalizability across individuals and studies.

## **MPVA analysis choices: Spatial scope and flexibility**

**Spatial scope.** One fundamental choice in any MVPA analysis concerns the spatial scope of the analysis. How many voxels or regions should be included, and how broadly should they be distributed across the brain? Choices include everything from using a small local region of the brain to voxels across the entire brain.

Many of the first uses of MPVA were applied within individual brain regions, particularly in the visual system, to ‘decode’ object features based on local topography. An advantage of this region of interest (ROI) based procedure is that one can compare ROIs to test which regions contain information about (are predictive of) a particular psychological process, like object feature representation. However, in the absence of high-quality definition of specific functional ROIs, this approach is limited.

An extension of this approach was developed that combined many of the benefits of ROI-based testing with the ‘brain mapping’ approach of estimating local effects across the brain. This came to be known as the ‘searchlight’ approach (Figure 26.1), in which all the voxels in a sphere around a center voxel are included in a local multivariate model. The ‘decoding’ accuracy in predicting the outcome is saved as a summary statistic for the center voxel, and the analysis is repeated for every voxel. This is typically referred to as ‘information-based mapping’, and an advantage is that

it can localize areas of the brain with high local predictive accuracy, which can be significant even if the level of activity in the region is not very predictive. However, it is not sensitive to patterns distributed across many brain areas, and it essentially tests thousands of local models rather than building a single integrated predictive model across all available brain measures.

Finally, for translational purposes, it has become increasingly popular to build models that include features distributed across many brain regions, even across the whole brain (Figure 26.1, right). This integrates all the measures available across the brain, and sometimes even across multiple types of images ('multi-modal' imaging), into a single predictive model that can be prospectively tested on new observations. This provides an easy way to validate its overall performance. This type of model is sensitive to brain information and organization captured in relative levels of activity (or other measures) across brain regions. Whole-brain predictive maps assess effects of local brain regions while *controlling for other regions*.

### Spatial scale: Searchlight vs. whole-brain MVPA analysis

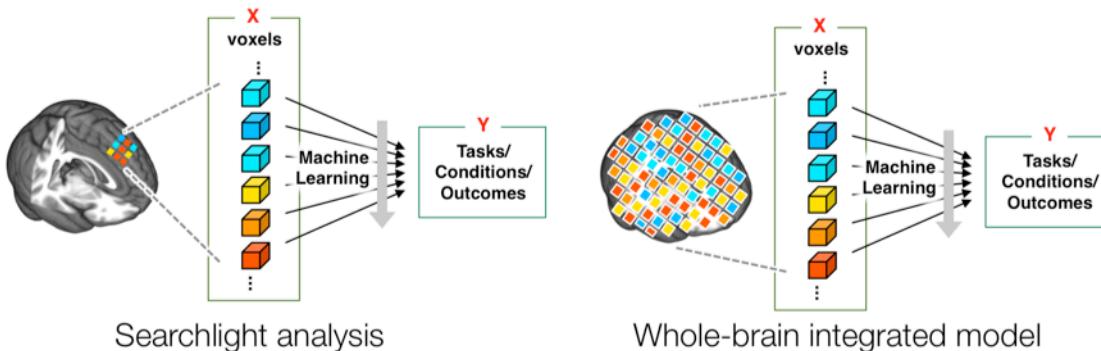


Figure 26.1. Two types of MPVA analyses with different spatial scopes. The left panel shows a searchlight analysis, which tests a series of thousands of local multivariate predictive models in an attempt to map which areas are predictive above chance. The right panel shows a whole-brain predictive map, which develops one predictive model within and across brain regions (or a subset thereof), and attempts to predict the outcome as accurately as possible. The latter type of model can be tested prospectively in new datasets to quantify its accuracy and generalizability. Image credit: Wani Woo.

The traditional wisdom has been that constructing such maps is not feasible because the number of variables (voxels) exceeds the number of observations (subjects, trials, or time points), causing problems with model over-fitting and interpretability. However, techniques from machine learning, including kernel form regression or classification, dimension reduction, and penalization, stabilize the maps even when large numbers of voxels are included in the model. Also, techniques like *cross-validation* and multi-study prospective testing, permit valid and essentially unbiased tests of the predictive model's performance. If effect sizes are large and brain activity or related measures are robustly related to the outcome, then predictive maps with high accuracy can be estimated even using small samples.

In principle, any model's accuracy in predicting new data can be tested. Training a model involves specifying its form and estimating its associated parameters. Testing involves applying the same model, with the exact same parameter estimates (e.g., regression slopes, voxel weights) and applying

it to new observations. However, some models are more amenable to out-of-sample testing than others. If models are integrated predictive models - whether they use the whole brain or a subset defined a priori - prospective tests are relatively straightforward. If the model is linear, it boils down to a set of weights (the parameter estimates) that can be applied to a new test dataset (a new brain image) to yield a single predicted outcome value. If the model includes voxels only in primary visual cortex, the test describes how well visual cortex predicts the outcome. If the model includes areas across the entire brain, the test describes how predictive the brain map is as a whole, integrating information across individual regions. An example of this is shown in Figure 26.2.

