FUNDAMENTALS OF FUNCTIONAL NEUROIMAGING

Stephan Geuter[1]
Martin A. Lindquist[2]
Tor D. Wager[1,3*]

[1]University of Colorado Boulder, Institute of Cognitive Science
[2]Johns Hopkins University, Department of Biostatistics
[3]University of Colorado Boulder, Department of Psychology and Neuroscience

Summary:
21303 words (text, without references)
26058 total words (with references)
3 tables
11 figures

Running head: FUNCTIONAL NEUROIMAGING

* Address correspondence to:
Dr. Tor D. Wager
Department of Psychology and Neuroscience
University of Colorado Boulder
345 UCB
Boulder, CO 80309
E-mail: tor.wager@ colorado.edu.

Acknowledgments

We would like to thank Jessica Andrews-Hanna for helpful comments on the manuscript. Parts of this chapter are adapted from Wager. T. D., Hernandez, L., and Lindquist, M.A. (2009). Essentials of functional neuroimaging. In: G. G. Berntson and J. T. Cacioppo (Eds.), Handbook of Neuroscience for the Behavioral Sciences. (pp. 152-97). Hoboken, NJ: John Wiley & Sons.

I. INTRODUCTION

Functional neuroimaging techniques have become a central research tool for psychologists, cognitive scientists and neuroscientists. The use of neuroimaging data from functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) studies is central to the fields of cognitive neuroscience, affective neuroscience, social cognitive neuroscience, neuroeconomics, and related disciplines. FMRI and PET data are being combined with data on human performance, psychophysiology, genetics, and computational models of performance and neural function in increasingly sophisticated ways. The result is enhanced models of human brain function in relation to thought, emotion, and behavior, which can be used to both understand the mind and guide applied research on performance enhancement and clinical assessment and treatment. The best such models are informed by the rich histories of cognitive psychology and psychophysiology, and—due largely to the integration of neuroimaging data—are increasingly grounded in brain physiology. This grounding permits stronger and more specific connections with the neurosciences and biomedical sciences, allowing behavioral scientists to leverage a vast and growing literature on brain systems developed in these fields.

All neuroscience methods have limitations, and neuroimaging is no exception. The current trend is towards increasingly multidisciplinary approaches that use multiple methodologies to overcome some of the limitations of each method used in isolation. For example, currently available techniques allow electroencephalography (EEG) and fMRI data to be collected simultaneously (Goldman et al., 2000), which provides improved temporal precision, among other benefits. Neuroimaging data are also being combined with transcranial magnetic stimulation, combining the ability of neuroimaging to observe brain activity with the ability of TMS to manipulate brain function and examine causal effects (Bohning et al., 1997).

The rapid pace of development and interdisciplinary nature of the neuro-behavioral sciences presents an enormous challenge to researchers. Moving this kind of science forward requires a collaborative team with expertise in psychology, neuroanatomy, neurophysiology, physics, biomedical engineering, statistics, signal

processing, and other disciplines. Having a successful team requires that individuals push beyond the boundaries of their disciplines and develop expertise in multiple areas, so that there is enough overlap that the team can work well together. Hence, the goal of this chapter is to review the basic techniques involved in the acquisition and analysis of neuroimaging data—and some recent developments—in enough detail to highlight the most important issues and concerns. We also intend to provide an overall roadmap of study design and analysis options and some of their limitations.

The various aspects of PET and fMRI methodology are organized here into four sections. Section II deals with what several neuroimaging techniques measure, including a comparison of PET and fMRI. Section III covers the basics of fMRI data acquisition and the relationship between brain activity and observed fMRI signals. Section IV describes how fMRI data are used to make psychological inferences and how inference relates to study designs. We emphasize two kinds of inferences: *forward* inferences about brain activity given a psychological experimental manipulation, and *reverse* inferences about psychology given patterns of brain activation. Section V deals with neuroimaging data analysis and signal processing before analysis ("pre-processing"), the general linear model (GLM), and methods for investigating brain connectivity. This section also deals with the localization of results from functional neuroimaging studies.

II. Overview of Neuroimaging Techniques

There are many ways to measure brain function, including fMRI, PET, single positron emission computerized tomography (SPECT), electroencephalography (EEG), magnetoencephalography (MEG), and near-infrared spectroscopy (NIRS). Each of these techniques provides a unique window into the functions of mind and brain (Figure 1).

In this chapter we will mainly focus on PET and fMRI, which provide the most anatomically specific information across the entire brain. The relatively high spatial resolution of PET and fMRI complement the precise timing information provided by EEG and MEG (*CITE EEG CHAPTER*). In addition, the ability of simultaneous multi-slice fMRI to measure activity over the entire brain every 500 msec or so is providing enhanced temporal resolution and resistance to some of the artifacts that have plagued

fMRI.

Whole-brain imaging techniques like PET and fMRI offer great potential for synergy with animal research. Whereas animal electrophysiology and lesion experiments are often focused on a single region, neuroimaging can assess global function and interactions across large-scale brain systems. Emerging neuroscientific techniques, including calcium imaging with two-photon or light sheet microscopy, are also able to examine fields of interacting neurons (at the single-neuron level, several orders of magnitude finer than fMRI). However, there is a critical gulf between what we learn in animals and what we can infer about the human brain and behavior. Comparing results from parallel experiments performed in humans and animals is thus critical. Before we discuss fMRI data acquisition, analysis, and inference in more detail, we provide a short overview of the most important measures available on MR and PET scanners (Table 1).

Measures available on MR and PET scanners

Structural images

MRI can provide detailed anatomical scans of gray and white matter with a spatial resolution below 1 mm^3. These images are used to localize functional results in individual or group-averaged brains, and are widely used to analyze measures of brain structure in relation to psychological or clinical variables of interest—e.g., practice or development, effects of aging, and differences between healthy individuals and clinical populations (see Figure 2A for examples).

A popular way of analyzing gray-matter density is the voxel-based morphometry (VBM) method (Ashburner and Friston, 2000; Good et al., 2001), which uses structural image intensity to measure gray- and white-matter density. Other methods use measures of cortical thickness derived from surface reconstruction and unfolding (Fischl et al., 1999; Van Essen and Dierker, 2007), or the volume of anatomically defined structures. One example using structural scans is a classic study reporting that London taxi drivers, who had developed extensive expertise in spatial navigation, had larger posterior hippocampi than controls (Maguire et al., 2000).

Diffusion weighted imaging (DWI)

Another type of structural imaging attempts to quantify white-matter tracts. MRI pulse sequences can be tuned to be sensitive to directional (anisotropic) patterns of water diffusion, called *diffusion-weighted* imaging. Water diffuses more readily along the axons that make up the brain's white matter than across them, and thus diffusion-weighted images be used to track the course of white-matter tracts (Figure 2C). Diffusion tensor imaging (DTI) is a popular technique for measuring directional diffusion and reconstructing the fiber tracts of the brain (Denis Le Bihan et al., 2001). DTI provides relatively low-resolution directional information, but many acquisition and analysis techniques for enhancing assessment and directionality of diffusion are being developed. Tractography analyses allow the quantification of the thickness and connectivity of these tracts (Behrens et al., 2007).

Such tools allow researchers to analyze the relationships between structural connectivity and neuro-psychological processes such as development, training, aging, cognitive and emotional function, and psychopathology (Johansen-Berg and Behrens, 2006). DWI can be combined with other techniques, such as fMRI (including measures of functional connectivity) or other anatomical and neurochemical measures. For example, one study used DWI to define adjacent sub-regions of the medial prefrontal cortex, and then used fMRI to show that the sub-regions responded differentially to different tasks (Johansen-Berg et al., 2004).

Functional MRI

Functional MRI allows the investigation of brain function during tasks or rest. It is by far the most popular MR technique and most often based on the blood oxygenation level dependent (BOLD) contrast (see *MR physics and BOLD basics*) that measures relative levels of oxygenated blood across the brain (Figure 2B).

Task-based fMRI studies use experimental approaches to relate brain activity or functional connectivity measures to experimentally induced psychological states and/or measured performance variables. *Resting state* fMRI studies measure BOLD activity during rest, i.e. without any specific task. The signal covariation across different regions

or voxels is then assessed to identify brain networks (Biswal et al., 1995; Fox et al., 2007; Yeo et al., 2011). Both task-related and resting-state fMRI signals can be related to performance, clinical outcomes, and other variables of interest, within a single group or across groups (e.g., patients and controls, older and younger individuals, etc.).

Though BOLD fMRI is the most popular type of functional MRI signal, there are other promising techniques as well, based on different pulse sequences and/or use of radiofrequency coils. Another popular fMRI technique is arterial spin labeling (ASL), which allows for quantitative measurement of regional cerebral blood flow (rCBF) (Detre et al., 1994), in many cases across long time scales (e.g., before and after cognitive training or a clinical intervention, Figure 2B). By magnetically labeling water molecules entering the brain through the carotid arteries and then comparing the MR images with labeled molecules to the MR images without labeling, one can estimate local blood flow throughout the brain (Buxton et al., 1998). 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. ASL can be used to test the same types of functional effects as BOLD, including task-induced activation and connectivity, resting-state connectivity, and relationships between brain activity and performance (or other outcomes).

Measures of brain activity using PET

Perhaps the most frequent use of both PET and fMRI is the study of metabolic and vascular changes that accompany changes in neural activity. With PET, one may separately measure glucose metabolism, oxygen consumption, and rCBF. Each of these techniques allows one to make inferences about the localization of neural activity based on the assumption that neural activity is accompanied by a change in metabolism, in oxygen consumption, or in blood flow.

The PET camera provides images by detecting positrons emitted by a radioactive tracer, the frequencies of which are reconstructed into three-dimensional volumes. Positrons are subatomic particles having the same mass but opposite charge as an electron. The most common radioactive tracers are ^{15}O, "oxygen-15," commonly used in blood-flow studies, ^{18}F (fluorine), used in deoxyglucose mapping, and ^{13}C (carbon) or ^{123}I

(iodine), used to label raclopride and other receptor agonists and antagonists. The decay rate of such isotopes is quite fast, and their half-lives vary from a couple of minutes to a few hours, which means that a cyclotron must be available nearby in order to synthesize the radioactive tracer minutes before each PET scan.

The tracer is injected into the subject's bloodstream in either a bolus or a constant infusion that produces a steady-state concentration of tracer in the brain. As the tracer decays within the blood vessels and tissue of the brain, positrons are emitted. The positrons collide with nearby electrons (being oppositely charged, they attract), annihilating both particles and emitting two photons that shoot off in opposite directions from one another. Photoreceptors positioned in an array around the participant's head detect the photons. The fact that matched pairs of photons travel in exactly opposite directions and reach the detectors simultaneously are important for the tomographic reconstruction of the 3-D locations where the particles were annihilated. Note that the scanner does not directly detect the positrons themselves; rather it detects the energy that results from their annihilation.

Depending on the design, most PET scanners are made up of an array of detectors arranged in a circle around the patient's head, or in two separate flat arrays that are rotated around the patient's head by a gantry. To detect simultaneously occurring pairs of photons, each pair of detectors on opposite sides of the participant's head must be wired to a "coincidence detector" circuit**.**

The injected tracer will be distributed throughout the blood vessels and tissue of the brain and body. Each pair of detectors counts photons emitted within the tissue between them. The density of photons emitted at each location in three-dimensional space can be estimated mathematically from the number of counts across the multiple detectors. The resulting, reconstructed PET images are maps of how many positron annihilation events occurred in the slice of interest. A more complete explanation of PET image formation, including a discussion of filtered backprojection and other methods, can be found in several good texts (Bendriem, 1998; Sandler, 2003).

What do PET counts reflect? The answer depends on what type of molecule the label is attached to and where that molecule goes in the brain. Ideally, for ^{15}O PET,

counts reflect the rate of water uptake into tissue. 18-fluorodeoxyglucose (FDG) PET measures glucose uptake, whereas ^{13}C Raclopride PET measures dopamine binding. However, in practice the observed level of signal depends on a number of factors, including the concentration of the radiolabeled substance in the blood, the blood flow and volume, the presence of other endogenous chemicals that compete with the labeled substance, and *kinetic properties*. Kinetic properties refer to the dynamics of interactions between the ligand (i.e., the radiolabeled molecule), the receptors, and the tissue types they move through. Important variables include the binding affinity of the substance to receptors, the rate of dissociation of the substance from receptors, and the rate at which the substance is broken down by endogenous chemicals.

Accurate quantification of binding requires study of the kinetic properties of the substance in animals and the use of this information in *kinetic models*, which use differential equations to estimate the biological parameters of interest (e.g., ligand bound specifically to the receptor type of interest). Different kinetic models estimate ligand concentrations in different numbers of *compartments*, or tissue types; for example, a two-compartment model estimates how much of the ligand is in the vasculature as opposed to in the brain. A three-compartment model often used in receptor binding studies estimates tracer quantities in blood, 'free' tracer in tissue, and label bound to receptors. Often a reference region with few or no receptors (e.g., the cerebellum for dopamine) is used to model the separation of free from bound tracer; this requires the assumption that none of the signal in the reference region comes from 'bound' tracer. A four-compartment model additionally separates tracer bound to receptors of a specific type (called specific binding) from those bound to other receptors (called nonspecific binding). For more details, we refer the reader to (Frey, 1999).

Measures of functional neurochemistry using PET

The affinity of particular pharmacological agents for certain types of neurotransmitter receptors, such as raclopride for dopamine D2 receptors, provides a way to investigate the functional neurochemistry of the human brain. Radioactive labels such as C-11, a radioactive isotope of carbon, are attached to the pharmacological agent.

Labeled compounds are then injected into the arteries by either a bolus or continuous infusion, typically until the brain concentrations reach steady state. This method can be used to image task-dependent neurotransmitter release. As radioactively labeled neurotransmitters binds to receptors, the label degrades and gamma rays are emitted that are detected by the PET camera. When endogenous neurotransmitters are released in the brain, there is greater competition at receptors, and less binding of the labeled substance (referred to as 'specific binding'). Thus, neurotransmitter release generally results in a reduction in radioactivity detected by the PET camera.

The most common radioligands and transmitter systems studied are dopamine (particularly D2 receptors) using [^{11}C]raclopride or [^{123}I]iodobenzamide, muscarinic cholinergic receptors using [^{11}C]scopolamine, opioids using [^{11}C]carfentanil, and benzodiazepines using [^{11}C]flumazenil. In addition, radioactive compounds that bind to serotonin, opioid, and several other receptors have been developed. In recent years, ligands for many other substances and cell markers have been developed as well, such as those related to neuroinflammation and glial-cell activity (Brown et al., 2007; Loggia et al., 2015). These types of molecular imaging may be very useful for examining specific forms of neuropathology in clinical disorders. For example, PET imaging with a compound called "Pittsburgh Compound B" or "PIB" is sensitive to molecules found in neurofibrillary tangles characteristic of Alzheimer's disease, and is now used clinically as a marker for early-onset Alzheimer's (Klunk et al., 2004).

Comparison of PET and fMRI

PET and fMRI can be used in different ways to measure a number of biological processes related to brain activity. Measures are generally obtained for each of a large number of local regions of brain tissue called "voxels" (three-dimensional pixels), providing 3-D brain maps. Popular techniques include measures of both brain structure and function. Structural measures may be divided into measures related to gray- and white-matter volume and density, and measures related to neurochemical receptors and other biomarkers.

The most frequently used functional measures are those that measure processes related to overall neuronal and/or glial activity, referred to here as "activation." These measures include measures of glucose metabolism, blood flow or perfusion in PET and arterial spin labeling (ASL) and the Blood Oxygen Level Dependent (BOLD) signal in fMRI. Activation and deactivation in both PET and fMRI reflect changes in neural activity only indirectly, and they measure different biological processes related to brain activity, which may be broadly defined as the energy-consuming activity of neurons and glia, and the electrical and chemical signals they produce. Thus, both PET and fMRI can be used to measure brain activity, though each has unique advantages over alternative techniques and one another. These are summarized in Table 2, which lists some of the strengths and weaknesses of PET and fMRI in terms of acquisition, signal types and interpretability, resolution, accessibility, and 'multimodal potential'—potential for combination with other techniques.

As one might expect, both PET and fMRI have their share of limitations as well. One should consider the limitations of each technique not only when designing experiments, but also when interpreting the results of studies and reading the neuroimaging literature as a whole. One should always ask the following question: "Are the activations caused by the experimental paradigm or by other unwanted sources?" and "What are the plausible psychological or physiological explanations for the reported activity?" Conversely one should also ask: "Were there other active regions that were likely missed by the experimental paradigm?" Together, the answers to these questions constitute an interpretation of both positive and negative findings. Errors of both commission and omission may occur because of the spatial or temporal limitations of the technique, image artifacts, task confounds, or mischaracterized noise.

Spatial limitations of PET and fMRI

The upper bound on spatial resolution of PET is on the order of 1-1.5 cm^3, though it varies across types of PET scanners and is likely quite a bit lower in practice. The upper bound of fMRI resolution is around 1 mm^3 in high-field imaging in humans or animals, but is typically on the order of 8-36 mm^3 for human studies. The limiting factors

in fMRI include signal strength and the point-spread function of BOLD imaging, which tends to extend beyond neural activation sites into draining veins (Duong et al., 2002). Estimates of the point-spread function of BOLD at 3 Tesla, a limit on the effective resolution based on the fact that BOLD samples oxygenation and flow in local vasculature, are around 3 mm—no matter how small the voxels are (Chaimow et al., 2011). Thus, separating out information encoded in brain features such as cortical columns and even major sub-nuclei (e.g., there are 30 or so in each of the amygdala and thalamus) require high-resolution techniques, often with customized acquisition parameters, to achieve the necessary resolution. Even if the BOLD point-spread function is limited, it is possible to obtain differential information encoded in brain structures with a spatial frequency of around 1-2 mm. For example, careful work in individual participants has demonstrated the imaging of ocular dominance columns in humans (Cheng et al., 2001).

