ESSENTIALS OF FUNCTIONAL NEUROIMAGING

Tor D. Wager[1*]
Luis Hernandez[2]
Martin Lindquist[3]

[1]Columbia University, Department of Psychology
[2]The University of Michigan, Department of Engineering
[3]Columbia University, Department of Statistics

Summary:
27497 words (text, without references)
32097 total words (with references)
4 tables
14 figures

Draft of a chapter to appear in G. G. Berntson and J. T. Cacioppo (Eds.), *Handbook of neuroscience for the behavioral sciences*. New York: Wiley.

Running head: FUNCTIONAL NEUROIMAGING

* Address correspondence to:
Dr. Tor D. Wager
Department of Psychology
Columbia University
1190 Amsterdam Ave.
New York, NY 10025
Phone: 212-854-5318
E-mail: tor@psych.columbia.edu

Acknowledgements

Parts of this chapter are adapted from Wager, T. D., Hernandez, L., Jonides, J., & Lindquist, M. (2007). Elements of functional neuroimaging. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), Handbook of Psychophysiology (4th ed., pp. 19-55). Cambridge: Cambridge University Press. We would like to thank Dr. Doug Noll for providing Figure 3, and Matthew Davidson, Damon Abraham, Katherine Dahl, and Bryan Denny for helpful comments on the manuscript.

There has been explosive interest in the use of brain imaging to study cognitive and affective processes in recent years (T. D. Wager, Hernandez, Jonides, & Lindquist, 2007). The use of neuroimaging data from functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) studies is central to the emerging fields of cognitive neuroscience, affective neuroscience, social cognitive neuroscience, neuroeconomics, and related neuro-behavioral disciplines. fMRI and PET data are being combined with data on human performance and psychophysiology in increasingly sophisticated ways to yield models of human thought, emotion, and behavior. 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 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 methods used in the human neuro-behavioral sciences have limitations, and neuroimaging is no exception. The current trend is towards increasingly interdisciplinary approaches that use multiple methodologies to overcome some of the limitations of each method used in isolation. For example, recent advances in engineering and signal processing allow electroencephalography (EEG) and fMRI data to be collected simultaneously (Goldman, Stern, Engel, & Cohen, 2000), which provides improved temporal precision, among other benefits. Combined fMRI and EEG/magnetoencephalography (MEG) analyses are being developed that can provide better spatio-temporal resolution than either method alone (A. M. Dale et al., 2000; V. Menon, Ford, Lim, Glover, & Pfefferbaum, 1997). Neuroimaging data are also being combined with transcranial magnetic stimulation to integrate 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 depending on the research questions. True interdisciplinary collaboration is exceedingly challenging, because team members must know enough about the other disciplines to be able to talk intelligently with experts in each field. Lead researchers on neuroimaging projects must know when to ask for help with various aspects of the project and what kind of expertise is needed. Supporting researchers must understand enough about the research questions and possibilities to bring their knowledge to bear in an optimal way. 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 road map of what kinds of study design and analysis options are available and

some of their important limitations.

The various aspects of PET and fMRI methodology are organized here into four sections. Section I deals with what neuroimaging techniques measure, including the essentials of PET and fMRI data acquisition and the relationship between brain activity and observed signals in each modality. Section II describes the hierarchical structure of neuroimaging data and how these data are used to make psychological inferences. 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. This section also deals with statistical inferences about populations and the localization of results from functional neuroimaging studies. Section III discusses experimental design for neuroimaging experiments, including some considerations that are particular to neuroimaging data. Section IV deals with neuroimaging data analysis, including sections on artifacts and signal processing before analysis ("pre-processing"), the general linear model (GLM) and brain-behavior and brain-physiology relationships, and methods for investigating brain connectivity.

I. WHAT NEUROIMAGING TECHNIQUES MEASURE

There are many ways to measure brain function, including fMRI, PET, single positron emission computerized tomography (SPECT), electroencephalography (EEG) with analysis of event-related potentials (ERP) (Fabiani, Gratton, & Federmeier, 2007; Pizzagalli, 2007), magnetoencephalography (MEG) (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993), and near-infrared spectroscopy (Villringer & Chance, 1997). Each of these techniques provides a unique window into the functions of mind and brain.

In this chapter we will mainly focus on PET and fMRI, because they are the most widely used and they provide the most anatomically specific information across the entire brain. The relatively good spatial resolution of PET and fMRI complement the precise timing information provided by EEG and MEG. In addition, the ability of fMRI to measure activity over the entire brain every 2 s or so offers 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.

[Insert Table 1 about here.]

PET and fMRI can be used in different ways, depending on the software and type of imaging used, 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; imagine little cubes stacked together), providing 3-D brain maps. A partial list of popular measures and techniques is summarized in Table 1, and includes 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/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 both be used to measure brain activity, though each has unique advantages. These are summarized in Table 2.

[Insert Table 2 about here.]

I.A. Measures of brain structure

I.A.1 Structural scans. MRI can provide detailed anatomical scans of gray and white matter with a spatial resolution well below 1 mm^3. These images are used to localize functional results in individual or group-averaged brains. A growing set of measures related to brain structure allows for the analysis of changes with practice or development, effects of aging, and differences between healthy individuals and those with a number of psychological disorders. A popular way of analyzing gray-matter density is the voxel-based morphometry (VBM) method (Ashburner & Friston, 2000; Good et al., 2001), which uses structural image intensity to measure of gray- and white-matter density. Other methods use measures of cortical thickness derived from surface reconstruction and unfolding (Fischl, Sereno, & Dale, 1999; Van Essen & Dierker, 2007) , or the volume of anatomically defined structures. For example, a recent study reported that London taxi drivers, who had developed extensive expertise in spatial navigation, had larger posterior hippocampi (Maguire et al., 2000).

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 (manipulations of the magnetic field's shape), are used to acquire data that can be reconstructed into a map of the MR signal sources, i.e., an image of the brain. Pulse sequence programming is the province of physicists and bioengineers; such divisions of labor among physicists, psychologists, neuroscientists, and statisticians are a hallmark of neuroimaging, which is highly interdisciplinary in nature. For more in-depth information, we recommend two very approachable texts (Elster, 1994; Huettel, Song, & McCarthy, 2004), and more detailed texts for the advanced reader (Bernstein, 2004; Haacke, 1999).

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. 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).

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 otherwise constant, T_1- and T_2-weighted images can produce detailed representations of the boundaries between gray matter (mostly cell bodies), white matter (mostly axons), and cerebrospinal fluid (CSF). Because T_2^* is sensitive to flow and oxygenation, T_2^*-weighting is used to create images of brain function. An example of the same slice of tissue imaged with T_1 and T_2 weighting can be seen in Figure 1A and 1B. 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.

[Insert Figure 1 about here.]

I.A.2 Anatomical connectivity. MRI pulse sequences may also be tuned to be sensitive to directional (anisotropic) patterns of water diffusion, which may be used to track the course of axon (fiber) tracts. Water diffuses more readily along the axons that make up the brain's white matter than across them. Diffusion tensor imaging (DTI) is an increasingly popular technique for measuring directional diffusion and reconstructing the fiber tracts of the brain (Figure 1C) (Denis Le Bihan et al., 2001). New tractography analyses for quantifying the thickness and connectivity of these tracts are being rapidly developed (Behrens, Berg, Jbabdi, Rushworth, & Woolrich, 2007). Such tools will increasingly 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 & Behrens, 2006). DTI can be combined with other techniques, such as fMRI or other anatomical and neurochemical measures. For example, one study used DTI 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).

I.A.3 Other anatomical measures. PET imaging is complementary to MRI in a very important way: It permits estimation of the density of a variety of neurochemical receptors across the brain. A radioactive label is chemically attached to a pharmacological agent and injected into the bloodstream. The agent is transported into the brain, where it binds to a specific class of receptors, depending on its biochemical nature. The PET camera detects the radiation emitted when the radioactive label decays, and so provides a 3-D map of the distribution of labeled substance across the brain. Kinetic models, which use systems of differential equations in conjunction with known kinetic properties of the pharmacological agent, can be used to quantify the label in extravascular space (tissue) and that bound to receptors. Related neurochemical measures, such as the rate of dopamine synthesis, can be obtained as well. This method is often used to study changes in endogenous neurochemical release as well, and we describe it more fully below. In addition, MR spectroscopy provides a way of testing for the presence of biochemicals and some kinds of gene expression in a brain volume of interest, though this has not been widely applied yet in the cognitive neurosciences. Certain compounds produce well-defined peaks in the measured frequency spectrum, and can be readily detected, but many compounds of interest in neuroscience cannot.

I.B. 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 regional cerebral blood flow (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—they are "anti-matter electrons." 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. Photoreceptive cells 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; it detects the energy that results from their annihilation.

Depending on the design, most PET scanners are made up of an array of detectors that are 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, as illustrated in Figure 2**.** Small tubes (called "septa" or "collimators") are placed around the detectors to shield them from radiation from the sides and help prevent coincidences due to background radiation.

[Insert Figure 2 about here.]

The injected tracer will be distributed throughout the blood vessels and tissue of the brain (indeed, throughout the rest of the body as well). Each pair of detectors counts photons emitted within the column of tissue between them. The density of photons that were emitted at each location can be calculated mathematically from the number of counts at each position or "projection". PET images are simply 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, of course, on what 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 such as 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).

Kinetic models have been developed to estimate how much tracer is contained in different categories, or *compartments*, of blood and tissue. Different forms of kinetic modeling have different numbers of compartments; for example, a two-compartment model estimates how much of the radiolabeled compound is in the vasculature as opposed to in the brain. A three-compartment model 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 (i.e., 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).

I.C Measures of brain activity using fMRI

Unlike PET, which can provide measures of overall 'activity' or specific neurochemical systems, fMRI is principally used to obtain measures of regional brain activity (see Table 1). The most popular method is currently the Blood Oxygenation Level Dependent (BOLD) signal (Kwong et al., 1992; Ogawa et al., 1992), which is obtained using T_2^*-weighted images. Other methods are available but less widely used, including several varieties of Arterial Spin Labeling (ASL) (Williams, Detre, Leigh, & Koretsky, 1992), which use pulse sequences sensitive to blood volume or cerebral perfusion. We focus here on BOLD physiology because it is overwhelmingly the most common method in current use.

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 (Shulman, Rothman, Behar, & Hyder, 2004; Sibson et al., 1997). As oxygen is extracted from the blood, the hemoglobin becomes paramagnetic—iron atoms are more exposed to the surrounding water—which creates small distortions in the $\mathbf{B_0}$ field that cause a T_2^* decrease (i.e. a faster decay of the signal). Increases in deoxyhemoglobin can lead to a decrease in BOLD signal, often referred to as the "initial dip." The initial decrease in signal (whose existence is controversial) is followed by an increase, due to an over-compensation in blood flow that tips the balance towards oxygenated hemoglobin (and less signal loss due to dephasing), which leads to a higher BOLD signal. Initially, fMRI was performed by injection of contrast agents (such as iron) with paramagnetic properties, but the discovery that the T_2^* relaxation rate of oxygenated hemoglobin was longer than that of deoxygenated hemoglobin led to BOLD imaging as it is currently used with humans, without contrast agents (Kwong et al., 1992; Ogawa, Lee, Kay, & Tank, 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 & Frank, 1997; Buxton, Uludag, Dubowitz, & Liu, 2004; Heeger & Ress, 2002; Vazquez & Noll, 1998). Some relationships among factors that contribute to BOLD signal are summarized in Figure 3.

[Insert Figure 3 about here]

Essentially, the BOLD signal corresponds relatively closely to the local electrical field potential surrounding a group of cells—which is itself likely to reflect changes in post-synaptic activity—under many conditions. Demonstrations by Logothetis and colleagues have shown that high-field BOLD activity closely tracks the position of neural firing and local field potentials in cat visual cortex, even to the locations of specific columns of cells responding to particular line orientations (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). However, under other conditions, neural activity and BOLD signal may become decoupled (Disbrow, Slutsky, Roberts, & Krubitzer, 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. Many regions may show changes in neural activity that is missed because they do not change the net metabolic demand of the region.

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. Shulman and Rothman (Shulman & Rothman, 1998) suggest that increased glucose uptake is controlled by astrocytes, whose end-feet contact the endothelial cells lining the walls of blood vessels. Glutamate, the primary excitatory neurotransmitter in the brain, is released by 60-90% of the brain's neurons. When glutamate is released into synapses, it is taken up by astrocytes and transformed into glutamine. When glutamate activates the uptake transporters in an astrocyte, it may signal the astrocyte to increase glucose uptake from the blood vessels. Although it remains plausible that some metabolic (and BOLD) increases could be caused by increased *inhibition* of a region, in many tasks where both BOLD studies and neuronal recordings have been made, BOLD increases are found in regions in which many cells increase their activity. This is true in studies of visual processing, eye movements, task switching, working memory, food reward, pain, and other domains.

I.D 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 synthesized in a cyclotron and attached to the pharmacological agent. Labeled compounds are injected into the arteries by either a bolus (a single injection) 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. As described above, because the dynamics of radioligands are complex, pharmacological agents must be carefully selected and tested in animals. Parameters from these studies are used in kinetic models to aid in quantifying how much labeled substance is bound to the receptor type of interest (Frey, 1999).

I.E Limitations of PET and fMRI

As one might expect, both PET and MRI have their share of pitfalls. One should consider the limitations of each technique not only when designing experiments, but also when examining the neuroimaging literature. One should always ask the following question: "Are the activations caused by the experimental paradigm or by other unwanted sources?" Conversely one should also ask: "Were there other active regions that were missed by the experimental paradigm?" Some of these errors may have occurred because of the spatial or temporal limitations of the technique, or they may be due to image artifacts or mischaracterized noise.

I.E.1 Spatial limitations. Neither PET nor fMRI is well-suited for imaging small subcortical nuclei or cortical microcircuitry, though advances in high-field imaging and parallel acquisition methods are helping. The spatial resolution of PET is on the order of 1-1.5 cm^3. fMRI resolution can be less than 1 mm^3 in high-field imaging in animals, but is typically on the order of 27-36 mm^3 or more for human studies. Thus, features such as cortical columns and even major sub-nuclei (e.g., there are 30 or so in each of the amygdala and thalamus) cannot typically be identified. 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). Careful work in individual participants has demonstrated the imaging of ocular dominance columns in humans (Cheng, Waggoner, & Tanaka, 2001).

While this resolution does not sound all that bad, there is another factor that seriously limits the spatial resolution in most studies. That is the fact that making inferences about populations of subjects requires analyzing groups of individuals, each with a different brain. Usually, individual brains are aligned to one another through a registration or warping process (see Section IV), which introduces substantial 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 (T. D. Wager, Reading, & Jonides, 2004).

Overcoming these limitations with high-resolution fMRI imaging is a challenging and developing research area. By focusing on particular regions and omitting data collection in much of the brain, voxels on the order of 1.5 mm per side can be acquired, 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). This technique provides several advantages over standard mapping techniques. Resolution can potentially be considerably enhanced, particularly when using high-field imaging and analysis techniques that remove some spread in fMRI signal due to draining veins (R. S. Menon, 2002). Secondly, collecting thinner slices can reduce susceptibility artifacts and improve imaging around the base of the brain (Morawetz et al., 2008). 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 (Zeineh, Engel, Thompson, & Bookheimer, 2003). However, there are costs as well. There is a substantial loss in signal due to the smaller volume of each voxel. In addition, coregistration techniques that ensure structure-to-function correspondence and normalization techniques typically used to provide inter-subject registration in group studies do not work very well when only a portion of the brain is imaged, because there are fewer functional landmarks for registration. Ultimately, high-resolution studies are very promising when a small set of subcortical nuclei or nearby cortical regions are of primary interest.

I.E.2 Acquisition artifacts. Artifactual activations (i.e., patterns of apparent activation arising from non-neural sources) and image distortions may arise from a number of sources, some unexpected. An early study, for example, found a prominent PET activation related to anticipation of a painful electric shock in the temporal pole (Reiman, Fusselman, Fox, & Raichle, 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!

Important types of artifacts include those related to magnetic susceptibility, reconstruction, head movement, heart-beat and breathing, instability in magnetic gradients used to acquire images, and radio-frequency interference from outside sources. Many of these artifacts apply only to or are more pronounced with fMRI, and we provide more details on dealing with artifacts in analysis in Section IV.

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 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 & Spencer, 1999) or spiral in/out sequences (Glover & Law, 2001) and/or using a physical magnetic shim held in the mouth of the participant (Wilson & Jezzard, 2003). Signal loss and distortion may be further minimized by using improved reconstruction algorithms (Noll, Fessler, & Sutton, 2005) and "unwarping" algorithms that measure and attempt to correct EPI distortion (Andersson, Hutton, Ashburner, Turner, & Friston, 2001).

Functional MRI also contains more sources of signal variation due to noise than does PET, including 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 (M. Lindquist & Wager, in press; Martin A. Lindquist, Waugh, & Wager, 2007). The vast majority of fMRI designs use discrete events that can be repeated many times over the course of the experiment—for example, the most common method for studying "emotion" in fMRI is to repeatedly present pictures with emotional content.

I.E.3 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 in subsequent sections, but it is important to note here that PET and fMRI measure very 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 processes. 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 limitation due largely to the latency and duration of the hemodynamic response to a neural event. Typically, changes in blood flow do not reach their peak until several seconds after local neuronal and metabolic activity has occurred. Thus, the locking of neural events to the vascular response is not very tight. Because of this limitation, a promising current direction is the estimation of the onset and peak latency of fMRI responses, and other parameters, averaged over many trials (M. A. Lindquist & Wager, 2007; R. S. Menon, Luknowsky, & Gati, 1998; Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000). We provide a more thorough discussion of this and related issues in Section IV.

II. FROM DATA TO PSYCHOLOGICAL INFERENCE

II.A Goals of data analysis: Prediction and Inference

A fundamental question in neuroimaging research is determining what one hopes to achieve with the chosen method. Successful research requires a solid grasp of what kinds of imaging results constitute evidence for a 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, Polyn, Detre, & Haxby, 2006). More often, however, the psychologist would like to infer something about the structure of mental processes from imaging data. Making inferences about psychological states has been termed *reverse inference*, because it involves estimating the relative probabilities of different psychological hypotheses given the data, whereas what is observed in imaging studies is the probability of the data given a psychological state.

Chapter 1 of this Handbook (Cacioppo & Berntson, in press) deals extensively with psychological inference from physiological data. In addition, several excellent papers review this issue in brain imaging (Poldrack, 2006; Sarter, Berntson, & Cacioppo, 1996) and physiological data generally (Cacioppo & Tassinary, 1990). Though we do not recapitulate this discussion here, we note that psychological inferences based on activation in single brain regions is particularly problematic. For example, researchers have inferred that romantic love and retribution involve "reward system" activation because these conditions activate the caudate nucleus (Aron et al., 2005; de Quervain et al., 2004), that social rejection is like physical pain because it activates the anterior cingulate (Eisenberger, Lieberman, & Williams, 2003), among countless similar conclusions in the literature. 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, Luu, & Posner, 2000; Kastner & Ungerleider, 2000; Paus, 2001; Van Snellenberg & Wager, in press; T. D. Wager, Jonides, & Reading, 2004; T. D. Wager, Jonides, Smith, & Nichols, 2005a).

Fortunately, there are other types of reverse inference that are less specific about the localization of psychological functions in the brain but more defensible. These inferences fall into two major categories: those based on dissociations in activation among tasks, and those based on activation overlap across tasks. Both types involve 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.

A recent study in our laboratory illustrates this approach. We found that different types of task switching, or switching attention from one feature or object to another, differentially activate a set of regions thought to be involved in the control of attention (T. D. Wager, Jonides, Smith, & Nichols, 2005b). Four types of switches were dissociable—each produced higher brain activity in some regions

than the others—paralleling behavioral findings that performance switch costs are more highly correlated for similar types of switches (see Figure 4). The implication from this converging evidence is that different types of attention switching involve unique processes.

[Insert Figure 4 about here.]

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.

A second type of psychological inference is based on the overlap in activation among tasks, which is often taken as evidence that the tasks share common processes (Sylvester et al., 2003). In the task switching study shown in Figure 4, even though there were quantitative dissociations in activation magnitude, all regions responded to at least two types of task switch, and some responded to all four. This implies at least some common processes across switch types, paralleled by significant performance correlations across types (T.D. Wager, Jonides, & Smith, 2006).

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 one million neurons, and it is entirely possible that different subsets of neurons in the same voxel are activated by different tasks. Paton et al. (Paton, Belova, Morrison, & Salzman, 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. Wang et al. (Wang, Tanaka, & Tanifuji, 1996), using optical imaging, found topographical maps of perceived head orientation in areas of temporal cortex that spanned only about 1 mm of cortex.

fMRI-adaptation designs. These issues have led to another method for assessing the utilization of common neural substrates across tasks. This method relies on repetition-suppression effects, or adaptation of fMRI responses to repeated events. One can take advantage of this effect to tell whether two stimulus types (A and B) activate the same or different populations of neurons within a voxel (Grill-Spector & Malach, 2001). If a stimulus of type A is presented, then subsequent presentations of A will result in reduced signal (adaptation). The logic is that other stimuli, say of type B, that engage the same set of neurons will also evoke a reduced signal ($[B_{alone} - B_{afterA}]$, cross-adaptation), whereas those that engage different neurons (even within the same voxel) will evoke a larger signal. Thus, small cross-adaptation effects may provide evidence that B engages different populations of neurons, whereas large cross-adaptation effects may be evidence that the circuitry for B and A overlap. However, caution in interpretation is in order, because habituation ($[B_{alone} - B_{afterA}]$) can be caused by the mechanical properties of the vascular bed (Vazquez et al., 2006; T. D. Wager, Vazquez, Hernandez, & Noll, 2005), and not to a neuronal habituation process. In fact, the response to A immediately after B is always likely to produce a reduced response compared with A alone because of the time it takes the vessels to regain their original shape after a BOLD response. This complicates the inference that similar adaptation and cross-adaptation

implies overlapping neuronal populations. Another issue is that a recent electrophysiological study designed to test the validity of this paradigm reported differential habituation in single-cell recording to two stimuli, even though they both activated the same neuron (Sawamura, Orban, & Vogels, 2006). This finding challenges the inference that different adaptation and cross-adaptation effects implies different populations of neurons. Finally, though interpretations of fMRI-adaptation effects are often cast in terms of neuronal firing, more global processes related to memory may play an important role as well (Henson, 2003).