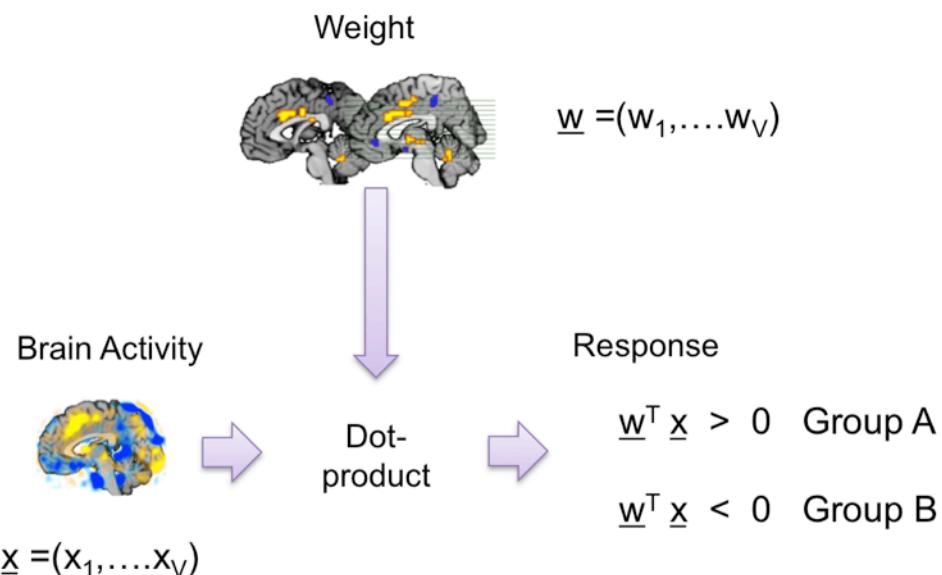


Figure 26.2. An MVPA example with binary classification. The input features are measurements over all  $V$  voxels in the brain contained in the vector  $x$ . These voxels are weighted by the vector  $w$  of length  $V$ , so that if  $x^T w > 0$  we categorize a subject as belonging to group  $A$  and if  $x^T w < 0$  we categorize them as belonging to group  $B$ .

### ***Level of analysis: Idiographic and between-subject***

Another important choice in MVPA analysis, independent of the algorithm used, is whether to allow the model to be flexible in identifying different patterns for each subject, or whether to constrain it to capture effects that are generalizable across individual participants.

The former method was the most widely used in early MVPA literature, and it uses training data only from one individual to develop an idiographic map for that person. Summary statistics like accuracy are used to test the mean effect and its significance across a population, but there is no single pattern or model that is predictive of the outcome for every participant. As a result, the model can only be applied to new observations from that individual. This was the starting point for MVPA because it was assumed that the neural topography that encodes functional information is often

not conserved across individuals. This is true for topographical features such as ocular dominance columns - the locations of cortical areas that map input from each eye are not located in the same place across individuals. However, in addition to this idiographic topography, larger spatial-scale organization may often be detectable across participants.

An alternative is to develop group-level models that explicitly capture topography that is consistent across individuals. Such patterns cannot capture neural features that are ideographically organized, but they can capture meso-scale organization that is below the scale at which we conceptually define and discuss regions. For example, the dorsal anterior cingulate (dACC) cortex is a large piece of real estate, and it's typically referred to as a unit; researchers discuss the overlap in tasks that activate the dACC and what its common, underlying function might be across these tasks. However, there appear to be topographical patterns of activity within the dACC that are conserved across participants and encode different types of motivational processes.

While group-level MVPA maps are less flexible (there is only one set of predictive weights for all subjects) and spatial resolution is reduced, there are a number of advantages that make them particularly appealing for translational work, and for efforts to develop a cumulative basic science of brain function. First, group-level patterns constitute a true predictive model that can be applied to new individuals, permitting direct replications of the predictions made by a model in new datasets, in new studies. This also permits examining the properties of a pattern across studies, including its robustness in different populations and its generalizability across task paradigms and related psychological processes.

Second, group-level maps are less susceptible to confounding variables than idiographic, within-participant patterns. An important issue is that because of the flexibility of idiographic models, it is possible that confounds that are typically controlled for in a group analysis can contaminate the results from some individuals. MVPA algorithms are ‘greedy’, using all available information to explain the outcome, so uncontrolled confounds will tend to inflate the overall accuracy. Furthermore, different participants may be influenced by different confounds (e.g., failures of randomization in task or trial order for one participant, head movement for another). This can lead to spurious findings that a brain area or areas ‘encodes’ a process, particularly when predictive accuracy is weak and the researchers are relying on results to be just above chance for this inference. Group-level maps include fewer parameters than idiosyncratically trained models, and it is possible to balance factors such as task order and stimulus sets across individuals, thus eliminating many sources of confounds.