While this resolution does not sound all that bad, there is another factor that seriously limits the effective spatial resolution in most studies. That is the fact that making inferences about populations of subjects requires analyzing groups of individuals, each with differing brain shapes. Usually, individual brains are aligned to one another through a registration or warping process (see *Preprocessing*), which introduces spatial blurring and noise in the group average. Thus, the effective resolution for group fMRI and PET studies is about the same. One estimate based on meta-analysis is that the spatial variation in the location of an activation peak among comparable group studies is 2-3 cm (Wager et al., 2004a).

Overcoming these limitations with high-resolution fMRI imaging is a challenging and rapidly developing research area. By focusing on particular regions and omitting data collection in much of the brain, it is possible to acquire voxels on the order of 1.5 mm per side, yielding fMRI maps with resolution closer to the physical size of functional sub-regions (e.g., cortical fields within the hippocampus, or nuclei in the brainstem). Resolution can potentially be considerably enhanced using high-field imaging and analysis techniques that remove some spread in fMRI signal due to draining veins (Menon, 2002). Secondly, collecting thinner slices can reduce susceptibility artifacts and

improve imaging around the base of the brain (Morawetz et al., 2008). However, there are costs as well. There is a substantial loss in signal due to the smaller volume of each voxel. Ultimately, high-resolution studies are very promising when a small set of subcortical nuclei or nearby cortical regions are of primary interest.

Finally, limitations in group studies related to inter-individual variability can be partially overcome using identification of regions of interest on individual participants' anatomical images or by advanced cortical unfolding and inter-subject warping techniques (Ashburner, 2007). Another interesting idea, called 'hyperalignment', is to match voxel response profiles between subjects to align them in a 'representational' space instead of anatomical space (Haxby et al., 2011). These techniques are making it increasingly possible to do group studies at higher effective spatial resolution, and thus make population inferences about performance, clinical status, and other outcomes.

Temporal resolution and trial structure

Another important limitation of scanning with PET and fMRI is the temporal resolution of data acquisition. The details of this are discussed below, but it is important to note here that PET and fMRI measure different things, over different time scales. Because PET computes the amount of radioactivity emitted from a brain region, at least 30 seconds of scanning must pass before a sufficient sample of radioactive counts is collected. This limits the temporal resolution to blocks of time of at least 30 seconds, well longer than the temporal resolution of most cognitive events, but more suitable for examining mental states or 'mindsets'. For glucose imaging (FDG) and receptor mapping using radiolabeled ligands, the period of data collection for a single condition is much longer, on the order of 30-40 minutes.

Functional MRI has its own temporal limitations, due largely to the latency and duration of the hemodynamic response to a neural event. Typically, even very brief events (e.g., 16 msec) induce measurable changes in BOLD signal, but the BOLD response does not reach its peak until 5-6 seconds after local neuronal and metabolic activity has occurred. Thus, the locking of neural events to the vascular response is not very tight. Current fMRI designs (see *Event-related fMRI*) use a General Linear Model

(GLM) to link BOLD activity to specific mental events. By examining differences in average event-related activity across conditions, it is possible to make inferences about the relative timing and duration of brain responses across different mental processes (Lindquist et al., 2008; Waugh et al., 2010).

Acquisition artifacts

Artifacts and image distortions may arise from a number of sources. An early study, for example, found a prominent PET activation related to anticipation of a painful electric shock in the temporal pole (Reiman et al., 1989). However, it was discovered some time later that this temporal activation was actually located in the *jaw* – the subjects were clenching their teeth in anticipation of the shock!

'Artifacts' refer to both (a) deviations in the spatial pattern and/or intensity of an image from the true, underlying values, and (b) spurious results related to confounding processes. Artifacts can be introduced or mitigated at virtually all stages of acquisition and analysis. Acquisition-related artifacts include those related to magnetic susceptibility, instability in magnetic gradients used to acquire images, and radio-frequency interference from outside sources. They also include distortions related to reconstruction and, importantly, interactions between the magnetic field gradients and physiological processes, mainly head movement, heartbeat and breathing (including induced motion and carbon dioxide levels, which affects BOLD signal).

Susceptibility artifacts in fMRI occur because magnetic gradients near air and fluid sinuses and at the edges of the brain cause local inhomogeneities in the magnetic field that affects the signal, causing distortion in echo-planar imaging (EPI) sequences and blurring and dropout (reduced signal intensity) in spiral sequences. These problems increase at higher field strengths and provide a significant barrier in performing effective high-field fMRI studies. Not all scanner/sequence combinations can reliably detect BOLD activity near these sinuses—which affects regions including the orbitofrontal cortex, inferior temporal cortex, hypothalamus, and amygdala. Signal may be recovered by using optimized sequences such as "z-shimming" (Constable and Spencer, 1999) or spiral in/out sequences (Glover and Law, 2001) and/or using a physical magnetic shim

held in the mouth of the participant (Wilson and Jezzard, 2003). Signal loss and distortion may be further minimized by using improved reconstruction algorithms (Noll et al., 2005) and "unwarping" algorithms that measure and attempt to correct EPI distortion (Andersson et al., 2001). Collecting thinner slices can reduce susceptibility artifacts and improve imaging around the base of the brain (Morawetz et al., 2008), which is now increasingly possible with simultaneous multi-slice or 'multiband' imaging (Feinberg et al., 2010; Setsompop et al., 2012).

Functional MRI also contains more sources of signal variation due to a substantial slow drift of the signal across time and higher frequency changes in the signal due to physiological processes accompanying heart rate and respiration. The low-frequency noise component in fMRI can obscure results related to a psychological process of interest and it can produce false positive results, so it is usually removed statistically prior to analysis. A consequence of slow drift is that it is often impractical to use fMRI for designs in which a process of interest only happens once or unfolds slowly over time, such as drug highs or the experience of strong emotions, though some experimental/analysis approaches have been developed to facilitate such studies (Lindquist et al., 2007). As scanners have become more stable, low-frequency drift has become less of a problem—though it is still important to consider—and many published studies analyze BOLD responses across periods of several minutes. These include, for example, BOLD imaging in response to pharmacological challenges (Wise et al., 2002; Atlas et al., 2012) and stressors (Sinha et al., 2004).

Combining Techniques: fMRI, EEG, TMS, Genetics

One option to overcome some of the temporal limitations of fMRI is the integration of multiple methodologies with low (fMRI) and high temporal resolution (EEG). Such multimodal imaging is associated with a number of technical challenges, but it is increasingly popular as more integrated solutions to some of these challenges become available. Figure 3 visualizes some potential synergies between MR measures and other methods.

The simplest way to combine fMRI with EEG or MEG is to repeat the same experiment once in the MR scanner and once outside using EEG or MEG. Structural MR images can also be used to improve source localization in EEG/MEG datasets. A more integrated approach is the concurrent acquisition of fMRI and EEG data inside the scanner. This enhances the analysis of fMRI by making more direct links, based on fMRI-EEG co-variation across time, trials, and/or conditions. This combination enhances the temporal resolution of fMRI and can support more informed temporal modeling choices. It can also enhance the spatial resolution of EEG analysis by constraining source localization (Phillips et al., 2002) or by testing covariation between EEG signals and activity in specific MRI voxels (Scheibe et al., 2010).

Simultaneous acquisition of fMRI and EEG data poses several technical challenges; among them, radiofrequency pulses during MRI scanning induce large artifacts in EEG recording by inducing currents in the EEG leads. However, in addition to hardware enhancements minimizing artifacts, the regular timing and waveform of the MR artifacts allow them to be subtracted out from the EEG recordings, as long as the timing synchronization is extremely precise (see (Laufs et al., 2008) for a review).

Neuroimaging is also being combined with transcranial magnetic stimulation (TMS) to integrate neuroimaging of brain activity with the ability afforded by TMS to manipulate brain function and examine causal effects (Bohning et al., 1997; Leitao et al., 2015). Different TMS protocols can be applied either before the fMRI session to investigate more tonic effects, or interleaved between acquisition of single fMRI volumes (Bohning et al., 1999; Ruff et al., 2006).

Finally, integrating genetics with brain imaging is seen as a way to study how genetic polymorphisms and other genetic characteristics may affect functional brain activity. For example, an early study found that prefrontal activation related to reward anticipation was dependent on a polymorphism in the Catechol-O-methyltransferase (COMT) gene, which regulates a transporter critical for the reupdake of dopamine, norepinephrine, and epinephrine (Yacubian et al., 2007). A hope for the field of imaging genetics is that quantitative indicators of brain function could facilitate the identification of the genetic determinants of complex brain-related disorders such as autism, dementia

and schizophrenia (Glahn et al., 2007b; Glahn et al., 2007a). Most studies look at (a) associations between brain activity and candidate genes or genome-wide single nucleotide polymorphisms (GWAS); (b) moderation of task- or performance-related brain responses by gene variants (a type of gene-by-brain or gene-by-brain-by-performance interactions); or (c) the heritability of structural and functional brain patterns in twin samples.

However, as usual in science, there are substantial challenges to be overcome. A fundamental issue with both brain imaging and candidate gene studies is the large number of tests that can potentially be performed to screen for significant effects. The more tests are conducted for a sample of a given size, the less likely the results are to replicate: The 'winning' tests might either be purely due to chance, or, if this possibility is minimized using appropriate multiple comparisons correction, their importance (i.e., effect size) is typically dramatically overestimated. With many voxels and many potential genes, this problem increases multiplicatively. Now imagine multiplying that number by the number of potential task effects and ways of defining outcomes in a given study. The flexibility can become astronomical if the space of hypotheses tested is not carefully constrained, with corresponding increases in false findings. To provide reasonably powered tests, very large sample sizes are needed—depending on the underlying effect size, thousands or tens of thousands of subjects may be required, therefore requiring huge investments and extensive collaborative efforts. This is what motivates consortia such as ENIGMA, a collaborative data-sharing project that includes summary measures of brain imaging and genetic data for over 50,000 individuals (Thompson et al., 2014).

Each of these multi-modal approaches promise to be important topics of future research, and to fully realize their promise, novel statistical techniques will be needed. Ultimately, combining information from different modalities is challenging to data analysts, if for no other reason than that the amount and variety of data will significantly increase. In addition, since different modalities are measuring fundamentally different quantities, it is not immediately clear how to best combine the information. However, clearly, this is an extremely important problem that has already started to become a major area of research.

III. fMRI MEASURES: SIGNAL ACQUISITION AND PHYSIOLOGY

MR physics and BOLD basics

Both structural and functional MRI images are obtained using the same scanner; the only difference is in how the scanner is programmed. A brief overview of the image acquisition process is as follows. A sample (e.g., a brain) is placed in a strong magnetic field and exposed to a radiofrequency (RF) electromagnetic field pulse. The nuclei absorb the energy only at a particular frequency band, which is strongly dependent on their electromagnetic environment, and 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. Pulse sequences, or software programs that implement particular patterns of RF and gradient magnetic field manipulations, are used to acquire data that can be reconstructed into a map of the MR signal sources, i.e., an image of the brain. For more in-depth information, we recommend two very approachable texts (Elster, 1994; Huettel et al., 2004), and more detailed texts for the advanced reader (Haacke, 1999; Bernstein, 2004).

The relaxation process can be described by three values: T_1, T_2, and T_2^*. T_1 and T_2 are constants determined by the spin frequency, field strength, and tissue type (largely based on the hydrogen content, which depends in turn on how much water is in the tissue). T_1 refers to the rate at which spins relax back to alignment with the main magnetic field, and T_2 refers to the rate of attenuation of the magnetic field applied by the RF pulse. T_2^* is like T_2, but depends additionally on local inhomogeneities in magnetic susceptibility that are caused by changes in blood flow and oxygenation, among other factors.

Different pulse sequences—patterns of RF excitations and data collection periods—produce images that are 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 can produce high-resolution depictions of the boundaries between gray matter (mostly cell bodies), white matter (mostly axons), and cerebrospinal fluid (CSF, mostly water). An example of the same slice of tissue imaged with T_1 and T_2 weighting can be seen in Figure 2A. The images look strikingly different.

Changing the contrast mechanism can be very useful in differentiating brain structures or lesions, since some structures will be apparent in some kind of images but not in others. For example, multiple sclerosis lesions are virtually invisible in T_1 weighted images, but appear very brightly in T_2 weighted images.

Because T_2^* is sensitive to flow and oxygenation, unlike T_1- and T_2-weighted images, T_2^*-weighting is used to create images of brain function. T_2^*-weighted images form the basis of functional MRI.

BOLD Physiology

Unlike PET, which can provide measures of both (a) overall activation related to metabolism and blood flow or (b) specific neurochemical systems, fMRI is principally used to obtain measures of regional brain activity. The most popular type of functional signal, which we focus on, is the Blood Oxygenation Level Dependent (BOLD) signal (Kwong et al., 1992; Ogawa et al., 1992), which is obtained using T_2^*-weighted images.

BOLD imaging takes advantage of the difference in T_2^* between oxygenated and deoxygenated hemoglobin. As neural activity increases, so does metabolic demand for oxygen and nutrients. Capillaries in the brain containing oxygen and nutrient-rich blood are separated from brain tissue by a lining of endothelial cells, which are connected to astroglia, a major type of glial cell that provides metabolic and neurochemical-recycling support for neurons. Neural firing signals the extraction of oxygen from hemoglobin in the blood, likely through glial processing pathways (Sibson et al., 1997; Shulman et al., 2004). As oxygen is extracted from the blood, the hemoglobin becomes paramagnetic, which creates small distortions in the magnetic field that cause dephasing of the protons or 'spins,' resulting in a faster decay of the signal and a lower T_2^*. Initial increases in deoxyhemoglobin can lead to a decrease in BOLD signal, often referred to as the "initial dip." The initial decrease in signal is followed by an increase, due to an over-compensation in blood flow that results in an increase in oxygenated hemoglobin (Figure 9A). The inflow of diamagnetic oxygenated blood leads to less local field inhomogenities, less dephasing of spins, and hence longer T_2^* and more measured signal. The longer T_2^* relaxation time of oxygenated compared to deoxygenated blood is the basis for the BOLD signal (Ogawa et al., 1990).

How well does BOLD signal reflect increases in neural firing? The answer to this important question is complex, and understanding the physiological basis of the BOLD response is currently a topic of intense research (Buxton and Frank, 1997; Vazquez and Noll, 1998; Heeger and Ress, 2002; Buxton et al., 2004).

Essentially, the BOLD signal corresponds relatively closely to the local electrical field potential surrounding a group of cells—which in turn is likely to reflect changes in post-synaptic activity under many conditions. Demonstrations by Logothetis and colleagues have shown that BOLD activity closely tracks the position of neural firing and local field potentials in monkey visual cortex, even to the locations of specific columns of cells responding to particular line orientations (Logothetis et al., 2001). However, under other conditions, neural activity and BOLD signal may become decoupled (Disbrow et al., 2000). Thus, for these reasons and others, BOLD signal is only likely to reflect a portion of the changes in neural activity in response to a task or psychological state.

Another important question is whether BOLD signal increases reflect neural excitation or inhibition. Some research supports the idea that much of the glucose and oxygen extraction from the blood is driven by glutamate metabolism, a major (usually) excitatory transmitter in the brain released by 60-90% of the brain's neurons (Shulman and Rothman, 1998). This is because glutamate is thought to be involved in generating the signals that trigger glucose uptake from blood vessels. However, this is not the whole story, and in some cases BOLD increases may be caused by activation of inhibitory interneurons as well.

Given these ambiguities, one might reasonably ask whether BOLD signal increases linearly with increases in cognitive effort, which we define for present purposes as the metabolic demand involved in engaging in a mental process. In addition to issues of what physiological processes BOLD signals sample, floor and ceiling effects could result in insensitivity to task/mental state demands, resulting in null findings. The answer to this question depends on the precise task, mental state, experiment, the subject's expertise, and brain region(s) tested. A helpful distinction between cognitive effort and cognitive work (what has been accomplished by the cognitive effort) illuminates two sets of findings. First, experts are able to achieve the same outcome with less cognitive effort

compared to novices. For example, expert drummers easily process congruent visual and auditory presentations of drumming, with BOLD responses in the cerebellum lower than those of novices (Petrini et al., 2011). Second, in *repetition suppression* experiments, stimulus repetition can lead to lower BOLD responses in sensory cortices while subjects still perceive the stimulus – the same percept is accompanied by reduced fMRI signal (Henson et al., 2000; Summerfield et al., 2008). These examples illustrate instances in which the BOLD signal may not increase linearly with task demands. Fortunately, and perhaps surprisingly, BOLD signal does go up approximately linearly in appropriate brain regions with increasing demand on visual processing (Boynton et al., 1996), reaction time (Grinband et al., 2011), subjective value (Hare et al., 2009), pain (Bornhovd et al., 2002; Buchel et al., 2002; Atlas et al., 2014), and other conditions. Such demonstrations that the BOLD signal is sensitive to particular mental processes in a specific psychological intensity range are important, because they help ensure that brain measures will be sensitive subsequent tests, e.g., those that try to augment or inhibit the mental state.

Practical considerations (acquisition)

There are a number of critical determinations that go into designing an fMRI study (for aspects regarding experimental design see *Practical considerations)*. One set of decisions concerns the desired spatial and temporal resolution of the study. The temporal resolution determines our ability to separate brain events in time. In fMRI its value depends upon how quickly each individual image is acquired, i.e. the TR. In contrast, the spatial resolution determines our ability to distinguish changes in an image across different spatial locations. The manner in which fMRI data is collected makes it impossible to simultaneously increase both, as increases in temporal resolution limit the number of measurements that can be made in the allocated sampling window and thereby directly influence the spatial resolution of the image (Figure 4). Therefore there are inherent trade-offs required when determining the appropriate spatial and temporal resolutions used in an fMRI experiment. A major exception to this trade-off is multi-slice sequences that simultaneously acquire date from multiple slices of a volume and thereby drastically increase the temporal resolution.