II.B The hierarchical structure of neuroimaging data

Whichever type of inference is desired, inference is based on data, usually from multiple individuals. This section describes the structure of neuroimaging data, and the following sections describe some conceptual essentials of the steps that lead to psychological inference: valid group analysis, thresholding techniques, and localization of activated regions. Proper analysis of multi-subject data in each voxel yields a statistical parametric map (SPM) of the reliability of contrast values—in other words, images that contain test statistic values (e.g., t-values) and p-values for the group analysis at each voxel. These statistic images are thresholded, with some provision for correcting for multiple comparisons across the many brain voxels tested, to obtain maps of suprathreshold or 'activated' regions. Activated regions are localized relative to standard brain landmarks, often with the aid of brain atlases and norms, and interpreted in the context of other human and animal literature.

Imaging data typically involves repeated observations over time—in fMRI as many as two thousand brain images can be collected in the course of a single imaging session for each participant. These images are nested within task conditions (e.g., tasks A and B, or "switch attention" [for a particular switch type] and "do not switch," in our example study). Task conditions, in turn, are crossed with participant, meaning that they are assessed for each participant. Participants may be additionally nested within groups (e.g., patients vs. controls, young vs. elderly). Most often, a statistical model is specified for each participant that estimates the average response to each task condition of interest.

Responses to different task conditions are compared by calculating *contrasts* across two or more conditions. Those measures are called *contrast values*, and they usually reflect a comparison of the activity levels between task conditions of interest (e.g., A minus B, "switch attention" minus "no switch") that yields a single number for each participant. Contrast values for each voxel yield *contrast images,* three-dimensional maps of activation difference values for each participant. T-tests or comparable analyses can be performed for each voxel to discover where in the brain the difference is reliable. More detail on contrasts is provided in Section IV.

Analyzing contrast values has been referred to as the 'subtraction method,' the logic of which is this: If one tests two experimental conditions that differ by only one process, then a subtraction of the activations of one condition from those of the other should reveal the brain regions associated with the target process. Subtraction logic rests on a critical assumption, what has been called the assumption of "pure insertion" (Sternberg, 1969). According to this assumption, changing one process does not change the way other processes are performed. Thus, by this assumption, the process of interest may be 'purely'

inserted into the sequence of operations without altering any other processes. Violations of subtraction logic have been demonstrated (Zarahn, Aguirre, & D'Esposito, 1997), and evoked activation depends on baseline cerebral blood flow in an area and other factors (Vazquez et al., 2006). However, subtractions remain widely used because comparisons among relative activity levels are central to the inference-making process. The assumption of pure insertion underlies the inference that more observed activity implies more intense neural and/or metabolic processes. However, in defense of the subtraction method, pure insertion need not be quantitatively or strictly true in all cases to yield useful comparisons across conditions.

The contrast method applies to many comparisons other than the simple Task A – Task B subtraction, including incremental variations in task difficulty and factorial designs. It also applies to brain-performance correlation designs, in which activation contrast values are correlated with performance contrast values. These designs may employ multiple control or comparison conditions to strengthen the case for a relationship between activity in a particular brain region and a psychological process. They also extend beyond imaging of "activation" to studies that image neurochemical activity and other signals.

II.C Principles of population inference

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 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).

II.C.1. Mixed vs. 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. 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 complete 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 popular packages such as Hierarchical Linear Modeling (HLM; (Raudenbush & Bryk, 2002)), R, and MLwiN (Rasbash, 2002). Neuroimaging data-friendly mixed-effects models are implemented in FSL (Beckmann, Jenkinson, & Smith, 2003; Woolrich, Behrens, Beckmann, Jenkinson, & Smith, 2004) and FMRISTAT (Worsley et al., 2002) software and are potentially implementable in SPM5.

II.D. Thresholding and multiple comparisons

The results of neuroimaging studies are often summarized as a set of 'activated regions,' such as those shown in Fig. 4. 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, with 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 any method that is used 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 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, which may distort our inferences and the usefulness of the experiment.

[Insert Figure 5 about here.]

[Insert Figure 6 about here.]

Most published PET and fMRI studies do not use either of these corrections; instead, they use arbitrary uncorrected thresholds, as shown in Figure 6, 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. Figure 5B shows the same activation map with spatially correlated noise thresholded at three different P-value levels. Due to the smoothness, the false-positive activation blobs (outside of the squares) are contiguous regions of multiple voxels.

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 (T. D. Wager, Lindquist, & Kaplan, 2007), 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.

II.D.1 FWE correction**.** The simplest way of controlling the FWER is to use Bonferroni correction. Here 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, Taylor, Tomaiuolo, & Lerch, 2004), used in SPM, FMRISTAT, and BRAINSTAT software (Taylor & Worsley, 2006), is another (more theoretically complicated) approach towards controlling the FWER. If the image is smooth and the number of subjects is relatively high (around 20), RFT is less conservative and provides control closer to the true false positive rate than the Bonferroni method. However, with small samples, RFT is often more conservative than the Bonferroni method. It is acceptable to use the more lenient of the two, as they both control the FWER, which is what SPM currently does. 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. Nichols and Hayasaka (T. Nichols & Hayasaka, 2003) provide an excellent review of FWER correction methods, and they find that while RFT is overly conservative at the voxel level, it is somewhat liberal at the cluster level with small sample sizes.

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 in group analyses, SnPM or "Statistical Non-Parametric Mapping" (T. E. Nichols & Holmes, 2002), is based on the use of permutation tests.

II.D.2 FDR control**.** The false discovery rate (FDR) is a recent development in multiple comparison problems developed by Benjamini and Hochberg (Benjamini, 1995). While the FWER controls the probability of any false positives, the FDR controls the proportion of false positives among all rejected tests. 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. Hence, any procedure that controls the FDR can only be less stringent 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.

II.D.3 ROI analysis**.** Because of the difficulty in preserving both false positive control and power in experiments with few subjects, researchers often specify regions-of-interest (ROIs) in which activation

is expected before the study is conducted. ROI analyses are conducted variously over the average signal within a region, the peak activation voxel within a region, or preferably on individually defined anatomical or functional ROIs. Another technique involves testing every voxel within an ROI (e.g., the amygdala) and correcting for the number of voxels in the search volume. This is often referred to as a "small volume correction."

Two important cautions must be mentioned. First, conducting multiple ROI analyses increases the false positive rate. While it may be philosophically sound to independently test a small number of areas in which activation is expected, testing many such regions violates the spirit of *a priori* ROI specification and leads to an increased false positive rate. Small volume corrections in multiple ROIs also do not preserve the false positive rate across ROIs. Second, although activated regions can be used as ROIs for subsequent tests, the test used to define the region must be *independent* of the test conducted in that region. Acceptable examples include defining a region based on a main effect and then testing to see whether activity in that region is correlated with performance, or using the main effect of (A+B) to define a region and then testing for a difference (A – B). Problematic examples are defining a region activating in older subjects and then testing to see 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.

II.E Functional localization and atlases

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 (Thompson, Schwartz, Lin, Khan, & Toga, 1996; Vogt, Nimchinsky, Vogt, & Hof, 1995). 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) (Table 3), and another useful list can be found at http://www.cma.mgh.harvard.edu/iatr/.

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 now standard in research on the visual system (Tootell, Dale, Sereno, & Malach, 1996).

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 (Duvernoy, 1995; Haines, 2000; Mai, Assheuer, & Paxinos, 2004), 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 & 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. The Talairach region labels were digitized, and a popular software program, the Talairach Daemon (Lancaster et al., 2000), allows researchers to map neuroimaging results onto Talairach's labels. In addition, at least two popular software packages, AFNI (Cox, 1996) and BrainVoyager (Brain Innovation, Maastricht, Netherlands), allow researchers to align brains from neuroimaging studies to "Talairach space" using a few key landmarks identified on the brain and on the atlas. The alignment is performed by estimating 12 linear transformation parameters, which include translation, rotation, zooms, and shears. 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, Johnsrude, & Owen, 2002; Devlin & Poldrack, 2007) for discussion) as better alternatives are now available.

Modern digital atlases based on group-averaged anatomy have largely replaced the use of the Talairach brain. A current standard in the field is the Montreal Neurologic Institute's (MNI's) 305-brain average[1] (Collins, Neelin, Peters, & Evans, 1994), shown in Figure 7A, which is the standard reference brain for two of the most popular software packages, SPM and FSL (S. M. 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

[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.

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*. *In this process,* intensity values in a brain image are matched to a reference atlas image (template) by deforming the brain image in linear or nonlinear ways and using search algorithms to find the deformations that yield the best match. One preferred intensity-based method is the "unified segmentation and normalization" algorithm in SPM5 (Ashburner & Friston, 2005).

A recent and very promising 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 & 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, Sereno, Tootell, & Dale, 1999; Van Essen & Dierker, 2007). Several free packages implement surface-based normalization to templates registered to MNI space, including FreeSurfer (Table 3), Caret/SureFit software (Van Essen et al., 2001), and BrainVoyager. AFNI, using SUMA software (Saad, Reynolds, Argall, Japee, & Cox, 2004), and FSL have facilities for viewing and analyzing surface-based data with FreeSurfer and SureFit. Surface-based add-ons in these packages permit surface-based registration to be performed after gross registration to the Talairach landmarks.

Tor Wager 1/6/08 1:58 PM
Comment: True? For all packages?

Because the original BAs were not precisely or rigorously defined in a group, reporting of BAs using the Talairach atlas is not recommended (Devlin & Poldrack, 2007). However, modern probabilistic cytoarchitectural atlases are being developed (Amunts, Schleicher, & Zilles, 2007), and some of these are available digitally either from the researchers or within FSL and SPM (as part of the SPM Anatomy Toolbox (Simon B. Eickhoff et al., 2005) (Figure 7B and 7C). In addition, software packages increasingly provide tools for visualizing activations relative to known functional and structural landmarks. Caret software, for example, allows study results to be mapped to a variety of atlases, including atlas brains included with SPM2, SPM99, and the Van Essen Lab's surface-based PALS atlas (see Figure 7F). Brain sections, surfaces, and flattened maps can be visualized, and digital overlays include probabilistic maps of visuotopic regions, orbitofrontal regions from a recent anatomical study (Ongur, Ferry, & Price, 2003), structural and functional landmarks, and a database of previous studies and reported peaks. The associated SumsDB database is a repository for study maps and peak coordinates (Table 3).

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. Some meta-analysis maps are available on the SumsDB and BrainMap databases

(Table 3), and a number are available on the web from individual researchers. Our lab currently has images from a number of meta-analyses available on the web (Table 3), and these can be loaded into SPM, FSL, BrainVoyager, Caret, or other packages for visualization.

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; see Figure 7E). 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.

III. EXPERIMENTAL DESIGN FOR NEUROIMAGING EXPERIMENTS

III.A Types of experimental designs

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 blocked design, 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 in certain types of trials, but suffer from decreased power. 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.

III.A.1 Blocked 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 blocked design. A blocked 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., the activation due to attention switching) using a blocked 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 blocked structure of PET designs (and blocked 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 blocked 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 blocked design is that it typically offers increased statistical power to detect a change. Under ideal conditions, it has been shown that blocked designs can be over 6 times as efficient as randomized event-related designs (T. D. Wager & 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 (Liu, 2004; Skudlarski, Constable, & Gore, 1999; T. D. Wager & Nichols, 2003). However, it is worth noting that this is not always true as the relative power of a blocked 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.

[Insert Figure 8 about here.]

III.A.2 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, Buckner, & Dale, 1998). In fMRI the whole brain can be measured every 2-3 seconds (the “TR”, or repetition time of image acquisition), depending on the type of data acquisition and the spatial resolution of the images. 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 (Aguirre, Zarahn, & D'Esposito, 1998; K. J. Friston, Frith, Turner, & Frackowiak, 1995). Figure 8 shows the canonical HRF used in SPM software.

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 (Figure 8; see also Section IV). The canonical estimates typically come from studies of brief visual and motor events. In practice, however, the timing and shape of the HRF are known to vary across the brain, within an individual and across individuals (Aguirre et al., 1998; Schacter, Buckner, Koutstaal, Dale, & Rosen, 1997; 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.

Blocked 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, Veltman, Ashburner, Josephs, & Friston, 1999).

In a single-trial event-related design, events are spaced at least 20-30 s apart in time. FMRI signal can be observed on single trials if the eliciting stimulus is very strong (Duann et al., 2002), permitting the possibility of fitting models at the level of an individual trial (Rissman, Gazzaley, & D'Esposito, 2004). This promising technique enables the testing of relationships between brain activity and trial-level performance measures such as reaction time and emotion ratings for particular stimuli (Phan et al., 2004).

Early studies frequently employed selective averaging of activity following onsets of a particular type (Aguirre, Singh, & D'Esposito, 1999; Buckner et al., 1998)(Menon, Luknowsky, & Gati, 1998). However, even brief events (e.g., a 125 ms visual checkerboard display) have been shown to affect fMRI signal more than 30 s later (T. D. Wager, Vazquez et al., 2005). Because the selective averaging procedure does not take the stimulus history into account, it must be used with caution when responses to different events overlap in time. Because of this, the majority of analyses, including those that estimate the shapes of HRFs, are currently done within the GLM framework (see Section IV).

Reports that the fMRI BOLD response is linear with respect to stimulus history (Boynton, Engel, Glover, & Heeger, 1996) encouraged the use of more rapidly-paced trials (Zarahn et al., 1997), spaced less than 1 s apart in the most extreme cases (Burock, Buckner, Woldorff, Rosen, & Dale, 1998; Anders M. Dale & Buckner, 1997). Here linearity implies that the magnitude and shape of the HRF does 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)(Birn, Saad, & Bandettini, 2001; K. J. Friston, Mechelli, Turner, & Price, 2000; T. D. Wager, Vazquez et al., 2005), 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 & Henson, 1999; T. D. Wager & 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 (Figure 4). 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 sometimes increased statistical efficiency—must be weighed against potential problems with nonlinearity, multicolinearity, 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.

III.A.3 Optimized experimental designs. What constitutes an optimal experimental design depends on the psychological nature of the task as well as on 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 IV. For a full treatment, however, we refer the reader to several excellent papers (Josephs & Henson, 1999; Liu, 2004; S. Smith, Jenkinson, Beckmann, Miller, & Woolrich, 2007; T. D. Wager & Nichols, 2003). 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 & Boynton, 2002), an one based on a genetic algorithm (T. D. Wager & Nichols, 2003), which incorporates m-sequence designs as a starting point and considers the relative importance of various contrasts to the study goals in calculating optimality.

III.B Design strategies for enhanced psychological inference

Thus far, we have alluded to a simple contrast between two conditions, the subtraction of a control condition (B) from an experimental one (A), or [A – B]. Such contrasts are critical because any task, performed alone, produces activation in huge portions of the brain. Though contrasts in event-related designs can usually be more readily interpreted as being evoked by specific psychological or physical events than those in blocked designs, a single contrast leaves much room for incorrect inference. This is because there may be multiple psychological and physical differences between task conditions A and B. Imagine a study that compares a difficult version of a working memory task (A) to an easy one (B). The more difficult task not only requires greater use of working memory, it may also elicit increases in heart rate, more frustration, more error-detection and correction processes, and more monitoring and adjustment of performance. The result is that the [A – B] contrast does not reveal activations associated only with working memory demand.

III.B.1 Parametric modulation designs. One way to constrain interpretation and strengthen the credibility of subtraction logic is to incrementally vary a parameter of interest across several levels (e.g., working memory demand), and perform multiple subtractions or linear contrasts across levels. An example is a study of the Tower of London task (Dagher, Owen, Boecker, & Brooks, 1999), which requires subjects to make a sequence of moves to transfer a stack of colored balls from one post to another in the correct order. The experimenters varied the number of moves incrementally from 1 to 6. Their results showed linear increases in activity in dorsolateral prefrontal cortex across all 6 conditions, suggesting that this area subserved the planning operations critical for good performance.

III.B.2 Multiple control conditions and conjunctions. Another fruitful approach is to include multiple control conditions matched for various aspects of a target task of interest. In our working memory example, this might amount to including a control condition that produces comparable increase in heart rate without involving working memory, and another that is frustrating without involving

working memory, and so on. If a brain region is more activated in the working memory task than each of the control tasks, then the case that the region subserves working memory is strengthened.

One productive line of research using this approach is that of Kanwisher and colleagues in the study of face recognition (Kanwisher, McDermott, & Chun, 1997). In a long series of studies, they identified an area in the fusiform gyrus that responded to pictures of faces and drawings of faces, but not to houses, scrambled faces, partial faces, facial features, animal faces, and other control stimuli. By presenting a large number of control stimuli of various types, Kanwisher et al. were able to rule out many confounding variables and infer that the brain area they studied, which they called the Fusiform Face Area (FFA), was specific to the perception of faces. Though the interpretation of these results as evidencing a face-selective "module" in the cortex is still being debated, this line of research is an excellent example of using multiple control conditions to rule out various alternative hypotheses for the cause of activation of a region. The fact that the ultimate implications for neuroscience are debated is a testament to the difficulty of conceptualizing and ruling out all the plausible confounds, and of making reverse inferences in general.

A natural way of making comparisons using multiple control conditions is to use conjunction analysis, which is a logical 'and' operator across multiple contrasts. One might want to identify voxels active in a [task A – task B] contrast *and* in a [task A – task C] contrast. In general this question is approached by first calculating a statistical map for each contrast of interest, and then selecting those voxels that meet a chosen statistical threshold in both (or all) maps. In effect, the minimum statistic is compared to the *conjunction null hypothesis*, which specifies that all the contrasts must have significant effects for the conjunction to hold (T. Nichols, Brett, Andersson, Wager, & Poline, 2005). This logic holds generally for all kinds of conjunctions, e.g., [A-B] *and* [C-D] *and* [E-F], whether or not they are independent.

Care must be taken when considering the selection of a significance threshold for a conjunction of contrasts [A-B] and [A-C]: Earlier versions of conjunction analysis in SPM99 and SPM2 software (Price & Friston, 1997), for example, tested the *global null hypothesis* that *none* of the effects are truly present. Rejecting this hypothesis implies a true effect in *at least one contrast*, which is actually an 'or' rule: a significant conjunction result in this case implies true activation for contrast [A-B] *or* contrast [A-C] (T. Nichols et al., 2005). The current version of SPM offers the user a choice of which null hypothesis to test, and also offers a range of intermediate alternatives, e.g., the hypothesis that *2 or fewer* of a series of contrasts have true effects (K. J. Friston, Penny, & Glaser, 2005). Unlike the other tests described above, this hypothesis requires the assumption of independence among the contrasts, which is clearly violated in our example conjunction with two control conditions [A-B] *and* [A-C] because they share a baseline. Overall, if one wishes to test for the intersection (logical *and*) of multiple effects, then the *conjunction null* is the proper null hypothesis. In reporting results, the precise procedures and null hypothesis should always be stated; as with other aspects of data analysis, it is *not sufficient* to merely state that one

performed a conjunction analysis with a particular software package.

A note on baselines. Whether a task produces "activation" or "deactivation" depends on the baseline condition with which it is compared. Over the past decade or so, Raichle and colleagues have argued for the idea that a quiet resting state provides a natural baseline condition against which to evaluate task-related activation (Gusnard, Raichle, & Raichle, 2001; Raichle et al., 2001). One source of support is that the oxygen extraction fraction, the ratio of oxygen use to oxygen supplied by blood, is relatively constant across the resting brain. The argument is that this ratio is one that we are equipped to maintain over long time periods, so it provides a natural physiological baseline. Due in large part to the evidence that Raichle has garnered, many researchers compare tasks to an open-eyed fixation or closed-eye resting baseline condition. The inter-trial intervals in an event-related design, if enough rest and temporal 'jitter' is provided, can provide an estimate of task-evoked activation relative to baseline activity (though the baseline level itself cannot be quantified with BOLD fMRI); however, it must be noted that tasks may also elicit *sustained* activity during the inter-trial intervals as well (Visscher et al., 2003).

However, others argue that the 'baseline' state is just another type of cognitive state, albeit one that is poorly experimentally controlled or characterized. Stark and Squire (Stark & Squire, 2001), for example, found that activity in the medial temporal lobes was substantially higher during rest than during some low-level cognitive tasks. Whether a task of interest "activated" or "deactivated" the medial temporal lobes depended on the choice of baseline, begging the question of exactly what kind of mnemonic or other cognitive activity is happening during "rest." Thus, a number of researchers choose to compare tasks of interest to low-level baseline tasks during which mental activity can be more precisely experimentally controlled (Johnson et al., 2005).

Ultimately, the comparison between task states, including rest, is a comparison of activity evoked by different kinds of mental representations. These comparisons can only be psychologically meaningful if the mental processes involved in each task can be specified. However, this does not preclude the resting state as a baseline condition of interest. Proponents of the 'baseline' state recognize it as an active state, and theories of mental activity during rest include simulation of situations, contingencies, and associated thoughts and feelings generally focused on the self (and likely involving memory retrieval and medial temporal lobe activation) (Gusnard et al., 2001). Each investigator must consider these issues in relation to the particular goals of the study when designing the tasks and comparisons.

III.B.3 Factorial designs**.** Another extension of subtraction logic is the factorial design. The study of task switching presented in the introduction to this chapter serves as an example (T. D. Wager, Jonides et al., 2005b). A subset of conditions in the study compared switch vs. non-switch trials for each of two different types: switches among object attributes and switches among objects. This design is a simple 2 X 2 factorial, with 2 types of trials (switch vs. no switch) crossed with 2 types of judgments (object/attribute). This design permits the testing of three contrasts: a) a main effect of switch vs. no switch; b) a main effect of task type; and c) the interaction between the two, which tests whether the switch vs. non-switch difference is larger for one task-type than the other. Factors whose measurements

and statistical comparisons are made within subjects, as are those described above, are *within-subjects* factors, and those whose levels contain data from different individuals (e.g., depressed patients vs. controls) are *between-subjects* factors. Within-subjects factors generally offer substantially more power and have fewer confounding issues (e.g., differences in brain structure and HRF shapes) than between-subjects factors.

Factorial designs allow one to investigate the effects of several variables on brain activations. They also permit a more detailed characterization of the range of processes that activate a particular brain region – e.g., attention switching in general, or switching more for one task-type than the other. Factorial designs also permit one to discover double dissociations of functions within a single experiment. In our example (Figure 4), a factorial design was required in order to infer that a manipulation (e.g. object-switching) affected dorsolateral prefrontal cortex, but a second manipulation (e.g. attribute switching) did not.