Third, with group-level maps, it is possible to provide group inferences on the significance of weights on voxels or other brain features. This allows one to establish that the pattern itself is meaningful and examine its relationship with findings from other tasks. This is important as MVPA can capture both effects that are meaningful in relation to other neuroscientific findings and those that capitalize on artifacts, such as large signal at the boundaries of brain and fluid space created by head movement. Like other findings, predictive maps should be compared with findings from electrophysiology, lesion studies, optogenetics, structural connectivity studies, and other methods in humans and animals. This can establish bridges across methods and species, and advance our understanding of the brain. However, patterns developed in idiographic maps, especially using searchlight algorithms,

are rarely examined or tested for neuroscientific meaningfulness and interpretability.

## Sensitivity to neural topography

A powerful advantage of MVPA over standard brain mapping is that multi-voxel patterns derived from machine learning can circumvent some of the limitations in spatial resolution of fMRI and be sensitive to functional brain topography at finer spatial scales. This occurs, because patterns can be sensitive to the local structure of functional columns and differential functional organization across ‘clumps’ or neurons and microvascular beds. Pattern information about functional topography is often preserved at a mesoscopic level, and is detectable even when the voxels sampled are larger than the columns themselves.

In addition, mesoscopic organization is often preserved across individuals, permitting accurate inferences about individuals based on normative group-level data. This means that distributed local patterns of activity can be sensitive to the differential distribution of different types of neurons across a brain region. Thus, whereas any one voxel in a region such as the anterior cingulate may not be highly diagnostic of pain, a particular local pattern across voxels within the anterior cingulate may be diagnostic.

This is quite important, because much of the functional information about which stimulus, feeling, or psychological state a person is experiencing is likely to be contained in fine-scale topographical maps and circuits comprised of intermixed populations of different types of neurons with different functional properties. Decades of electrophysiology and drug microinjections, and rapid advances in optogenetics, 2-photon and similar forms of microscopy, designer receptors, and other neuroscience techniques strongly reinforce this view. Figure 26.3 shows four types of functional organization in the brain, at four different spatial scales. fMRI has inherently limited ability to separate activity in functional circuits that are interdigitated and whose neurons are distributed in a relatively homogenous way. However, idiographic MVPA maps can be sensitive to functional topography organized at the columnar level, with spatial frequencies of  $< 1\text{mm}$  to several  $\text{mm}$ . Group-level maps can be sensitive to functional topography that is conserved across participants and has spatial frequencies of  $2 - 3\text{mm}$  or greater, which includes many types of established functional maps. And finally, all kinds of integrated predictive maps can be sensitive to information distributed across macroscopic brain regions and networks, though searchlight mapping is not sensitive to this large-spatial-scale information.