The main limitation in terms of temporal resolution for standard sequences – besides the slowness of the hemodynamic response – is the time T2* time that has to be considered for each slice. If the whole brain is to be imaged, at least 38 slices at about 4 mm thickness are needed. This typically takes about 2 or more seconds. It is possible to reduce brain coverage by measuring fewer slices per volume and achieving a faster TR. Another option is to decrease slice thickness, and thus increase spatial resolution, while keep the number of slices and TR constant, but also reducing brain coverage.

The spatial resolution of fMRI studies is typically on the order of 3x3x4mm, corresponding roughly to image dimensions on the order of 64x64x38 voxels. However, combining higher field strengths and new acquisition techniques allows for much higher spatial and temporal resolution. For example, multi-slice sequences excite multiple slices (typically 2-8 slices) at the same time, thereby drastically reducing sampling rates. Currently it is possible to acquire 2 x 2 x 2 mm data across the brain in less than 1 second. An important aspect to keep in mind is that smaller voxels are less prone to susceptibility artifacts, but have lower signal-to-noise ratio (SNR). The SNR is critical for detecting changes in the signal induced by the experiment. Table 3 summarizes some of the acquisition parameter choices for fMRI experiments.

As previously mentioned, respiration and cardiac pulsation induce artifacts in functional images. Almost all MR systems offer the possibility to record respiratory and finger pulse data during acquisition of functional images. These data can be used to reduce artifacts related to these physiological processes. This can be done on the acquired images (Glover et al., 2000) or within the subject-level GLM (Deckers et al., 2006; Brooks et al., 2008). Correcting for physiological noise has been shown to be beneficial for amygdala imaging (Boubela et al., 2015) and is necessary for spinal cord fMRI (Kong et al., 2012).

IV. USING FMRI TO MAKE INFERENCES ABOUT BRAIN AND MIND

Interpretation of fMRI studies

Forward inference and reverse inference

A fundamental question in neuroimaging research, and a good starting point for planning a study, is determining what question one hopes to answer with the study. Successful research requires a solid grasp of how neuroimaging results can and cannot bear on psychological or physiological theory, and a grounded understanding of what kinds of results are likely to be obtainable.

There are several potential inferential goals in neuroimaging studies. One goal is prediction of a psychological or disease state using neuroimaging data, which can be accomplished using regression or classification techniques (Norman et al., 2006). 'Prediction' can literally refer to predicting the future, e.g., to detecting early-onset Alzheimer's disease before other measures. But it can also be much more, including developing measures that track mental events or other outcomes (e.g., performance) so that brain-based measures can support or even replace those outcomes when they are suspect or unavailable. For example, pain and emotion are typically measured in terms of self-reports, which are appropriate in many circumstances; but self-report has fundamental limitations and biases, and progress in research may require complementary, objective measures. Neuroimaging-based measures are unique in this respect because they are close to the neurophysiological mechanisms that generate pain and emotion, and so can provide more clues about their mechanisms.

Another important goal, and the traditional one in cognitive neuroscience, is to infer something about the structure of mental processes from imaging data—i.e., to use neuroimaging to inform *psychological* theory. This is a difficult goal, and some psychologists have argued that it has not been achieved in any domain (Coltheart, 2006). However, it is possible under some circumstances. Making inferences about psychological states has been termed *reverse inference*, because it involves inferences about the state of the mind given some observed measures of the brain. Several excellent papers review some of the issues related to reverse inference in brain imaging (Sarter et

al., 1996; Poldrack, 2011) and physiological data generally (Cacioppo and Tassinary, 1990).

Valid reverse inference requires strong knowledge of what types of mental events a physiological measure can and cannot reflect. For example, let's say we apply a new drug to the skin and want to know if it affects *pain* – a subjective mental state (Figure 5). Let's say we measure signaling in pain-specific neurons "nociceptors" demonstrated to respond only to painful events, and we find that the drug suppresses their firing. We might then be justified in inferring that the neurophysiological mechanisms that give rise to pain (or at least some forms of pain) have been disrupted. Of course, we would still want to know whether people reported less pain (a behavior). But there may be circumstances in which people are not able to communicate their pain effectively, and we may want to know if the drug effects occur in neurophysiological systems that normatively give rise to pain, or those associated with emotions beyond pain specifically or other decision-making or social cognitive processes.

Often, reverse inference is done improperly and heuristically, leading to the impression that it cannot be done at all. However, reverse inference can also be done formally, with conclusions valid for the range of hypotheses considered (see Chapter 1 for an in-depth treatment of valid inference and Bayesian approaches).

In experimental studies, a psychological state is typically manipulated, and we calculate statistics related to the probability of observing the brain data given the psychological state. This probability is called *sensitivity* in testing theory and *forward inference* in the neuroimaging literature. To make reverse inferences about psychological states, we must estimate the relative probabilities of a defined set of psychological hypotheses given the data, typically using Bayes Rule. This requires assessing both sensitivity and *specificity*—the probability of not observing a brain pattern in the absence of a particular psychological state—across a range of potential states. If sensitivity and specificity are high enough, a brain measure may have high *positive predictive value*—that is, a high likelihood of implying a particular mental state or class of states (Poldrack, 2011).

As noted above, reverse inference based on activation in single brain regions is

problematic. For example, researchers have inferred that romantic love and retribution involve "reward system" activation because these conditions activate the caudate nucleus (de Quervain et al., 2004; Aron et al., 2005), that social rejection is like physical pain because it activates the anterior cingulate (Eisenberger et al., 2003), and many other, similar conclusions. These inferences are problematic because both these regions are involved in a wide range of tasks, including shifting of attention, working memory, and inhibition of simple motor responses, so their activation is not indicative of any particular psychological state (Bush et al., 2000; Kastner and Ungerleider, 2000; Paus, 2001; Wager et al., 2004b; Wager et al., 2005b). That is, the specificity of activation in these regions is low, and so they have little positive predictive value for any particular state.

These examples demonstrate the broader principle that overlapping brain activity is not sufficient to infer overlap in neural or mental processes.

From regions to patterns: Enhanced potential for inference

The type of 'reverse inference' discussed so far relates to inferences based on activation of a particular brain region (e.g., the primary visual cortex, the hippocampus, etc.). Strong inferences based on activation of a region are rarely valid, because (a) specificity of activation of a region is too low; (b) the definition of what constitutes activation is not precise—exactly which voxels and their relative activity levels should be specified; and (c) many psychological processes are distributed across brain networks, and activity in one region is insufficient to characterize them even in principle.

However, it is possible to apply the same logic for making reverse inferences to other types of brain measures as well. Rather than focusing on single regions, recent studies have identified *patterns* of activity across brain regions. Increasingly, the sensitivity and specificity of these patterns are being assessed, in a few cases across studies and laboratories, and they are thus being evaluated as markers for particular mental processes. For example, our lab is interested in the development identifying whole brain patterns that have positive predictive power for pain perception. This is currently possible for cutaneous heat pain (Wager et al., 2013); a distributed pattern-based marker called the Neurologic Pain Signature has high sensitivity and specificity in discriminating

painful heat from non-painful warmth, pain anticipation, and pain recall (> 90%). This pattern is also able to discriminate painful heat from social rejection, which has been claimed to involve the same processes as pain experience.

Dissociation logic

Another type of reverse inference is less specific about the localization of psychological functions in the brain but in some cases more defensible. Based on dissociations in activation among tasks on can learn about representations of mental states. This is used in studies that test two or more tasks in the same experiment. Dissociation occurs when a brain region is more active in Task A than Task B. A double dissociation occurs when each task activates one region more than the other task. Double dissociations are a powerful tool because they imply that the two tasks utilize different processes, and that one task is not a subset of the other. These kinds of inferences can both be answered using classical univariate approaches or multivariate approaches (see *Comparison of univariate and multivariate techniques*).

Though double dissociations are potentially powerful, they have been criticized on several counts. For one thing, nonlinear relationships between task demands and activation can produce a double dissociation even if there are no processes unique to each task. Sternberg (Sternberg, 2001) has proposed a stronger criterion for task separability called 'separate modifiability', which entails finding outcomes that are affected by each task but *not* the other task, which is a stronger criterion for the separability of two brain processes. Secondly, even if double dissociations or separately modifiable brain measures demonstrate that there are unique brain processes involved in each task, it does not strongly imply that the brain processes are those the investigators are interested in. Therefore, if the brain processes that are dissociated are also predictive of psychological or behavior outcomes of interest, we can make stronger inferences that the brain processes involved in the behaviors are separable.

For example, consider a recent study that looked at the overlap between physical pain and romantic rejection (Woo et al., 2014b). Physical pain was induced by noxious heat (somatic stimulation), and feelings of rejection were induced by showing

participants pictures of their ex-romantic partners. The research team identified separate brain patterns that were responsive to physical pain but not rejection and vice versa, demonstrating separate modifiability. But does that mean that the brain processes that give rise to feelings of pain and rejection are separable? This evidence alone is not enough, because the separately modifiable brain patterns could simply be related to the stimulus modality (touch vs. vision) rather than the feelings involved. In this case, further demonstration of separate modifiability in all the brain regions thought to encode physical pain—including the anterior cingulate and insula, thought to be the seat of shared representations—provided stronger evidence that the *relevant* brain processes were dissociable. Even stronger evidence would be provided if the brain patterns involved were demonstrated to be sensitive and specific to pain across studies. This was the case for the physical pain pattern in our example study, but the sensitivity and specificity of rejection-related brain patterns remains to be assessed.

Interpretation of overlapping brain signals

The complement to dissociations, which argue for separability of brain processes, is inferences based on overlap in patterns of activity, which is often taken as evidence that the tasks share common processes (Sylvester et al., 2003). Though the logic that *activation overlap* equals *process overlap* is commonly used, it provides weak support for shared neuronal processes: A single voxel in a neuroimaging study typically contains on the order of 5.5 million neurons, and it is entirely possible that different subsets of neurons in the same voxel are activated by different tasks (Logothetis, 2008). Paton et al. (Paton et al., 2006), for example, found different cells in the monkey amygdala that respond to either positive or negative predictions about upcoming rewards within the volume of a single neuroimaging voxel. Recent optogenetic studies, which can experimentally manipulate the firing of specific, genetically tagged subpopulations of neurons with light, are increasingly identifying microcircuits with different, and often opposing, functional properties (Tye et al., 2011; Kvitsiani et al., 2013). Activation of distinct microcircuits is likely to produce similar profiles of activation in fMRI and PET studies. Thus, two tasks that activate any given brain region might do so for very

different reasons. The difference in activation patterns elicited by functionally distinct neuronal circuits may not be evident in univariate analyses. Multivariate techniques that analyze multiple voxels at a time may be able to pick up the subtle differences on a voxel level. In addition, neurons involved in different functional microcircuits often project to different areas of the brain, suggesting that patterns of long-range fMRI functional connectivity may be useful in disentangling them in some cases. We discuss these techniques in more detail below.

Comparison of univariate and multivariate techniques

During the last few years, multivariate analyses methods have gained enormous popularity. While this chapter provides a general overview of functional neuroimaging, we refer the reader to one of the several excellent papers covering multivariate fMRI methods in more depth (e.g., (Kriegeskorte et al., 2009; Haynes, 2015).

In the univariate, statistical parametric mapping (SPM), type of analysis, brain responses are modeled using a GLM separately for every voxel. The subject level GLM is often defined in a way that allows the parameter estimates in each voxel to be interpreted as the amplitude of the response to a specific experimental condition. The group statistic computed in every voxel then indicates how likely it is that a response of this magnitude occurred by chance. This approach tries to explain the fMRI data by the experimental condition, say viewing pictures of faces or houses. It is sometimes referred to as ‘encoding’ analysis.

Multivariate analyses, also referred to as ‘decoding’ approaches, use data from multiple voxels at the same time (hence the term multivariate) to model experimental conditions (say, looking at faces or houses). Now, the brain data constitute a set of predictors, and the experimental variable the outcome. The term multivariate analysis or multivariate pattern analysis (MVPA) does not refer to a single method, but rather a large family of multivariate techniques. These kinds of analyses can be executed on raw data, selective trial averages, or on parameter estimates from subject level GLMs. In addition, multivariate analysis can be extended to cover many types of both continuous and categorical outcomes, both within- and between-persons, including the perceptual

characteristics during natural viewing, performance, emotional experiences, age, clinical symptoms or status, and more.

Put another way, 'encoding' models and 'massively univariate' analyses are typically univariate in the brain (analyzing one voxel at a time) and multivariate in psychological/behavioral space, and 'decoding' models are typically multivariate in brain space and univariate in psychological space. Some techniques, such as partial least squares and canonical correlation, are multivariate in both brain and psychological space.

Multivariate techniques differ from univariate approaches in that they (i) simultaneously analyze multiple voxels, and (ii) use brain data as predictors of outcomes of intrinsic interest (switching the predictors and predicted variables). A major benefit of analyzing multiple voxels at a time is that it takes into account the spatial interdependencies across voxels. Each voxel's response is analyzed while controlling for other voxels in a set. Common choices of voxel sets are spheres of voxels (searchlight), anatomical regions of interest, or whole brain (e.g., all gray-matter voxels). The simultaneous analysis of multiple voxels can thus pick up on patterns across brain space—i.e., the *relative* activity across a set of voxels—that a univariate analysis cannot. In many cases, multivariate techniques likely offer enhanced sensitivity. For example, using a multivariate technique called support vector machine (SVM), allowed researchers to discriminate different emotions from auditory stimuli which was not possible using univariate methods (Ethofer et al., 2009).

Common choices of multivariate algorithms are SVM's and linear discriminant analysis (LDA) to distinguish between categorical variables of interest (e.g., perception of left vs. rightward oriented gratings (Kamitani and Tong, 2005)). Principal component regression (PCR) or support vector regression are often chosen to model continuous variables (e.g., perceived pain, (Wager et al., 2013). It is important to note that there is no single best algorithm for all questions. The best algorithm for a given dataset is the one that best matches the process that generated the data – i.e. whose assumptions are most correct for the process of interest.

Due to their high sensitivity for difference in activation patterns, multivariate methods are able to predict many variables that do not correlate with univariate voxel

responses (Ethofer et al., 2009). However, the term prediction is often used in a misleading way. The switch of independent and dependent variable does not imply changes in causality. The causal nature of the analysis still depends on the experimental design. The mere possibility of 'prediction' does not entail causality (Friston, 2009). A prediction can also be made using univariate analysis, but the term is not regularly used in this context. For example, by evaluating the response magnitude of a single voxel in the fusiform face area, one can make predictions of whether the subject saw a face or a house during a given trial. Since this area is particularly sensitive to faces, one would simply predict that the subject was looking at a face for all responses of this voxel above a certain threshold. Nevertheless, the multivariate model should outperform the univariate prediction in most cases because it based on more data.

Designs for fMRI studies

Experimental designs

Designing a neuroimaging study involves a series of tradeoffs between experimental power and the ability to make strong inferences from the results. Some types of designs, such as block designs, typically yield high experimental power, but provide imprecise information about the particular psychological processes that activate a brain region. Event-related designs, on the other hand, allow brain activation to be related more precisely to the particular cognitive processes engaged by particular mental events, but often are reduced in power to detect activation, depending on the process being studied. Researchers may also choose to focus intensively on testing one comparison of interest, and maximizing the power to detect this particular effect, or they may test multiple conditions in order to draw inferences about the generality of a brain region's involvement in a class of similar psychological processes. Below we describe several types of experimental designs and provide some discussion of the applications for which they are best suited.

Block designs

Because long intervals of time (30 seconds or more) are required to obtain good PET images, the standard experimental design used in PET studies is the block design. A block design is one in which different conditions in the experiment are presented as separate blocks of trials. For example, to image a briefly occurring psychological process (e.g., activation due to attention switching) using 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 (A minus B) comparison is the most basic type of contrast for this design. The block structure of PET designs (and block fMRI designs) imposes limitations on the interpretability of results. While activations related to slowly changing factors such as task-set or general motivation are well captured by block designs, they are not well suited if one wishes to image the neural responses to individual stimuli. In addition, the A – B contrast does not allow researchers to determine whether a region is activated solely in A, deactivated solely in B, or some combination of both effects. Multiple controls and comparison conditions can ameliorate this problem to some degree.

The main advantage to using a block design is that it typically offers increased statistical power to detect a change. Under ideal conditions, it has been shown that block designs can be over 6 times as efficient as randomized event-related designs (Wager and Nichols, 2003). Generally, theory and simulations designed to assess experimental power in fMRI designs point to a 16-18 s task / 16-18 s control alternating-block design as being optimal with respect to statistical power (Skudlarski et al., 1999; Wager and Nichols, 2003; Liu, 2004).

However, it is worth noting that this is not always true, as the relative power of a block design depends on whether the target mental process is engaged continuously in A and not at all in B, and whether imposing a block structure changes the nature of the task. For example, the updating of internal predictions based on visual information elicits so-called ‘prediction errors’ whose associated neuronal firing lasts for only a very short time. Here, a block design is a bad choice, because the signal of interest will decay quickly and the block design will mis-model the neuronal responses. An event-related design will yield higher power and better interpretability in this case and similar ones. In

the end, it is important to consider how the temporal structure influences the magnitude of the underlying psychological and neural events being studied as well as its impact on the ability to detect signals in the fMRI environment.

