

Interpretation of clinical functional neuroimaging studies

2

Geoffrey K Aguirre

INTRODUCTION

The first functional magnetic resonance imaging (fMRI) scans of a human being were obtained on a clinical MRI scanner.¹ This is not surprising as, at the time, virtually every MRI machine into which a human could fit was in clinical practice, reaping the benefits of the previous decade of rapid development of imaging technology. Bellivue's groundbreaking observation of visual cortex activation was performed using bolus injections of an MRI contrast agent, but subsequent studies quickly adopted Ogawa's technique² of measuring the endogenous effects of deoxygenated hemoglobin. In the years that have followed, blood oxygen level-dependent (BOLD) fMRI has become a pervasive method in the normative study of the human brain. In contrast to the rapidity with which BOLD fMRI has transformed cognitive neuroscience, clinical applications have come more slowly. As the chapters in this book attest, however, the time is now ripe to reap the benefits of clinical fMRI.

This chapter provides an overview of fMRI concepts and approaches, with a particular focus upon the clinical uses of fMRI. After a brief overview of the physics and physiology that provide the basis of the BOLD fMRI technique, the properties of the BOLD fMRI system that impact paradigm design are considered, particularly the sluggish nature of the hemodynamic response and the presence of slow drifts and fluctuations in the fMRI signal over time. This is followed by a discussion of how classic 'blocked' and 'event-related' approaches are simply extreme points

within a continuous space of possible designs that exchange statistical power for task predictability. The chapter concludes with a consideration of the types of clinically relevant information that might be gleaned from a functional neuroimaging study, and provides an inferential framework for such approaches.

THE ORIGIN OF THE fMRI SIGNAL

The existence of fMRI depends upon two rather fortuitous properties of physics and physiology. The first is that hemoglobin, the primary oxygen-carrying molecule in the blood, has different magnetic properties when bound and unbound to oxygen. The second is that there is an exquisite coupling of local neuronal activity and blood flow within the brain. As a result, changes in neural activity result in perturbations of the local magnetic field via changes in blood flow. Although the mechanisms that mediate neurovascular coupling are still very much under investigation, the properties of this relationship are fairly well described. Increases in neural activity are accompanied by increased metabolic demands, leading to a transient decrease in local oxygen content.³ A compensatory vascular response follows within 1–2 s. It appears that the aspect of neuronal activity that drives this delayed hemodynamic response is the local field potential: a measure of synchronous dendritic activity over a population of neurons.⁴ The increases in local blood flow and volume produce an overabundance of oxygenated hemoglobin, decreasing the deoxyhemoglobin concentration.⁵ This is sometimes

referred to as a paradoxical change, as increased metabolic activity leads to decreased deoxyhemoglobin.

Deoxyhemoglobin has stronger magnetic properties than oxyhemoglobin. Therefore, a decrease in the deoxyhemoglobin concentration results in a decreased perturbation of the local magnetic field (referred to as a susceptibility gradient). The local perturbation of the magnetic field can be measured using MRI techniques, specifically the T2* relaxation time.⁶ fMRI data, then, are images of the brain over time that measure T2* and reflect, through a chain of associations, local neuronal activity. It is because imaging contrast is mediated by blood flow and oxygen concentration that this method of fMRI is called blood oxygen level-dependent (BOLD). There are several excellent reviews of these details for the interested reader (see e.g. Monnen and Bandettini⁷).

BOLD fMRI data in raw form are volumetric images of the brain, obtained every couple of seconds, over the course of minutes to a couple of hours. Just as the image on a television screen is composed of small dots called 'pixels', the three-dimensional images of the brain provided by fMRI are composed of small, three-dimensional dots called 'voxels' (as they have volume). Typical BOLD fMRI experiments obtain a complete image of the brain every 2–3 s, and are composed of voxels that are 3 mm × 3 mm × 3 mm in size (requiring approximately 40 000 voxels to cover the entire volume of the brain).