## Sensitivity of MVPA to brain topography at different spatial scales

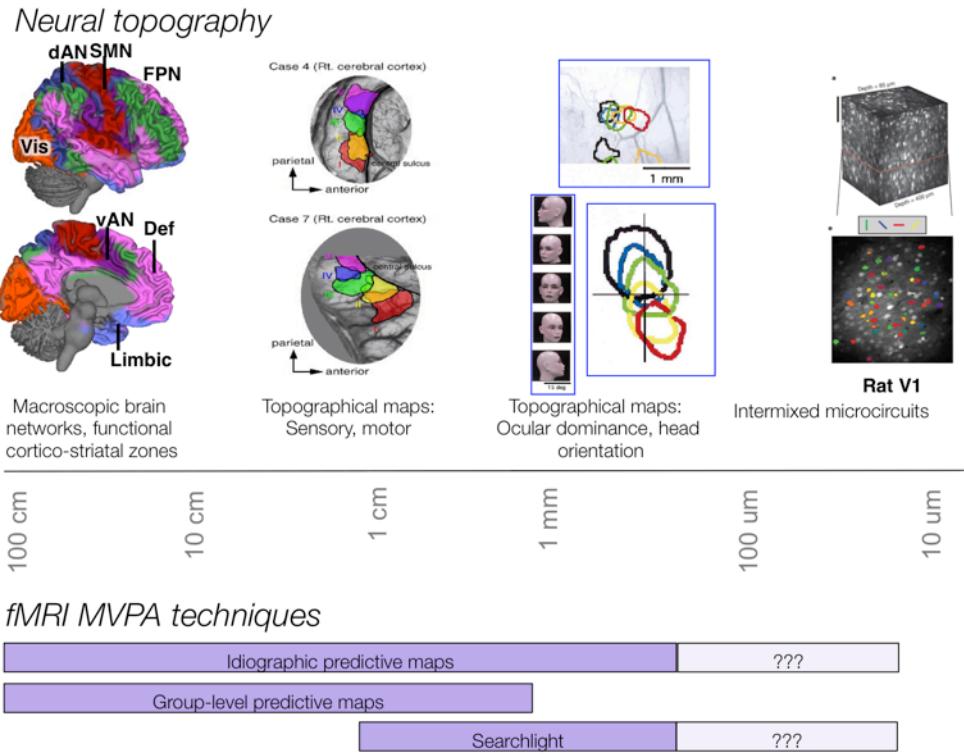


Figure 26.3. Sensitivity of MVPA measures to functional topography at different spatial scales. Functional topography in the brain is multi-scale, and information about functional distinctions related to types of sensation, perception, emotion, behavior, and other contents of thought is reflected in patterns of activity at whole-brain, map-level, column-level, and neuron-level scales. The sensitivity of MVPA varies depending on the outcome and system, and also on the technique. Idiographic patterns developed for one individual can potentially be sensitive to information at many scales, though ‘searchlight’ analyses are sensitive only to information within individual local regions. Group-based maps have reduced spatial acuity but can capture functional topography that is preserved across individuals. The lower bound of the spatial scale that fMRI can be sensitive to is unknown, and depends on the precise system(s) and outcome, and the distribution of functional neurons with different properties across the systems.

It is even possible that MVPA analyses can capture functional topography that occurs at smaller spatial scales than the voxels that constitute the unit of analysis in neuroimaging studies. This has come to be known as ‘hyperacuity’. Figure 26.4 shows a simulated example of how signal from two intermixed sets of neurons that encode two different functional tasks might result in differential fMRI activity patterns. In the simulation, the two types of neurons are randomly intermixed, with many neurons of both types (Type A and Type B) contained within each voxel. A key principle is that ‘random’ does not mean ‘homogenous’, and the precise mix of neurons of each type will vary from voxel to voxel. The neurons here are distributed using a homogenous Poisson process; though the average number of neurons of Type A and Type B is homogenous in each voxel, the actual number varies due to random variation. This produces maps of simulated local field potentials that vary randomly across space, and are uncorrelated for Type A and Type B neurons. BOLD fMRI often

tracks local field potentials, but values are sampled on a grid of voxels. Here, down-sampling the local field potential map results in fMRI patterns that are still spatially distinct and uncorrelated for Type A and Type B neurons. Thus, this simulation shows how activation of two different sets of functional neurons can result in two distinct patterns of fMRI activity across voxels, even if the neurons are randomly intermixed with no large-scale topography and fMRI acts as a low-pass spatial smoothing filter.

## FMRI hyperacuity with randomly intermixed neurons

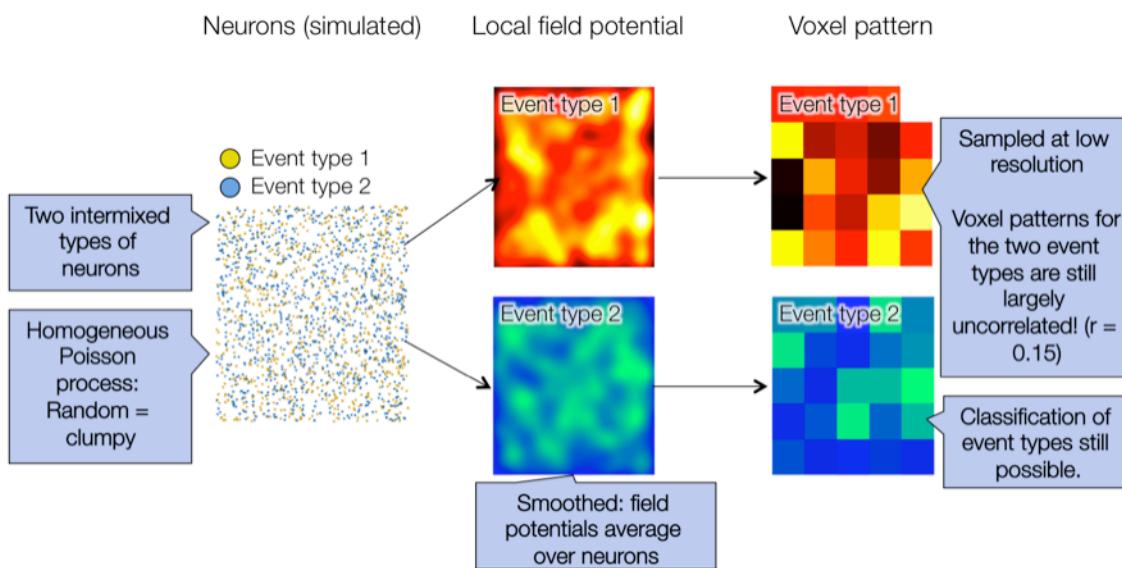


Figure 26.4. Hyperacuity with randomly intermixed neurons. Even if neurons of different functional types are randomly intermixed, they can produce uncorrelated patterns of local field potential activity and fMRI activity. Thus, patterns of activity across voxels may be sensitive to fine-grained differences in neural firing in some cases.