Event-related fMRI

Event-related fMRI designs take advantage of the rapid data-acquisition capabilities of fMRI. They provide the ability to estimate the fMRI response evoked by specific stimuli or cognitive events within a trial (Rosen et al., 1998). With modern multi-slice sequences the whole brain can be measured with standard spatial resolution (e.g., 3 x 3 x 3 mm voxels) every 0.5 seconds. The limiting factor in the temporal resolution of fMRI is generally not the speed of data acquisition, but rather the speed of the underlying evoked hemodynamic response to a neural event, referred to as the hemodynamic response function (HRF). A typical HRF begins within a second after neural activity occurs, and peaks 5-8 seconds after that neural activity has peaked (Friston et al., 1995; Aguirre et al., 1998).

While event-related designs are attractive because of their flexibility and the information they provide about individual responses, they rely more strongly on assumptions about the time course of both evoked neural activity and the HRF. It is common to assume a near-instantaneous neural response for brief events and a canonical HRF shape in order to generate linear models for statistical analyses. In practice, however, the timing and shape of the HRF are known to vary across the brain, within an individual, and across individuals (Schacter et al., 1997; Aguirre et al., 1998; Summerfield et al., 2006). Part of the variability is due to the underlying configuration of the vascular bed, which may cause differences in the HRF across brain regions in the same task for purely physiological reasons (Vazquez et al., 2006). Another source of variability is differences in the pattern of evoked neural activity in regions performing different functions related to the same task.

Block designs are less sensitive to the variability of the HRF because they depend on the total activation caused by a *train* of stimulus events, which makes the overall predicted response less sensitive to variations in the shape of responses to individual

events. However, predicted responses in block designs may still be quite inaccurate if the HRF model is very inaccurate or if the density and time-course of neural activity is not appropriately modeled (Price et al., 1999), or if complex responses cause signals from different events to cancel each other out (Gonzalez-Castillo et al., 2012).

Event-related designs rely on the response estimation of voxels to single trials or brief events. The underlying assumption is that the magnitude and shape of the BOLD response do not change depending on the preceding stimuli. Studies have found that nonlinear effects in rapid sequences (1 or 2 s) can be quite large (Vazquez & Noll, 1998)(Friston et al., 2000; Birn et al., 2001; Wager et al., 2005a), but that responses are roughly linear if events are spaced at least 4 s – 5 s apart (Miezin et al., 2000). If they are properly designed, rapid designs still allow one to discriminate the effects of different conditions. One key is incorporating 'jitter,' or variable inter-stimulus interval (ISI), between events, which is critical for comparing event-related responses to an implicit resting baseline—i.e., determining 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 because the hemodynamic response to closely spaced events sums in a roughly linear fashion, the expected response to that trial-type will build to a high peak. Introducing jitter allows peaks and valleys in activation to develop that are specific to particular experimental conditions. If one cares only about comparing event types (e.g., A – B), randomizing the order of events creates optimal rise and fall without additionally jittering the ISI. However, jittered ISIs are critical for comparing events to baseline activity and thus determining whether events *activate* or *deactivate* a voxel relative to that baseline (Josephs and Henson, 1999; Wager and Nichols, 2003). Suppose, for example, you have a rapid sequence with two types of trials—say, attention-switch trials (S) and no-switch trials (N) as in the task switching experiment described above. Randomly intermixing the trials with an ISI of 2 s will allow you to estimate the difference S – N. However, you will not be able to tell if S and N *activate* or *deactivate* relative to some other baseline. If you vary the inter-stimulus intervals randomly between 2 and 16 s, you'll be able to compare S – N (with less power because there are fewer

trials), but you'll also be able to test whether S and N show positive or negative activation responses. This ability comes from the inclusion of inter-trial rest intervals against which to compare S and N, and the relatively unique signature of predicted responses to both S and N afforded by the random variation in ISIs.

The advantages of rapid pacing—including faster trials and possible increased statistical efficiency—must be weighed against potential problems with nonlinearity, multicolinearity of regressors, and model mis-fitting. A current popular choice is to use 'jittered' designs with inter-stimulus intervals of at least 4 s, with exponentially decreasing frequencies of delays up to 16 s.

Optimized experimental designs

What constitutes an optimal experimental design depends on the psychological nature of the task as well as the ability of the fMRI signal to track changes introduced by the task manipulations over time. It also depends on the specific comparisons (contrasts) of interest in the study. And to make matters worse, the delay and shape of the BOLD response (and ASL signals, and other blood flow-based methods), scanner drift and nuisance factors such as physiological noise, and other factors conspire to make experimental design for fMRI more complicated than for experiments that measure behavior alone. Not all designs with the same number of trials of a given set of conditions are equal, and the spacing and ordering of events is critical.

Some intuitions and tests of design optimality follow from a deeper understanding of the statistical analysis of fMRI data and are elaborated on in Section 0. For a full treatment, however, we refer the reader to several excellent papers (Josephs and Henson, 1999; Wager and Nichols, 2003; Liu, 2004; Smith et al., 2007). We also note that several computer algorithms are available for constructing statistically optimized designs, including an approach based on m-sequences - mathematical sequences which are near-optimal for certain types of designs (Buracas and Boynton, 2002), and approaches based on genetic algorithms (Wager and Nichols, 2003; Kao et al., 2009), that incorporate m-sequence designs as a starting point and considers the relative importance of various contrasts to the study goals in calculating optimality.

Figure 6 plots the power of different designs based on effect sizes estimated from visual cortex data (Wager et al., 2005a). Block designs have large power for estimating contrast effects that are based on amplitude differences, whereas event-related and m-sequence designs have more power in HRF shape estimation. Optimized designs offer a balance between the two.

Resting state

The majority of fMRI (and EEG/MEG) studies are still studying brain activations related to cognitive tasks, perception, and action. However, some years ago Biswal et al. (1995) observed that the BOLD time-courses in left and right sensorimotor cortices were highly correlated at rest (Biswal et al., 1995), suggesting that much of the 'noise' in these regions, and possibly the rest of the brain, was not noise at all but rather coherent spontaneous activity. Further studies identified a set of large-scale networks that show correlated activity during rest in the absence of any task (Raichle et al., 2001; Fox et al., 2005; Buckner et al., 2008). These networks are most often identified using clustering approaches on pairwise correlations or data-decomposition algorithms such as Independent Components Analysis (ICA) or Principal Components Analysis (PCA); voxels that load highly on the same component are thought to comprise the 'network.' Often, voxels are additionally assigned to discrete, non-overlapping 'networks' using clustering algorithms. For instance, a study based on a large sample of 1000 subjects grouped brain regions in the cortex into 7 and 17 large-scale networks (the choice of how many to extract is to some degree arbitrary) (Yeo et al., 2011). Other studies have found that 'networks' derived from resting state scans are in broad agreement with clusters obtained from structural connectivity measures (Honey et al., 2009; Wiech et al., 2014).

These networks can be reliably identified in different samples, and they are often labeled with psychological terms and used as units of analysis in other studies. The 'default mode network' (DMN) (Raichle et al., 2001), includes the ventromedial and dorsomedial prefrontal cortices (vmPFC/dmPFC), posterior cingulate, medial temporal lobe, superior temporal cortices, and several other areas. The name is based on observations that many of its regions show high metabolic activity when a person is 'at

rest' (not doing a task) and decrease during the performance of many cognitive tasks. However, so-called DMN regions are activated above resting levels by a number of tasks focused on reflection on internal states, including retrieval of semantic memories (Binder et al., 2009), imagining the future (Schacter et al., 1997), experiencing psychological stress (Wager et al., 2009; Gianaros and Wager, 2015), experiencing emotion (Kober et al., 2008; Lindquist et al., 2012a), reflection on one's self (Northoff et al., 2006; Denny et al., 2012), reflecting on others' minds (Denny et al., 2012), and 'mind-wandering,' a mix of often self-focused thoughts and memories (Andrews-Hanna et al., 2010).

Many other networks have been identified and labeled with terms that imply they implement specific functions. The 'salience network,' for example, includes regions activated by many cognitive and affective states, including the dorsal anterior cingulate, anterior insula, and amygdala (Seeley et al., 2007). Regions in this 'network' certainly respond to many kinds of salient events, but it would be a mistake to make the fallacious *reverse inference* that a task activates the network because it is 'salient.' As we discussed above, specific neurons in these regions participate in micro-circuits that encode specific, and diverse, types of information and behavior.

Resting-state studies have become increasingly popular, and there is much hope that they will provide markers for characteristics related to aging, psychopathology, performance, and clinical symptoms. These studies do not employ a specific task or experimental manipulation, but rather acquire fMRI data while the subjects rest in the scanner. Most studies display a fixation cross during the measurement and ask subjects to look at the crosshair. Another approach is to minimize visual input and have subjects close their eyes during the scan. Typical scan durations are 5 – 12 minutes per subject, making it easy and cost-effective to acquire data in many subjects.

The analysis of resting state data is different from experimental fMRI studies. Since there is no experimental manipulation, a conventional GLM analysis is impossible. Instead, most of the techniques are analyzing the correlational structures among voxels. The analysis of resting state involves first estimating brain connectivity measures—using 'seed' regions, ICA, or voxel-by-voxel pairwise inter-correlation matrices across the brain. Then, those connectivity metrics are correlated with outcomes of interest—for

example, clinical symptom scores. For an overview of connectivity and correlation based analysis see *Connectivity analyses in fMRI*.

Though increasingly popular, resting state analyses are not without serious pitfalls. One is ambiguity, and person-to-person variability, regarding what mental states and physiological processes are actually being imaged. A large amount of research funding is currently dedicated to exploring the idea that resting state connectivity will be able to tell us about depression, anxiety, dementia, cognitive and emotional development, and a host of other outcomes of interest. However, at least some of the coherent brain activity observed at rest is demonstrably due to physiological noise, including artifacts related to head movement, respiration (which affects fMRI signal via inducing head movement, magnetic field currents, and changes in carbon dioxide levels), pulsatile motion, and vascular oxygenation due to heartbeat. In addition, though it is often implicitly assumed that participants are complying with task instructions and are all equally awake and alert, this is clearly not the case. A recent study found that 50% of participants in resting state studies are asleep after 10 minutes (Tagliazucchi and Laufs, 2014). Since activity patterns and neuronal oscillations change drastically during the transition from wakefulness to sleep, it is important to control for wakefulness during the scan and carefully check potential group differences. In addition, activity patterns consistent with resting state networks are present even in anesthetized animals (Vincent et al., 2007). And finally, different patterns of resting state connectivity are related to different types of spontaneous thought (Andrews-Hanna et al., 2010; Doucet et al., 2012). Whereas the goal of experimental paradigms is to explicitly control the types of mental processes in which a participant engages and study brain activity in relation to those processes, resting state studies do not control the types of mental processes that a participant engages in.

Thus, for some researchers, resting state scans are viewed as a window into the intrinsic architecture of the brain; for others, they are windows into mental states or mental status, or physiological artifacts to be discarded. The trouble is that it is hard to tell how much of the brain connectivity patterns at rest are related to which of these three alternatives. Even if outcomes are reliably associated with resting state networks, it may

not be clear why, or whether the associations have interesting implications for neuroscience or are merely physiological or image artifacts. The utility of resting-state fMRI, like all areas of scientific inquiry, is ultimately an empirical question that is being asked now in myriad ways.

Non-experimental designs

The fast growth in computing power together with the introduction of multivariate techniques into fMRI paved the way for large-scale decoding studies. The aim of these studies is to study brain processes of natural vision. In order to achieve higher external validity as in natural conditions, experimental control is reduced. However, compared with traditional experiments, these designs have the potential to establish profiles of brain activity, and their specificity to particular mental states, across a wide range of more naturalistic conditions.

Early approaches used quasi-experimental designs to search for brain regions whose activity tracks conscious perception. These studies use multi-stable visual stimuli (e.g. a Necker cube) that lead to fairly regular, spontaneous switches in conscious percepts. Subjects are asked to report the perceptual switches via button-presses and researchers can analyze responses following perceptual switches. An early univariate fMRI study reported phasic positive responses in the fusiform gyrus and negative responses in the thalamus (Kleinschmidt et al., 1998). A later study using multivariate analyses was able to predict the current percept from activity in the lateral geniculate nucleus, an early visual processing nucleus in the thalamus (Haynes et al., 2005).

To achieve even more natural viewing conditions across a wide range of stimuli, it is increasingly common to present movies or podcasts to their subjects while measuring fMRI data. Studies aimed at mapping responses within individuals can include data collected over 10 hours or more, across multiple sessions. The enormous amount of data is then used to predict current perceptions from brain activity by exploiting the unique covariation patterns between brain activity and features of the current stimulus composition (Haxby et al., 2011; Huth et al., 2012; Horikawa et al., 2013).

Practical considerations (design, power)

Designing a neuroimaging study involves a tradeoff between experimental power and the ability to make strong inferences from the results. Some types of designs, such as the block design, typically yield high experimental power, but provide imprecise information about the particular psychological processes that activate a brain region. They also rely on the ability of the task to activate neuronal populations for the duration of a whole block (see *Block designs*). Event-related designs, on the other hand, allow brain activation to be related more precisely to the particular cognitive processes engaged in certain types of trials, but often suffer from decreased power. The choice of the design should thus be guided by the research question, the underlying psychological model, and estimated effect sizes. For valid inference it is necessary that task is appropriate to isolate the psychological process of interest. Increasing the sample size can often compensate a relative loss in power. Sometimes technical constraints limit the choice of the design; for example, heat pain studies are typically done using sustained heat epochs, essentially like block designs, because many heat stimulation devices were unable to change the temperature fast enough for event related designs.

Another major aspect of planning a neuroimaging study is the desired statistical power and the question of how to best achieve it. Statistical power depends on having either a large effect size (high contrast values) or a small standard error. The standard error in a group analysis is determined by both σ^2_W and σ^2_B. At the group level, σ^2_B can be reduced and power increased by increasing the sample size, more accurate normalization or more informed ROI selection, and increased control of strategies used and individual psychological responses to the task. σ^2_W can be reduced by improving modeling procedures and reducing acquisition-related scanner noise and physiological noise.

A key question when beginning to design a group study is determining an adequate sample size. The answer to this question ultimately depends on the effect size in the group, the amount of scanner noise, and signal optimization. It will be different for each task and each brain voxel (Zarahn and Slifstein, 2001; Desmond and Glover, 2002). Power analysis is difficult in fMRI because power depends on so many factors relating to

psychology, task design and analysis, and hardware—however, by referring to standard effect sizes, one can obtain estimates of what sample sizes are needed in a group analysis. There are several tools for estimating power in fMRI studies. For example, Mumford and Nichols (2008) developed a website and software to estimate group statistical power for the average voxel in regions of interest (http://fmripower.org).

With reduced scanning costs, the sample sizes and statistical power of fMRI studies have increased over the last years. However, many studies still have low power to detect small or medium size effects due to small sample sizes. Some have argued that this is not a real concern, because small sample studies can detect only large effects that are presumably strong enough to be of interest (Friston, 2012). However, such analyses neglect to consider that because of fMRI noise, not all regions identified in small studies actually have large effects! Thus, this view neglects the large confidence intervals and associated uncertainty about the true effect size (Lindquist et al., 2013). Because of the large sampling error associated with estimates from small studies, significant results from small studies are more likely to be inflated by voxel selection bias and thus capitalize on chance. Hence, many positive results from underpowered studies will overestimate the true effect size, giving rise to problems with replication of the results (Button et al., 2013).

One way to consolidate findings and estimate true effect sizes is to use meta-analytic techniques to aggregate across studies (Wager et al., 2007). For these meta-analyses to be unbiased, it is important to also report fMRI results as completely as possible, even non-significant results (e.g., those not surviving multiple comparison correction, but $p < 0.001$, uncorrected) should be reported in supplemental tables when possible.

Figure 7 shows an example of power calculation and variance component estimation from a working memory study. Figure 7A shows the main effect for working memory (an N-back task vs. rest), which we used to identify voxels of interest. We calculated power averaged across these voxels of interest shown in (A) in a different contrast, the more difficult 3-back vs. easier 2-back condition in the N-back. This analysis is illustrative; we note that for a truly unbiased power analysis, the selection of

voxels must be independent of the data used to calculate power. Figure 7B shows plots of power (*y*-axis) as a function of sample size (*x*-axis) for three different significance thresholds. Power will always increase with larger sample sizes, but sample size is always limited in practice. Thus, this analysis assumes a fixed number of scan hours available for a replication study—in this case, 40 total hours. With a few other assumptions, such as a maximum session time of 90 minutes and a 30 min startup cost (for anatomical images, etc.) for the first session and 15 min startup cost for additional sessions (for scanner placement), we can calculate the power as a function of number of subjects and scan time per subject. With a total of only 40 scan hours, the U-shaped function suggests that the optimal allocation is to run 38 people in a session just under 1-hour in length, with about 35 minutes of functional time. This is a typical case with moderately strong activation. The within- and between-subjects noise is roughly balanced (shown in the Venn diagram), and voxel-wise power with 40 hours to allocate is around 15% with family-wise error rate (FWER) multiple comparisons correction control at $p < .05$ corrected. There are many active voxels to detect, so this power level might be acceptable or not, depending on the study goals. This is a sobering analysis however: If one wants to detect *most* of the active voxels with only a 5% chance of a false positive anywhere in the map (FWER control), then large numbers of subjects are needed. Using less stringent forms of control (e.g., False Discovery Rate, discussed below) and specifying precise *a priori* hypotheses can increase power dramatically.

As we said above, the optimal balance of numbers of subjects vs. scan time per subject depends on the ratio of between-subject and within-subject variances. In contrast to the example above, with extremely strong effects and little within-subject error, 80% power is achievable with 15 subjects and about two hours per subject. This type of effect size and error distribution is more typical of visual cortical stimulation (e.g., retinotopic mapping). If you cannot easily estimate this ratio and perform power calculations, then scanning as many subjects as possible with about 30 min of functional time per subject for cognitive studies, and fewer subjects with more time per subject for visual psychophysical studies, is a reasonable rule of thumb.