The spatial and temporal resolution of BOLD fMRI is limited by the neurovascular coupling that is the source of contrast. While MRI images can readily be obtained every 100 ms, and with spatial resolution of the order of tenths of a millimeter, this fine resolution has little practical advantage. Changes in neural activity give rise to a change in BOLD fMRI signal that evolves over seconds (described in detail below). As a result, BOLD fMRI images are seldom acquired more frequently than once a second. Additionally, a point of neural activity engenders a change in BOLD signal that

spreads over several millimeters;⁸ thus, BOLD images are typically composed of voxels (the smallest volume 'pixel' of which the image is composed) no smaller than 1 mm on a side.

An important property of the BOLD fMRI signal is that it has no simple, absolute interpretation. This is because the particular signal value obtained is not exactly a measure of deoxyhemoglobin concentration, but is instead a measure weighted by this concentration (i.e. it is T2*-weighted) and is also influenced by a number of other factors that can vary from voxel to voxel, scan to scan, and subject to subject. As a result, experiments conducted with BOLD fMRI generally test for differences in the magnitude of the signal between different conditions within a scan. One could not, for example, directly contrast the absolute level of the BOLD fMRI signal obtained within the temporal lobe of schizophrenic patients with that from controls with much hope of obtaining a useful statistical test. Notably, this is not necessarily true of all fMRI methods. For example, arterial spin-labeled (ASL) perfusion MRI⁹ can provide a quantitative measure of cerebral blood flow that is in absolute units (e.g. cm³ of blood/100 g of tissue/min). Because of this property, and others mentioned in greater detail below, perfusion fMRI may have particular application to clinical fMRI.

THE PROTOTYPICAL fMRI STUDY

There is now a bewildering array of paradigm designs and analysis approaches that are applied to BOLD fMRI. Later in this chapter, a heuristic is provided for considering different experimental designs. Let us start, however, by considering a 'prototypical' fMRI study to introduce several concepts.

Suppose that we wish to measure the magnitude of the neural response within the primary visual cortex to a standardized flash of light in a group of subjects. Perhaps we ultimately wish to compare the evoked response between patients with optic nerve damage and those without. It might be the case that evoked neural activity is consistently