MVPA ‘hyperacuity’ is important because neurons that encode different functional properties or behaviors are intermixed in many brain areas, including intermixed neurons in the amygdala<sup>64</sup> and nucleus accumbens (e.g., revealed with TAI-FISH) that encode positive and negative events, intermixed neurons in visual cortex that encode orientation, and many more. As suggested in Figure 26.3, the lower bound for what types of topography fMRI can be sensitive to is unknown, and depends on the distribution of neural and glial functions, the layout of the microvascular beds involved, the extent of the neural fields involved, and other factors related to the specific outcome domain studied.

<sup>64</sup><http://www.ncbi.nlm.nih.gov/pubmed/16482160>

## Sensitivity to distributed representations

An advantage of MVPA models that include voxels across brain systems is that they can capture information encoded in large scale systems, including patterns of activity across macroscopic regions and in networks connected by long-range anatomical projections. Many animal models and early human neuroimaging and lesion studies alike focused on measurements and representations in local regions, and large-scale networks were relatively neglected until recently. For example, it is not uncommon for a neuroscientist to make a career of studying one nucleus (e.g., the dorsal raphe nucleus) or local circuit (e.g., the lateral geniculate nucleus-primary visual cortical circuit). However, a great deal of human neuroimaging evidence suggests that large-scale networks are a major organizing principle of brain function, which complements information encoded in fine-grained topographical organization. This information can be captured by whole-brain predictive models, though searchlight analyses are insensitive to information at this scale.

Thus, which type of MVPA model is appropriate depends strongly on whether the underlying neural representations relevance for an outcome are contained solely within one local region or are distributed across regions or networks. Many of the ‘success stories’ in neuroscience have focused on local representations. For example, face recognition appears to depend critically on an area of inferior temporal cortex called the ‘[fusiform face area](#)<sup>65</sup>’. Local representations produce focal deficits in patients, which makes them easier to study and discuss. However, the vast majority of processes of psychological interest, from emotion to empathy to cognitive control, are accomplished not by isolated brain regions but by distributed networks. Even processes that are relatively localized can be influenced by processes across the brain, making the study of large-scale brain interactions essential for the study of mental phenomenology. For example, face perception is influenced by goal- and context-related processing in the [prefrontal cortex](#)<sup>66</sup>. Theorists studying memory, emotion and many other domains have focused on distributed representations.

A simulated example of distributed representation is shown in Figure 26.5. In this simulation, voxels across many areas of the brain are important for predicting an outcome, but the true effect size in each of these regions is low, and the true brain-outcome correlation is only  $r = 0.18$ . Because this is a simulation, we know that this is the ‘ground truth’. When we do a local ‘searchlight’ analysis, the maximum correlation is  $r = 0.57$ . We know this is an over-estimate, because the true correlation in each voxel is  $r = 0.18$  plus or minus some amount due to noise. The searchlight analysis capitalizes on chance by picking the region with the best-looking noise profile, the one that happens to provide the strongest correlation post hoc. However, just because local information is not sufficient it does not mean that the signal is not there - it is just distributed. In this case, a whole-brain predictive analysis provides an unbiased accuracy of  $r = 0.98$ . Averaging across many regions that are weakly correlated with the outcome, but are largely independent from one another, averages out the noise and results in very accurate prediction.