Factorial designs can also be used to test for violations of the critical assumption of pure insertion, and for a number of other processes. If the baseline process (e.g., task difficulty) can be manipulated independently of the target process (task switching requirement), then researchers can test for interactions between task difficulty and switching, and test the notion that the switch process produces an additive increase in activation beyond the processes involved in the basic task.

IV. DATA ANALYSIS: IMPLEMENTATION

IV.A Data Preprocessing

IV.A.1 Artifacts, assumptions, and the need for preprocessing. PET and fMRI studies yield data in a format that requires substantial pre-processing before statistical analysis and inference can be performed in a valid and optimal way. The goals of preprocessing are a) to minimize the influence of data acquisition and physiological artifacts; b) to check statistical assumptions and transform data to meet the assumptions; c) to standardize the locations of brain regions across subjects to achieve validity and sensitivity in group analysis.

Most analyses are based on the assumption that all the voxels in any given image were acquired at the same time. Second, it is assumed that each data point in the time series from a given voxel was collected from that voxel only (i.e., that the participant did not move in between measurements). Third, it is assumed that the residual variance will be constant over time and have a white noise distribution. Additionally, when performing group analysis and making population inference, all individual brains are assumed to be in register, so that each voxel is located in the same anatomical region for all subjects. Without any pre-processing, none of these assumptions hold and statistical analysis would not yield valid or interpretable results.

In addition, as noted in Section I.E.2, neuroimaging data contain artifacts that arise from a number of sources, including head movement, brain movement and vascular effects related to periodic physiological fluctuations, and reconstruction and interpolation processes. fMRI data in particular often contains transient spike artifacts and slow drift over time related to a variety of sources, including magnetic gradient instability, RF interference, and movement-induced inhomogeneities in the magnetic field. An example of transient artifacts as visualized in AFNI is shown in Figure 9. Spikes in the data during isolated volume acquisitions are apparent in some entire slices but not others, as shown by the bright bands in the saggital slices at the bottom of Figure 9. This pattern suggests that gradient performance was affected during acquisition of some echo-planar images, which were acquired slice-by-slice in interleaved order in this experiment.

[Insert Figure 9 about here.]

These artifacts likely constitute a violation of the assumptions of normally and identically distributed errors; unless they are dealt with, the consequences include reduced power in group analysis and potentially increased false positives in single-subject inference. A first line of defense is, as with any kind of data analysis, to examine the data—in as raw a form as possible—and diagnose problems. This can be challenging given the massive proportions of neuroimaging data, and different packages provide different ways of looking at the data. As shown in Figure 9, AFNI provides an excellent facility for viewing time-courses and images from one or more voxels (see Table 3 for a list of packages and websites). Spike artifacts are often identified and problematic images removed prior to or in the course of analysis, or minimized using trimming procedures, as in FIASCO software. VoxBo software also has good ‘data-surfing’ capabilities. A popular approach implemented in FSL, FMRISTAT, and specialized packages such as GIFT (see Section IV.C) is to extract principal components or independent components from the whole-brain timeseries and visualize them. These components are increasingly used for artifact removal (Nakamura et al., 2006; Tohka et al., 2007), though care must be taken if single-subject inference is desired not to bias the results by removing variance from the data without accounting for it in the statistical analysis.

Apart from using the procedures described above, the effects of slow drift, the problem of inter-subject registration, and some other kinds of artifacts can be minimized using preprocessing and analysis techniques described below. In the following sections, we focus on fMRI analysis and briefly describe common preprocessing steps. Other neuroimaging methods, including PET, will require some different steps than those described here.

IV.A.2 Preprocessing steps for fMRI. 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. Single-subject analyses do not require the warping step, which introduce spatial uncertainty in terms of anatomical locations, and thus can provide much 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. 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 Analyze, also known as AVW, which uses a separate header file and image file for each brain volume acquired. Other formats, such as NIFTI, are also gaining popularity. A series of images describes the pattern of activity over the course of the experiment. It is also common to store images in a 4-D matrix, where the fourth dimension is time.

Slice Timing. Statistical analysis 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 “target” 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 a target image and the reference image. Usually, the matching process is done by minimizing sums of squared differences between the two images.

Martin Lindquist 1/6/08 2:00 PM

Comment: I usually think of the target image and the reference image to be synonyms, both referring to the image one is matching all the other images to. Here you refer to the target image as being the image that is being transformed. Is this common practice? If it is, this formulation doesn't make sense to me……

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.

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. There are no hard and fast rules for how much movement is too much, but more than 1.5 mm displacement within a scanning session (while the scanner is running continuously) is typically considered problematic, and can usually be avoided with proper instructions to subjects and head restraints.

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 (see Figure 10). 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 (see Section II.E). Here, we describe nonlinear intensity-based normalization as implemented in SPM software.

[Insert Figure 10 about here.]

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 pixels 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, as shown in Figure 10B. 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). Figure 10C shows a process of checking normalization for one subject. We have identified control points on the MNI ICBM152 template brain (left) that correspond to easily identifiable features. Then, we have taken those points and overlaid them on the subject's normalized T1 image. For this subject, unlike the pathological case in Figure 10B, each of the control points matches with the corresponding anatomical feature on the subject's brain quite well. Such checking can be done in a number of ways, and though there are unfortunately no hard and fast rules for how to check and how much error is too much, each lab should develop a set of standardized procedures.

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! It is optimal in terms of sensitivity to acquire images at the desired resolution and *not* employ smoothing. Some recent acquisition schemes acquire images at the final functional resolution desired, which also permits much more rapid image acquisition as time is not spent acquiring information that would be discarded in analysis (M. Lindquist, Glover, & Shepp, in press).

Previously, many investigators applied *temporal smoothing* to the data as well as spatial smoothing. This procedure is another form of filtering like the high-pass filtering done in the course of model estimation; it removes high-frequency signals from the data, whereas high-pass filtering removes low-frequency signals. This procedure was implemented in SPM99 software (Table 3) primarily as a way of facilitating accurate estimation of the degrees of freedom, which was assumed after smoothing to equal that implied by the kernel. However, this approach has largely been replaced by more standard timeseries models (e.g., autoregressive modeling). There is no expected benefit to temporal smoothing on sensitivity, as it further decreases the temporal resolution of the data, and it is not recommended.

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 & Thomas, 1991). Typically, a single structural image is co-registered to the first or mean functional image.

IV.B Localizing task-related activations with the GLM

The GLM is the most common statistical method for assessing task – brain activity relationships in neuroimaging (Worsley & Friston, 1995). GLM 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.

Because of the hierarchical structure of the data, an appropriate analysis for multi-subject PET and fMRI studies is the mixed-effects GLM model. This is often approximated by performing a GLM model for each subject, and using the resulting activation parameter estimates in a 'second level' group analysis. We refer to this as the *unweighted summary statistic* approach. FSL software currently performs a mixed-effects analysis, whereas the most typical analysis in SPM, AFNI, BrainVoyager, VoxBo, and other packages is the *unweighted summary statistic* approach. We describe the mechanics of a single subject analysis and then the mixed-effects approach in the following sections.

IV.B.1 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. 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 predictors 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:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \textbf{(5)}$$

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 estimated $\hat{\beta}$s correspond to the *estimated magnitude of activation* for each psychological condition described in the columns of **X**.

One of the advantages of the GLM is that there exists an algebraic solution for $\hat{\beta}$ that minimizes the squared error:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \quad \textbf{(6)}$$

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:

$$se(\hat{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\hat{\sigma} \quad \textbf{(7)}$$

Notably, the error term is composed of two separate terms from different sources. σ is the residual error variance, which depends on many factors, including scanner noise. $(\mathbf{X}^T\mathbf{X})^{-1}$ depends on the design matrix itself, and reflects both the variability in the predicted signal and covariance among preditors (i.e., multicolinearity). Design optimization algorithms, described in Section III.A.3, 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. fMRI data are autocorrelated—signals are correlated with versions of 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. 5) 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:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y} \quad \textbf{(8)}$$

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. (6). The standard errors of each contrast are the diagonals of:

$$se(\mathbf{C}^T\hat{\beta}) = \mathbf{C}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}\hat{\sigma} \quad \textbf{(9)}$$

The whitening process is omitted here for simplicity, but can be readily incorporated. 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.

Assumptions. The model-fitting procedure assumes that the effects due to each of the predictors add linearly and do not change over time (i.e., the system is linear and time-invariant). The inferential process assumes that the observations are independent, that they all come from the same distribution, and that the residuals are distributed normally and with equal variance across the range of predicted values. All of these assumptions are violated to a degree in at least some brain regions in a typical imaging experiment, which has prompted the development of a number of important extensions, including diagnostic tools and robust model-fitting procedures (Loh, 2008; Luo & Nichols, 2003; T. D. Wager,

Keller, Lacey, & Jonides, 2005). Violations of the assumptions are not merely a theoretical nuisance. They can make the difference between a valid finding and a false positive result, or between finding meaningful activations in the brain and wasting substantial time and money.

Diagnostic tools have been developed for exploring the data, looking for artifacts, and checking a number of assumptions about the data and model (Loh, 2008; Luo & Nichols, 2003), and like many tools developed by members of the neuroimaging community, they are freely available on the internet. The quantity of data—e.g., 100,000 separate regressions on 1000 data points per subject x 20 subjects—and the software and data structures that support its analysis makes it very difficult to examine assumptions and check the data, which makes such diagnostic tools all the more important.

Another active area of research concerns strategies for dealing with some known violations of assumptions, described below. Violations of independence can be handled in a limited way using generalized least squares. Violations of equality and normality can be dealt with by using nonparametric permutation tests to make statistical inferences (T. E. Nichols & Holmes, 2002), or, if they result from the presence of outliers, by robust regression techniques (T. D. Wager, Keller et al., 2005). Free implementations of each of these extensions are available (Table 3).

IV.B.2 GLM model-building in fMRI. Perhaps the most challenging task in linear regression analysis is the creation of *realistic* predictions of task-related signals for the columns of **X**. PET images integrate across many psychological events, obviating the need for accurate models but also limiting the specificity with which activation can be linked to specific events or time periods. As discussed in Section III.A.2, a popular method of forming predicted BOLD timeseries is to use a canonical HRF. The process is shown in Figure 8. To build the model, researchers start with an 'indicator' vector representing the neuronal activity for each condition sampled at the resolution of the fMRI experiment, shown at the left of Figure 8 for four hypothetical event types (A – D). This vector has zero value except during hypothesized neural activation periods, when the signal is assigned value of 1. Each indicator vector is convolved with the HRF to yield a predicted time course related to that event, which forms a column of the **X**. The rightmost panel shows **X** in image form, a common format for presentation in papers.

If the canonical HRF fits the shape of the BOLD response to psychological events, then using the canonical HRF simplifies the analysis and has great sensitivity to detect differences. Consider two psychological events A and B that both activate a voxel, but with different amplitudes, as shown in the top left panel of Figure 11. Empirical timecourses are shown in light lines, and the fitted responses (model fits) with the canonical HRF are shown in dark lines. The [A – B] contrast will appropriately reflect the different response amplitudes.

However, the canonical HRF is a double-edged sword. If the canonical HRF does not fit, there is

at best a drop in power, and at worst false positives and mis-interpretation of results (M. A. Lindquist & Wager, 2007). Consider an example in which two conditions A and B produce responses of equivalent amplitude, but at different delays. This is shown in the top center panel of Figure 11, where the response to B is delayed by 3 s. Since the HRF shape is fixed, any difference in model fits will produce a difference in the only free parameter, amplitude. In this example, the estimated amplitude for A will be greater than for B. Without some additional diagnostic tests, one might falsely infer that A activates the brain region more than B. This example illustrates the importance of visualizing the data and fits, rather than on simply interpreting a statistically significant result at face value.

Comparing groups of individuals (e.g., older versus younger adults, or patients and normal controls) can be especially problematic. If one finds [A – B] amplitude differences, are those differences caused by differences in neural activity amplitude or the timing and shape of the vascular component of the BOLD response? Elderly subjects have reduced and more variable shapes of their HRFs compared to younger subjects (D'Esposito, Zarahn, Aguirre, & Rypma, 1999), making direct comparisons with a canonical HRF problematic. Alternate approaches include a) measuring HRFs in visual and motor cortex for each individual subject using a separate task (Aguirre et al., 1998) or b) using a more flexible model of the HRF by using a *basis set*, which we describe next.

Basis sets. In the previous discussion, conditions are modeled by a single linear regressor, which allows one to estimate only the amplitude of the predicted response ($\hat{\beta}$) or contrast ($C^T\hat{\beta}$). Alternatively, the same 'neural' indicator vector can be convolved with *multiple* canonical waveforms and entered into multiple columns of X for a single event type. These reference waveforms are *basis functions*, and the predictors for an event type constructed using different basis functions can combine linearly to better fit the evoked BOLD responses. An example is shown in the second row of Figure 11, in which a linear combination of the canonical HRF and its temporal derivative provide better fits to responses that look similar to the HRF (left panel), are shifted in time (center panel), or have extended activation durations (right panel). This basis set is the most popular current alternative to the canonical HRF alone among users of SPM software (K. J. Friston, Glaser et al., 2002; K. J. Friston, Josephs, Rees, & Turner, 1998). Notice that the fits are better, but changes in delay and duration are far from perfectly modeled.

The ability of a basis set to capture variations in hemodynamic responses such as those depicted in Figure 11 depends on both the number and shape of the reference waveforms. There is a fundamental tradeoff between flexibility to model variations and power. This is because each parameter is estimated with error, and flexible models can tend to model noise and thus produce noisier parameter estimates.

One of the most flexible models, a finite impulse response (FIR) basis set, contains one free parameter for every time-point following stimulation in every cognitive event-type that is modeled

(Glover, 1999; Goutte, Nielsen, & Hansen, 2000; Ollinger, Shulman, & Corbetta, 2001). Using such a model makes minimal assumptions about the shape of the HRF because the $\hat{\beta}$s estimate the average response at each time point following the onset of an event. The FIR model is a preferred way to estimate and visualize the shape of BOLD responses, and it is implemented in major software packages including AFNI, SPM, and FSL. An example of model fits using a smooth FIR model, which is constrained to produce smooth response functions, is shown in the third row of Figure 11. The model fits (dark black lines) fit the data reasonably accurately in all conditions, including those shifted in time (center) and extended in duration (right).

Other choices of basis sets include those composed of principal components (Aguirre et al., 1998; Woolrich, Behrens, & Smith, 2004), cosine functions (Zarahn, 2002), radial basis functions (Riera et al., 2004), spectral basis sets (Liao et al., 2002), and other functions. The bottom row in Figure 11 shows fitted responses from a basis set recently developed in our lab that uses three superimposed inverse logit functions to model the rise, fall, and undershoot of the BOLD response (M. A. Lindquist & Wager, 2007). The model can handle both delays and variations in duration, making a single model appropriate for both brief events and prolonged epochs of stimulation. In addition, fits are as accurate as the FIR model fits for these data, and simulations showed that the model compares favorably to a range of other models in terms of statistical power. The model is freely available (see Table 3).

Basis sets offer a major advantage—more accurate modeling of the HRF across subjects and across the brain—but they pose additional technical difficulties that make their use less common than perhaps it should be. First, it is not straightforward to calculate contrasts across conditions when there are multiple parameter estimates per condition. Leaving out some basis functions when calculating contrasts, though it is often done, is *not* generally advised. An alternative is to calculate one contrast per basis function for each contrast of interest. Group analysis can then be done using repeated measures analyses at the second level (in group analysis) rather than the usual one-sample t-test. However, there is a cost in power when basis functions are added, and in general whenever more parameter estimates are compared.

Physiological noise and covariates of no interest. In both PET and fMRI designs, additional predictors are typically added to account for known sources of noise in the data. These nuisance covariates are included to reduce noise and to prevent signal changes related to head movement and physiological (e.g., respiration) artifacts from influencing the contrast estimates. In addition, covariates that implement high-pass filtering, or removal of signal frequencies below a specified cutoff, can also be added at this stage; this is the standard approach in SPM software. In PET, a common covariate is the global (whole-brain) mean signal value for each subject, included to control for differences in amount of radioactive tracer in circulation.

In fMRI, the signal typically drifts slowly over time, so that the most power is in the lowest temporal frequencies. This characteristic has prompted the widespread use of *high-pass filters* that

removes fluctuations below a specified frequency cutoff from the data. High-pass filtering is often performed in the GLM analysis by adding covariates of no interest (e.g. low-frequency cosines). Of course, care must be taken to ensure that the fluctuations induced by the task design are not in the range of frequencies removed by the filter! Design optimization algorithms can take this into account when constructing trial sequences (T. D. Wager & Nichols, 2003).

Much of the autocorrelated noise and other noise variance in fMRI may come from aliased physiological artifacts (Lund, Madsen, Sidaros, Luo, & Nichols, 2005). Thus, it is increasingly popular to measure heart beat and respiration during scanning and to use pre-processing algorithms for removing signals related to measured physiological fluctuations from the data prior to analysis (Glover, Li, & Ress, 2000). Programs for doing this are typically available from authors of research articles, but have not yet been incorporated as standard tools in neuroimaging analysis packages.

IV.B.3. Group analysis

The analysis described so far has been, for fMRI datasets, an analysis of data from a single subject. However, researchers are often interested in making inferences about a population, not just about a single subject or even a set of individual subjects, which requires a group analysis. Both PET and fMRI studies nearly always involve collecting more than one image per subject, and testing for the significance of effects in a group of subjects. In fMRI, typically, separate GLM analyses are conducted on the time series data for each subject at each voxel in the brain to estimate the magnitude of activation evoked by the task. This is called a “first level” analysis. These estimates are carried forward and tested for reliability across subjects in a “second level” group analysis. In PET, the first level analysis often consists of simple image subtractions, followed by the same type of second level analysis as for fMRI.

The unweighted summary statistics approach referred to in Section II consists of a simple one-sample t-test across contrast estimates for each subject. This analysis, like others discussed so far, is repeated at each voxel. It can be specified in the GLM framework, so Eqs. 5-7 hold, and independence is typically assumed across subjects so no prewhitening is needed. The one-sample t-test for overall activation corresponds to a test of the model intercept in a GLM model. Additional covariates across subjects (e.g., average performance scores) can be specified and tested in simple or multiple regression. Two-sample and ANOVA designs to compare groups and related GLM variants can also be specified. Including covariates can improve statistical power for the test of overall activation, though care must be taken: the significance of the intercept can only be assessed if all other covariates are transformed to have a mean of zero.

The unweighted summary statistic approach is valid if the contrast standard error is the same across all subjects, which implies identical design matrices and residual variances. This is rarely if ever

true in practice, though the cost is mostly in the statistical power of the analysis and it is still widely used. Full mixed-effects models relax those stringent assumptions by considering the standard errors within each subject as well as contrast estimates. Mixed-effects analyses are standard in FSL and FMRISTAT software (see Section II.C.1 and Table 3).

Mixed-effects analyses essentially weight subjects when calculating group statistics. The larger a subject's standard error, the less reliable their estimate, and the less that subject should contribute to the group results. This requires estimating *variance components*: One component is variance related to within-subject measurement error and model mis-fitting (σ^2_W), and another component is variance related to true inter-individual differences among subjects (σ^2_B). Accurate estimation of the relative contribution of error within- and between-subjects allows for appropriate weighting. Restricted maximum likelihood (ReML) is a popular type of estimate of variance components based on the residuals. Since variance estimates and model fits ($\hat{\beta}$s) are inter-dependent, iterative algorithms such as EM are used to estimate ReML variance components.

IV.B.4 Statistical power and sample size. 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 acqusition-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, and it will be different for each task and each brain voxel (Zarahn & Slifstein, 2001)(Desmond & 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.

Figure 12 shows plots of power (*y*-axes) as a function of sample size (*x*-axes) for three effect sizes in two kinds of analysis. The effect sizes are Cohen's *d* values, which is defined as mean activation magnitude divided by its standard deviation, for a simple one-sample t-test in group analysis. In behavioral sciences, $d = 0.3$, 0.5, and 1 are considered small, medium, and large effect sizes, respectively. Most activations reported in neuroimaging have effect sizes that are substantially larger—$d = 2$ or more. However, this is partly because voxel-wise mapping capitalizes on chance due to selection bias: Voxels in which chance favors the evidence for activation have large effect sizes and tend to be reported. Whereas

observed effect sizes in published reports are usually over-estimated due to selection bias, the problem is exacerbated when many tests are performed. Here, we show power curves here for effect sizes of 0.5, 1, and 2. Figure 12A shows results for a whole-brain search with 200,000 voxels, a typical number depending on acquisition and analysis choices, and FWE correction at $p < .05$ using the Bonferroni method. To achieve 80% power with a reasonable sample size, the effect size must be larger than 0.5, and around 40 subjects are required for $d = 1$ and 18 subjects for $d = 2$. Figure 12B shows the same results using nonparametric permutation testing, which takes into account the spatial smoothness in the data. We used nonparametric thresholds from 10 analyses from various studies reported in (T. Nichols & Hayasaka, 2003) to estimate the effective number of independent comparisons and thus power. With nonparametric analysis, around 25 subjects for $d = 1$ and 11 subjects for $d = 2$ provides 80% power.

[Insert Figure 12 about here.]

Design optimization procedures can be employed before data is ever colleted to increase the effect size. For a fixed effect size and sample size, power depends on the within-subject standard error, $(se(\mathbf{C}^T\beta))$, which depends on both the design matrix, $\mathbf{X}$, and the residual standard deviation, σ (Equation 9). The latter can be reduced by optimizing data collection (e.g., pulse sequences and hardware) and in the study design by ensuring the engagement of subjects in the tasks. Error related to X can be minimized during experimental design by carefully choosing the number, sequence, and spacing of events to minimize the design-related component of the standard error, $\mathbf{C}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}$. Effective minimization increases predictor variance and reduces predictor covariance (i.e., multicolinearity), and is particularly critical in event-related fMRI. It is possible to build an event-related fMRI design in which even large neuronal effects cannot be detected! For this reason, computer-aided design optimization can be very useful (Buracas & Boynton, 2002; T. D. Wager & Nichols, 2003).

[Insert Figure 13 about here.]