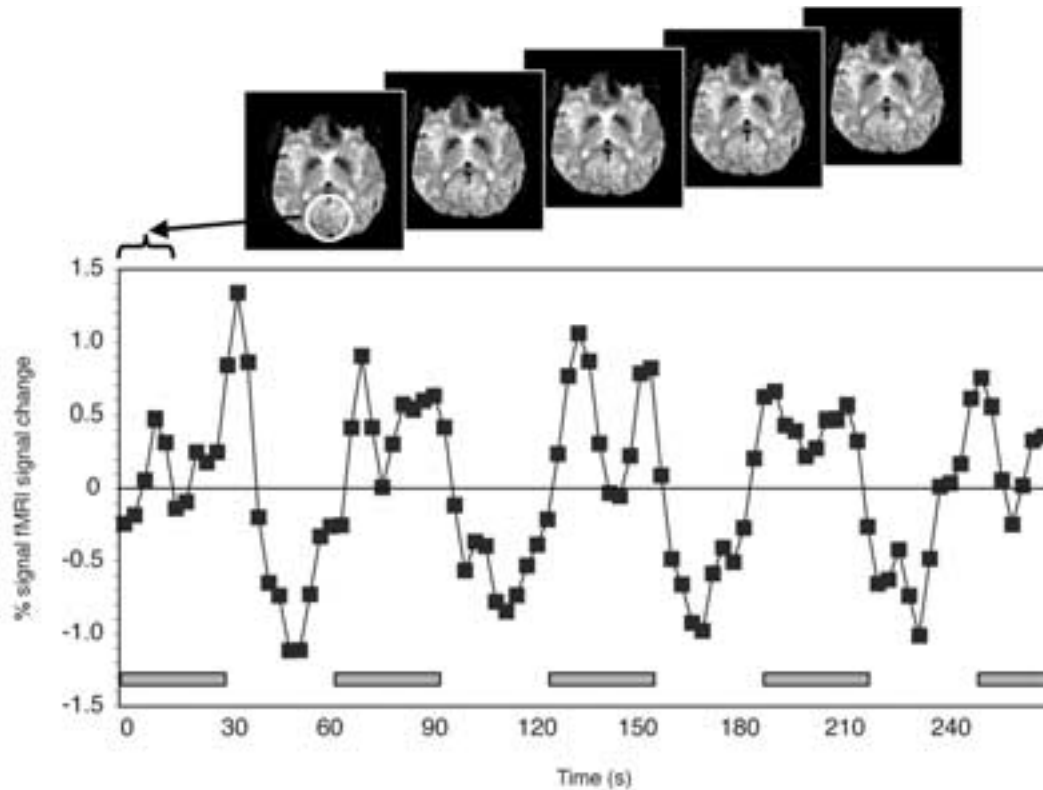


Figure 2.1 Example of a BOLD fMRI time series that might be obtained from the visual cortex during 30 s periods of visual stimulation alternating with 30 s of darkness. A single, axial slice through the raw, echo-planar images is shown for the first five time points.

less in those patients who have been afflicted with optic neuritis, so that measurement of such a response in future, unselected patients may have diagnostic value (replacing the visual evoked potential study that measures surface electrical potentials).

Consider the data that we might obtain from one, normal control subject in this study. The subject is placed in the scanner and, while whole-brain BOLD fMRI images are collected every 3 s, the subject is presented with alternating periods of 30 s of flashing lights followed by 30 s of darkness. After the data have been collected, we anatomically define the location of the primary visual cortex, and then examine the BOLD fMRI signal that was obtained from this region. Figure 2.1 shows one axial slice through the first five BOLD fMRI images

across time, and the corresponding average signal that was obtained from within the primary visual cortex. As can be seen, there was a rise and fall of the BOLD fMRI signal over time, corresponding to the periods when the lights were on and off. We could perform a statistical analysis to confirm the impression that the BOLD fMRI signal responded to the presentation of lights. We could further measure the magnitude of the signal change between the light and dark conditions, and compare that with other subjects and between patient and control groups.

This simple paradigm illustrates several basic points about BOLD fMRI data. First, the effect of flashing lights upon the neuroimaging signal can only be assessed by comparison with the periods of darkness. If the subject had been presented with a continuous period

of flashing lights, then no modulation of the signal would be obtained and there would be nothing to measure. Second, this paradigm makes use of a ‘blocked’ alternation between stimulus conditions – a block of 30 s of flashing light is alternated with a block of 30 s of darkness. This is in contrast to other approaches (so-called event-related designs) where stimuli are presented more rapidly and in a less predictable order. In this simple example, the two conditions compared were fairly elementary. Other types of fMRI designs might compare more complicated mental operations, such as the cognitive process of working memory and an appropriate control condition. Finally, from inspection of the graph of the BOLD fMRI signal, it can be seen that the changes in the signal seem to follow the changes in light stimulation after a delay of a few seconds.

SYSTEMS, HEMODYNAMIC SIGNALS AND TEMPORAL NOISE

In the preceding example, we measured the BOLD fMRI signal that followed the presentation of a light stimulus to a subject. One useful way of considering the study is in terms of systems theory. Simply defined, a system is something that takes input and provides output. There are many examples of systems, such as a stereo speaker that takes electrical input through a wire and provides acoustic output. What is the system under study in the example experiment provided above? A useful way to answer this question is to consider that two, separate, systems are at work in an fMRI study. The first system is that of cognition, in which the inputs are the instructions, stimuli, and tasks presented to the subject, and the output is the pattern of neural activity evoked within the brain. The second system is the domain of physiology and physics, and mediates the transformation of neural activity inputs into blood flow responses and imaging signal. For our example, the first, neural, system converted the light stimulation into neural activity in primary visual cortex. The properties of this system are determined by

the connections between, and dynamic properties of, the retina, lateral geniculate nucleus, and visual cortex. The second system is that of BOLD fMRI, which transforms local neural activity into the T2*-weighted BOLD signal. In this example, and in most clinical applications, it is the response of the first system that is principally under study. An exception might be studies of stroke and compromised vascular states (see Chapter 8), in which the properties of the transformation of neural activity into a change in blood flow may be of interest.