In addition to aspects of experimental design and statistical power, practical

considerations like session length and subject alertness and focus are important. Most participants feel increasingly uncomfortable as the duration of the imaging session progresses beyond one hour total, with corresponding increases in head movement, pain, and fatigue, and likely reductions in data quality.

V. FUNDAMENTALS OF FMRI SIGNAL PROCESSING AND ANALYSIS

Preprocessing

The major steps in fMRI preprocessing are reconstruction, slice acquisition timing correction, realignment, coregistration of structural and functional images, registration or nonlinear warping to a template (also called normalization), and smoothing (Figure 8). Single-subject analyses do not require the warping step, which introduce spatial uncertainty in terms of anatomical locations, and thus can provide higher anatomical resolution. Group studies, however, largely preclude false positives due to fMRI time series artifacts, and permit population inference. Some group studies do not employ smoothing in order to increase spatial resolution.

Reconstruction. Images must be first reconstructed from the raw MR signal. Reconstruction is commonly automated directly at the scanner site. Raw and reconstructed data are stored in a variety of formats, but reconstructed images are generally composed of a 3-D matrix of data, containing the signal intensity at each "voxel" or cube of brain tissue sampled in an evenly-spaced grid, and a *header* that contains information about the dimensionality, voxel size, and other image parameters. A popular format is the nifti-format, which can hold single or multiple 3-D volumes per file. The format allows storing multiple images in a 4-D matrix, where the fourth dimension is time.

Slice Timing. Statistical analysis at the subject level using a single hemodynamic reference function assumes that all the voxels in an image are acquired simultaneously. In reality, the data from different slices are shifted in time relative to each other—because most BOLD pulse sequences collect data slice-by-slice, some slices

are collected later during the volume acquisition than others. Thus, we need to estimate the signal intensity in all voxels at the same moment in the acquisition period. This can be done by interpolating the signal intensity at the chosen time point from the same voxel in previous and subsequent acquisitions. A number of interpolation techniques exist, from bilinear to sinc interpolations, with varying degrees of accuracy and speed. Sinc interpolation is the slowest, but generally the most accurate. Some researchers do not use slice timing, as it adds interpolation error to the data, and instead use more flexible hemodynamic models to account for variations in acquisition time.

Realignment. A major problem in most time-series experiments is movement of the subject's head during acquisition of the time series. When this happens, the image voxels' signal intensity gets "contaminated" by the signal from its neighbors. Thus, one must rotate and translate each individual image to compensate for the subject's movements. Realignment is typically performed by choosing a reference image (popular choices are the first image or the mean image) and using a *rigid body transformation* of all the other images in the time series to match it, which allows the image to be translated (shifted in the x, y, and z directions) and rotated (altered roll, pitch, and yaw) to match the reference. The transformation can be expressed as a pre-multiplication of the image spatial coordinates to be altered by a 3 x 3 affine matrix. The elements of this matrix are parameters to be estimated, and an iterative algorithm is used to search for the parameter estimates that provide the best match between an image and the reference image. Usually, the matching process is done by minimizing sums of squared differences between the two images.

Realignment corrects adequately for small movements of the head, but it does not correct for the more complex spin-history artifacts created by the motion. The parameters at each time point are saved for later inspection and are often included in the analysis as covariates of no interest; however, even this additional step does not completely remove the artifacts created by head motion. Residual artifacts remain in the data and contribute to noise. Sometimes this noise is correlated with task contrasts of interest, which poses a problem, and can create false results in single-subject analyses. However, because these artifacts are expected to (and typically do) differ in sign and magnitude across subjects,

group analysis is valid. Group analyses are usually robust to such artifacts in terms of false positives, but power can be severely compromised if large movement artifacts are present. An exception is task-correlated motion. When all subjects move their head at the same time as the events of interest, it is not possible to dissociate task from motion artifacts.

Because of these issues, it is typical to exclude subjects that move their heads substantially during the scan. Subject motion in each of the 6 directions can be estimated using the magnitudes of the transformation required for each image during the realignment process, and time series of displacements are standard output for realignment algorithms.

Coregistration. Often, high-resolution structural images (T_1 and/or T_2) are used for warping and localization. The same transformations (warps) are applied to the functional images, which produce the activation statistics, so accurate registration of structural and functional images is critical. Coregistration aligns structural and functional images, or in general, different types of images of the same brain. Because functional and structural images are collected with different sequences and different tissue classes have different average intensities, using a least squares difference method to match images is often not appropriate. For example, the signal intensity in gray matter (G), white matter (W), and ventricles are ordered W > G > V in functional T_2^* images, and V > G > W in structural T_2 images (Figure 1). In such cases, an affine transformation matrix can be estimated by maximizing the *mutual information* among the two images, or the degree that knowing the intensity of one can be used to predict the intensity of the other (Cover and Thomas, 1991). Typically, a single structural image is co-registered to the first or mean functional image.

Warping to atlas (normalization). For group analysis, each voxel must lie within the same brain structure in each individual subject. Individual brains have different shapes and features, but there are regularities shared by every non-pathological brain, and normalization attempts to register each subject's anatomy with a standardized atlas space defined by a *template* brain. Normalization can be linear, involving simple registration of the gross shape of the brain, or nonlinear, involving warping to match local

features. In intensity-based normalization, matching is done using image intensities corresponding to gray/white matter/fluid tissue classes. Surface-based normalization uses extracted features such as gyral and sulcal boundaries explicitly. Here, we describe nonlinear intensity-based normalization as implemented in SPM software.

Whereas the realignment and co-registration procedures perform a *rigid body* rotation, normalization can stretch and shrink different regions of the image to achieve the closest match. This warping consists of shifting the locations of voxels by different amounts depending on their original location. The function that describes how much to shift the voxels is unknown, but can be described by a set of cosine basis functions. The task is then to search for a set of coefficients (weights of each basis function) that minimize the least squares difference between the transformed image and the template. How closely the algorithm attempts to match the local features of the template depends on the number and spatial frequency of basis functions used. Often, warping that is too flexible (using many basis functions) can produce gross distortions in the brain, as local features are matched at the expense of getting the right overall shape. This happens essentially because the problem space is too complex, and the algorithm can settle into a "local minimum" solution that is not close to the global optimal solution. 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 analysis. Thus, it is important to inspect each normalized brain and, if necessary, take remedial measures. These include manually improving the initial alignment, 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).

Smoothing. Currently, many investigators apply a spatial smoothing kernel to the functional data, blurring the image intensities in space. This is ironic, given the push for higher spatial resolutions and smaller voxels—so why does anyone do it? One reason is to improve inter-subject registration. A second reason is that Gaussian Random Field Theory, a popular multiple-comparisons correction procedure, assumes

that the variations across space are continuous and normally distributed. However, images are sampled on a grid of voxels, and neither assumption is likely to hold; smoothing can help to meet these assumptions. Smoothing typically involves convolution 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. One estimate of the amount of smoothing required to meet the assumption is a FWHM of 3 times the voxel size (e.g., 9 mm for 3 mm voxels).

An important consideration is that acquiring an image with large voxels and acquiring with small voxels and smoothing an image are not the same thing. The signal-to-noise ratio during acquisition increases as the square of the voxel volume, so acquiring small voxels means that much signal is lost that can never be recovered!

Researchers using multivariate analyses methods often choose not to smooth the functional 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 subject. When the aim of the study is to accurately predict variables across subjects, e.g. from new fMRI data sets, some smoothing can increase inter-subject alignment and predictive performance.

General linear model

Localizing task-related activations with the GLM

The GLM is the most common statistical method for assessing task – brain activity relationships in neuroimaging (Worsley and Friston, 1995). It is a linear analysis method that subsumes many basic analysis techniques, including t-tests, ANOVA, and multiple regression. The GLM can be used to estimate whether the brain responds to a single type of event, to compare different types of events, to assess correlations between brain activity and behavioral performance or other psychological variables, and for other tests.

The GLM is appropriate when multiple predictor variables—which together constitute a simplified *model* of the sources of variability in a set of data—are used to

explain variability in a single, continuously distributed outcome variable. In a typical neuroimaging experiment, the predictors are related to psychological events, and the outcome variable is signal in a brain voxel or region of interest. Analysis is typically 'massively univariate,' meaning that the analyst performs a separate GLM analysis at every voxel in the brain, and summary statistics are saved in maps of statistic values across the brain.

It is usually advantageous to design studies and statistical analyses in a way that permits inferences about a population of participants. Population inference is typical in all kinds of studies; for example, when testing a new drug, researchers perform statistical tests that allow them to infer that the drug is likely to produce a benefit on average for individuals in a certain population. Even most studies of psychophysics and electrophysiology in monkeys, which often rely on only one or two participants for the entire study, need to be able to claim that their results apply beyond the particular individuals studied. They do so by invoking the additional assumption that all participants will behave the same way as the few observed in the study. In almost all domains of human neuro-psychology, this is not a safe assumption, and statistics should be performed that permit population inference in a standard way. This can be achieved by considering the multi-level nature of neuroimaging data.

A key to population inference (see *Interpretation of fMRI studies*) is to treat the variation across participants as an error term in a group statistical analysis, which leads to generalizability of the results to new participants drawn from the same population. The most popular group analysis is the one-sample t-test on contrast estimates (e.g., Task A – Task B) at each voxel. This analysis tests whether the contrast of interest is non-zero on average for the population from which the sample was drawn, and it provides a starting point for our discussion on population inference. The principle, however, applies to any kind of statistical model, including more complex ANOVA and regression models and multivariate analyses such as group independent components analysis (ICA).

Single-subject GLM model basics

For a single subject, the fMRI time course or series of PET values from one voxel is the outcome variable (**y**). Activity is modeled as the sum of a series of independent predictors (**x** variables, i.e., $\mathbf{x}_1$, $\mathbf{x}_2$, etc.) related to task conditions and other nuisance covariates of no interest (e.g., head movement estimates). In fMRI analysis, for each task condition or event type of interest, a time series of the predicted shape of the signal response is constructed, usually using prior information about the shape of the vascular response to a brief impulse of neural activity. Most often, a canonical hemodynamic response function (HRF) implemented in the respective software package is used (Figure 9A shows an example of an empirical HRF). The vectors of predicted time series values for each task condition are collated into the columns of the design matrix, **X**, which contains a row for each of *n* observations collected (observations over time) and a column for each of *k* predictors. The GLM fitting procedure estimates the best-fitting amplitude (scaling factor) for each column of **X**, so that the sums of fitted values across columns best fits the data. These amplitudes are regression slopes, and are denoted with the variable $\hat{\beta}$ (the "hat" denotes an estimate of a theoretical constant value). It also estimates a time series of error values, $\hat{\varepsilon}$, that cannot be explained by the model. The model is thus described by the equation:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \textbf{(1)}$$

where β is a *k* x 1 vector of regression slopes, **X** is an *n* x *k* model matrix, **y** is an *n* x 1 vector containing the observed data, and ε is an *n* x 1 vector of unexplained error values. The equation is in matrix notation, so that $\mathbf{X}\beta$ indicates the rise and fall in the data explained by the model, or the sum of each column of X multiplied by each element of β. Error values are assumed to be independent and to follow a normal distribution with mean 0 and standard deviation σ. The values of $\hat{\beta}$ correspond to the *estimated magnitude of activation* for each psychological condition described in the columns of **X**. An example for **X** is shown in Figure 9B.

One of the advantages of the GLM is that there exists an algebraic solution for $\hat{\beta}$ that minimizes the squared error, the *ordinary* least-squares solution:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \quad \textbf{(2)}$$

where T indicates the transpose operator.

Inference is generally conducted by calculating a t-statistic, which equals the $\hat{\beta}$s divided by their standard errors, and obtaining p-values using classical inference. The standard errors of the estimates are the diagonal elements of the matrix:

$$se(\hat{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\hat{\sigma} \quad \textbf{(3)}$$

Notably, the error term is composed of two separate terms from different sources. The term $\hat{\sigma}^2$ is the estimated residual error variance, which depends on many factors, including scanner noise. The term $(\mathbf{X}^T\mathbf{X})^{-1}$ depends on the design matrix itself, and reflects both the variability in the predicted signal and covariance among predictors (i.e., multicolinearity). It should be noted that the design optimization algorithms described in the section on *Optimized experimental designs*, work on minimizing the design-related component of the standard error, i.e. $(\mathbf{X}^T\mathbf{X})^{-1}$.

One important additional feature of the data requires a further extension of the model. Typically, fMRI data are autocorrelated—signals are correlated with themselves shifted in time and are not independent—and the autocorrelation must be removed for valid single-subject inference. This is typically done by estimating the autocorrelation in the residuals, after model fitting, and then removing the autocorrelation by 'prewhitening'. Prewhitening works by pre-multiplying both sides of the general linear model equation (Eq. 1) by the square root of a filtering matrix **W**, that will counteract the autocorrelation structure and create a new design matrix $\mathbf{W}^{1/2}\mathbf{X}$ and whitened data $\mathbf{W}^{1/2}\mathbf{y}$. This process is incorporated into what is known as the *generalized* least-squares

solution, so that:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y} \quad \textbf{(4)}$$

Note that the standard errors and degrees of freedom change as well due to the whitening process. Because the estimation of **W** depends on $\hat{\beta}$, and vice versa, a one-step algebraic solution is not available, and the parameters are estimated using an iterative algorithm. There are many ways of designing **W**, ranging from estimates that make strong simplifying assumptions about the form of the data, such as the one-parameter autoregressive AR(1) model, to empirical estimates that 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 fitting random noise patterns in the data.

Contrasts. Contrasts across conditions can be easily handled within the GLM framework. Mathematically, a contrast is a linear combination of predictors. The contrast (e.g., A – B in a simple comparison, or A + B – C – D for a main effect in a 2 x 2 factorial design) is coded as a *k* x 1 vector of contrast weights, which we denote with the letter **c**. For example, the contrast weights for a simple subtraction is $\mathbf{c} = [1\ {-1}]^T$., while a single contrast for a linear effect across four conditions might be $\mathbf{c} = [-3\ {-1}\ 1\ 3]^T$. Concatenating multiple contrasts into a matrix can simultaneously test a whole set. Thus, the main effects and interaction contrasts in a 2 x 2 factorial design can be specified with the following matrix:

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix};$$

Columns 1 and 2 test main effects, and the third tests their interaction. In order to

test contrast values against a null hypothesis of zero—the most typical inferential procedure—contrast weights must sum to zero. If the weights do not sum to zero, then the contrast values partially reflect overall scanner signal intensity, and the resulting t-statistics are invalid. The analyst must take care to specify contrasts correctly, as contrast weights in neuroimaging analysis packages are often specified by the analyst, rather than being created automatically as in SPSS, SAS, and other popular statistical packages. The true contrast values $\mathbf{C}^T\beta$ can be estimated using $\mathbf{C}^T\hat{\beta}$, where $\hat{\beta}$ is obtained using Eq. (2).

Most imaging statistics packages write a series of images to disk containing the betas for each condition throughout the brain, and another set of *contrast images* containing the values of $\mathbf{C}^T\hat{\beta}$ throughout the brain. Contrast images are typically used in a group analysis. A third set of images contains t-statistics, or the ratio of contrast estimates to their standard errors.

Mixed and fixed effects.

The one-sample t-test across contrast values treats the value of that contrast as a random variable with a normal distribution over subjects, and hence the error term in the statistical test is based on the variance across participants. Such an analysis has come to be known as a "random effects" analysis in the neuroimaging literature. Many early studies performed incorrect statistical analyses by lumping data from different participants together into one "super subject" and analyzing the data using a single statistical model. This is called a "fixed effects" analysis because it treats participant as a fixed effect, and assumes the only noise is due to measurement error within subjects. It is not appropriate for population inference because it does not account for individual differences (Figure 10). For example, collecting five hundred images each (250 of Task A and 250 of Task B) on two participants would be treated as the equivalent of collecting two images each (Task A and B) on 500 participants. Some researchers have argued that the fixed analysis allows researchers to make inferences about the brains of participants in the study, but not to a broader population. While this is technically true, inferences about particular individuals are seldom useful; such a lack of generalizability would be

unacceptable in virtually any field, and we do not consider it appropriate for neuroimaging studies either.

A more correct analysis is the "*mixed effects* analysis," so termed because it estimates multiple sources of error, including measurement error within subjects and inter-individual differences between subjects. The one-sample t-test on contrast estimates described above is actually a simplified mixed-effects analysis that is valid if the standard errors of contrast estimates are the same for all participants. Full mixed-effects analyses use iterative techniques (such as the Expectation-Maximization (EM) algorithm) to obtain separate estimates of measurement noise and individual differences. They are implemented in packages such as Hierarchical Linear Modeling (HLM; (Raudenbush and Bryk, 2002)), R packages, such as LME4 (Bates et al., 2013), for Matlab (Lindquist et al., 2012b), and MLwiN (Rasbash, 2002). Neuroimaging data-friendly mixed-effects models are implemented in FSL (Beckmann et al., 2003; Woolrich et al., 2004) and another implementation is available via the command line in SPM8 and via the batch editor in SPM12.

Thresholding and multiple comparisons

The results of neuroimaging studies are often summarized as a set of 'activated regions' or statistical maps. Such summaries describe brain activation by color-coding voxels whose t-values or comparable statistics (z or F) exceed a certain statistical threshold for significance. The implication is that these voxels are activated by the experimental task. A crucial decision is the choice of threshold to use in deciding whether voxels are 'active.' In many fields, test statistics whose p-values are below 0.05 are considered sufficient evidence to reject the null hypothesis, with an acceptable false positive rate (alpha) of 0.05. However, in brain imaging we often test on the order of 100,000 hypothesis tests (one for each voxel) at a single time. Hence, using a voxel-wise alpha of 0.05 means that 5% of the voxels *on average* will show false positive results. This implies that we actually *expect* on the order of 5,000 false positive results. Thus, even if an experiment produces *no true activation*, there is a good chance that without a more conservative correction for multiple comparisons, the activation map will show a

number of activated regions, which would lead to erroneous conclusions.