Finally, both theory and simulations show that there is a substantial tradeoff in power between *detecting* activation differences between conditions using an assumed HRF shape and estimating the *shape* of evoked activations with a more flexible model (Liu, Frank, Wong, & Buxton, 2001). This tradeoff is shown in Figure 13, in which shape-estimation power is shown on the x-axis and contrast-detection power is shown on the y-axis. The points in the model represent designs with different sequences and timing of events. Blocked designs have the highest [A – B] contrast detection power when the canonical HRF is used, but provide little information about the shape of the HRF. M-sequences, or sequences which are orthogonal to themselves shifted in time, provide optimal shape estimation power (the non-optimality in the figure is due to truncation of the m-sequences - so they are not perfect), but low detection power (Buracas & Boynton, 2002). Random event-related designs fall somewhere in between.

As the Figure shows, designs optimized with a genetic algorithm (T. D. Wager & Nichols, 2003) can produce substantially better results than random designs on both measures.

IV.B.5 Bayesian inference. Recently, Bayesian methods have received a great deal of attention in fMRI literature. Bayesian inferential methods are now key components in several major fMRI analysis software packages (e.g. SPM and FSL). A full treatment of Bayesian methods is beyond the scope of this chapter, but an excellent overview can be found in Gelman et. al. (2004). A key difference from the 'frequentist' approach discussed above (which subsumes classical inference in the GLM and its extensions) is that Bayesian analysis combines evidence from the data through priors—beliefs about the data specified as probabilities prior to data collection—to yield posterior probability values. This can be a big advantage in that estimates from data (e.g., of HRF shapes) can be easily regularized based on known information from other studies. Such prior constraints are also possible in frequentist analyses, though they require modifications and/or special procedures; lasso, ridge regression, and robust regression are examples.

If one does not want to impose strong prior beliefs, then it is possible to use *non-informative priors,* which is implemented in the Bayesian approach in FSL software (Woolrich, Behrens, Beckmann et al., 2004). For the single-level model this leads to parameter estimates that are equivalent to those obtained using classical inference. Another way to choose prior beliefs is by estimating them from data. This is the 'empirical Bayes' approach. It is a hybrid between classical and Bayesian inference which can provide some regularization without biasing the results of hypothesis tests, and is used in SPM software (K. J. Friston, Glaser et al., 2002; K. J. Friston, Penny et al., 2002).

Martin Lindquist 1/6/08 2:00 PM

Comment: Bayesians are typically not particularly fond of hypothesis testing. Rather they like to compare models by comparing the odds on one model over the other (Bayes factors). Doing this allows one to avoid some of the pitfalls in traditional hypothesis testing. I don't know how much we want to go into this here since space is limited, but I don't think we can include the deleted sentences below without having a more in depth discussion. Therefore I vote to delete them……

IV.C Assessing brain connectivity

Human brain mapping has been primarily used to provide maps that show which regions of the brain are activated by specific tasks. Recently, there has been an increased interest in augmenting this type of analysis with connectivity studies that describe how various brain regions interact and how these interactions depend on experimental conditions. It is common practice in the analysis of neuroimaging data to make the distinction between functional and effective connectivity (K. Friston, 1994). Functional connectivity is defined as the undirected association between two or more fMRI time series, while effective connectivity is the directed influence of one brain region on the physiological activity recorded in other brain regions; it implies both causality and directness. It implies causality because the models used to assess effective connectivity are usually directional, and directness in the sense that effective connectivity measures attempt to partial out indirect influences from other regions.

Functional connectivity is a statement about observed associations among regions and/or other performance and physiological variables—for example, the correlation between time series in two regions (bivariate connectivity). Simple functional connectivity analyses usually compare correlations between ROIs, sometimes in a task-dependent fashion, or between a 'seed' region of interest and voxels throughout the brain. Multivariate analysis methods are also used to reveal networks of multiple

interconnected regions. Popular methods include Principal Components Analysis (PCA) (Andersen AH, 1999), Partial Least Squares (PLS) (A. R. McIntosh, Bookstein, F.L., Haxby, J.V., Grady, C.L., 1996) and Independent Components Analysis (ICA) (V. D. Calhoun, Adali, T., Pearlson, G.D. and Pekar, J.J., 2001; McKeown, 1998). Connectivity between two or more regions may result from *direct* influences (i.e., functional links between regions) or *indirect* effects due to common input from a third variable. None of these methods are able to address issues of causality or the common influences of other variables.

[Insert Figure 14 about here.]

Functional connectivity methods can be applied at different levels of analysis, with different interpretations at each level (See Figure 14). Connectivity across time series data can reveal networks that are dynamically co-activated over time (either 'intrinsically,' regardless of task state, or in a task-dependent fashion), and is closest to the concept of communication among regions, though it does not conclusively demonstrate that. Connectivity across single-trial response estimates (Rissman et al., 2004) can identify coherent networks of task-related activations. Whereas these levels are only accessible to fMRI and EEG/MEG, which provide relatively rich time series data, other levels of analysis may be examined in PET studies as well. Connectivity across subjects can reveal patterns of coherent individual differences, which may result from communication among regions but also from differences in strategy use or other genetically determined or learned differences among individuals. Finally, connectivity across studies can reveal tendencies for studies to co-activate within sets of regions, which may be influenced by any of the factors mentioned above, and also differences among tasks or other study-level variables. An example is the finding that studies in which post-traumatic stress disorder (PTSD) patients showed increased amygdala activity tended to be the same studies in which patients showed decreased activation of the medial frontal cortex (Etkin & Wager, 2007). Regardless of the level of analysis, functional connectivity analyses can be useful for understanding which brain activations are part of coherent patterns and which are separate, independent effects of task manipulations.

Activation is generally only informative if it's restricted to specific brain regions (activation of the insula, for example, means little if every other brain region is activated to the same degree). Likewise, demonstrating that connectivity is greater within a set of regions than among other regions (e.g., for the 'cognitive control network' of Cole and Schneider (Cole & Schneider, 2007) or demonstrating two or more separable sets of interconnected regions (such as the multiple separate networks of coherent opioid release reported by Wager (T. D. Wager, Scott, & Zubieta, 2007) can provide valuable information about how brain regions function together. Demonstrating specificity of functional connectivity to a particular task state, as the psychophysiological interaction (PPI)/moderation analysis described below is designed

to do, can be informative about how functional connectivity relates to psychological states. Reporting reciprocal activity (negative correlations) between ventromedial PFC and amygdala, for example, may be of limited usefulness if such correlations can be found in any task state; in that case, they may be a general feature of BOLD physiology or vasculature rather than an interesting instance of communication among brain regions.

Effective connectivity analysis, on the other hand, is model-dependent. Typically, a small set of regions and a proposed set of connections are specified *a priori*, and tests of fit are used to compare a small number of alternative models and assess the statistical significance of individual connections. Because connections may be specified directionally (with hypothesized causal influences of one area on another), the model implies causal relationships. Because there are many possible models, the choice of regions and connections must be anatomically motivated. Most effective connectivity depends on two models: a neuroanatomical model that describes which areas are connected, and a mathematical model that describes how areas are connected. Common methods include Structural Equation Modeling (SEM) (A. McIntosh, Gonzalez-Lima, F, 1994) and Dynamic Causal Modeling (DCM) (K. Friston, Harrison, L, Penny, W, 2003). While 'effective connectivity' methods have become increasingly popular, it is important to keep in mind that the conclusions about direct influences and causality obtained using these models are only as good as the specified models. Any misspecification of the underlying model will almost certainly lead to erroneous conclusions. In particular, the exclusion of important lurking variables (brain regions involved in the network but not included in the model) can completely change the fit of the model and thereby affect both the direction and strength of the connections. Great care always needs to be taken when interpreting the results of these methods.

The distinction between functional and effective connectivity is not entirely clear (Horwitz, 2003). If the discriminating features are a) a directional model in which causal influences are specified; and b) the willingness to make claims about direct vs. indirect connections, then many analyses, including multiple regression, might count as effective connectivity. Indeed, the PPI analysis referred to above is typically described as an effective connectivity model, but it tests an interaction effect using linear regression (whether the slope of the linear association between two variables depends on the level of a third, moderating variable). The three-variable PPI model is actually a very simple SEM, though the criterion of assessing direct effects is not met, since no common indirect influences are accounted for. Thus, in the end, the difference between this model and more complicated SEMs is one of scale, and direct effects in any SEM can only be properly assessed if all relevant "3rd variables" have been included in the model and their connections modeled appropriately.

While the reason many researchers use both SEM and DCM is to obtain the goal of ascribing

causality between different brain regions, it is important to keep in mind that the tests performed in both techniques are based on model fit rather than on the causality of the effect. Similarly, Granger causality (Roebroeck, Formisano, & Goebel, 2005) is another approach that is typically considered to test effective connectivity, though neither causal influences nor direct vs. indirect effects are tested within the basic model framework. Causality is tested strictly in the sense of temporal relationships, rather than on whether activity in a brain region is necessary or sufficient for activity in another. In the end, it is not the label of "functional" or "effective" that is important, but the specific assumptions and robustness and validity of inference afforded by each method.

When performing connectivity and correlation studies it is tempting to make statements regarding causal links between different brain regions. The idea of causality is a very deep and important philosophical issue (Pearl, 2000; Rubin, 1974). Often a cavalier attitude is taken in attributing causal effects and the differentiation between explanation and causation is often blurred. Properly randomized experimental designs permit causal inferences of task manipulations on brain activity. However, in neuroimaging and EEG/MEG studies, all the brain variables are observed, and none are manipulated. Therefore, we do not recommend making strong conclusions about causality and 'direct' influences among brain regions using these methods, because the validity of such conclusions is very difficult to verify. The combination of neuroimaging and TMS or related forms of brain stimulation (Bohning et al., 1997) may provide more reliable causal inferences about the effects of activating one brain region on another. By stimulating the brain, experimental manipulation of one brain area can be achieved and its causal effects on other brain regions thus examined. However, the problem remains of assessing which effects are 'direct' as opposed to mediated by other intervening regions.

IV.C.1 Bivariate connectivity. Functional connectivity is a statement about the observed associations among regions and/or other performance and physiological variables. The simplest approach towards functional connectivity is to simply calculate the cross-correlation between time series from two separate brain regions. The results can be used to determine whether the changes in activity in these regions are related to each other in a linear manner. This idea is expanded upon in seed analysis (Cordes et al., 2000; Della-Maggiore et al., 2000), where the cross-correlation between the time course from a predetermined region or cluster (the seed region) and all other regions of the brain is calculated. This allows researchers to search the brain for other regions that are positively (or negatively) correlated with the activity pattern found in the seed region.

In addition to standard statistical assumptions, time series connectivity typically assumes that the connectivity is instantaneous, meaning that the time constants for neuronal and vascular effects are the same for each pair of regions, and the impulse response functions are thus the same. This assumption is

often likely to be violated, and several approaches have been taken to account for variability in the neuronal activity—fMRI signal coupling, such as multivariate autoregressive modeling (Harrison, Penny, & Friston, 2003; Kim, Zhu, Chang, Bentler, & Ernst, 2007). Granger causality, a kind of autoregressive model discussed in more detail below, is a promising approach towards relaxing this assumption. Whatever method is used, functional connectivity is meaningful only to the degree that it is not driven by artifacts related to image acquisition and physiological noise; some artifactual influences are listed in Figure 14.

Another approach which helps to minimize issues of inter-region neuro-vascular coupling differences and artifacts (but does not eliminate them) is the beta series approach (Rissman et al., 2004). In this technique, correlations are not estimated directly from the time series data. Instead one obtains trial-by-trial estimates of event-related activity within the standard GLM framework. These trial-level activation parameter estimates (called beta values) are correlated across regions to obtain a measure of functional connectivity during each of the individual task components.

IV.C.2 Component analysis: PCA, ICA, and PLS. Multivariate methods model brain imaging data by decomposing a large dataset (e.g. 1000 time points x 100,000 voxels x 20 subjects) into a smaller set of *components* and a series of *weights*. The components may be canonical patterns of activity across time and the weights their distribution across brain space, or vice versa. PCA, ICA, and PLS are variations on this theme. These and related multivariate methods—Canonical Variates Analysis (CVA), Factor Analysis, Ordinal Trends Analysis (Habeck et al., 2005), and the Multivariate Linear Model (MLM) (Kherif, 2002)—are becoming an increasingly important part of the neuroimaging analyst's toolbox. They all share the common core idea of decomposing the data into simpler components that maximize the amount of variability explained by the model. Ultimately, the approaches differ in the criteria used to select components, and in whether or not the experimental design is included as part of the data to be modeled (inclusion is a defining feature of PLS).

Each technique described in this section involves decomposing a data matrix, $\mathbf{Y}$, into a set of spatial and temporal components. Let us define $\mathbf{Y}$ to be a $t \times v$ matrix, where t is the number of time points and v the number of voxels. Each column of $\mathbf{Y}$ is therefore a time series corresponding to one voxel in the brain, and each row is the collection of voxels that make up an image at a specific time point.

Principal Components Analysis (PCA) decomposes the data matrix, $\mathbf{Y}$, by finding linear combinations of time series, each of which make up a column in a matrix $\mathbf{U}$ (also of dimension $t \times v$), such that each column of $\mathbf{U}$ is uncorrelated with every other column of $\mathbf{U}$. The columns of $\mathbf{U}$, called components, are arranged in order of variance explained: the first component explains the most variance possible in $\mathbf{Y}$, the second component explains the maximal amount of remaining variance, and so forth.

Together with their associated spatial maps and variances (described below) these v components perfectly reproduce the data, but most of the total variance is usually captured in just the first few components of **U**. Thus, the first components can be considered a 'compressed' representation of the data.

Because each component is a weighted sum across time series of different voxels, another matrix **V** (of dimension voxel x component, [$v \times v$]) contains columns of voxel weights used to create each component in **U**. For example, the first column of **V** shows how to weight each of the v voxel time series in order to capture the most variance in **Y**, and represents the spatial distribution of the first component. Thus, the columns of **U** are the temporal components (the 'canonical' time series) and those of **V** are the spatial components (the maps across brain voxels) of these time series.

In neuroimaging, the components are usually calculated through singular value decomposition (SVD) of the centered (mean-zero) data. SVD is a numerical technique that decomposes a data matrix, **Y**, into three simpler matrices (i.e. – zeros make up at least half of the new matrices), while still representing the original data. In the case of neuroimaging data, these matrices can be interpreted as temporal components **U** and spatial components **V** such that:

$$\mathbf{Y} = \mathbf{USV}^{\mathbf{T}} \tag{10}$$

With centered (mean-zero) data, **S** is a diagonal matrix (only the diagonal elements are non-zero) whose entries are the 'singular values,' the sums of squared deviations explained by each component. These are related to the eigenvalues such that $\lambda = S^2/(t-1)$. The columns of **V** are the eigenvectors, as in the eigendecomposition described above, and **US** are the component scores (components scaled by the amount of variability they explain), equal to **YV** in the eigendecomposition. The power of this technique lies in that the eigenvectors are orthogonal to each other. In other words, by decomposing the data into its eigenvectors and eigenvalues, we obtain a set of components (whether temporal or spatial) that are uncorrelated with each other. Furthermore, we also obtain coefficients of how heavily those components are represented in the original data. A thorough treatment of eigenvectors, eigenvalues and SVD is provided by Strang (1988).

Once one grasps the central idea of data decomposition into spatial and temporal components, many other techniques, such as ICA, can be understood as variations on this theme. Rather than maximizing the variance explained by each additional, orthogonal component, ICA components are chosen to maximize the statistical independence of the components in a more general sense. The components are not required to be orthogonal; rather, the constraint is that they be independent, i.e., the distribution of one component cannot be predicted from the values of the other, or more formally the joint

probability P(A,B) of components A and B is equal to P(A)P(B). In the Infomax variant of ICA, mutual information between components—a general measure of dependence that does not require the relationships between components to be linear or monotonic—is minimized (McKeown 1998a). ICA assumes that the data, **Y**, are a weighted sum of a number of source signals (timeseries) contained in the source matrix **X**. The data **Y** is a linear mixture of these source components described by the weighting or *mixing matrix* of spatial weights **M**:

$$\mathbf{Y} = \mathbf{MX} \tag{11}$$

Since both **M** and **X** are both unknown, there is no algebraic solution, so iterative search algorithms are used to estimate both **M** and **X**. An alternative decomposition is to transpose the data matrix and treat the spatial components as sources and the temporal components as mixing weights. For more details, we refer the reader to (Bell & Sejnowski, 1995; McKeown & Sejnowski, 1998; McKeown et al., 1998; Petersson, Nichols, Poline, & Holmes, 1999b; Petersson, Nichols, Poline, & Holmes, 1999a).

At first glance, it appears close to impossible to solve Equation 11 for both **M** and **X** simultaneously. However, ICA makes a number of crucial assumptions that allow one to obtain a solution. The main assumptions are that the data set consists of p statistically independent components, where at most one component is Gaussian. The independence assumption entails that the activations do not have a systematic overlap in time or space, while the non-Gaussiantity assumption is required for the problem to be well defined. In addition it is assumed that the mixing matrix, **M**, is both square and invertible which implies that the independent components can be expressed as a linear combination of the data matrix.

Both PCA and ICA reduce the data to a simpler (lower-dimension than that of the v voxels) space by capturing the most prominent variations across the set of voxels. The components may reflect signals of interest or they may alternatively be dominated by artifacts, and it is up to the user to determine which are ‘of interest’ (e.g., task-related). Both ICA and PCA assume all variability results from signal, as noise is not included in the model formulation. In ICA, one issue involved with interpreting the results of an ICA analysis is that the sign of the independent components cannot be determined. In addition, the order of importance of the independent components cannot be determined either. Therefore it is necessary to sift through all of the components to search for ones that are task-related or otherwise of interest. There is also no guarantee that a specific number of components can be used to explain most of the variation as is the case in PCA.

A popular variant in the social sciences literature is factor analysis, which additionally fits a

parameter for the noise variance at each voxel. A disadvantage of factor analysis is that the solution is *rotationally indeterminate*, and thus a number of combinations of spatial and temporal components can explain the same variability in the data. While both ICA and PCA are not rotationally indeterminate, there is some question as to what the 'right' rotation is (in PCA it is determined by the amount of variance explained, which is not an index of meaningfulness since artifacts can create much variance). Interpreting thresholded component maps, as is commonly done, depends critically on establishing a rotation that is meaningful and reliable across studies.

Multi-subject extensions. These techniques as described so far model only a single subject's data. In a group study there is the additional complexity of making population inference. It is not correct to treat all the data as coming from one 'super-subject' and decomposing the group data matrix, for the same reasons that fixed effects analyses in the GLM are not appropriate. One approach is to decompose the group matrix, and subsequently 'back-reconstruct' or estimate spatial weights for each subject for a component of interest (V. D. Calhoun, Adali, Pearlson, & Pekar, 2001). The spatial weights at each voxel across subjects are treated as random variables, and one-sample t-test is conducted to test whether that voxel loaded significantly on that component in the group. This approach is implemented in the Group Analysis of Functional Imaging toolbox (GIFT; Table 3). Another approach, called *tensor ICA*, is to use a 3-way data decomposition, using the group data to estimate temporal components and weights for each subject and each voxel (Beckmann & Smith, 2005). The subject weights at each voxel are then tested for significance. This approach is similar to related PCA-based techniques of PARAFAC (Bro, 1997) and INDSCAL/ALSCAL (Young, Takane, & Lewyckyj, 1978). It is implemented in the ICA tool (called MELODIC) in FSL software (Table 3).

IV.C.3 Structural Equation Modeling. Structural equation Modeling (SEM) has a rich history in the social sciences literature (Bollen, 1989). It was first applied to imaging data by McIntosh and Gonzalez-Lima (A. McIntosh, Gonzalez-Lima, F, 1994). In SEM the emphasis lies on explaining the variance-covariance structure of the data. While SEM allows for the inclusion of latent variables (which is one of its major selling points in the social sciences), this option is not typically used by the neuroimaging community. It should be noted that an SEM without latent variables is typically called path analysis – but we will in the continuation refer to methodology by the name structural equation modeling as this is the common practice in the neuroimaging literature.

Structural Equation Models comprise a set of *a priori* determined regions and directed connections between these regions. A causal relationship is attributed *a priori* to the connections where an arrow from A to B implies that A causes B. Further path coefficients are defined corresponding to each link that represents the expected change in activity of one region given a unit change in the region

influencing it. The path coefficient indicates the average influence across the time interval measured. Algebraically, we can express an SEM model as

$$\mathbf{Y} = \mathbf{MY} + \boldsymbol{\varepsilon} \tag{12}$$

where $\mathbf{Y}$ is the data matrix, $\mathbf{M}$ is a matrix of coefficients that reflect the linear relationship between regions and $\boldsymbol{\varepsilon}$ is independent and identically distributed normal noise. Typically this model is rewritten

$$\mathbf{Y} = (\mathbf{I}-\mathbf{M})^{-1}\boldsymbol{\varepsilon} \tag{13}$$

where $\mathbf{I}$ represents the identity matrix. The solution of the unknown coefficients in contained in $\mathbf{M}$ is obtained by studying the empirical covariance matrix of $\mathbf{Y}$. Like ICA, this model is also not straightforward to solve, and typically one resorts to iterative techniques. The covariance of the data represents how the activities in two or more regions are related. In SEM we seek to minimize the difference between the observed covariance matrix and the one implied by the structure of the model. The parameters of the model are adjusted to minimize the difference between the observed and modeled covariance matrix.

All inferences regarding the path coefficients rest on the use of nested or stacked models. A hypothesis test on a single path coefficient may be performed by comparing the full model, with all path coefficients estimated, with a 'nested' model in which the coefficient of interest is constrained to be zero[2]. The two models are compared using a likelihood ratio test (LRT)—a statistical test of the goodness-of-fit between two models—to test whether a non-zero coefficient results in a significantly better model fit, and thus whether the coefficient is reliably different from zero. The LRT is only valid if it is used to compare nested models, i.e. the more complex model must differ from the simple model only by the addition of one or more parameters.

A similar approach can be taken when making inference about changes in connectivity between different experimental conditions. This is done by first partitioning the data according to the different experimental conditions. Next, two models are specified. In the *null model*, path coefficients are constrained to be equal across conditions, and in the *alternative model,* coefficients of interest are allowed to vary. The LRT is used to test whether there is any significant difference between the models. If a significant difference exists we reject the hypothesis that the path coefficients are equal in both conditions and a condition dependent effect is declared.

[2] Or another test value of interest.