If it were the case that the properties of both systems were unknown, then the relationship between cognition and BOLD fMRI signal would be underdefined: one would be unable to assign a given change in imaging signal to cognition or neurovascular coupling. Fortunately, the properties of the second system (the BOLD fMRI system) are lawful and well described, even if the exact mechanisms of the transformation are still not well understood. Indeed, one of the advantages of viewing BOLD fMRI in terms of systems theory is that we gain the ability to describe the transformation of neural activity to imaging signal without having to consider the mechanistic details that provide the transformation.

As it turns out, BOLD fMRI behaves very much like a linear system. Linear systems have properties that make them particularly amenable to study. For example, doubling the size of the input to a linear system doubles the size of the corresponding output. Properties such as this make it possible to accurately predict what the output of a linear system is to any particular input. The predictive abilities of a linear system can be completely characterized by a property called the impulse response function (IRF), which is the output of the system to an infinitely brief, infinitely intense input. In the context of BOLD fMRI, the hemodynamic response function (HRF) is taken as an estimate of the IRF of the BOLD fMRI system, and is the change in BOLD fMRI signal that results from a brief (<1 s) period of neural activity. Knowledge of the shape of the HRF allows one to predict the

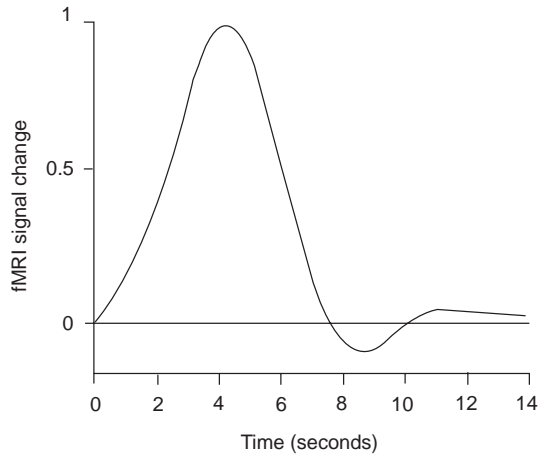


Figure 2.2 The BOLD hemodynamic response. This is the average BOLD signal change that follows a brief period of neuronal activity.

BOLD fMRI signal that will result from any pattern of neural activity.

The HRF can be measured empirically from human subjects by obtaining the BOLD fMRI signal that is evoked by experimentally induced, brief periods of neural activity in known cortical areas (e.g. neural activity in the primary motor cortex in response to a button press). Figure 2.2 presents the shape of the average HRF that might be found within the motor cortex across a population of healthy, young subjects. As can be seen, the shape of the HRF reflects the relatively slow changes in vascular physiology that follow changes in neural activity, and the response rises and falls smoothly over a period of about 16 s. While the shape of the HRF varies significantly across subjects, it is very consistent within a subject, even across days to months.¹⁰ There is some evidence that the shape of the HRF varies from one region of the brain to another (perhaps from variations in neurovascular coupling), but this is a difficult notion to test as it is necessary to create evoked patterns of neural activity in disparate areas of the brain that can be guaranteed to be very similar.

Although neural activity can rise and fall rapidly (on the order of milliseconds), the

shape of the HRF tells us that changes in blood flow respond much more slowly (on the order of seconds). One consequence of this is that rapid changes in neural activity are not well represented in the BOLD fMRI signal. The ‘temporal blurring’ induced by the HRF limits the patterns of neural activity that might be detected using BOLD fMRI. Specifically, the smooth shape of the HRF makes it difficult to discriminate closely spaced neural events. As a consequence, many fMRI paradigms are designed to evoke relatively prolonged changes in neural activity, as was the case in our paradigmatic example above. With clever experimental design, however, it is still possible to use BOLD fMRI to detect (1) brief periods of neural activity, (2) differences between neural events in a fixed order, spaced as closely as 4 s apart, (3) differences between neural events, randomly ordered, closely spaced (e.g. every second or less), and (4) neural onset asynchronies on the order of 100 ms. As is described in greater detail below, there is a cost in statistical power that accompanies these designs.