The traditional way to deal with this problem of multiple comparisons is to adjust the threshold so that the probability of obtaining a false positive is simultaneously controlled for every voxel (i.e., statistical test) in the brain. In neuroimaging, a variety of different approaches towards controlling the false positive rate are commonly used – we will discuss them in detail below. The fundamental difference between methods is whether they control for the family-wise error rate (FWER) or the false discovery rate (FDR). The FWER is the probability of obtaining any false positives in the brain, whereas the FDR is the proportion of false positives among all rejected tests.

To illustrate the difference between FWER and FDR, imagine that we conduct a study on 100,000 brain voxels at alpha = .001 uncorrected, and we find 300 'significant' voxels. According to theory we would expect that 100 (or 33%) of our significant 'discoveries,' to be false positives, but which ones we cannot tell. Since 33% is a significant proportion of all active voxels, we may have low confidence that the activated regions are true results. Thus, it may be advantageous to set a threshold that limits the expected number of false positives to 5%. This is referred to as FDR control at the $q = 0.05$ level. In this case, we might argue that most of the results are likely to be true activations; however, we will still not be able to tell which voxels are truly activated and which are false positives. FWER, by contrast, is a stronger method for controlling false positives. Controlling the FWER at 5% implies that we set a threshold so that, if we were to repeat the above-mentioned experiment 100 times, only 5 out of the 100 experiments will result in one or more false positive voxels. Therefore when controlling the FWER at 5% we can be fairly certain that all voxels that are deemed active are truly active. However, the thresholds will typically be quite conservative, leading to problems with false negatives, or truly active voxels that are now deemed inactive. For example, in our example perhaps only 50 out of the 200 truly active voxels will give significant results. While we can be fairly confident that all 50 are true activations, we have still 'lost' 150 active voxels, most of the true activity.

Many published PET and fMRI studies do not use either of these corrections; instead, they use arbitrary uncorrected thresholds, with a modal threshold of $p < .001$. A

likely reason is because with the sample sizes typically available, corrected thresholds are so high that power is extremely low. This is, of course, extremely problematic when interpreting conclusions from individual studies, as many of the activated regions may simply be false positives. Imposing an arbitrary 'extent threshold' for reporting based on the number of contiguous activated voxels does not necessarily correct the problem because imaging data are spatially smooth, and thus corrected thresholds should be reported whenever possible.

However, because achieving sufficient power is often not possible, it does make sense to report results at an uncorrected threshold and use meta-analysis or a comparable replication strategy to identify consistent results (Wager et al., 2007; Yarkoni et al., 2011) with the caveat that uncorrected results from individual studies cannot be strongly interpreted. Ideally, a study would report both corrected results and results at a reasonable uncorrected threshold (e.g., $p < .001$ and 10 contiguous voxels) for archival purposes.

Methods controlling for multiple comparisons can be applied to the whole brain, gray matter masks, or other regions of interest (ROI). It is reasonable to define regions of interest based on *a priori* hypotheses. Such hypotheses regarding regions of interest can be based on functional (e.g., functional localizer for face sensitive areas) or anatomical constraints (e.g., mask of V1 and V2). The important issue is that the definition of the ROI must be *independent* from the statistical test conducted in that ROI (see (Kriegeskorte et al., 2009; Vul et al., 2009; Kriegeskorte et al., 2010). Problematic examples are defining a region activated in older subjects and then testing if its activity is reduced in younger subjects or defining a region based on activity in the first run of an experiment and then testing whether it shows less activity in subsequent runs. Both of these are not valid tests because they do not control for regression to the mean.

FWE correction

The simplest way of controlling the FWER is to use Bonferroni correction in which the alpha value is divided by the total number of statistical tests performed (i.e., voxels). However, if there is spatial dependence in the data—which is almost always the case, because the natural resolution and applied smoothing both lead to spatial

smoothness in imaging data—this is an unnecessarily conservative correction that leads to a decrease in power to detect truly active voxels. Gaussian Random Field Theory (RFT) (Worsley et al., 2004), used in SPM software (Taylor and Worsley, 2006), is another (more theoretically complicated) approach towards controlling the FWER. If the image is smooth and the number of subjects is high enough (around 20), RFT is less conservative and provides control closer to the true false positive rate than the Bonferroni method.

In addition, RFT is used to assess the probability that k contiguous voxels exceeding the threshold under the null hypothesis, leading to a "cluster-level" correction. The probability that a cluster of size k is found under the null hypothesis is specific to an initial, uncorrected significance threshold. It is much more likely to obtain a cluster of $k = 300$ at an initial threshold of $p < 0.05$ than using $p < 0.001$ as initial threshold, simply because more voxels will survive a more liberal threshold. Recent analyses have shown that a liberal initial threshold (higher than $p < 0.001$) inflates the number of false positives above the nominal level of 5% (Woo et al., 2014a). Nichols and Hayasaka (Nichols and Hayasaka, 2003) provide an excellent review of FWER correction methods. Their conclusions are that while RFT is overly conservative at the voxel level, it is liberal at the cluster level with small sample sizes. Another aspect to keep in mind when using cluster-level correction is that inference is also on the cluster level. Inference is only valid for the whole cluster. It is thus not possible to make inferences about single voxels within that cluster, rather the interpretation is that 'there is true signal somewhere in the cluster' (Woo et al., 2014a). For large clusters spanning multiple anatomical or functional regions, it is thus impossible to state in which of these regions activation is present. This problem is particularly prominent with liberal initial thresholds, since more voxels are considered active and form larger clusters. Cluster-level inference with liberal initial threshold hence reduces the spatial resolution of fMRI.

Both methods described above for controlling the FWER assume that the error values are normally distributed, and that the variance of the errors is equal across all values of the predictors. As an alternative, nonparametric methods instead use the data themselves to find the appropriate distribution. Using such methods can provide

substantial improvements in power and validity, particularly with small sample sizes, and we regard them as the "gold standard" for use in imaging analyses. Thus, these tests can be used to verify the validity of the less computationally expensive parametric approaches. A popular package for doing non-parametric tests, SnPM or "Statistical Non-Parametric Mapping" (Nichols and Holmes, 2002) (http://warwick.ac.uk/snpm), is based on the use of permutation tests. FSL also offers permutations tests via its 'randomise' function (Winkler et al., 2014).

FDR control

The false discovery rate (FDR) is a relatively recent development in multiple comparison correction developed by Benjamini and Hochberg (Benjamini, 1995) While the FWER controls the probability of any false positives occurring in a family of tests (e.g., a statistical brian map), the FDR controls the expected proportion of false positives among significant tests. In a brain map, this means that approximately 95% of the voxels reported at $q < .05$ FDR-corrected (q is used instead of p) are expected to show some true effect. The FDR controlling procedure is adaptive in the sense that the larger the signal, the lower the threshold. If all of the null hypotheses are true, the FDR will be equivalent to the FWER. Any procedure that controls the FWER will also control the FDR. Conversely, any procedure that controls the FDR can only be less stringent than FWER and lead to increased power. A major advantage is that since FDR controlling procedures work only on the p-values and not on the actual test statistics, it can be applied to any valid statistical test.

Anatomical localization and inference

Accurately identifying the anatomical locations of activated regions is critical to making inferences about the meaning of brain imaging data. Knowing where activated areas lie permits comparisons with animal and human lesion and electrophysiology studies. It is also critical for accumulating knowledge across many neuroimaging studies.

Localization is challenging for several reasons; first among them is the problem of

variety: Each brain is different, and it is not always possible to identify the 'same' piece of brain tissue across different individuals (Vogt et al., 1995; Thompson et al., 1996). Likewise, names for the same structures vary: The same section of the inferior frontal gyrus (IFG) can be referred to as IFG, inferior frontal convexity, Brodmann's Area 47, ventrolateral prefrontal cortex, the pars orbitalis, or simply the lateral frontal cortex. Standard anatomical atlas brains differ as well, as do the algorithms used to match brains to these atlases. There is currently a wide and expanding array of available tools for localization and analysis. A database of tools is available from the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC), and another useful list can be found at http://www.nitrc.org.

The most accurate way to localize brain activity is to overlay functional activations on a co-registered, high-resolution individual anatomical image. Many groups avoid issues of variability by defining anatomical regions of interest (ROIs) within individual participants and testing averaged activity in each ROI. The use of functional localizers—separate tasks or contrasts designed to locate functional regions in individuals—is also a widely used approach, and functional and structural localizers can be combined to yield individualized ROIs. For example, structural ROIs are often used in detailed analysis of medial temporal regions in memory research; and the use of retinotopic mapping, a functional localization procedure, to define individual visual-processing regions (V1, V2, V4, etc.) is standard in research on the visual system.

However, the vast majority of studies are analyzed using voxel-wise analysis over much of the brain. In most applications, precise locations are difficult to define *a priori* within individuals, and often many regions as well as their connectivity are of interest. In such cases, atlas-based localization is used. Such localization can be performed using paper-based atlases (Haines, 2000; Mai et al., 2007; Duvernoy, 2012), and there is no substitute for a deep knowledge of neuroanatomy. However, a range of automated atlases and digital tools are becoming increasingly integrated with analysis software. Some of the major ones are described below.

Early approaches to atlas-based localization were based on the Talairach atlas (Talairach and Tournoux, 1988), a hand-drawn illustration of major structures and

Brodmann's Areas (BAs)—cortical regions demarcated according to their cytoarchitecture by Brodmann in 1909—from the left hemisphere of an elderly French woman. The brain is superimposed on a 3-D Cartesian reference grid whose origin is located at the anterior commissure. This allows brain structures to be identified by their coordinate locations. This stereotactic convention remains a standard today. Peak or center-of-mass coordinates from neuroimaging activations are reported in left to right (*x*), posterior to anterior (*y*), and inferior to superior (*z*) dimensions. Negative values on each dimension indicate locations at left, posterior, and inferior positions, respectively. However, because the Talairach brain is not representative of any population and is not complete—only the left hemisphere was studied, and no histology was performed to accurately map BAs—'Talairach' coordinates and their corresponding BA labels should not be used (see (Brett et al., 2002; Devlin and Poldrack, 2007) for discussion) as better alternatives are now available. A current standard in the field is the Montreal Neurologic Institute's (MNI's) 305-brain average[1] (Collins et al., 1994), which is the standard reference brain for two of the most popular software packages, SPM and FSL (Smith et al., 2004) and the International Consortium for Brain Mapping project.

Digital atlases, including the MNI-305 template (not the Talairach template!), permit fine-grained nonlinear warping of brain images to the template and can (if data quality is adequate) match the locations of gyri, sulci, and other local features across brains. A popular approach implemented in SPM software is *intensity-based normalization* (see *Preprocessing*).

An alternative to intensity-based approaches is *surface-based normalization*, in which brain surfaces are reconstructed from segmented gray-matter maps and inflated to a spherical shape or flattened (reviewed in (Van Essen and Dierker, 2007). Features (e.g., gyri and sulci) are identified on structurally simpler 2-D or spherical brains, and the inflated brain is warped to an average spherical atlas brain. This approach has yielded better matches across individuals in comparison studies (Fischl et al., 1999; Van Essen

[1] Called avg305T1 in SPM software. A higher-resolution template in the same space, called the ICBM-152 and named avg152T1 in SPM, is also available. It was created from the average of the 152 most prototypical images in the 305-brain set.

and Dierker, 2007). Several free packages implement surface-based normalization to templates, including FreeSurfer, Caret/SureFit software (Van Essen et al., 2001), and BrainVoyager. AFNI, using SUMA software (Saad et al., 2004), and FSL have facilities for viewing and analyzing surface-based data with FreeSurfer and SureFit.

Because the original BAs were not precisely or rigorously defined in a group, reporting of BAs using the Talairach atlas is not recommended (Devlin and Poldrack, 2007). However, modern probabilistic cytoarchitectural atlases are being developed (Amunts et al., 2007), and some of these are available digitally either from the researchers or within FSL (Juelich Atlas) and SPM (as part of the SPM Anatomy Toolbox (Eickhoff et al., 2005).

Another way to localize functional activations is to compare them with the results of meta-analyses of other neuroimaging studies. Comparison with meta-analytic results can help to identify functional landmarks and provide information on the kinds of different tasks that have produced similar activation patterns. Whereas it was typical in early neuroimaging studies to claim consistency with previous studies based on activation in the same gross anatomical regions (e.g., activation of the anterior cingulate cortex), it is now recognized that many such regions are very large, and more precise correspondence is required to establish consistency across studies. Quantitative meta-analyses identify the precise locations that are most consistently activated across studies, and they thus provide excellent functional landmarks.

The variety and heterogeneity of tools that are currently available is both a strength and an obstacle to effective localization. A few guidelines may aid in the process. First, it is preferable to overlay functional activations on an average of the actual anatomical brains from the study sample, after normalization (registration and/or warping) to a chosen template, rather than relying solely on an atlas brain. Normalization cannot be achieved perfectly in every region, and showing results on the subject's actual anatomy is more accurate than assuming the template is a perfect representation. In addition, viewing the average warped brain can be very informative about whether the normalization process yielded high co-registration of anatomical landmarks across participants, and can help identify problem areas. Single-subject atlases should not be

taken as precise indicators of activation location in a study sample, and while they make attractive underlay images for activations, they should not be used for this purpose. Second, it is important to remember that atlas brains are different, and different algorithms used with the same atlas produce different results. Therefore, it is important to report which algorithm and which atlas was used. Also, it would be highly misleading to use a probabilistic atlas such as those in the SPM anatomy toolbox if the study brains were normalized to a different template (and/or with different procedures) than the one used to create the atlas (e.g., the SPM anatomy toolbox should not be used when normalizing to the ICBM-452 atlas). Regardless of the tools used, identifying functional activations on individual and group-averaged anatomy, collaborating with neuroanatomists when possible, and using print atlases to identify activations relative to structural landmarks are all essential components of the localization and interpretation process.

Connectivity analyses in fMRI

Most analysis techniques discussed so far focus on questions of functional specialization. The kinds of questions that fMRI can answer with regard to specialized functions are inherently limited by the spatial resolution of fMRI. A different type of question asks how cognitive functions are integrated across brain regions or how neuronal populations work together. To this end, it is necessary to study multiple regions at the same time and investigate their relationships. The commonality of all these techniques is that they build on time-series data from voxels or ROI's. There are many ways of extracting measures of brain connectivity data, and the literature is now replete with a huge, and growing, variety of possibilities (Figure 11). We can only provide a short overview here and refer the reader to some of excellent specialized reviews (Friston, 2011; Smith, 2012; Calhoun et al., 2014).

Time-series values can be used in structured, *hypothesis-driven models* of connectivity, including path models, Granger causal models, Dynamic Causal Models (DCM), and related state-space models. Some of these are discussed below. Large-scale connectivity matrices can be used to estimate higher-order, *graph theoretic properties* of

the networks as a whole, which can then be related to outcomes. There is currently a proliferation of such measures, including 'small worldness,' path length, betweenness-centrality, 'rich club' indices, and metrics of degree distribution (Sporns, 2014). These describe, in various ways, organizational properties concerning how all of the 'objects' (in this case, brain voxels or regions) relate to the others. Spectral measures, which summarize connectivity based on its temporal frequencies, include voxel-wise amplitude of low-frequency fluctuations (ALFF) and measures derived from time-frequency analysis.

Two very popular techniques for connectivity analysis are psycho-physiological interaction (PPI) analysis (Friston et al., 1997) and Dynamic causal modeling (DCM) (Friston, 2003). PPI correlate the time-series from a ROI (seed-region) with all other voxels' time-series. The question of interest is then, where in the brain the correlation with the seed region is effected by a psychological moderator. The term PPI is used because the test is formulated as interaction between the seed time-series and the time-course of the psychological variable within the GLM framework.

PPI belongs to class of techniques often labeled 'functional connectivity' that do not imply and directionality of the estimated connections. DCM and Granger causal models assume directionality and thus explicitly model whether the influence is from A to B or from B to A.

DCM also includes a neuronal network model and links the observed fMRI to its underlying generative model via a model of neurovascular coupling. The nodes and connections between nodes are explicitly specified in DCM and can include psychological moderator variables affecting connections or nodes. This explicit formulation of hypothesis is one of the strengths of DCM because it forces the researcher to clearly define hypothetical models of brain function. After a set of candidate models has been specified and estimated, DCM uses Bayesian model selection to choose the model that best explains the observed data (Friston et al., 2003; Stephan et al., 2009).

While most of the literature has focused on stationary correlations that are constant across time, researchers are increasingly interested in *time-varying correlations* (Cribben et al., 2012; Calhoun et al., 2014), which provide expanded measures of how

correlations change across time and can be used to estimate time-varying graph or network structures.

Hypothesis-driven models of connectivity (e.g., path models and DCM), graph theoretic measures, spectral measures, and time-varying connectivity metrics are all brain-derived measures that can be used to learn how brain activity maps into mental states, performance, experiences and clinical symptoms, behavior, and other outcomes. We think of them as part of a "grand search" for the critical levels and type of brain measures that will predict and eventually explain how the brain shapes those outcomes.

VI. CONCLUSIONS

In this chapter we have reviewed the basics of functional neuroimaging with a focus on PET and fMRI. We have covered data acquisition, experimental design, analysis of the data, and covered principles of inference in neuroimaging studies. We hope that this brief introduction provides some practical advice for conducting, analyzing, and interpreting fMRI studies and encourages the reader to study these topics in more depth.

The field has seen a marked increase in the data quality of fMRI over the past decade, and at the same time the options for data analyses have multiplied. Together with the marked increases in sample size due to collaborative efforts and the new ease of sharing data, these developments open exciting avenues for increasing our knowledge about brain function.