SEM makes a number of assumptions in setting up the model formulation. The data is assumed to be normally distributed and independent from sample to sample. An important consequence of the assumptions is that SEM discounts temporal information. Consequently permuted data sets produce the same path coefficients as the original data, which is a weakness. The assumption of independence is clearly violated in the analysis of a single subject. However, when looking at the individual differences level this assumption is more reasonable.

IV.C.4 Dynamic Causal Modeling. It is important to note that the measurements used in each of the connectivity approaches described so far are hemodynamic in nature and this limits the scope of the interpretation that can be made at the neuronal level. Dynamic Casual Modeling (K. J. Friston, Harrison, & Penny, 2003) is an attempt to move the connectivity analysis from the hemodynamic to the neuronal level. DCM uses standard linear systems analyses techniques, namely state-space design (Franklin, Workman, & Powell, 1997), and treats the brain as a deterministic nonlinear dynamic system that is subject to inputs and produces outputs. It makes inference about the coupling among brain areas and how the coupling is influenced by changes in experimental context. DCM models interactions at the neuronal rather than the hemodynamic level and is therefore more biologically accurate than many other models. However, the hemodynamic properties of the system must also be taken into account, as they can confound the measurements (e.g., a vascular delay could be interpreted as a neuronal delay).

DCM is based on a neuronal model of interacting cortical regions, supplemented with a forward model describing how neuronal activity is transformed into the measured hemodynamic response. Effective connectivity is parameterized in terms of the coupling among unobserved neuronal activity in different regions. We can estimate these parameters by perturbing the system and measuring the response. Experimental inputs cause changes in effective connectivity at the neuronal level which in turn causes changes in the observed hemodynamics.

DCM uses a bilinear model for the neuronal level and an extended Balloon model (Buxton, Wong, & Frank, 1998) for the hemodynamic level. In a DCM model the user specifies a set of experimental inputs (the stimuli) and *a set of* outputs (the activity in each region for each region). The task of the algorithm is then to estimate the parameters of the system, in this case, the “state variables”. Each region has five state variables, four which correspond to the hemodynamic model and the fifth that corresponds to neuronal activity. The estimation process is then carried out using Bayesian statistics: Normal priors are placed on the model parameters and an optimization scheme is used to estimate parameters that maximize the posterior probability. The posterior density is than used to make inferences about the significance of the connections between various brain regions. It should be noted that DCM is quite computationally demanding and is limited to 8 regions in the current implementation of SPM.

IV.C.5 Granger Causality. As mentioned above, the main problem with methods such as SEM and DCM is that any misspecification of the underlying model will lead to erroneous conclusions. Granger causality takes a very different approach to the problem. The technique was originally developed in economics (Granger, 1969) that has recently been applied to connectivity studies (Roebroeck, Formisano, & Goebel, 2005). The benefit of Granger causality is that it does not rely on any *a priori* specification of a structural model, but rather is an approach for quantifying the usefulness of past values from various brain regions in predicting values in other regions. Granger causality provides information about the *temporal precedence* of relationships among two regions, but it is in some sense a misnomer because it does not actually provide information about causality. It is true that one variable (x) may precede a correlated variable (y) because x causes y. For example, hitting a baseball causes flight. However, there may be no causal relationship at all: a rooster may crow (x) every morning just before the sun rises (y), but it does not cause the sun to rise. For purposes of economic forecasting for which the technique was developed—or for making predictions based on fMRI data—the actual causal relationships may not matter, and Granger "causality" may be sufficient to be informative. However, it should not be taken as a measure of true causality.

To illustrate the method let **x** and **y** be two time courses of length N extracted from two brain regions or voxels. Each time course is modeled using a linear autoregressive model[3] of the M^{th} order (where $M \leq N$-1), i.e.

$$x[n] = \sum_{m=1}^{M} a[i]x[n-m] + \varepsilon_x[n] \tag{14}$$

$$y[n] = \sum_{m=1}^{M} b[i]y[n-m] + \varepsilon_y[n] \tag{15}$$

where both ε_x and ε_y are defined to be white noise. The vectors **a** and **b** are coefficients that describe how the current values of the time course depends on its past, and therefore it is clear from this formulation that both time courses depend immediately on their own past M values.

As a second step of the analysis, one can expand each time course's model using the autoregressive terms from the other signal. These additional autoregressive terms correspond to the directed influence (previous history) and not to the instantaneous signal, i.e. they can be written on the format:

$$value\_now = self\_history + other\_history + error$$

More formally, the equations in our example can be expressed as:

[3] Autoregressive models are used to represent processes whose "current" values can be written as a function of their own past values. The order of the model specifies how many steps back into the past the specified function goes.

$$x[n] = \sum_{m=1}^{M} a[i]x[n-m] + \sum_{m=1}^{M} b[i]y[n-m] + \varepsilon_x[n] \quad (16)$$

$$y[n] = \sum_{m=1}^{M} b[i]y[n-m] + \sum_{m=1}^{M} a[i]x[n-m] \varepsilon_y[n] \quad (17)$$

In this formulation the current value of both time courses are assumed to depend both on the past *M* values of its own time course, but also the past *M* values of the other time course.

By fitting each of these models (Equations 14-17), one can perform tests to determine whether the previous history of **x** has predictive value of the time course **y** (and vice versa). If the model fit is *significantly* improved by the inclusion of the cross-autoregressive terms, it provides evidence that the history of one of the time courses can be used to predict the current value of the other and a "Granger-causal" relationship is inferred. To test the influence between the two regions, one compares the fits to the model for each time course both with and without the additional "cross-autoregressive" terms (Roebroeck, Formisano, & Goebel, 2005). The ratio of error sums of squares obtained from these fits are used to define a measure of the linear directed influence from **x** to **y**, which is denoted $F_{x \to y}$. If past values of **x** improve upon the prediction of the current value of y, then $F_{x \to y}$ is large. A similar interpretation, but in the opposite direction, holds for $F_{y \to x}$, which is defined in an analogous manner. The difference between these two terms can be used to infer which region's history is more influential on the other. This difference is referred to as "Granger Causality". From this definition it is clear that the idea of temporal precedence is used to identify the direction and strength of "causality" from information in the data. However, while it can reasonably be argued that temporal precedence is a necessary condition for causation, it is certainly not a sufficient condition. Therefore to directly equate Granger causality and causality is a large leap of faith.

References

Aguirre, G. K., Singh, R., & D'Esposito, M. (1999). Stimulus inversion and the responses of face and object-sensitive cortical areas. *Neuroreport, 10*(1), 189-194.

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *Neuroimage, 8*(4), 360-369.

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N. J., et al. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat Embryol (Berl), 210*(5-6), 343-352.

Amunts, K., Schleicher, A., & Zilles, K. (2007). Cytoarchitecture of the cerebral cortex--More than localization. *NeuroImage, 37*(4), 1061-1065.

Andersen AH, G. D., Avison MJ. (1999). Principal component analysis of the dynamic response measured by fMRI: a generalized linear systems framework. *Magnetic Resonance in Medicine, 17*(6), 785-815.

Andersson, J. L., Hutton, C., Ashburner, J., Turner, R., & Friston, K. (2001). Modeling geometric deformations in EPI time series. *Neuroimage, 13*(5), 903-919.

Aron, A., Fisher, H., Mashek, D. J., Strong, G., Li, H., & Brown, L. L. (2005). Reward, motivation, and emotion systems associated with early-stage intense romantic love. *J Neurophysiol, 94*(1), 327-337.

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry--the methods. *Neuroimage, 11*(6 Pt 1), 805-821.

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *Neuroimage, 26*(3), 839-851.

Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *Neuroimage, 20*(2), 1052-1063.

Beckmann, C. F., & Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject FMRI analysis. *Neuroimage, 25*(1), 294-311.

Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S., & Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage, 34*(1), 144-155.

Bendriem, B., Townsend, D.W. (1998). *The theory and practice of 3D PET*. (Vol. 32). Boston: Dordrecht; Boston: Kluwer Academic, 1998.

Benjamini, Y. a. H., 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, M. A., King, K.F., & Zhou, Z.J. (2004). *Handbook of MRI pulse sequences*. Burlington, MA.: Elsevier Academic Press.

Birn, R. M., Saad, Z. S., & Bandettini, P. A. (2001). Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. *Neuroimage, 14*(4), 817-826.

Bohning, D. E., Pecheny, A. P., Epstein, C. M., Speer, A. M., Vincent, D. J., Dannels, W., et al. (1997). Mapping transcranial magnetic stimulation (TMS) fields in vivo with MRI. *Neuroreport, 8*(11), 2535-2538.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nat Rev Neurosci, 3*(3), 243-249.

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems, 38*(2), 149-171.

Buckner, R. L., Koutstaal, W., Schacter, D. L., Dale, A. M., Rotte, M., & Rosen, B. R. (1998). Functional-anatomic study of episodic retrieval. II. Selective averaging of event-related fMRI trials to test the retrieval success hypothesis. *Neuroimage, 7*(3), 163-175.

Buracas, G. T., & Boynton, G. M. (2002). Efficient design of event-related fMRI experiments using M-sequences. *Neuroimage, 16*(3 Pt 1), 801-813.

Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R., & Dale, A. M. (1998). Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *Neuroreport, 9*(16), 3735-3739.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences, 4*(6), 215-222. [Record as supplied by publisher].

Buxton, R. B., & Frank, L. R. (1997). A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J Cereb Blood Flow Metab, 17*(1), 64-72.

Buxton, R. B., Uludag, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *Neuroimage, 23 Suppl 1*, S220-233.

Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med, 39*(6), 855-864.

Cacioppo, J. T., & Berntson, G. G. (in press). Integrative Neuroscience for the Behavioral Sciences:

Implications for Inductive Inference. In *Handbook of Neuroscience for the Behavioral Sciences*.

Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *Am Psychol, 45*(1), 16-28.

Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp, 14*(3), 140-151.

Calhoun, V. D., Adali, T., Pearlson, G.D. and Pekar, J.J. (2001). Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Human Brain Mapping, 13*, 43-53.

Cheng, K., Waggoner, R. A., & Tanaka, K. (2001). Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron, 32*(2), 359-374.

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *Neuroimage, 37*(1), 343-360.

Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr, 18*(2), 192-205.

Constable, R. T., & Spencer, D. D. (1999). Composite image formation in z-shimmed functional MR imaging. *Magn Reson Med, 42*(1), 110-117.

Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. A., Moritz, C. H., et al. (2000). Mapping functionally related regions of brain with functional connectivity MR imaging. *AJNR Am J Neuroradiol, 21*(9), 1636-1644.

Cover, T. M., & Thomas, J. A. (1991). Elements of Information Theory. In (pp. 18-26). New York: Wiley.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res, 29*(3), 162-173.

D'Esposito, M., Zarahn, E., Aguirre, G. K., & Rypma, B. (1999). The effect of normal aging on the coupling of neural activity to the bold hemodynamic response. *Neuroimage, 10*(1), 6-14.

Dagher, A., Owen, A. M., Boecker, H., & Brooks, D. J. (1999). Mapping the network for planning: a correlational PET activation study with the Tower of London task. *Brain, 122*(Pt 10), 1973-1987.

Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping, 5*, 329-340.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic Statistical Parametric Mapping Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. *Neuron, 26*(1), 55-67.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science, 305*(5688), 1254-1258.

Della-Maggiore, V., Sekuler, A. B., Grady, C. L., Bennett, P. J., Sekuler, R., & McIntosh, A. R. (2000). Corticolimbic interactions associated with performance on a short-term memory task are modified by age. *J Neurosci, 20*(22), 8410-8416.

Denis Le Bihan, M. D., Mangin, J. F., Poupon, C., Clark, C. A., Pappata, S., Molko, N., et al. (2001). Diffusion Tensor Imaging: Concepts and Applications. *JOURNAL OF MAGNETIC RESONANCE IMAGING, 13*, 534-546.

Devlin, J. T., & Poldrack, R. A. (2007). In praise of tedious anatomy. *NeuroImage, 37*(4), 1033-1041; discussion 1050-1038.

Disbrow, E. A., Slutsky, D. A., Roberts, T. P., & Krubitzer, L. A. (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*(17), 9718-9723.

Duann, J. R., Jung, T. P., Kuo, W. J., Yeh, T. C., Makeig, S., Hsieh, J. C., et al. (2002). Single-trial variability in event-related BOLD signals. *Neuroimage, 15*(4), 823-835.

Duong, T. Q., Yacoub, E., Adriany, G., Hu, X., Ugurbil, K., Vaughan, J. T., et al. (2002). High-resolution, spin-echo BOLD, and CBF fMRI at 4 and 7 T. *Magn Reson Med, 48*(4), 589-593.

Duvernoy, H. M. (1995). *The Human Brain Stem and Cerebellum: Surface, Structure, Vascularization, and Three-dimensional Sectional Anatomy with MRI*: Springer-Verlag Wien.

Eickhoff, S. B., Amunts, K., Mohlberg, H., & Zilles, K. (2006). The human parietal operculum. II. Stereotaxic maps and correlation with functional imaging results. *Cereb Cortex, 16*(2), 268-279.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage, 25*(4), 1325-1335.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An FMRI study of social exclusion. *Science, 302*(5643), 290-292.

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

Etkin, A., & Wager, T. D. (2007). Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am J Psychiatry, 164*(10), 1476-1488.

Fabiani, M., Gratton, G., & Federmeier, K. D. (2007). Event-related brain potentials: Methods, theory, and applications. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 85-119). Cambridge: Cambridge University Press.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage, 9*(2), 195-207.

Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp, 8*(4), 272-284.

Franklin, G. F., Workman, M. L., & Powell, D. (1997). *Digital Control of Dynamic Systems*: Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.

Frey, K. A. (1999). Positron Emission Tomography. In G. J. Siegel, B. W. Agranoff, R. W. Albers, S. K. Fisher & M. D. Uhler (Eds.), *Basic Neurochemistry* (6 ed., pp. 1109-1131). Philadelphia: Lippincott, Williams, & Wilkins.

Friston, K. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping, 2*, 56-78.

Friston, K., Harrison, L, Penny, W. (2003). Dynamic causal modelling. *Neuroimage, 19*,

1273-1302.
Friston, K. J., Frith, C. D., Turner, R., & Frackowiak, R. S. (1995). Characterizing evoked hemodynamics with fMRI. *Neuroimage, 2*(2), 157-165.
Friston, K. J., Glaser, D. E., Henson, R. N., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: applications. *Neuroimage, 16*(2), 484-512.
Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage, 19*(4), 1273-1302.
Friston, K. J., Josephs, O., Rees, G., & Turner, R. (1998). Nonlinear event-related responses in fMRI. *Magn Reson Med, 39*(1), 41-52.
Friston, K. J., Mechelli, A., Turner, R., & Price, C. J. (2000). Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *Neuroimage, 12*(4), 466-477.
Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: theory. *Neuroimage, 16*(2), 465-483.
Friston, K. J., Penny, W. D., & Glaser, D. E. (2005). Conjunction revisited. *Neuroimage, 25*(3), 661-667.
Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage, 9*(4), 416-429.
Glover, G. H., & Law, C. S. (2001). Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magn Reson Med, 46*(3), 515-522.
Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn Reson Med, 44*(1), 162-167.
Goldman, R. I., Stern, J. M., Engel, J., Jr., & Cohen, M. S. (2000). Acquiring simultaneous EEG and functional MRI. *Clin Neurophysiol, 111*(11), 1974-1980.
Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N. A., Friston, K. J., & Frackowiak, R. S. J. (2001). A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains. *Neuroimage, 14*(1), 21-36.
Goutte, C., Nielsen, F. A., & Hansen, L. K. (2000). Modeling the haemodynamic response in fMRI using smooth FIR filters. *IEEE Trans Med Imaging, 19*(12), 1188-1201.
Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*, 424-438.
Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst), 107*(1-3), 293-321.
Gusnard, D. A., Raichle, M. E., & Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci, 2*(10), 685-694.
Haacke, E. M. (1999). *Magnetic resonance imaging : physical principles and sequence design*. New York: Wiley.
Habeck, C., Krakauer, J. W., Ghez, C., Sackeim, H. A., Eidelberg, D., Stern, Y., et al. (2005). A new approach to spatial covariance modeling of functional brain imaging data: ordinal trend analysis. *Neural Comput, 17*(7), 1602-1645.
Haines, D. E. (2000). *Neuroanatomy: An Atlas of Structures, Sections, and Systems*. Philadelphia: Lippincott Williams & Wilkins.
Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics, 65*(2), 413-497.
Harrison, L., Penny, W. D., & Friston, K. (2003). Multivariate autoregressive modeling of fMRI time series. *Neuroimage, 19*(4), 1477-1491.

Heeger, D. J., & Ress, D. (2002). What does fMRI tell us about neuronal activity? *Nat Rev Neurosci, 3*(2), 142-151.
Henson, R. N. (2003). Neuroimaging studies of priming. *Prog Neurobiol, 70*(1), 53-81.
Horwitz, B. (2003). The elusive concept of brain connectivity. *Neuroimage, 19*, 466-470.
Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging*. Sunderland, Mass.: Sinauer Associates, Publishers.
Johansen-Berg, H., & Behrens, T. E. (2006). Just pretty pictures? What diffusion tractography can add in clinical neuroscience. *Curr Opin Neurol, 19*(4), 379-385.
Johansen-Berg, H., Behrens, T. E., Robson, M. D., Drobnjak, I., Rushworth, M. F., Brady, J. M., et al. (2004). Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proc Natl Acad Sci U S A, 101*(36), 13335-13340.
Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., Cunningham, W. A., & Sanislow, C. A. (2005). Using fMRI to investigate a component process of reflection: prefrontal correlates of refreshing a just-activated representation. *Cogn Affect Behav Neurosci, 5*(3), 339-361.
Josephs, O., & Henson, R. N. (1999). Event-related functional magnetic resonance imaging: modelling, inference and optimization. *Philos Trans R Soc Lond B Biol Sci, 354*(1387), 1215-1228.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.
Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annu Rev Neurosci, 23*, 315-341.
Kherif, F., Poline J.-B., Flandin G., Benali H., Dehaene S., and Worsley K.J. (2002). Multivariate model specification for fMRI data. *NeuroImage, 16*(4), 795-815.
Kim, J., Zhu, W., Chang, L., Bentler, P. M., & Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Hum Brain Mapp, 28*(2), 85-93.
Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., et al. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci U S A, 89*(12), 5675-5679.
Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping, 10*(3), 120-131.
Liao, C. H., Worsley, K. J., Poline, J. B., Aston, J. A., Duncan, G. H., & Evans, A. C. (2002). Estimating the delay of the fMRI response. *Neuroimage, 16*(3 Pt 1), 593-606.
Lindquist, M., Glover, G. H., & Shepp, L. (in press). Rapid acquisition of functional MRI images.
Lindquist, M., & Wager, T. D. (in press). Application of change-point theory to modeling state-related activity in fMRI. *Applied Data Analytic Techniques for "Turning Points Research"*.
Lindquist, M. A., & Wager, T. D. (2007). Validity and power in hemodynamic response modeling: a comparison study and a new approach. *Hum Brain Mapp, 28*(8), 764-784.
Lindquist, M. A., Waugh, C., & Wager, T. D. (2007). Modeling state-related fMRI activity using change-point theory. *NeuroImage, 35*(3), 1125-1141.
Liu, T. T. (2004). Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part II: design of experiments. *Neuroimage, 21*(1), 401-413.
Liu, T. T., Frank, L. R., Wong, E. C., & Buxton, R. B. (2001). Detection power, estimation efficiency, and predictability in event-related fMRI. *Neuroimage, 13*(4), 759-773.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*(6843), 150-157.

Loh, J. M., Lindquist, M.A., Wager, T.D. (2008). Residual Analysis for Detecting Mis-modeling in fMRI. *Statistica Sinica, To appear*.

Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W. L., & Nichols, T. E. (2005). Non-white noise in fMRI: Does modelling have an impact? *Neuroimage*.

Luo, W. L., & Nichols, T. E. (2003). Diagnosis and exploration of massively univariate neuroimaging models. *Neuroimage, 19*(3), 1014-1032.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., et al. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proc Natl Acad Sci U S A, 97*(8), 4398-4403.

Mai, J. K., Assheuer, J., & Paxinos, G. (2004). *Atlas of the human brain* (2nd ed.). San Diego, Calif.: Elsevier Academic Press.

McIntosh, A., Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapping, 2*, 2-22.

McIntosh, A. R., Bookstein, F.L., Haxby, J.V., Grady, C.L. (1996). Spatial Pattern Analysis of Functional Brain Images Using Partial Least Squares. *NeuroImage, 3*, 143-157.

McKeown, M. J., Makeig, S. (1998). Analysis of fMRI data by blind separation into independant spatial components. *Human Brain Mapping, 6*, 160-188.

Menon, R. S. (2002). Postacquisition suppression of large-vessel BOLD signals in high-resolution fMRI. *Magnetic Resonance in Medicine, 47*(1), 1-9.

Menon, R. S., Luknowsky, D. C., & Gati, J. S. (1998). Mental chronometry using latency-resolved functional MRI. *Proc Natl Acad Sci U S A, 95*(18), 10902-10907.

Menon, V., Ford, J. M., Lim, K. O., Glover, G. H., & Pfefferbaum, A. (1997). Combined event-related fMRI and EEG evidence for temporal-parietal cortex activation during target detection. *Neuroreport, 8*(14), 3029-3037.

Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., & Buckner, R. L. (2000). Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage, 11*(6 Pt 1), 735-759.

Morawetz, C., Holz, P., Lange, C., Baudewig, J., Weniger, G., Irle, E., et al. (2008). Improved functional mapping of the human amygdala using a standard functional magnetic resonance imaging sequence with simple modifications. *Magn Reson Imaging, 26*(1), 45-53.

Nakamura, W., Anami, K., Mori, T., Saitoh, O., Cichocki, A., & Amari, S. (2006). Removal of ballistocardiogram artifacts from simultaneously recorded EEG and fMRI data using independent component analysis. *IEEE Trans Biomed Eng, 53*(7), 1294-1308.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage, 25*(3), 653-660.

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat Methods Med Res, 12*(5), 419-446.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp, 15*(1), 1-25.