Another important property of BOLD fMRI data is that the signal over time is rather unstable. Regardless of whether the subject is engaged in the performance of a task or resting quietly, the BOLD fMRI data contain large drifts and surges over time. This noise in the data becomes more and more prominent at longer and longer timescales (and can be termed low-frequency noise). In addition to complicating statistical analysis of such data,¹¹ the presence of the low-frequency noise in BOLD fMRI data renders slow changes in neural activity difficult to detect. Notably, the effects of the slow hemodynamic response and the noise properties of BOLD fMRI are in opposition. The shape of the HRF would tend to favor paradigms that induce slow changes in neural activity, while the presence of low-frequency noise would argue for experimental designs that produce more rapid alterations in neural activity. As it happens, knowledge of the shape of the HRF and the distribution of the noise is sufficient to

provide a principled answer as how best to balance these two conflicting forces.

Perfusion fMRI was mentioned before as an alternative fMRI method that is able to provide quantitative measures of cerebral blood flow over time. Perfusion imaging is also notable for the absence of the long-timescale noise that plagues BOLD fMRI studies.¹² As a consequence, perfusion fMRI is well suited to the study of slow changes in neural activity, and might find application in the study of (for example) rehabilitation or emotional states.

CLINICAL CONSIDERATIONS OF THE HRF

As has been described, with knowledge of the shape of the HRF, one can make accurate predictions regarding the transformation of neural activity into BOLD fMRI signal. These predictions can be used for optimizing paradigm design, by asking which patterns of induced neural activity would produce the largest and most easily detected changes in BOLD fMRI signal. Knowledge of the HRF can also be used in a statistical fashion to identify cortical areas where predicted changes in neural activity have taken place. If the task paradigm is expected to produce a particular pattern of neural activity, then it is possible to identify cortical locations where the predicted pattern of BOLD fMRI signal was observed. It is this predictability of the BOLD fMRI system in its mapping of neural activity to BOLD signal that makes possible the use of the technique to test ideas about the brain and behavior.

Because of the central role played by the HRF in dictating optimal paradigm design and the analysis of BOLD fMRI data, it is important to ask if aging and disease pathology can alter the properties of the HRF. This is a critical issue if one wishes to compare BOLD fMRI responses between, for example, young and elderly subjects, or between a patient population and age-matched controls. Consider again our prototypical fMRI experiment. There, we measured the neural

response within the primary visual cortex to a standardized flash of light. Suppose we had measured this response within a group of 10 college-age students and a group of 10 healthy, elderly subjects. If we found that there was a smaller BOLD fMRI signal change in response to the light in the elderly group as compared with the young group, we would conclude that less neural activity is evoked in the visual cortex in the elderly group. An alternative explanation, however, is also available. If it were the case that the coupling of neural activity and blood flow response is altered in the elderly as compared with young controls, then the same magnitude of neural activity in the two populations might produce different BOLD fMRI signals.¹³ Clinical studies that make use of fMRI are rather susceptible to this confound, where differences in the population HRF might be mistaken for differences in the neural response (for a comprehensive review, see D'Esposito et al¹⁴). It is possible that drug treatments or pathological states might alter neurovascular coupling and produce these effects.

A general approach to addressing concerns regarding neurovascular coupling is to measure a hemodynamic response in each of the two populations to be compared using a paradigm that is not expected to differ between the groups. For example, one might wish to examine the effect of drug administration upon the working memory response of the prefrontal cortex in children with attention deficit hyperactivity disorder. To ensure that any difference between the treatment and control populations is the result of a difference in neural activity in the prefrontal cortex, and not the effect of the drug upon neurovascular coupling, a simple control task, such as paced finger movements, can be performed. If there is no difference between the populations in the neural activity evoked by this control task within the primary motor area, then it is less plausible that a global change in neurovascular coupling can explain the effects seen in the prefrontal cortex during the working memory challenge.

PARADIGM DESIGN

Control of mental operations

In our prototypical fMRI experiment, we compared the neural activity obtained during presentation of flashing lights and darkness. In other clinical applications, it is often the case that more complex mental states are to be examined. Instead of the effects of flashing light upon neural activity, one might wish to measure the effects of holding information in memory or making judgments about the appearance of objects. Many studies have as a requirement that some aspect of the stimulus or mental operation be varied so that the neurocomputational correlate of its processing can be studied. In the following, two broad classes of experimental manipulation of mental states are described.