VII. REFERENCES

Aguirre GK, Zarahn E, D'Esposito M (1998) The variability of human, BOLD hemodynamic responses. NeuroImage 8:360-369.

Amunts K, Schleicher A, Zilles K (2007) Cytoarchitecture of the cerebral cortex--More than localization. NeuroImage 37:1061-1065.

Andersson JL, Hutton C, Ashburner J, Turner R, Friston K (2001) Modeling geometric deformations in EPI time series. NeuroImage 13:903-919.

Andrews-Hanna JR, Reidler JS, Huang C, Buckner RL (2010) Evidence for the default network's role in spontaneous cognition. J Neurophysiol 104:322-335.

Aron A, Fisher H, Mashek DJ, Strong G, Li H, Brown LL (2005) Reward, motivation, and emotion systems associated with early-stage intense romantic love. J Neurophysiol 94:327-337.

Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95-113.

Ashburner J, Friston KJ (2000) Voxel-based morphometry--the methods. Neuroimage 11:805-821.

Atlas LY, Lindquist MA, Bolger N, Wager TD (2014) Brain mediators of the effects of noxious heat on pain. Pain 155:1632-1648.

Atlas LY, Whittington RA, Lindquist MA, Wielgosz J, Sonty N, Wager TD (2012) Dissociable influences of opiates and expectations on pain. J Neurosci 32:8053-8064.

Bates D, Maechler M, Bolker B, Walker S (2013) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.

Beckmann CF, Jenkinson M, Smith SM (2003) General multilevel linear modeling for group analysis in FMRI. Neuroimage 20:1052-1063.

Behrens TEJ, Berg HJ, Jbabdi S, Rushworth MFS, Woolrich MW (2007) Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage 34:144-155.

Bendriem B, Townsend, D.W. (1998) The theory and practice of 3D PET. Boston: Dordrecht; Boston: Kluwer Academic, 1998.

Benjamini YaH, Y. (1995) Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society B 57:289 -300.

Bernstein MA, King, K.F., & Zhou, Z.J. (2004) Handbook of MRI pulse sequences. Burlington, MA.: Elsevier Academic Press.

Binder JR, Desai RH, Graves WW, Conant LL (2009) Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. Cerebral Cortex 19:2767-2796.

Birn RM, Saad ZS, Bandettini PA (2001) Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. NeuroImage 14:817-826.

Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34:537-541.

Bohning DE, Pecheny AP, Epstein CM, Speer AM, Vincent DJ, Dannels W, George MS (1997) Mapping transcranial magnetic stimulation (TMS) fields in vivo with MRI. Neuroreport 8:2535-2538.

Bohning DE, Shastri A, McConnell KA, Nahas Z, Lorberbaum JP, Roberts DR, Teneback C, Vincent DJ, George MS (1999) A combined TMS/fMRI study of intensity-dependent TMS over motor cortex. Biol Psychiatry 45:385-394.

Bornhovd K, Quante M, Glauche V, Bromm B, Weiller C, Buchel C (2002) Painful stimuli evoke different stimulus-response functions in the amygdala, prefrontal, insula and somatosensory cortex: a single-trial fMRI study. Brain 125:1326-1336.

Boubela RN, Kalcher K, Huf W, Seidel EM, Derntl B, Pezawas L, Nasel C, Moser E (2015) fMRI measurements of amygdala activation are confounded by stimulus

correlated signal fluctuation in nearby veins draining distant brain regions. Sci Rep 5:10499.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207-4221.

Brett M, Johnsrude IS, Owen AM (2002) The problem of functional localization in the human brain. Nat Rev Neurosci 3:243-249.

Brooks JC, Beckmann CF, Miller KL, Wise RG, Porro CA, Tracey I, Jenkinson M (2008) Physiological noise modelling for spinal functional magnetic resonance imaging studies. Neuroimage 39:680-692.

Brown AK, Fujita M, Fujimura Y, Liow JS, Stabin M, Ryu YH, Imaizumi M, Hong J, Pike VW, Innis RB (2007) Radiation dosimetry and biodistribution in monkey and man of 11C-PBR28: a PET radioligand to image inflammation. J Nucl Med 48:2072-2079.

Buchel C, Bornhovd K, Quante M, Glauche V, Bromm B, Weiller C (2002) Dissociable neural responses related to pain intensity, stimulus intensity, and stimulus awareness within the anterior cingulate cortex: a parametric single-trial laser functional magnetic resonance imaging study. J Neurosci 22:970-976.

Buckner RL, Andrews-Hanna JR, Schacter DL (2008) The brain's default network. Annals of the New York Academy of Sciences 1124:1-38.

Buracas GT, Boynton GM (2002) Efficient design of event-related fMRI experiments using M-sequences. NeuroImage 16:801-813.

Bush G, Luu P, Posner MI (2000) Cognitive and emotional influences in anterior cingulate cortex. Trends in Cognitive Sciences 4:215-222. [Record as supplied by publisher].

Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafo MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14:365-376.

Buxton RB, Frank LR (1997) A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. J Cereb Blood Flow Metab 17:64-72.

Buxton RB, Uludag K, Dubowitz DJ, Liu TT (2004) Modeling the hemodynamic response to brain activation. NeuroImage 23 Suppl 1:S220-233.

Buxton RB, Frank LR, Wong EC, Siewert B, Warach S, Edelman RR (1998) A general kinetic model for quantitative perfusion imaging with arterial spin labeling. Magn Reson Med 40:383-396.

Cacioppo JT, Tassinary LG (1990) Inferring psychological significance from physiological signals. Am Psychol 45:16-28.

Calhoun VD, Miller R, Pearlson G, Adali T (2014) The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. Neuron 84:262-274.

Chaimow D, Yacoub E, Ugurbil K, Shmuel A (2011) Modeling and analysis of mechanisms underlying fMRI-based decoding of information conveyed in cortical columns. Neuroimage 56:627-642.

Cheng K, Waggoner RA, Tanaka K (2001) Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. Neuron 32:359-374.

Collins DL, Neelin P, Peters TM, Evans AC (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18:192-205.

Coltheart M (2006) What has functional neuroimaging told us about the mind (so far)? Cortex 42:323-331.

Constable RT, Spencer DD (1999) Composite image formation in z-shimmed functional MR imaging. Magn Reson Med 42:110-117.

Cover TM, Thomas JA (1991) Elements of Information Theory. In, pp 18-26. New York: Wiley.

Cribben I, Haraldsdottir R, Atlas LY, Wager TD, Lindquist MA (2012) Dynamic connectivity regression: determining state-related changes in brain connectivity. Neuroimage 61:907-920.

de Quervain DJ, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, Fehr E (2004) The neural basis of altruistic punishment. Science 305:1254-1258.

Deckers RH, van Gelderen P, Ries M, Barret O, Duyn JH, Ikonomidou VN, Fukunaga M, Glover GH, de Zwart JA (2006) An adaptive filter for suppression of cardiac and respiratory noise in MRI time series data. Neuroimage 33:1072-1081.

Denis Le Bihan MD, Mangin JF, Poupon C, Clark CA, Pappata S, Molko N, Chabriat H (2001) Diffusion Tensor Imaging: Concepts and Applications. JOURNAL OF MAGNETIC RESONANCE IMAGING 13:534-546.

Denny BT, Kober H, Wager TD, Ochsner KN (2012) A Meta-analysis of Functional Neuroimaging Studies of Self- and Other Judgments Reveals a Spatial Gradient for Mentalizing in Medial Prefrontal Cortex. Journal of Cognitive Neuroscience 24:1742-1752.

Desmond JE, Glover GH (2002) Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. Journal of neuroscience methods 118:115-128.

Detre JA, Zhang WG, Roberts DA, Silva AC, Williams DS, Grandis DJ, Koretsky AP, Leigh JS (1994) Tissue-Specific Perfusion Imaging Using Arterial Spin-Labeling. Nmr Biomed 7:75-82.

Devlin JT, Poldrack RA (2007) In praise of tedious anatomy. NeuroImage 37:1033-1041; discussion 1050-1038.

Disbrow EA, Slutsky DA, Roberts TP, Krubitzer LA (2000) Functional MRI at 1.5 tesla: a comparison of the blood oxygenation level-dependent signal and electrophysiology. Proc Natl Acad Sci U S A 97:9718-9723.

Doucet G, Naveau M, Petit L, Zago L, Crivello F, Jobard G, Delcroix N, Mellet E, Tzourio-Mazoyer N, Mazoyer B, Joliot M (2012) Patterns of hemodynamic low-frequency oscillations in the brain are modulated by the nature of free thought during rest. Neuroimage 59:3194-3200.

Duong TQ, Yacoub E, Adriany G, Hu X, Ugurbil K, Vaughan JT, Merkle H, Kim SG (2002) High-resolution, spin-echo BOLD, and CBF fMRI at 4 and 7 T. Magn Reson Med 48:589-593.

Duvernoy HM (2012) The Human Brain Stem and Cerebellum: Surface, Structure, Vascularization, and Three-Dimensional Sectional Anatomy, with MRI: Springer Science & Business Media.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage 25:1325-1335.

Eisenberger NI, Lieberman MD, Williams KD (2003) Does rejection hurt? An FMRI study of social exclusion. Science 302:290-292.

Elster AD (1994) Questions and answers in magnetic resonance imaging. St. Louis, Mo.: Mosby.

Ethofer T, Van De Ville D, Scherer K, Vuilleumier P (2009) Decoding of emotional information in voice-sensitive cortices. Curr Biol 19:1028-1033.

Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Gunther M, Glasser MF, Miller KL, Ugurbil K, Yacoub E (2010) Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PLoS One 5:e15710.

Finsterbusch J, Busch MG, Larson PE (2013) Signal scaling improves the signal-to-noise ratio of measurements with segmented 2D-selective radiofrequency excitations. Magn Reson Med 70:1491-1499.

Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195-207.

Fox MD, Snyder AZ, Vincent JL, Raichle ME (2007) Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. Neuron 56:171-184.

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A 102:9673-9678.

Frey KA (1999) Positron Emission Tomography. In: Basic Neurochemistry, 6 Edition (Siegel GJ, Agranoff BW, Albers RW, Fisher SK, Uhler MD, eds), pp 1109-1131. Philadelphia: Lippincott, Williams, & Wilkins.

Friston K (2012) Ten ironic rules for non-statistical reviewers. Neuroimage 61:1300-1310.

Friston K, Harrison, L, Penny, W (2003) Dynamic causal modelling. Neuroimage 19:1273-1302.

Friston KJ (2009) Modalities, modes, and models in functional neuroimaging. Science 326:399-403.

Friston KJ (2011) Functional and Effective Connectivity: A Review. Brain Connectivity 1:13-36.

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. NeuroImage 19:1273-1302.

Friston KJ, Frith CD, Turner R, Frackowiak RS (1995) Characterizing evoked hemodynamics with fMRI. NeuroImage 2:157-165.

Friston KJ, Mechelli A, Turner R, Price CJ (2000) Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. NeuroImage 12:466-477.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218-229.

Gianaros PJ, Wager TD (2015) Brain-Body Pathways Linking Psychological Stress and Physical Health. Curr Dir Psychol Sci 24:313-321.

Glahn DC, Thompson PM, Blangero J (2007a) Neuroimaging endophenotypes: strategies for finding genes influencing brain structure and function. Hum Brain Mapp 28:488-501.

Glahn DC, Paus T, Thompson PM (2007b) Imaging genomics: mapping the influence of genetics on brain structure and function. Hum Brain Mapp 28:461-463.

Glover GH, Law CS (2001) Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn Reson Med 46:515-522.

Glover GH, Li TQ, Ress D (2000) Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn Reson Med 44:162-167.

Goldman RI, Stern JM, Engel J, Jr., Cohen MS (2000) Acquiring simultaneous EEG and functional MRI. Clin Neurophysiol 111:1974-1980.

Gonzalez-Castillo J, Saad ZS, Handwerker DA, Inati SJ, Brenowitz N, Bandettini PA (2012) Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. Proc Natl Acad Sci U S A 109:5487-5492.

Good CD, Johnsrude IS, Ashburner J, Henson RNA, Friston KJ, Frackowiak RSJ (2001) A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains. Neuroimage 14:21-36.

Grinband J, Savitskaya J, Wager TD, Teichert T, Ferrera VP, Hirsch J (2011) The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. NeuroImage 57:303-311.

Haacke EM (1999) Magnetic resonance imaging : physical principles and sequence design. New York: Wiley.

Haines DE (2000) Neuroanatomy: An Atlas of Structures, Sections, and Systems. Philadelphia: Lippincott Williams & Wilkins.

Hare TA, Camerer CF, Rangel A (2009) Self-control in decision-making involves modulation of the vmPFC valuation system. Science 324:646-648.

Haxby James V, Guntupalli JS, Connolly Andrew C, Halchenko Yaroslav O, Conroy Bryan R, Gobbini MI, Hanke M, Ramadge Peter J (2011) A Common, High-Dimensional Model of the Representational Space in Human Ventral Temporal Cortex. Neuron 72:404-416.

Haynes JD (2015) A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. Neuron 87:257-270.

Haynes JD, Deichmann R, Rees G (2005) Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. Nature 438:496-499.

Heeger DJ, Ress D (2002) What does fMRI tell us about neuronal activity? Nat Rev Neurosci 3:142-151.

Henson R, Shallice T, Dolan R (2000) Neuroimaging evidence for dissociable forms of repetition priming. Science 287:1269-1272.

Honey CJ, Sporns O, Cammoun L, Gigandet X, Thiran JP, Meuli R, Hagmann P (2009) Predicting human resting-state functional connectivity from structural connectivity. Proc Natl Acad Sci U S A 106:2035-2040.

Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y (2013) Neural decoding of visual imagery during sleep. Science 340:639-642.

Huettel SA, Song AW, McCarthy G (2004) Functional magnetic resonance imaging. Sunderland, Mass.: Sinauer Associates, Publishers.

Huth Alexander G, Nishimoto S, Vu An T, Gallant Jack L (2012) A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. Neuron 76:1210-1224.

Johansen-Berg H, Behrens TE (2006) Just pretty pictures? What diffusion tractography can add in clinical neuroscience. Curr Opin Neurol 19:379-385.

Johansen-Berg H, Behrens TE, Robson MD, Drobnjak I, Rushworth MF, Brady JM, Smith SM, Higham DJ, Matthews PM (2004) Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. Proc Natl Acad Sci U S A 101:13335-13340.

Josephs O, Henson RN (1999) Event-related functional magnetic resonance imaging: modelling, inference and optimization. Philos Trans R Soc Lond B Biol Sci 354:1215-1228.

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679-685.

Kao MH, Mandal A, Lazar N, Stufken J (2009) Multi-objective optimal experimental designs for event-related fMRI studies. Neuroimage 44:849-856.

Kastner S, Ungerleider LG (2000) Mechanisms of visual attention in the human cortex. Annu Rev Neurosci 23:315-341.

Kleinschmidt A, Buchel C, Zeki S, Frackowiak RS (1998) Human brain activity during spontaneously reversing perception of ambiguous figures. Proc Biol Sci 265:2427-2433.

Klunk WE et al. (2004) Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. Ann Neurol 55:306-319.

Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD (2008) Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. Neuroimage 42:998-1031.

Kong Y, Jenkinson M, Andersson J, Tracey I, Brooks JC (2012) Assessment of physiological noise modelling methods for functional imaging of the spinal cord. Neuroimage 60:1538-1549.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535-540.

Kriegeskorte N, Lindquist MA, Nichols TE, Poldrack RA, Vul E (2010) Everything you never wanted to know about circular analysis, but were afraid to ask. Journal of Cerebral Blood Flow & Metabolism 30:1551-1557.

Kvitsiani D, Ranade S, Hangya B, Taniguchi H, Huang JZ, Kepecs A (2013) Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. Nature 498:363-366.

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R, et al. (1992) Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci U S A 89:5675-5679.

Laufs H, Daunizeau J, Carmichael DW, Kleinschmidt A (2008) Recent advances in recording electrophysiological data simultaneously with magnetic resonance imaging. NeuroImage 40:515-528.

Leitao J, Thielscher A, Tunnerhoff J, Noppeney U (2015) Concurrent TMS-fMRI Reveals Interactions between Dorsal and Ventral Attentional Systems. J Neurosci 35:11445-11457.

Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Barrett LF (2012a) The brain basis of emotion: a meta-analytic review. Behav Brain Sci 35:121-143.

Lindquist MA, Waugh C, Wager TD (2007) Modeling state-related fMRI activity using change-point theory. NeuroImage 35:1125-1141.

Lindquist MA, Caffo B, Crainiceanu C (2013) Ironing out the statistical wrinkles in "ten ironic rules". Neuroimage 81:499-502.

Lindquist MA, Zhang CH, Glover G, Shepp L (2008) Rapid three-dimensional functional magnetic resonance imaging of the initial negative BOLD response. J Magn Reson 191:100-111.

Lindquist MA, Spicer J, Asllani I, Wager TD (2012b) Estimating and testing variance components in a multi-level GLM. NeuroImage 59:490-501.

Liu TT (2004) Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part II: design of experiments. Neuroimage 21:401-413.

Loggia ML, Chonde DB, Akeju O, Arabasz G, Catana C, Edwards RR, Hill E, Hsu S, Izquierdo-Garcia D, Ji RR, Riley M, Wasan AD, Zurcher NR, Albrecht DS, Vangel MG, Rosen BR, Napadow V, Hooker JM (2015) Evidence for brain glial activation in chronic pain patients. Brain 138:604-615.

Logothetis NK (2008) What we can do and what we cannot do with fMRI. Nature 453:869-878.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150-157.

Maguire EA, Gadian DG, Johnsrude IS, Good CD, Ashburner J, Frackowiak RS, Frith CD (2000) Navigation-related structural change in the hippocampi of taxi drivers. Proc Natl Acad Sci U S A 97:4398-4403.

Mai JK, Paxinos G, Voss T (2007) Atlas of the Human Brain, 3 edition Edition: Academic Press.