Noll, D. C., Fessler, J. A., & Sutton, B. P. (2005). Conjugate phase MRI reconstruction with spatially variant sample density correction. *IEEE Trans Med Imaging, 24*(3), 325-336.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci, 10*(9), 424-430.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S*

A, 87(24), 9868-9872.
Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., et al. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A, 89*(13), 5951-5955.
Ollinger, J. M., Shulman, G. L., & Corbetta, M. (2001). Separating processes within a trial in event-related functional MRI. *Neuroimage, 13*(1), 210-217.
Ongur, D., Ferry, A. T., & Price, J. L. (2003). Architectonic subdivision of the human orbital and medial prefrontal cortex. *Journal of Comp Neurol, 460*(3), 425-449.
Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature, 439*(7078), 865-870.
Paus, T. (2001). Primate anterior cingulate cortex: where motor control, drive and cognition interface. *Nat Rev Neurosci, 2*(6), 417-424.
Pearl, J. (2000). *Causality : models, reasoning, and inference*. Cambridge, U.K. ; New York: Cambridge University Press.
Phan, K. L., Taylor, S. F., Welsh, R. C., Ho, S. H., Britton, J. C., & Liberzon, I. (2004). Neural correlates of individual ratings of emotional salience: a trial-related fMRI study. *Neuroimage, 21*(2), 768-780.
Pizzagalli, D. A. (2007). Electroencephalography and high-density electrophysiological source localization In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 56-84). Cambridge: Cambridge University Press.
Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn Sci, 10*(2), 59-63.
Price, C. J., & Friston, K. J. (1997). Cognitive conjunction: a new approach to brain activation experiments. *Neuroimage, 5*(4 Pt 1), 261-270.
Price, C. J., Veltman, D. J., Ashburner, J., Josephs, O., & Friston, K. J. (1999). The critical relationship between the timing of stimulus presentation and data acquisition in blocked designs with fMRI. *Neuroimage, 10*(1), 36-44.
Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proc Natl Acad Sci U S A, 98*(2), 676-682.
Rasbash, J. (2002). *A User's Guide to MLwiN*: Centre for Multilevel Modelling, University of London.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis* (Second ed.). Newbury Park, CA: Sage.
Reiman, E. M., Fusselman, M. J., Fox, P. T., & Raichle, M. E. (1989). Neuroanatomical correlates of anticipatory anxiety [published erratum appears in Science 1992 Jun 19;256(5064):1696]. *Science, 243*(4894 Pt 1), 1071-1074.
Riera, J. J., Watanabe, J., Kazuki, I., Naoki, M., Aubert, E., Ozaki, T., et al. (2004). A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. *Neuroimage, 21*(2), 547-567.
Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage, 23*(2), 752-763.
Roebroeck, A., Formisano, E., & Goebel, R. (2005). Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage, 25*(1), 230-242.
Rosen, B. R., Buckner, R. L., & Dale, A. M. (1998). Event-related functional MRI: past, present, and future. *Proc Natl Acad Sci U S A, 95*(3), 773-780.
Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology, 66*(5), 688-701.
Saad, Z. S., Reynolds, R. C., Argall, B., Japee, S., & Cox, R. W. (2004). *SUMA: an interface for surface-based intra- and inter-subject analysis with AFNI*. Paper presented at the Biomedical Imaging: Nano to Macro, 2004. IEEE International

Symposium on.
Sandler, M. P. (2003). *Diagnostic nuclear medicine*. Philadelphia, PA: Lippincott / Williams & Wilkins.
Sarter, M., Berntson, G. G., & Cacioppo, J. T. (1996). Brain imaging and cognitive neuroscience. Toward strong inference in attributing function to structure. *Am Psychol, 51*(1), 13-21.
Sawamura, H., Orban, G. A., & Vogels, R. (2006). Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the FMRI adaptation paradigm. *Neuron, 49*(2), 307-318.
Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M., & Rosen, B. R. (1997). Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. *Neuroimage, 6*(4), 259-269.
Shulman, R. G., & Rothman, D. L. (1998). Interpreting functional imaging studies in terms of neurotransmitter cycling. *Proc Natl Acad Sci U S A, 95*(20), 11993-11998.
Shulman, R. G., Rothman, D. L., Behar, K. L., & Hyder, F. (2004). Energetic basis of brain activity: implications for neuroimaging. *Trends Neurosci, 27*(8), 489-495.
Sibson, N. R., Dhankhar, A., Mason, G. F., Behar, K. L., Rothman, D. L., & Shulman, R. G. (1997). In vivo 13C NMR measurements of cerebral glutamine synthesis as evidence for glutamate-glutamine cycling. *Proc Natl Acad Sci U S A, 94*(6), 2699-2704.
Skudlarski, P., Constable, R. T., & Gore, J. C. (1999). ROC analysis of statistical methods used in functional MRI: individual subjects. *Neuroimage, 9*(3), 311-329.
Smith, S., Jenkinson, M., Beckmann, C., Miller, K., & Woolrich, M. (2007). Meaningful design and contrast estimability in FMRI. *Neuroimage, 34*(1), 127-136.
Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage, 23 Suppl 1*, S208-219.
Stark, C. E., & Squire, L. R. (2001). When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proc Natl Acad Sci U S A, 98*(22), 12760-12766.
Sternberg, S. (1969). Memory-scanning: mental processes revealed by reaction-time experiments. *Am Sci, 57*(4), 421-457.
Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychol (Amst), 106*(1-2), 147-246.
Summerfield, C., Greene, M., Wager, T., Egner, T., Hirsch, J., & Mangels, J. (2006). Neocortical connectivity during episodic memory formation. *PLoS Biol, 4*(5), e128.
Sylvester, C. Y., Wager, T. D., Lacey, S. C., Hernandez, L., Nichols, T. E., Smith, E. E., et al. (2003). Switching attention and resolving interference: fMRI measures of executive functions. *Neuropsychologia, 41*(3), 357-370.
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, J. E., & Worsley, K. J. (2006). Inference for magnitudes and delays of responses in the FIAC data using BRAINSTAT/FMRISTAT. *Hum Brain Mapp, 27*(5), 434-441.
Thompson, P. M., Schwartz, C., Lin, R. T., Khan, A. A., & Toga, A. W. (1996). Three-dimensional statistical analysis of sulcal variability in the human brain. *J Neurosci, 16*(13), 4261-4274.
Tohka, J., Foerde, K., Aron, A. R., Tom, S. M., Toga, A. W., & Poldrack, R. A. (2007). Automatic independent component labeling for artifact removal in fMRI. *Neuroimage*.
Tootell, R. B. H., Dale, A. M., Sereno, M. I., & Malach, R. (1996). New images from human visual cortex. *Trends in Neurosciences, 19*(11), 481-489.

Van Essen, D. C., & Dierker, D. L. (2007). Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron, 56*(2), 209-225.
Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., & Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *J Am Med Inform Assoc, 8*(5), 443-459.
Van Snellenberg, J. X., & Wager, T. D. (in press). Cognitive and motivational functions of the prefrontal cortex. In.
Vazquez, A. L., Cohen, E. R., Gulani, V., Hernandez-Garcia, L., Zheng, Y., Lee, G. R., et al. (2006). Vascular dynamics and BOLD fMRI: CBF level effects and analysis considerations. *Neuroimage, 32*(4), 1642-1655.
Vazquez, A. L., & Noll, D. C. (1998). Nonlinear aspects of the BOLD response in functional MRI. *Neuroimage, 7*(2), 108-118.
Villringer, A., & Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences, 20*(10), 435-442.
Visscher, K. M., Miezin, F. M., Kelly, J. E., Buckner, R. L., Donaldson, D. I., McAvoy, M. P., et al. (2003). Mixed blocked/event-related designs separate transient and sustained activity in fMRI. *Neuroimage, 19*(4), 1694-1708.
Vogt, B. A., Nimchinsky, E. A., Vogt, L. J., & Hof, P. R. (1995). Human cingulate cortex: surface features, flat maps, and cytoarchitecture. *J Comp Neurol, 359*(3), 490-506.
Wager, T. D., Hernandez, L., Jonides, J., & Lindquist, M. (2007). Elements of functional neuroimaging. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 19-55). Cambridge: Cambridge University Press.
Wager, T. D., Jonides, J., & Reading, S. (2004). Neuroimaging studies of shifting attention: a meta-analysis. *Neuroimage, 22*(4), 1679-1693.
Wager, T. D., Jonides, J., & Smith, E. E. (2006). Individual differences in multiple types of shifting attention. *Memory & Cognition, 34*(8), 1730-1743.
Wager, T. D., Jonides, J., Smith, E. E., & Nichols, T. E. (2005a). Toward a taxonomy of attention shifting: individual differences in fMRI during multiple shift types. *Cogn Affect Behav Neurosci, 5*(2), 127-143.
Wager, T. D., Jonides, J., Smith, E. E., & Nichols, T. E. (2005b). Towards a taxonomy of attention-shifting: Individual differences in fMRI during multiple shift types. *Cogn Affect Behav Neurosci, 5*(2), 127-143.
Wager, T. D., Keller, M. C., Lacey, S. C., & Jonides, J. (2005). Increased sensitivity in neuroimaging analyses using robust regression. *Neuroimage, 26*(1), 99-113.
Wager, T. D., Lindquist, M., & Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: Current and future directions. *Social, Cognitive, and Affective Neuroscience, 2*(2), 150-158.
Wager, T. D., & Nichols, T. E. (2003). Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *Neuroimage, 18*(2), 293-309.
Wager, T. D., Reading, S., & Jonides, J. (2004). Neuroimaging studies of shifting attention: a meta-analysis. *Neuroimage, 22*(4), 1679-1693.
Wager, T. D., Scott, D. J., & Zubieta, J. K. (2007). Placebo effects on human mu-opioid activity during pain. *Proc Natl Acad Sci U S A, 104*(26), 11056-11061.
Wager, T. D., Vazquez, A., Hernandez, L., & Noll, D. C. (2005). Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. *Neuroimage, 25*(1), 206-218.
Wang, G., Tanaka, K., & Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science, 272*(5268), 1665-1668.
Williams, D. S., Detre, J. A., Leigh, J. S., & Koretsky, A. P. (1992). Magnetic resonance imaging of perfusion using spin inversion of arterial water. *Proc Natl Acad Sci U S A, 89*(1), 212-216.
Wilson, J. L., & Jezzard, P. (2003). Utilization of an intra-oral diamagnetic passive shim

in functional MRI of the inferior frontal cortex. *Magn Reson Med, 50*(5), 1089-1094.
Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage, 21*(4), 1732-1747.
Woolrich, M. W., Behrens, T. E., & Smith, S. M. (2004). Constrained linear basis sets for HRF modelling using Variational Bayes. *Neuroimage, 21*(4), 1748-1761.
Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited--again. *Neuroimage, 2*(3), 173-181.
Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., et al. (2002). A general statistical analysis for fMRI data. *Neuroimage, 15*(1), 1-15.
Worsley, K. J., Taylor, J. E., Tomaiuolo, F., & Lerch, J. (2004). Unified univariate and multivariate random field theory. *Neuroimage, 23 Suppl 1*, S189-195.
Young, F. W., Takane, Y., & Lewyckyj, R. (1978). ALSCAL: A nonmetric multidimensional scaling program with several difference options. *Behavioral Research Methods and Instrumentation, 10*, 451-453.
Zarahn, E. (2002). Using larger dimensional signal subspaces to increase sensitivity in fMRI time series analyses. *Hum Brain Mapp, 17*(1), 13-16.
Zarahn, E., Aguirre, G., & D'Esposito, M. (1997). A trial-based experimental design for fMRI. *Neuroimage, 6*(2), 122-138.
Zarahn, E., & Slifstein, M. (2001). A reference effect approach for power analysis in fMRI. *Neuroimage, 14*(3), 768-779.
Zeineh, M. M., Engel, S. A., Thompson, P. M., & Bookheimer, S. Y. (2003). Dynamics of the Hippocampus During Encoding and Retrieval of Face-Name Pairs (Vol. 299, pp. 577-580).

Table 1. Summary of PET and fMRI Methods

Techniques for studying brain structure

What is imaged	Technique	Analysis
Gray/white matter/CSF distinctions	T1-weighted imaging (MRI)	Voxel-based morphometry (VBM), volume-based measures, surface-based measures (e.g., cortical thickness)
Gray/white matter/CSF distinctions	T2-weighted imaging (MRI)	Same as above
White-matter structure	Diffusion tensor imaging (DTI	Diffusion tractography
Receptor density	Radioligand binding (PET); GABA-A: [C-11] flumazenil; dopamine D2: [C-11 raclopride]; Mu-opioids, [C-11] carfentanil; acetylcholine: [F-18] epibatidine, [C-11] scopolamine, serotonin: [C-11] benzylamine; others	Kinetic modeling, Logan-plot analysis
Gene expression	PET radiolabeling; MR spectroscopy with kinetic modeling	
Metabolites and various biomarkers	MR spectroscopy	

Techniques for studying brain function

What is imaged	Technique	Analysis
Regional blood flow (perfusion)	[O-15] PET	Voxel-wise linear modeling; multivariate connectivity techniques
Relative Hb deoxygenation	Blood Oxygen Level Dependent (BOLD) signal, T2*-weighted image	Same as above
Glucose metabolism	[F-18]-fluorodeoxyglucose (FDG) PET	Same as above
Regional blood flow (perfusion)	Arterial spin labeling (ASL) fMRI	Same as above
Task-related neurochemistry	Radioligand binding (PET); see above	Kinetic modeling, Logan-plot analysis followed by linear modeling

Table 2. Relative advantages of fMRI and PET

Advantages of fMRI	
Cost and availability	fMRI has lower cost, more facilities available
Spatial resolution	fMRI has higher resolution, but new PET scanners can have same functional resolution for group studies
Temporal resolution	fMRI is superior, permitting event-related designs
Brain connectivity analyses	fMRI permits time series connectivity analysis; PET and fMRI both permit individual differences analysis
Combination with other measures	Simultaneous time-series acquisition of fMRI and EEG provides most detailed mapping of relationships
Single-subject studies	fMRI permits detailed high-resolution studies of individuals
Repeatability	fMRI does not use radioactive substances, so frequent scans are considered safe

Advantages of PET	
Measuring neurochemistry	PET is superior; can be used to directly investigate neurochemistry
Transparency of activation measures	PET provides more direct measures of blood flow or metabolism
Artifacts	PET does not suffer from magnetic susceptibility artifacts and gradient- or RF-related artifacts
Combination with other measures	PET is not magnetic and can be combined with simultaneous EEG, MEG, and TMS
Studying baseline activity	PET provides quantitative measure of baseline state; ASL fMRI also can, but is less commonly available
Naturalness of environment	PET is quieter and has more open physical environment; advantage for auditory and emotion tasks

Table 3. Current websites for key resources

Software registries	
Neuroimaging Informatics Tools and Resources	http://www.nitrc.org/
Internet Analysis Tools Registry	http://www.cma.mgh.harvard.edu/iatr/
Software packages	
SPM	http://www.fil.ion.ucl.ac.uk/spm/software/
FSL	http://www.fmrib.ox.ac.uk/fsl/
AFNI	http://afni.nimh.nih.gov/
BrainVoyager	http://www.brainvoyager.com/
FMRISTAT	http://www.math.mcgill.ca/keith/fmristat/
VoxBo	http://www.voxbo.org/
FIASCO	http://www.stat.cmu.edu/~fiasco/index.php?ref=FIASCO_home.shtml
Analysis toolboxes	
SnPM, nonparametric analysis	http://www.sph.umich.edu/ni-stat/SnPM/
SPMd, image diagnostics	http://www.sph.umich.edu/ni-stat/SPMd/
Robust regression toolbox	http://www.columbia.edu/cu/psychology/tor/software.htm
Mediation analysis toolbox	http://www.columbia.edu/cu/psychology/tor/software.htm
GIFT (Group ICA)	http://icatb.sourceforge.net
MVPA toolbox: Classification	http://www.csbmb.princeton.edu/mvpa/
Netlab: Pattern classification	http://www.ncrg.aston.ac.uk/netlab/
Inverse logit HRF model	http://www.columbia.edu/cu/psychology/tor/software.htm
Atlases and databases	
BrainMap	http://brainmap.org/
ICBM	http://www.loni.ucla.edu/ICBM/
SUMS DB	http://sumsdb.wustl.edu:8081/sums/index.jsp
SPM Anatomy Toolbox	http://www.fz-juelich.de/inb/inb-3//spm_anatomy_toolbox
Wager lab meta-analyses	http://www.columbia.edu/cu/psychology/tor/MetaAnalysis.htm
Surface-based normalisation/warping	
FreeSurfer	http://surfer.nmr.mgh.harvard.edu
Caret/SureFit	http://brainmap.wustl.edu/caret/
Design optimization	
Genetic Algorithm for fMRI	http://www.columbia.edu/cu/psychology/tor/software.htm
M-sequence toolbox	http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=3083

Figure Captions

Figure 1. (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 among tissue types, and shortening the repetition time emphasizes T_1 differences among tissue types. 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 2. A schematic diagram of the main components of a PET scanner.

Figure 3. Influences on T_2*-weighted signal in BOLD fMRI imaging. Courtesy Dr. Doug Noll.

Figure 4. Axial slices showing brain regions responsive to different types of switching and their overlap, from Wager et al. (2005). All voxels identified show significant switch costs in at least two switch-no switch contrasts ($p < .05$ Family-wise error rate corrected in each). Thus, many regions not shown here may also show brain switch costs at less stringent thresholds. Regions colored in red are *common activations* that show no significant differences among costs for different types of switch (at $p < .05$ uncorrected). Other regions show evidence for greater activation in some switch types than others, as indicated in the legend. I, internal; E, external; O, object; A, attribute switch types.

Figure 5. (A) An overview of the effects of various approaches towards dealing with multiple comparisons. (Top) Ten simulated t-maps were analyzed using an uncorrected threshold $p < .10$. True positives are indicated by white regions inside of the gray squares. False positives are white pixels outside of the gray square. The proportion of false positives is listed under each image. They average 10%, as expected. (Middle) The same images with the threshold designed to control the Familywise error rate (FWER) at 10% using Bonferroni correction. There is only one false positive in the 10 images, at the cost

of a significant increase in the number of false negatives. (Bottom) Similar results obtained using an FDR controlling procedure at the 10% level. The proportion of active voxels that are false positives is listed under each image. They average 10% as expected. (B) Imposing an arbitrary 'extent threshold' based on the number of contiguous activated voxels does not necessarily solve the problem of false positives. The same activation map, with spatially correlated noise, is thresholded at three different P-value levels. Due to the smoothness, the false-positive activation blobs (outside of the squares) are contiguous regions of multiple voxels which can easily be misinterpreted as regions of activity.

Figure 6. Common thresholds used in neuroimaging experiments. A midline saggital slice (left) shows the peak activations reported in 195 separate studies of long-term memory. The frequencies of P-value thresholds used for all statistical parametric maps in these studies are shown to the right. The most common threshold is $P < .001$, uncorrected for multiple comparisons. Corr: Corrected threshold. Adapted from (T. D. Wager, Lindquist et al., 2007).

Figure 7. Examples of atlas template images and a group-averaged, normalized structural image from a study. A) The Montreal Neurologic Institute (MNI) 305-brain average as included with Statistical Parametric Mapping (SPM99, SPM2 or SPM5) software. The atlas brain is the same across software versions, though the algorithms for normalizing brains to the template have changed. B) The International Consortium for Brain Mapping (ICBM) LBPA atlas, based on manually labeled region from the Center for Morphometric Analysis at Harvard University. Each color represents a gross brain structure based on a consensus among 40 individually labeled brains. C) The single-subject T1 brain coregistered with MNI space—the "colin brain" based on an average of 27 images of one individual—with overlaid consensus regions based on probabilistic cytoarchitecture. The probabilistic maps represented here are available in the SPM Anatomy toolbox, V1.5, and represent data from a series of studies on cytoarchitectural mapping of post-mortem brains registered to the single-subject MNI template (Amunts et al., 2005; S. B. Eickhoff, Amunts, Mohlberg, & Zilles, 2006); see Table 3. Note that because the underlay brain is only a single brain, it may not be representative of anatomical locations in a study sample (compare the midbrain with that in D) and thus is not an ideal underlay image for localization of new study data. D) A trimmed average of

18 subjects' T1 images initially warped to the MNI template using SPM and refined using a genetic algorithm based on custom code. This average brain shows good structural definition, indicating good inter-subject registration, and is a suitable underlay image for functional activations. E) The ICBM 452-brain MNI average, with 5th-order polynomial warping to the standard space. The structural definition is excellent for the average of many brains, but the space is different from the MNI 305 space (the brainstem, for example, is much more anterior in the 452 brain), illustrating the need to report the specific atlas and procedures used in neuroimaging studies. F) Activations from a task-switching paradigm (yellow; Wager et al. 2005) superimposed on results from a meta-analysis of executive working memory (blue; Wager & Smith, 2003). Surface reconstruction was done with Caret software (Table 3) and shows a partially inflated left hemisphere (left) and a flattened cortical map of that hemisphere (right). Red and green arrows show the medial frontal gyrus and inferior frontal junction on each rendering.

Figure 8. Construction of an event-related fMRI design matrix with four different event types (A-D), using the canonical SPM HRF. Indicator functions corresponding to the four event types are convolved with the canonical HRF to create the regressors that make up the design matrix. An image of the design matrix is shown to the far right.

Figure 9. Transient spike artifacts as visualized in the software package AFNI. Spikes in the data during isolated volume acquisitions are apparent in certain slices, as shown by the bright bands in the saggital slices (bottom). This suggests that gradient performance was affected during acquisition of some echo-planar images, which were acquired slice-by-slice in interleaved order in this experiment.

Figure 10. Normalization attempts to register each subject's anatomy with a standardized template brain using an intensity-based warping procedure. (A) A schematic overview of the warping process. High resolution T_1 images are warped onto a T_1 template to give normalized images in a standard space. (B) Incorrect warping can produce gross distortions in the brain, as local features are matched at the expense of getting the correct overall shape. (C) The normalization procedure can be checked by identifying control points on the MNI ICBM152 template brain (left) that correspond to easily identifiable

features and overlaying them on the subject's normalized T_1 image. For this subject each of the control points matches well with the corresponding anatomical feature.

Figure 11. Basis sets differ in the amount of flexibility they provide in modeling different HRF shapes. Each column in the figure shows HRF estimates for an experiment where two conditions A and B produce responses of: (left) different amplitudes; (center) equivalent amplitude, but at different delays; and (right) equivalent amplitude, but different durations. The ability of four different basis sets to estimate the hemodynamic response function (HRF) corresponding to each condition is shown in the four rows. The basis sets that were used are SPM software's canonical HRF, the canonical HRF + its temporal derivative, a smooth finite impulse response (FIR) model and the inverse logit model of Lindquist and Wager (2007). Adapted from (M. A. Lindquist & Wager, 2007).