<sup>65</sup><http://www.jneurosci.org/content/17/11/4302.full>

<sup>66</sup><http://www.sciencemag.org/content/314/5803/1311>

## Distributed prediction: A simulation

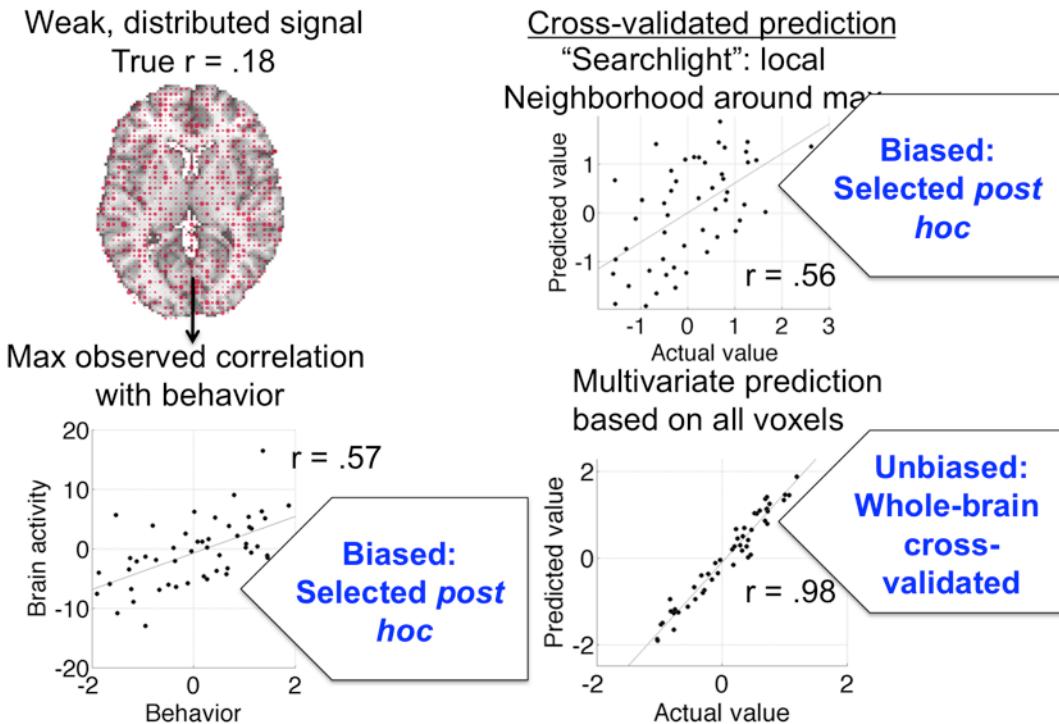


Figure 26.5. A simulation showing the power of whole-brain MVPA to capture distributed representations. This simulation shows a case where weak but real signals predicting an outcome are widely distributed across the brain. Local searchlight analysis produces a moderately strong, but over-optimistic prediction, because no single brain area is strongly related to the outcome. Whole-brain predictive analysis strongly predicts the outcome by averaging over regions and thus averaging out noise that varies across brain regions.

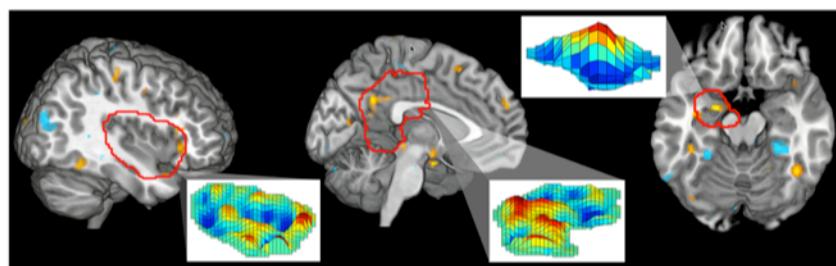
Whether the information required to predict (and ultimately explain) a psychological, behavioral, or clinical outcome is encoded in local or distributed circuits is ultimately an empirical question. Brain analyses have not focused much on this question because traditional univariate analyses do not provide the tools needed to address it. MVPA, however, can provide answers. In addition to capturing information encoded in large-scale systems, by comparing predictive accuracy across local and distributed models, MVPA models can be used to test whether spatially distributed information is important or not. In this way, it can be used to test the spatial basis that is necessary and sufficient to predict an outcome.

An empirical example of distributed representation is shown in Figure 26.6. In this study, brain images were used to predict the intensity of reported negative emotion. A distributed pattern across many brain areas, some strongly identified with negative emotion in the neuroscience literature (e.g., the amygdala) and others associated with processes such as theory of mind (e.g., dorsomedial prefrontal cortex), strongly predicted negative emotion, explaining over 70% of the variance in reports. Many previous theories have focused only on selective brain regions like the amygdala

as key generators of negative emotion. Here, however, MVPA patterns in the amygdala explained only 10% of the variance (the average amygdala signal was even lower), and no other region of the brain fared much better. In fact, the maximum variance explained in a searchlight analysis across the brain was less than 20%, and this value capitalizes on chance and is thus an over-estimate. This and related analyses in this study provides an empirical demonstration that activity distributed across multiple systems throughout the brain is required to capture the intensity of negative emotion.

## An empirical example of distributed representation

The Picture-Induced Negative Emotion (PINES) signature



Region of interest vs. whole-brain: Percentage of variance explained

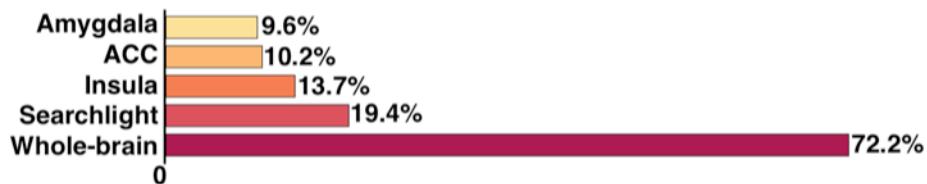


Figure 26.6. An empirical demonstration of distributed representation in the domain of emotion. In Chang et al. 2011, a whole-brain pattern strongly predicted the intensity of negative emotion ratings in response to viewing emotionally evocative pictures. The whole-brain pattern explained much of the variance in picture ratings (> 70%), but local activity in the amygdala, insula, anterior cingulate (ACC), and others (searchlight) explained relatively little variance. This comparison and related analyses demonstrate that the representation of negative emotion is distributed across systems.