Menon RS (2002) Postacquisition suppression of large-vessel BOLD signals in high-resolution fMRI. Magnetic Resonance in Medicine 47:1-9.

Miezin FM, Maccotta L, Ollinger JM, Petersen SE, Buckner RL (2000) Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. Neuroimage 11:735-759.

Morawetz C, Holz P, Lange C, Baudewig J, Weniger G, Irle E, Dechent P (2008) Improved functional mapping of the human amygdala using a standard functional

magnetic resonance imaging sequence with simple modifications. Magn Reson Imaging 26:45-53.

Mumford JA, Nichols TE (2008) Power calculation for group fMRI studies accounting for arbitrary design and temporal autocorrelation. Neuroimage 39:261-268.

Nichols T, Hayasaka S (2003) Controlling the familywise error rate in functional neuroimaging: a comparative review. Stat Methods Med Res 12:419-446.

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15:1-25.

Noll DC, Fessler JA, Sutton BP (2005) Conjugate phase MRI reconstruction with spatially variant sample density correction. IEEE Trans Med Imaging 24:325-336.

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424-430.

Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J (2006) Self-referential processing in our brain—A meta-analysis of imaging studies on the self. NeuroImage 31:440-457.

Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci U S A 87:9868-9872.

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci U S A 89:5951-5955.

Paton JJ, Belova MA, Morrison SE, Salzman CD (2006) The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature 439:865-870.

Paus T (2001) Primate anterior cingulate cortex: where motor control, drive and cognition interface. Nat Rev Neurosci 2:417-424.

Petrini K, Pollick FE, Dahl S, McAleer P, McKay LS, Rocchesso D, Waadeland CH, Love S, Avanzini F, Puce A (2011) Action expertise reduces brain activity for audiovisual matching actions: an fMRI study with expert drummers. Neuroimage 56:1480-1492.

Phillips C, Rugg MD, Friston KJ (2002) Anatomically Informed Basis Functions for EEG Source Localization: Combining Functional and Anatomical Constraints. NeuroImage 16:678-695.

Poldrack RA (2011) Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. Neuron 72:692-697.

Price CJ, Veltman DJ, Ashburner J, Josephs O, Friston KJ (1999) The critical relationship between the timing of stimulus presentation and data acquisition in blocked designs with fMRI. Neuroimage 10:36-44.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001) A default mode of brain function. Proc Natl Acad Sci U S A 98:676-682.

Rasbash J (2002) A User's Guide to MLwiN: Centre for Multilevel Modelling, University of London.

Raudenbush SW, Bryk AS (2002) Hierarchical Linear Models: Applications and Data Analysis Second Edition. Newbury Park, CA: Sage.

Reiman EM, Fusselman MJ, Fox PT, Raichle ME (1989) Neuroanatomical correlates of anticipatory anxiety [published erratum appears in Science 1992 Jun 19;256(5064):1696]. Science 243:1071-1074.

Rosen BR, Buckner RL, Dale AM (1998) Event-related functional MRI: past, present, and future. Proc Natl Acad Sci U S A 95:773-780.

Ruff CC, Blankenburg F, Bjoertomt O, Bestmann S, Freeman E, Haynes JD, Rees G, Josephs O, Deichmann R, Driver J (2006) Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex. Curr Biol 16:1479-1488.

Saad ZS, Reynolds RC, Argall B, Japee S, Cox RW (2004) SUMA: an interface for surface-based intra- and inter-subject analysis with AFNI. In: Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on, pp 1510-1513 Vol. 1512.

Sandler MP (2003) Diagnostic nuclear medicine. Philadelphia, PA: Lippincott / Williams & Wilkins.

Sarter M, Berntson GG, Cacioppo JT (1996) Brain imaging and cognitive neuroscience. Toward strong inference in attributing function to structure. Am Psychol 51:13-21.

Schacter DL, Buckner RL, Koutstaal W, Dale AM, Rosen BR (1997) Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. NeuroImage 6:259-269.

Scheibe C, Ullsperger M, Sommer W, Heekeren HR (2010) Effects of parametrical and trial-to-trial variation in prior probability processing revealed by simultaneous electroencephalogram/functional magnetic resonance imaging. J Neurosci 30:16709-16717.

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD (2007) Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. The Journal of Neuroscience 27:2349-2356.

Setsompop K, Gagoski BA, Polimeni JR, Witzel T, Wedeen VJ, Wald LL (2012) Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magn Reson Med 67:1210-1224.

Shulman RG, Rothman DL (1998) Interpreting functional imaging studies in terms of neurotransmitter cycling. Proc Natl Acad Sci U S A 95:11993-11998.

Shulman RG, Rothman DL, Behar KL, Hyder F (2004) Energetic basis of brain activity: implications for neuroimaging. Trends Neurosci 27:489-495.

Sibson NR, Dhankhar A, Mason GF, Behar KL, Rothman DL, Shulman RG (1997) In vivo 13C NMR measurements of cerebral glutamine synthesis as evidence for glutamate-glutamine cycling. Proc Natl Acad Sci U S A 94:2699-2704.

Sinha R, Lacadie C, Skudlarski P, Wexler BE (2004) Neural Circuits Underlying Emotional Distress in Humans. Annals of the New York Academy of Sciences 1032:254-257.

Skudlarski P, Constable RT, Gore JC (1999) ROC analysis of statistical methods used in functional MRI: individual subjects. Neuroimage 9:311-329.

Smith S, Jenkinson M, Beckmann C, Miller K, Woolrich M (2007) Meaningful design and contrast estimability in FMRI. Neuroimage 34:127-136.

Smith SM (2012) The future of FMRI connectivity. NeuroImage 62:1257-1266.
Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23 Suppl 1:S208-219.
Sporns O (2014) Contributions and challenges for network models in cognitive neuroscience. Nat Neurosci 17:652-660.
Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. Neuroimage 46:1004-1017.
Sternberg S (2001) Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. Acta Psychol (Amst) 106:147-246.
Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T (2008) Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci 11:1004-1006.
Summerfield C, Greene M, Wager T, Egner T, Hirsch J, Mangels J (2006) Neocortical connectivity during episodic memory formation. PLoS Biol 4:e128.
Sylvester CY, Wager TD, Lacey SC, Hernandez L, Nichols TE, Smith EE, Jonides J (2003) Switching attention and resolving interference: fMRI measures of executive functions. Neuropsychologia 41:357-370.
Tagliazucchi E, Laufs H (2014) Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. Neuron 82:695-708.
Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain : 3-dimensional proportional system : an approach to cerebral imaging. Stuttgart ; New York: G. Thieme ; New York : Thieme Medical Publishers.
Taylor JE, Worsley KJ (2006) Inference for magnitudes and delays of responses in the FIAC data using BRAINSTAT/FMRISTAT. Hum Brain Mapp 27:434-441.
Thompson PM, Schwartz C, Lin RT, Khan AA, Toga AW (1996) Three-dimensional statistical analysis of sulcal variability in the human brain. J Neurosci 16:4261-4274.
Thompson PM et al. (2014) The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav 8:153-182.
Tye KM, Prakash R, Kim S-Y, Fenno LE, Grosenick L, Zarabi H, Thompson KR, Gradinaru V, Ramakrishnan C, Deisseroth K (2011) Amygdala circuitry mediating reversible and bidirectional control of anxiety. Nature 471:358-362.
van Ast V, Spicer J, Smith E, Schmer-Galunder S, Liberzon I, Abelson J, Wager T (2014) Brain mechanisms of social threat effects on working memory. Cerebral Cortex:bhu206.
Van Essen DC, Dierker DL (2007) Surface-based and probabilistic atlases of primate cerebral cortex. Neuron 56:209-225.
Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH (2001) An integrated software suite for surface-based analyses of cerebral cortex. J Am Med Inform Assoc 8:443-459.

Vazquez AL, Noll DC (1998) Nonlinear aspects of the BOLD response in functional MRI. NeuroImage 7:108-118.

Vazquez AL, Cohen ER, Gulani V, Hernandez-Garcia L, Zheng Y, Lee GR, Kim SG, Grotberg JB, Noll DC (2006) Vascular dynamics and BOLD fMRI: CBF level effects and analysis considerations. Neuroimage 32:1642-1655.

Vincent JL, Patel GH, Fox MD, Snyder AZ, Baker JT, Van Essen DC, Zempel JM, Snyder LH, Corbetta M, Raichle ME (2007) Intrinsic functional architecture in the anaesthetized monkey brain. Nature 447:83-86.

Vogt BA, Nimchinsky EA, Vogt LJ, Hof PR (1995) Human cingulate cortex: surface features, flat maps, and cytoarchitecture. The Journal of comparative neurology 359:490-506.

Vul E, Harris C, Winkielman P, Pashler H (2009) Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspectives on psychological science 4:274-290.

Wager TD, Nichols TE (2003) Optimization of experimental design in fMRI: a general framework using a genetic algorithm. Neuroimage 18:293-309.

Wager TD, Reading S, Jonides J (2004a) Neuroimaging studies of shifting attention: a meta-analysis. Neuroimage 22:1679-1693.

Wager TD, Jonides J, Reading S (2004b) Neuroimaging studies of shifting attention: a meta-analysis. NeuroImage 22:1679-1693.

Wager TD, Lindquist M, Kaplan L (2007) Meta-analysis of functional neuroimaging data: Current and future directions. Social, Cognitive, and Affective Neuroscience 2:150-158.

Wager TD, Vazquez A, Hernandez L, Noll DC (2005a) Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. Neuroimage 25:206-218.

Wager TD, Jonides J, Smith EE, Nichols TE (2005b) Toward a taxonomy of attention shifting: individual differences in fMRI during multiple shift types. Cogn Affect Behav Neurosci 5:127-143.

Wager TD, Waugh CE, Lindquist M, Noll DC, Fredrickson BL, Taylor SF (2009) Brain mediators of cardiovascular responses to social threat: part I: Reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. Neuroimage 47:821-835.

Wager TD, Atlas LY, Lindquist MA, Roy M, Woo CW, Kross E (2013) An fMRI-based neurologic signature of physical pain. N Engl J Med 368:1388-1397.

Waugh CE, Hamilton JP, Gotlib IH (2010) The neural temporal dynamics of the intensity of emotional experience. Neuroimage 49:1699-1707.

Wiech K, Jbabdi S, Lin CS, Andersson J, Tracey I (2014) Differential structural and resting state connectivity between insular subdivisions and other pain-related brain regions. Pain 155:2047-2055.

Wilson JL, Jezzard P (2003) Utilization of an intra-oral diamagnetic passive shim in functional MRI of the inferior frontal cortex. Magn Reson Med 50:1089-1094.

Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE (2014) Permutation inference for the general linear model. NeuroImage 92:381-397.

Wise RG, Rogers R, Painter D, Bantick S, Ploghaus A, Williams P, Rapeport G, Tracey I (2002) Combining fMRI with a pharmacokinetic model to determine which brain areas activated by painful stimulation are specifically modulated by remifentanil. Neuroimage 16:999-1014.

Woo CW, Krishnan A, Wager TD (2014a) Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. Neuroimage 91:412-419.

Woo CW, Koban L, Kross E, Lindquist MA, Banich MT, Ruzic L, Andrews-Hanna JR, Wager TD (2014b) Separate neural representations for physical pain and social rejection. Nat Commun 5:5380.

Woolrich MW, Behrens TE, Beckmann CF, Jenkinson M, Smith SM (2004) Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage 21:1732-1747.

Worsley KJ, Friston KJ (1995) Analysis of fMRI time-series revisited--again. Neuroimage 2:173-181.

Worsley KJ, Taylor JE, Tomaiuolo F, Lerch J (2004) Unified univariate and multivariate random field theory. Neuroimage 23 Suppl 1:S189-195.

Yacubian J, Sommer T, Schroeder K, Glascher J, Kalisch R, Leuenberger B, Braus DF, Buchel C (2007) Gene-gene interaction associated with neural reward sensitivity. Proc Natl Acad Sci U S A 104:8125-8130.

Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. Nat Methods 8:665-670.

Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zollei L, Polimeni JR, Fischl B, Liu H, Buckner RL (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 106:1125-1165.

Zarahn E, Slifstein M (2001) A reference effect approach for power analysis in fMRI. Neuroimage 14:768-779.

VIII. FIGURE CAPTIONS

Figure 1. Overview of functional measures in human neuroscience. Temporal resolution (x-axis) is plotted against spatial resolution (y-axis). The spatial scale of selected entities of interest is indicated by the arrows.

Figure 2. Examples of Magnetic Resonance Imaging (MRI) data. (A-B) The same slice of brain tissue can appear very different, depending on which relaxation mechanism is emphasized as the source of contrast in the pulse sequence. Using long echo times emphasizes T_2 differences between tissues, and shortening the repetition time emphasizes T_1 differences in tissue. The same slice of the brain acquired as (A) a T_1-weighted image and (B) a T_2-weighted image. (C) Diffusion tensor imaging allows researchers to measure directional diffusion and reconstruct the fiber tracts of the brain. This provides a way to study how different brain areas are connected. Diffusion image is adapted from (Behrens et al., 2007).

Figure 3. Measures available on MR scanners. MRI provides structural (left) and functional (right) measures of the brain. These measures can be used to study psychological processes in relationship to the brain. The combination of MR measures with measures of peripheral processes allows the study of integrated physiological system, e.g. stress responses in the hypothalamic–pituitary–adrenal (HPA) axis, responses in the autonomic nervous system (ANS). Panel with task-related activity adapted from Huth et al. (2012) with permission from Elsevier. Spectrogram from gray matter voxel (inset) adapted from (Finsterbusch et al., 2013) with permission from John Wiley and Sons.

Figure 4. Space basic tradeoffs in fMRI. Choices of sequences (EPI vs. spiral, standard vs. multi-slice) and parameters (TR, TE, parallel vs. non-parallel acquisition) occupy different points in a space of basic tradeoffs. Multi-slice EPI acquisition is a special case, since it allows for high temporal resolution and high coverage at the same time.

Figure 5. Prediction from brain activity. Various measures of brain activity can be used to predict outcome variables. The example on the left used trained a voxel weight map on BOLD responses to painful stimuli in order to predict pain. The weight map can easily be applied to new subjects by computing the dot-product between the weight map and the brain activity (e.g. GLM parameter estimates). The predicted pain response can then be compared against actual pain ratings for this condition.

Figure 6. The tradeoff between contrast detection (y-axis) and hemodynamic response function (HRF) shape estimation power (x-axis), and the performance of different types of designs on each. Power on each axis is expressed here in terms of z-scores in a simulated group analysis (n = 10, effect sizes estimated from visual cortex data in Wager et al., 2005b). The double-circle shows a block design with roughly optimal task alternation frequency (16 s / task). The dark circles show power for a number of randomized event-related designs with roughly optimal parameters under linear modeling assumptions (randomized sequences with a stimulus every 2 s). The dark squares show truncated m-sequence designs with the same parameters as the randomized design. The open circles show results for genetic algorithm (GA) optimized designs with the same parameters. Each circle represents the results of one run of the optimization routine with different user-specified detection/shape estimation tradeoff settings.

Figure 7. Balancing scan time and participants. **A** Contrast for working memory task (N-back vs. Rest, N=21, data from (van Ast et al., 2014). Hot colors indicate positive activations, cool colors indicate negative effects. **B** Statistical power as a function of sample size with a maximum total scan time of 40 h (see text for details). Different plot power curves for three different significance thresholds. The maximum power is around 38 subjects, each scanned for approximately 1 h.

Figure 8. Schematic of fMRI data analyis steps. Information about the experimental design is necessary at all levels of the analysis pipeline. After data

acquisition on the scanner, images are reconstructed (usually automated onsite). Preprocessing can included various steps (slice-timing correction, motion correction, co-registration with the structural image, spatial normalization to a template space, and spatial smoothing). Data analysis itself can also include different steps and techniques. The standard approach is to estimate subject-level general linear models (GLM) and compute group statistics for contrast images obtained from all subjects. Other kinds of analyses include connectivity analysis, multivariate techniques, or prediction of other variables from brain data.

Figure 9. Hemodynamic responses and subject-level GLM. **A** Empirical hemodynamic response functions (HRFs) from primary visual and motor cortices, adapted from Lindquist et al. (2008), with permission of Elsevier. Data were sampled at a high time-resolution using a recently developed acquisition technique (100 ms, with whole-brain coverage at 12 mm spatial resolution), permitting visualization of fine-grained details of the HRF, including the initial dip in signal due to blood oxygenation decreases. Participants saw a contrast-reversing checkerboard (visual) for 100 ms and made a button-press response an average of 250 ms later. The signal in the visual cortex proceeds the signal in the motor cortex throughout the duration of the HRF. **B** In an fMRI experiment with four conditions (A to D), the stimulus function is convolved with a canonical HRF to obtain two sets of predicted BOLD responses. The responses are placed into the columns of a design matrix X and used to compute whether there is significant signal corresponding to the conditions in a particular BOLD time course.

Figure 10. The hierarchical structure of fMRI experiments. Each of the 1 to *N* subjects has participated in multiple sessions on the same or different days. Within each session the data acquisition (and task) is split into *K* different runs. In between runs the scanner is stopped, but the subjects remains inside the bore. SPM does not differentiate between runs and sessions. It refers to each run as ‘session’. Within each run *T* volumes are acquired, resulting in a time-series of *T* samples for each voxel. The inter-sample interval is the repetition time *TR*.

Figure 11. Varieties of connectivity. Connectivity analyses can be grouped into 3 classes; Functional connectivity analyses correlate time-series obtained from voxels or regions with time-series from other regions or voxels. Psychophysiological interactions (PPI) look for time-series correlations depend on a external, psychological, variable. Measures of effective connectivity estimate a directional of effect of one source onto other time-series. Multivariate connectivity techniques analyze multiple sources (time-series) simultaneously to estimate their connectivity graphs.