Figure 12. (A) Power curves—calculated for effect sizes of 0.5, 1, and 2—for a whole-brain search with 200,000 voxels and FWE correction at $p < .05$ using the Bonferroni method. The number of voxels would be typical of a whole-brain search through gray and white matter with a 2 x 2 x 2 mm sampling resolution. (B) The same power curves calculated based on the results of nonparametric permutation testing, which takes into account the spatial smoothness in the data. Based on the smoothness reported in Nichols and Hayasaka (2003) for 10 different statistic maps, we calculated an average of ~750 effective independent comparisons. Correction across this number of comparisons was used in calculating power.

Figure 13. 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., 2005). 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 14. Functional connectivity methods can be applied at different levels of analysis, with different interpretations at each level. Left: Connectivity across time series data can reveal networks that are dynamically co-activated over time. The solid, dotted, and dashed lines indicate activation time series from three different subjects on the left, and average activation magnitudes for the same subjects (shown by hemodynamic response function (HRF) curves) at the right. Alternatively, measures of single-trial activation amplitude (black dots) can be extracted and used to estimate connectivity, which avoids some ambiguity with respect to the source of connectivity (task-related vs. spontaneous). However, artifactual influences can make interpretation of both of these types of connectivity difficult (see list at the bottom). Distributed artifacts tend to create positive covariance, whereas neuro-vascular coupling differences—and resulting differences in HRF shapes between regions—tends to weaken covariance estimates. Right: At the subject level one can correlate magnitudes within condition or differences across conditions. This analysis is conducted on individual differences, rather than on time series data, and results may have different interpretations than time series connectivity data. Again, special care needs to be taken to limit the influence of artifacts, which are likely to be largely related to factors that create individual differences in model fits across the brain (see list).

Figure 1.

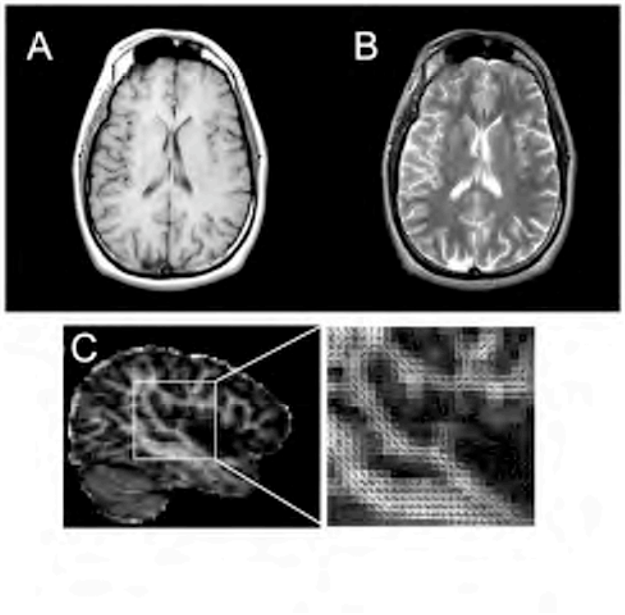


Figure 2.

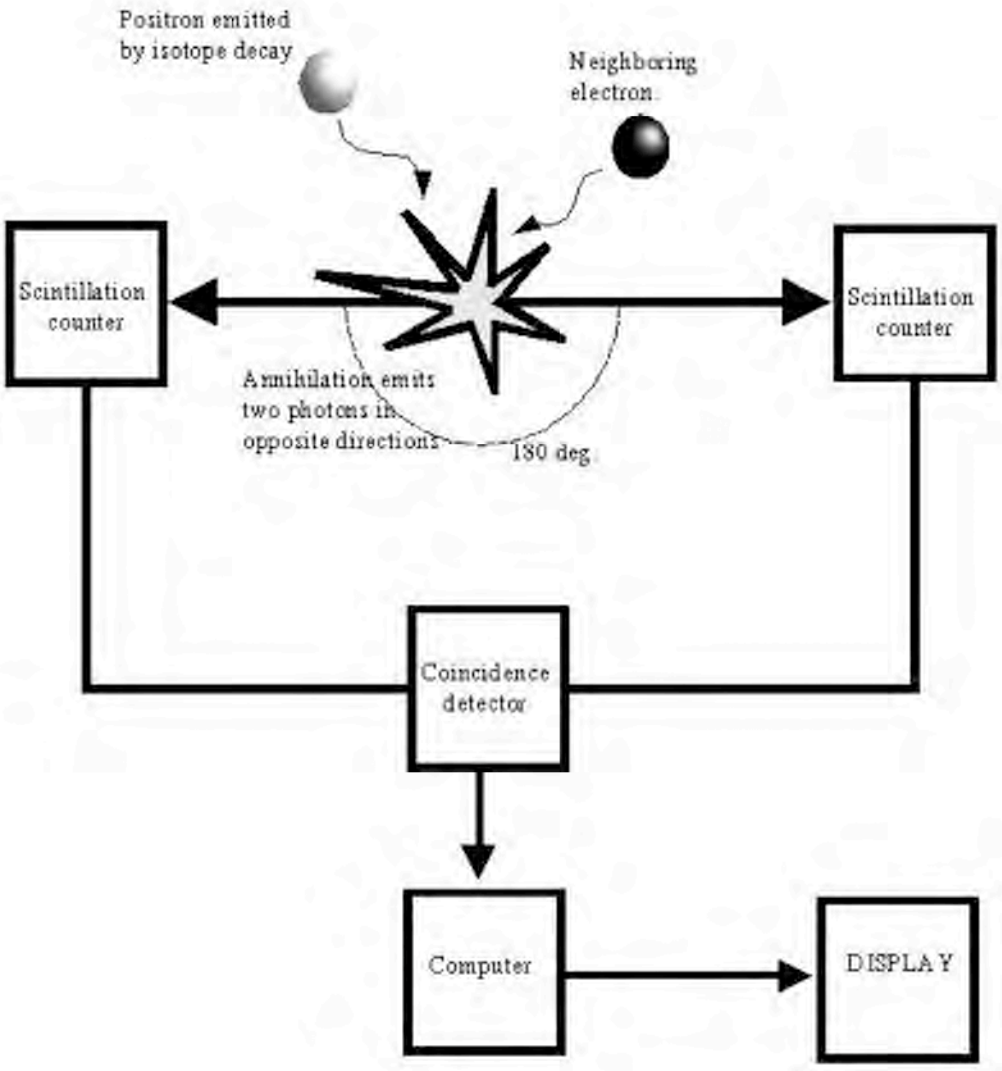

Positron emitted by isotope decay
Neighboring electron
Scintillation counter
Scintillation counter
Annihilation emits two photons in opposite directions
180 deg
Coincidence detector
Computer
DISPLAY

Figure 3.

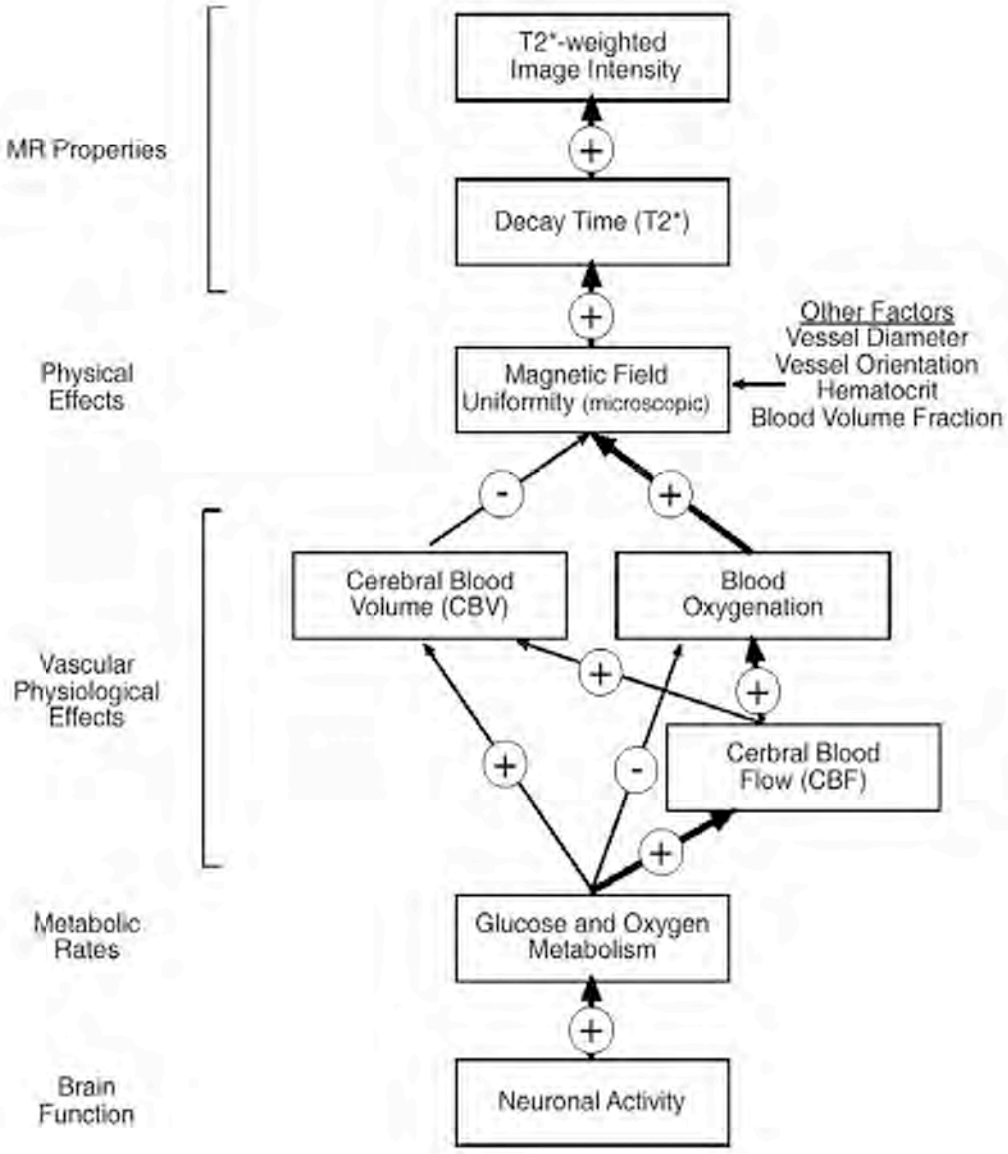

MR Properties
T2*-weighted Image Intensity
+
Decay Time (T2*)
+
Physical Effects
Magnetic Field Uniformity (microscopic)
Other Factors
Vessel Diameter
Vessel Orientation
Hematocrit
Blood Volume Fraction
-
+
Vascular Physiological Effects
Cerebral Blood Volume (CBV)
Blood Oxygenation
+
+
+
-
Cerbral Blood Flow (CBF)
+
Metabolic Rates
Glucose and Oxygen Metabolism
+
Brain Function
Neuronal Activity

Figure 4.

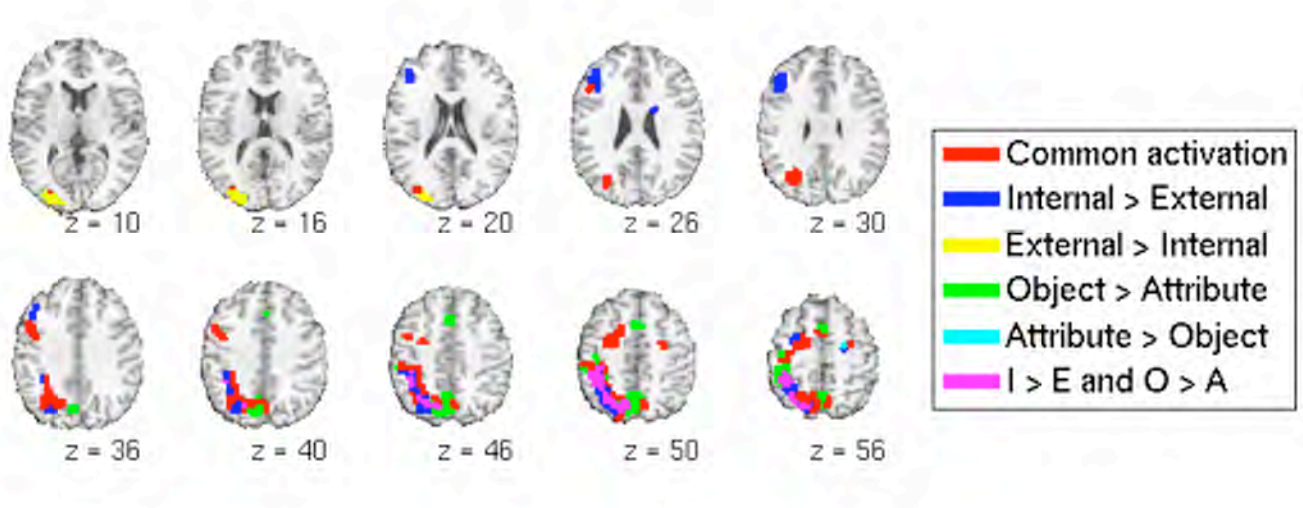

z = 10
z = 16
z = 20
z = 26
z = 30
z = 36
z = 40
z = 46
z = 50
z = 56
Common activation
Internal > External
External > Internal
Object > Attribute
Attribute > Object
I > E and O > A

Figure 5.

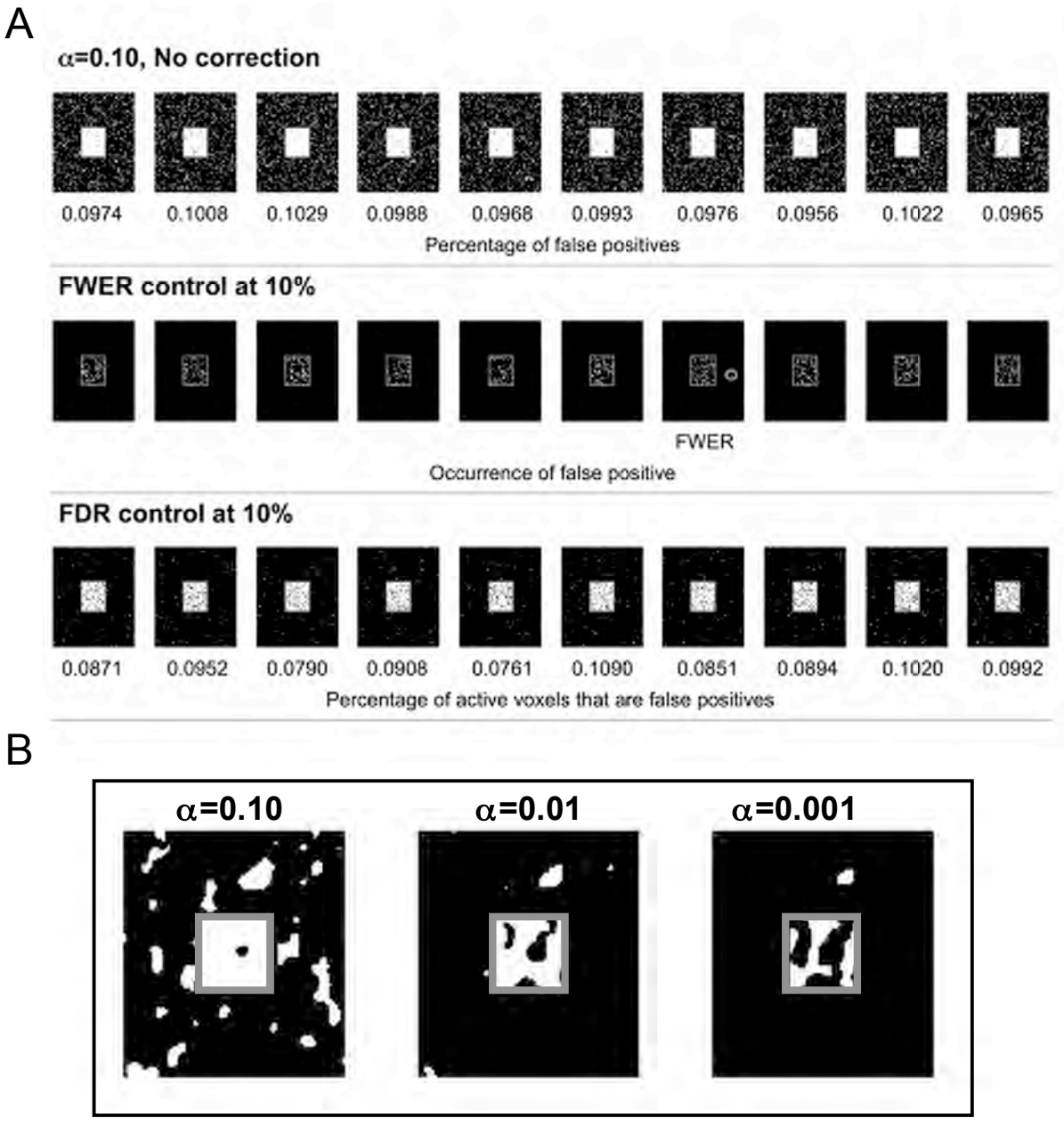

A
α=0.10, No correction
0.0974 0.1008 0.1029 0.0988 0.0968 0.0993 0.0976 0.0956 0.1022 0.0965
Percentage of false positives
FWER control at 10%
FWER
Occurrence of false positive
FDR control at 10%
0.0871 0.0952 0.0790 0.0908 0.0761 0.1090 0.0851 0.0894 0.1020 0.0992
Percentage of active voxels that are false positives
B
α=0.10
α=0.01
α=0.001

Figure 6.

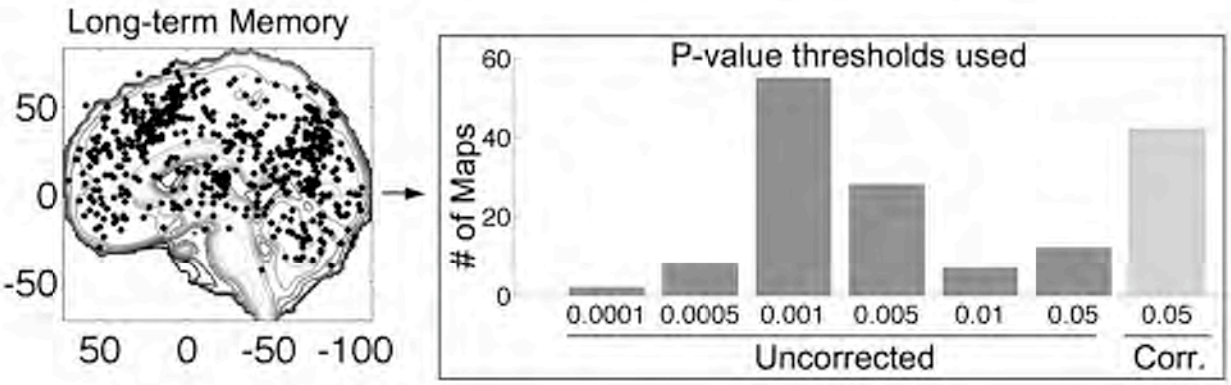

Long-term Memory
50
0
-50
50 0 -50 -100
P-value thresholds used
60
40
20
0
# of Maps
0.0001 0.0005 0.001 0.005 0.01 0.05 0.05
Uncorrected
Corr.

Figure 7.

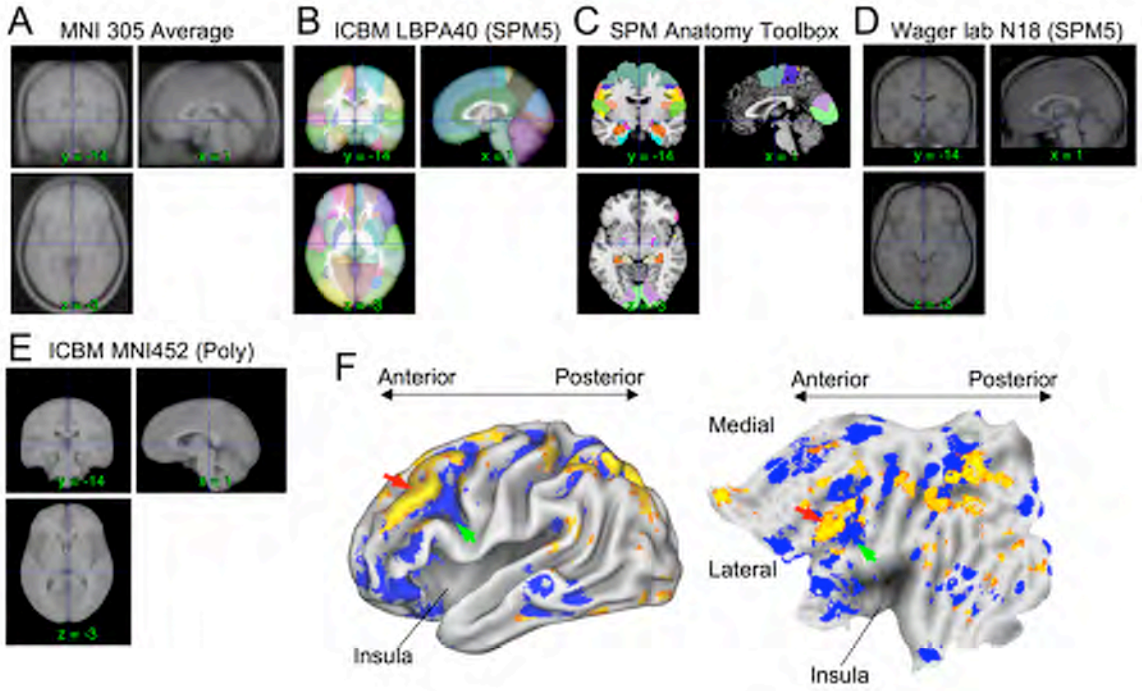

A MNI 305 Average
B ICBM LBPA40 (SPM5)
C SPM Anatomy Toolbox
D Wager lab N18 (SPM5)
E ICBM MNI452 (Poly)
F
Anterior
Posterior
Insula
Anterior
Posterior
Medial
Lateral
Insula

Figure 8.