## Benefits in testing generalizability across individuals and studies

Recently there has been a great deal of attention paid to the issue of replicability in science, and in psychology and neuroscience in particular. Many findings, once established, are not extensively tested in further empirical tests, and those that are have often not held up. In neuroimaging, our field has been marked by a rush to explore the space of interesting psychological effects that might be mapped to brain function, with many effects that are neither definitively replicated or refuted because the researchers have moved on to the next set of novel, potentially ‘high impact’ findings. This has resulted in an accumulation of interesting findings, but few brain maps or models that

can be used to build a cumulative science, and which have proven utility in helping to understand human performance, behavior, and clinical science. As [Kapur et al.](<http://www.ncbi.nlm.nih.gov/pubmed/22869033>) write, ‘Why has it taken so long for biological psychiatry to develop clinical tests?’

As neuroimaging is increasingly used in clinical and translational studies, there is increasing focus on the diagnostic value of brain patterns for clinical conditions and treatment responses. For maps to be useful in translation, they must be prospectively tested and used in new samples. This attempt to use neuroimaging results in an applied setting can both establish usefulness and validate neural models, weeding out those that work from those that do not.

MVPA models at the group level are explicitly constructed to generalize across participants. This also means that they can be used as biomarkers that can be tested prospectively in new studies. These patterns are not simply results - they are *research products* that, like biochemical assays and protocols, can be passed from laboratory to laboratory and either validated or disproven. The scope of their utility can also be defined and refined as additional tests are performed. Such tests can include:

1. Replicating the accuracy of predictions and the spatial weights in new studies.
2. Testing the generalizability of the marker in a new population (e.g., patients, diverse ethnic and racial groups).
3. Testing generalizability across scanners.
4. Testing generalizability across variants of the task and design.
5. Testing generalizability vs. specificity across similar psychological and behavioral constructs.

Thus, MPVA based predictive maps can become useful biomarkers or ‘signatures’ analogous to other medical and biological assays. This provides a way of deciding which results are useful and warrant larger-scale tests.

This approach provides a way of making decisions about how to pursue scientific development and translational work that will hopefully be more successful than efforts made over past decades. A central problem is that definitive tests require very large samples across diverse populations. Consider trials of drugs done by major pharmaceutical companies. These are often multinational, conducted in coordinated sites in multiple countries worldwide. They involve rigorous sampling frames, personnel to monitor the progress of the study and dedicated statistical analysts, external oversight, and hundreds or thousands of participants assessed at multiple time points in longitudinal designs. Individual trials can easily cost upwards of \$100 million dollars. Needless to say, not every neuroscientific hypothesis can be tested with such rigor and scope.

A complementary, and equally important, problem is that there are thousands of important neuroscientific problems to address, millions of ways of approaching these problems scientifically, and unlimited potential variations in the specific implementations involved. If we put ‘all our eggs in one basket’, so to speak, and focus neuroimaging efforts on one or a very few large-scale studies, we are necessarily limiting the scope of what is tested to a small set of questions, hypotheses, and

design choices. If these large studies are conducted before work establishing the promise of these hypotheses and designs is established, the results could be disastrous.

Recent work has established a path forward, proposing the widespread exploration of many research questions and hypotheses, with a narrowing of scope and increased resources progressively devoted to the most promising biomarkers and hypotheses. This progression is shown in Figure 26.7. Promising MVPA-derived brain patterns can serve as markers that can be tested and shared across studies. The more broadly their performance and generalizability are assessed, the more useful they are as representations of defined mental, behavioral, and clinical constructs.