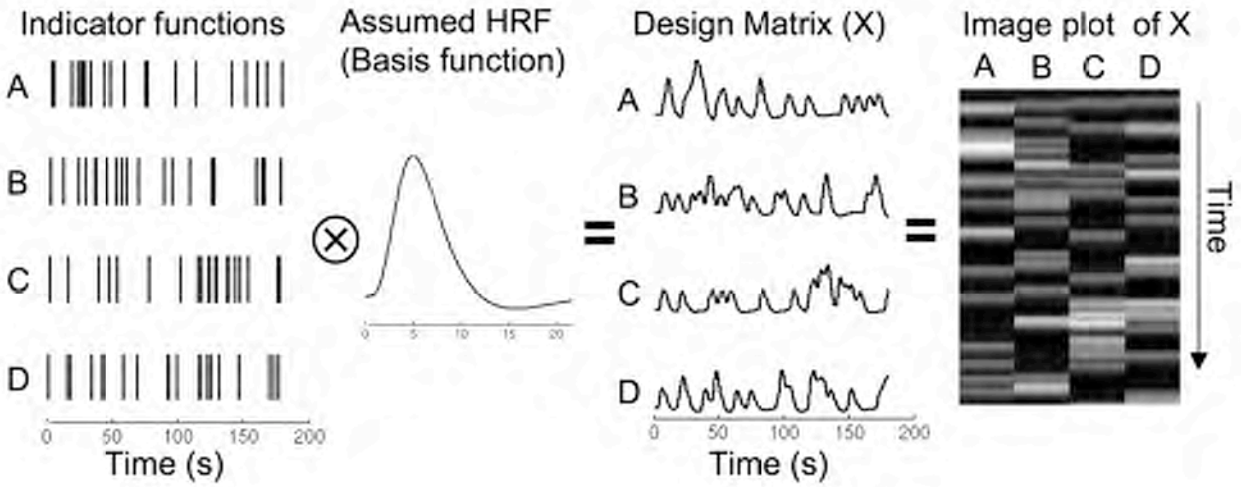

Indicator functions
A
B
C
D
0 50 100 150 200
Time (s)
Assumed HRF
(Basis function)
0 5 10 15 20
Design Matrix (X)
A
B
C
D
0 50 100 150 200
Time (s)
Image plot of X
A B C D
Time

Figure 9.

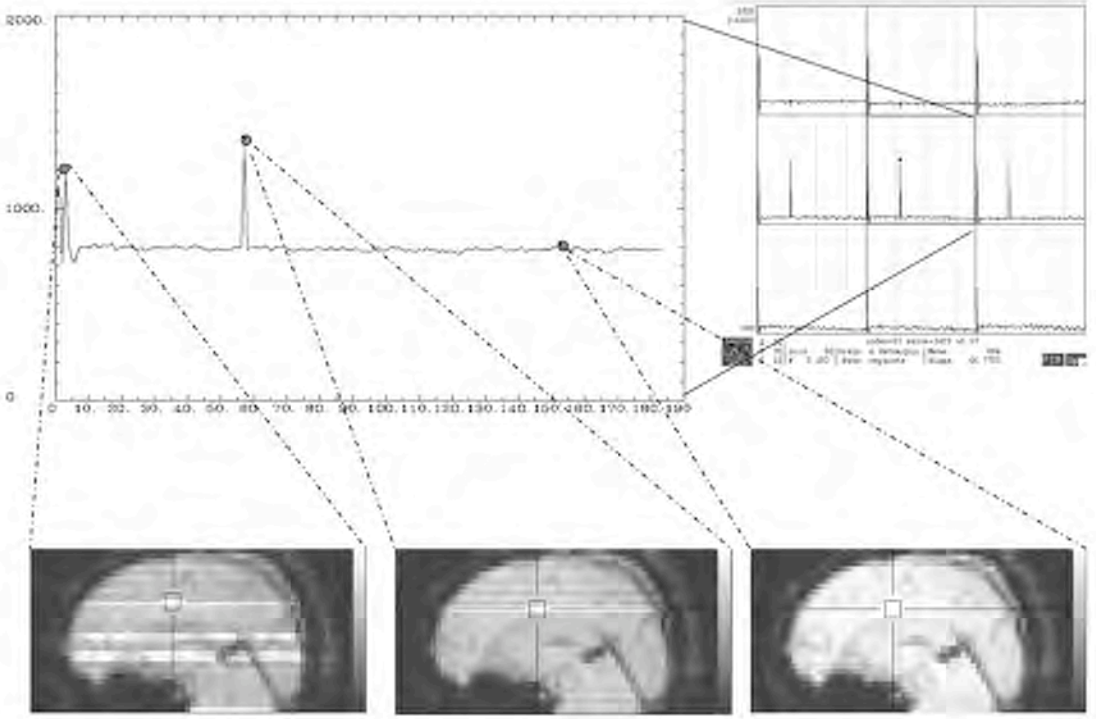


Figure 10.

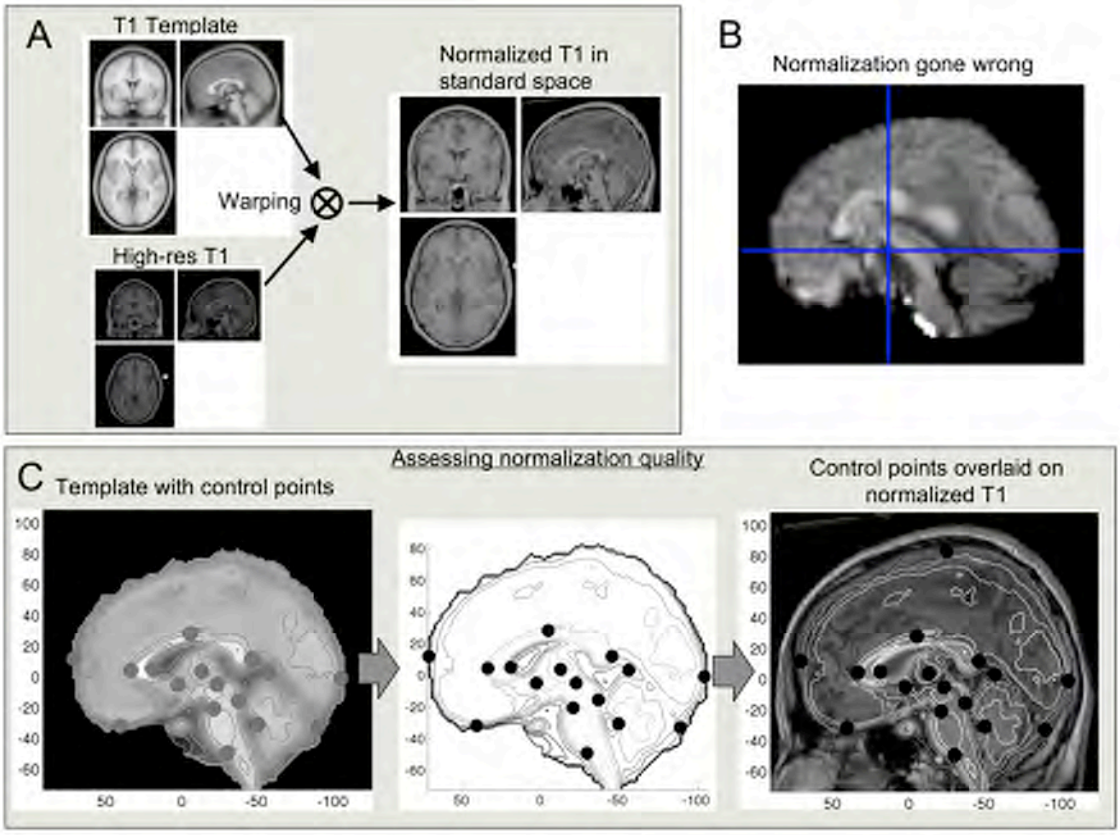

A
T1 Template
Normalized T1 in standard space
Warping
High-res T1
B
Normalization gone wrong
C
Template with control points
Assessing normalization quality
Control points overlaid on normalized T1

Figure 11.

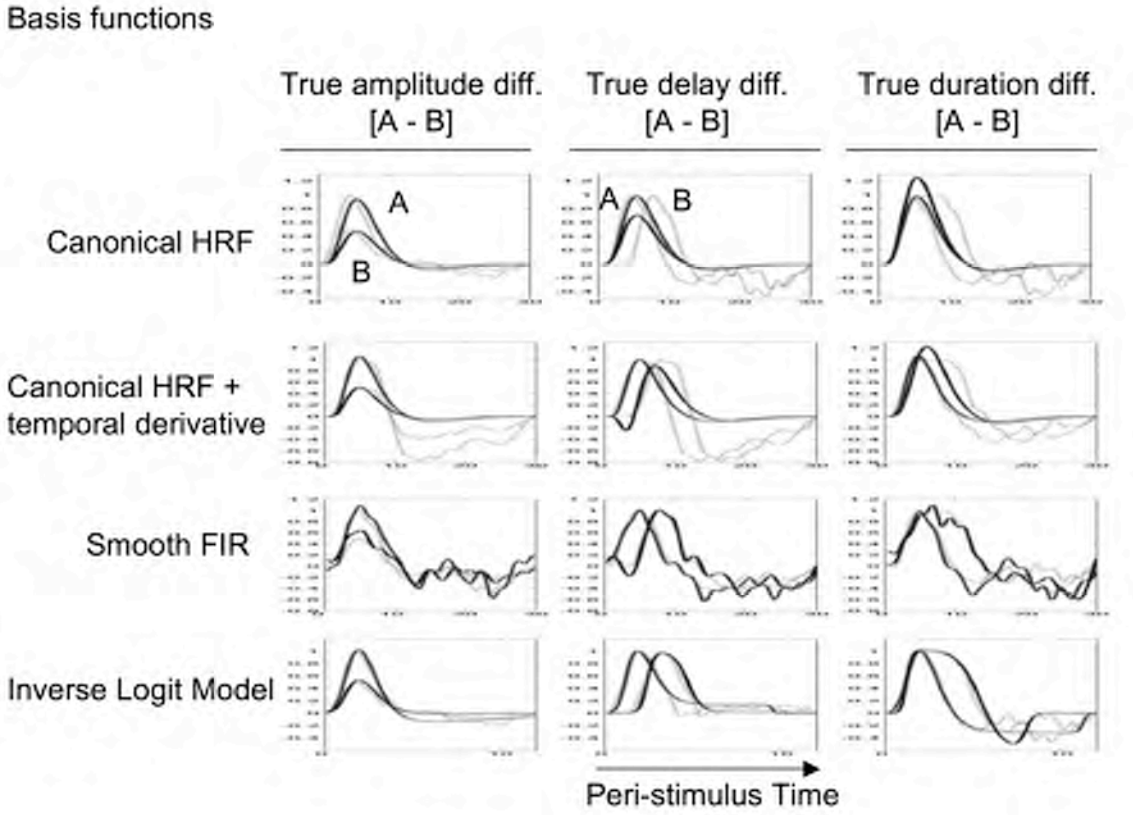

Basis functions
True amplitude diff. [A - B]
True delay diff. [A - B]
True duration diff. [A - B]
Canonical HRF
Canonical HRF + temporal derivative
Smooth FIR
Inverse Logit Model
A
B
A
B
Peri-stimulus Time

Figure 12.

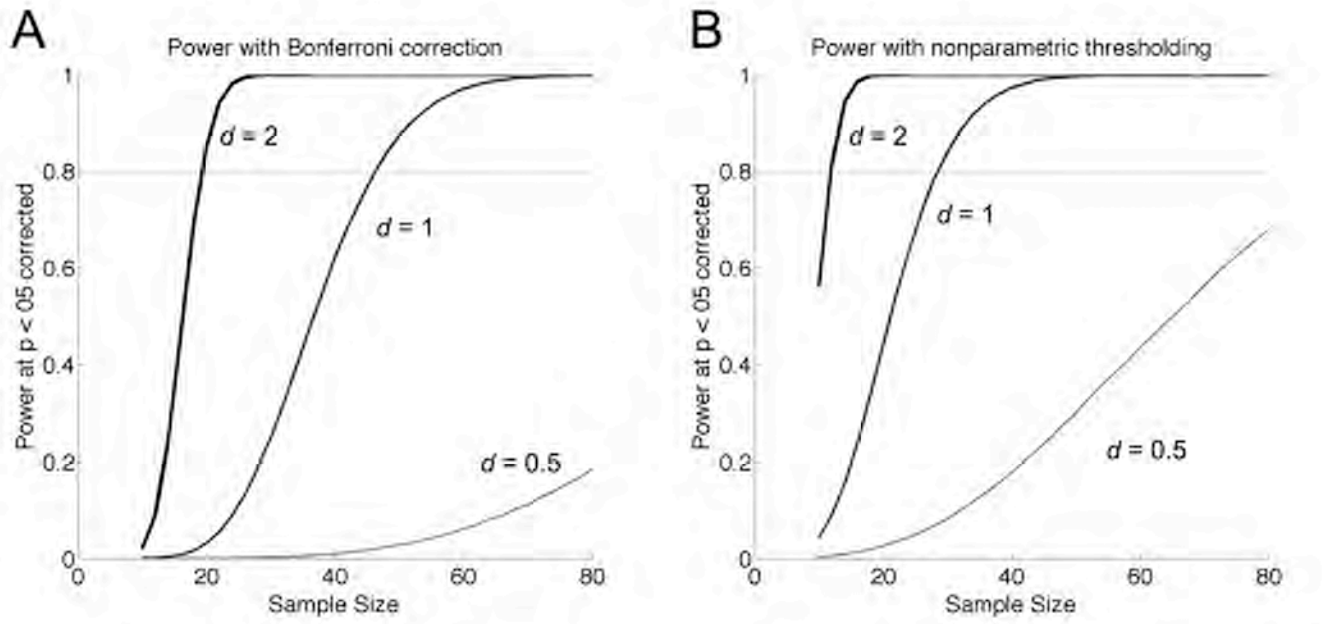

A
Power with Bonferroni correction
Power at p < 05 corrected
Sample Size
d = 2
d = 1
d = 0.5
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
B
Power with nonparametric thresholding
Power at p < 05 corrected
Sample Size
d = 2
d = 1
d = 0.5

Figure 13.

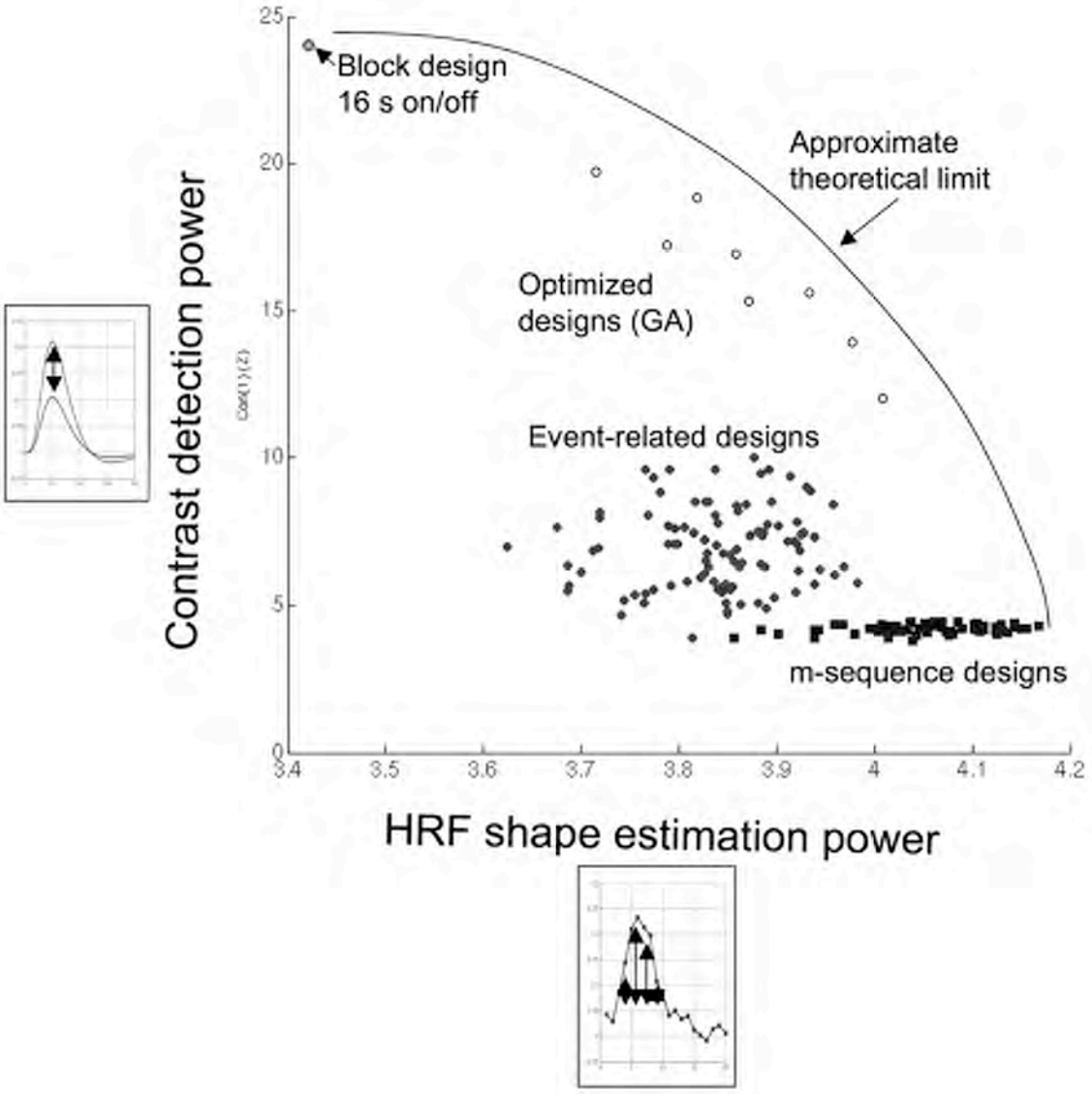
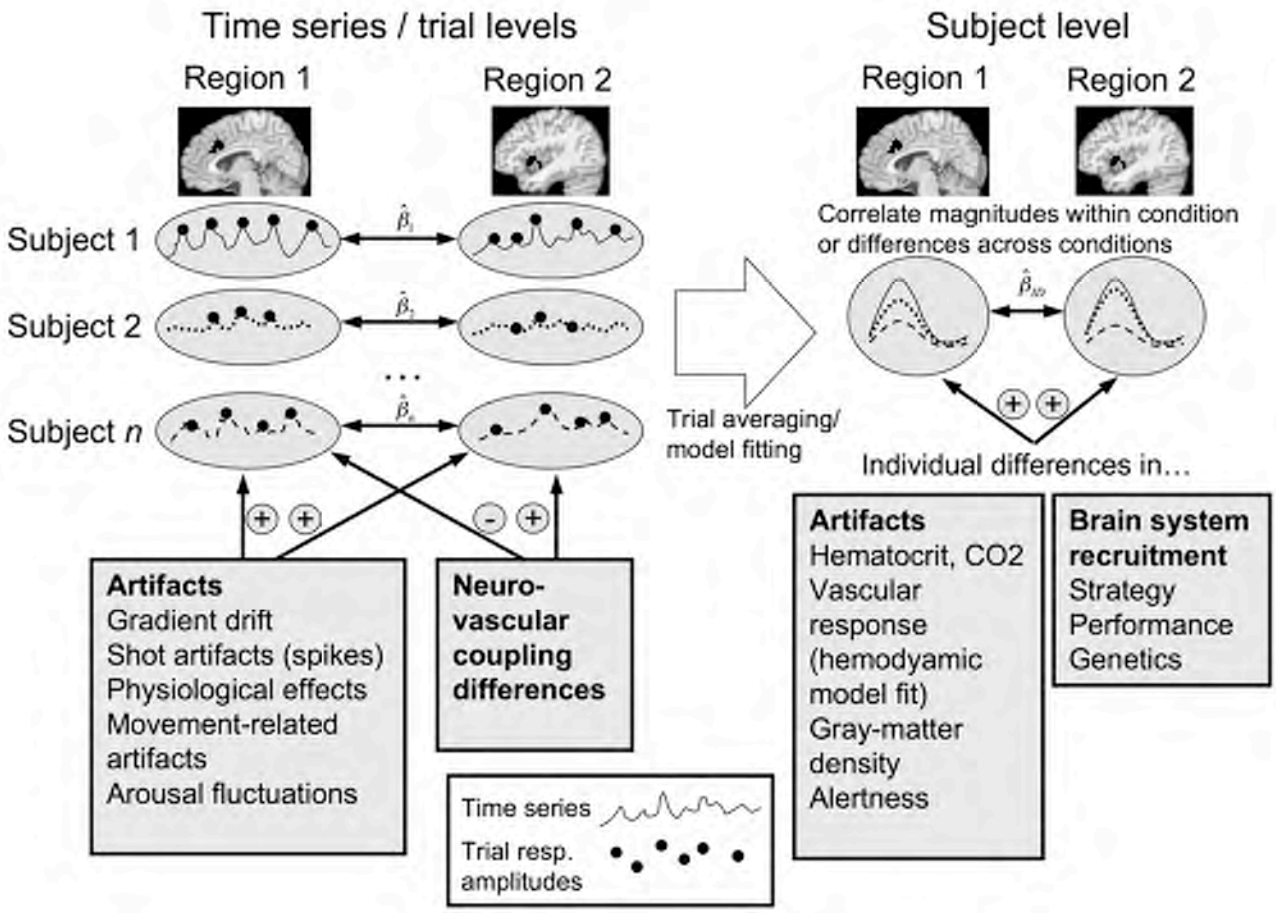

Block design
16 s on/off
Approximate
theoretical limit
Optimized
designs (GA)
Event-related designs
m-sequence designs
Contrast detection power
HRF shape estimation power
25
20
15
10
5
0
3.4
3.5
3.6
3.7
3.8
3.9
4
4.1
4.2

Figure 14.


Time series / trial levels
Region 1
Region 2
Subject 1
Subject 2
Subject n
Trial averaging/
model fitting
Subject level
Region 1
Region 2
Correlate magnitudes within condition
or differences across conditions
Individual differences in…
Artifacts
Gradient drift
Shot artifacts (spikes)
Physiological effects
Movement-related artifacts
Arousal fluctuations
Neuro-vascular coupling differences
Artifacts
Hematocrit, CO2
Vascular response (hemodyamic model fit)
Gray-matter density
Alertness
Brain system recruitment
Strategy
Performance
Genetics
Time series
Trial resp. amplitudes