### MVPA biomarkers as “research products” that can be shared and tested

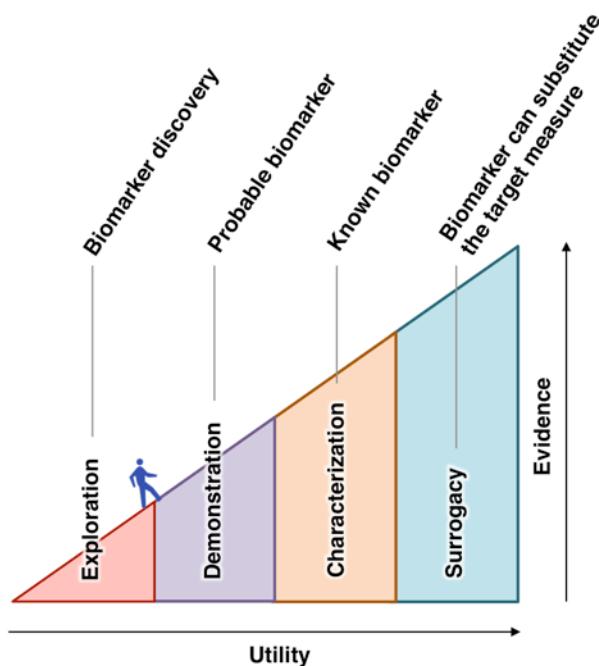


Figure 26.7. Progression of biomarker discovery, validation, and testing. Exploration and biomarker discovery should take place using small studies across many paradigms and types of experiments. The most promising biomarkers can be validated and tested in larger-scale studies. The utility of a brain model or pattern, both as a predictive marker and as a representation of a mental or behavioral construct, is proportional to the strength, replicability, and generalizability of the evidence across studies. This approach is explained in Borsook et al., 2011. Image credit: Wani Woo.

# Resources and further reading

These links provide resources to other books and overviews of fMRI.

## Other books about fMRI

Poldrack, R. A., Mumford, J. A., & Nichols, T. E. (2011). *Handbook of functional MRI data analysis*. Cambridge University Press.<sup>67</sup>

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging* (Vol. 1). Sunderland: Sinauer Associates.<sup>68</sup>

Buxton, R. B. (2009). *Introduction to functional magnetic resonance imaging: principles and techniques*. Cambridge university press.<sup>69</sup>

D'Esposito, M. (Ed.). (2006). *Functional MRI: Applications in clinical neurology and psychiatry*. CRC Press.<sup>70</sup>

Ashby, F. G. (2011). *Statistical analysis of fMRI data*. MIT press.<sup>71</sup>

Faro, S. H., & Mohamed, F. B. (Eds.). (2006). *Functional MRI: basic principles and clinical applications*. Springer Science & Business Media.<sup>72</sup>

Lazar, N. (2008). *The statistical analysis of functional MRI data*. Springer Science & Business Media.<sup>73</sup>

---

<sup>67</sup><http://www.cambridge.org/us/academic/subjects/statistics-probability/statistics-life-sciences-medicine-and-health/handbook-functional-mri-data-analysis?format=HB>

<sup>68</sup><http://www.amazon.com/Functional-Magnetic-Resonance-Imaging-Edition/dp/0878932860>

<sup>69</sup>[http://www.barnesandnoble.com/w/introduction-to-functional-magnetic-resonance-imaging-richard-b-buxton/1100295143?ean=9780521899956&st=PLA&sid=BNB\\_DRSCore+Shopping+Textbooks\\_00000000&2sid=Google\\_&sourceId=PLGoP120&k\\_clickid=3x120&kpid=9780521899956](http://www.barnesandnoble.com/w/introduction-to-functional-magnetic-resonance-imaging-richard-b-buxton/1100295143?ean=9780521899956&st=PLA&sid=BNB_DRSCore+Shopping+Textbooks_00000000&2sid=Google_&sourceId=PLGoP120&k_clickid=3x120&kpid=9780521899956)

<sup>70</sup>[https://play.google.com/store/books/details/Mark\\_D\\_Esposito\\_Functional\\_MRI?id=QqXqzQP88KMC](https://play.google.com/store/books/details/Mark_D_Esposito_Functional_MRI?id=QqXqzQP88KMC)

<sup>71</sup>[http://www.barnesandnoble.com/w/statistical-analysis-of-fmri-data-f-gregory-ashby/1100660490?ean=9780262015042&st=PLA&sid=BNB\\_DRSCore+Shopping+Textbooks\\_00000000&2sid=Google\\_&sourceId=PLGoP120&k\\_clickid=3x120&kpid=9780262015042](http://www.barnesandnoble.com/w/statistical-analysis-of-fmri-data-f-gregory-ashby/1100660490?ean=9780262015042&st=PLA&sid=BNB_DRSCore+Shopping+Textbooks_00000000&2sid=Google_&sourceId=PLGoP120&k_clickid=3x120&kpid=9780262015042)

<sup>72</sup>[https://play.google.com/store/books/details?id=MkjTO4wx-bkC&source=productsearch&utm\\_source=HA/Desktop\\_US&utm\\_medium=SEM&utm\\_campaign=PLA&pcampaignid=MKTAD0930BO1&gl=US&gclid=COiVgqyrtMcCFAohMgodg6sGfw&gclsrc=ds](https://play.google.com/store/books/details?id=MkjTO4wx-bkC&source=productsearch&utm_source=HA/Desktop_US&utm_medium=SEM&utm_campaign=PLA&pcampaignid=MKTAD0930BO1&gl=US&gclid=COiVgqyrtMcCFAohMgodg6sGfw&gclsrc=ds)

<sup>73</sup><http://www.springer.com/us/book/9780387781907>