Cognitive subtraction is the prototypical method of isolation of a cognitive processes. Typically, one condition of an experiment is designed to engage a particular cognitive process, such as face perception, working memory, or semantic recall. This “experimental” condition is contrasted with a “control” condition designed to evoke all of the cognitive processes present in the experimental period except for the cognitive process of interest. Differences in neural activity between the two conditions are then attributed to the cognitive process of interest. In essence, a mental state is isolated in an ‘all or none’ fashion. In our example experiment, periods of flashing light were contrasted with periods of darkness to isolate the neural response to light perception. In a study that sought to assess the neural correlates of working memory, periods during which the subject held information across a brief delay might be contrasted with periods that did not require retention of information.

While it is a widely applied approach, cognitive subtraction is prone to some failures of inference. For example, we do not have direct control over the mental states of the subject, so the danger is always present that the subject might engage in a confounding mental operation in addition to the one of

interest. Additionally, cognitive subtraction relies upon the assumption that a cognitive process can be added to a pre-existing set of cognitive processes without affecting them (an assumption termed pure insertion). This might fail if, for example, the act of pressing a button to signal a working memory judgment is different from pressing a button in response to a control task. Effects upon the imaging signal that result from this difference in motor output would be erroneously attributed to working memory *per se*.

Several other cognitive process manipulations have as their goal a reduction in the reliance upon the assumption of pure insertion. The cognitive conjunction design¹⁵ was developed for this purpose. The method uses a set of paired cognitive subtractions, each of which need not completely isolate the cognitive process of interest. The imaging data are then analyzed to find areas that have a significant, consistent response across subtractions. The identification of the same region across multiple pairs of subtractions strengthens the conclusion that the area is activated by the cognitive process that is isolated in each of the subtraction pairs.

Parametric designs offer another alternative to subtraction approaches. In a parametric design, the experimenter presents a range of different levels of some parameter, and seeks to identify relationships (linear or otherwise) between the imaging signal and the values that the parameter assumes. If we were to modify our prototypical experiment to employ a parametric approach, we might present flashing lights that varied in their degree of contrast. It would then be possible to measure the relationship between neural activity and stimulus contrast (a contrast response function). Particular disease states might be identified by an alteration in the shape of this function, as opposed to an absolute reduction in the magnitude of the neural response to maximal stimulation.

There are other types of clinical fMRI study that might not require experimental control of the behavioral state of the subject. For example, studies of seizure localization using

fMRI (see Chapter 6) do not attempt to induce particular patterns of neural activity in the subject, but instead detect neural activity that is spontaneously produced. Similarly, studies of visual hallucinations would measure the neural correlate of a brain state that is not directly under the control of the experimenter (or the subject!).

Timing of events

As BOLD fMRI experiments by necessity include multiple task conditions (e.g. an ‘experimental’ and a ‘control’ period), several ways of ordering the presentation of these conditions exist. In our prototypical experiment, periods of flashing lights and darkness were grouped into relatively long, alternating ‘blocks’ of 30 s each. In contrast to this ‘blocked’ design, we might have employed an ‘event-related’ approach, in which brief flashes of light would be presented every 10 or 15 s, or perhaps randomly intermixed with brief periods of darkness of different durations. While ‘blocked’ and ‘event-related’ approaches are often perceived as rather concrete categories, the distinction between these, and other sorts of designs, is fairly arbitrary. They are better considered as extremes along a continuum of arrangements of stimulus order. Consider every period of time during an experiment as a particular experimental condition. This includes the ‘intertrial interval’ or ‘baseline’ periods between stimulus presentations. In this setting, blocked and event-related designs are viewed simply as different ways of arranging periods of ‘rest’ (or no stimulus) with respect to other sorts of conditions. (For a more complete exploration of these concepts, see Friston et al.¹⁶).

The trade-offs between different experimental designs can be understood in terms of three factors: detection power, randomness, and estimation efficiency.¹⁷ Detection power is the statistical power that the design provides to detect induced changes in neural activity. The benefit of greater detection power is largely self-evident – the greater the

detection power, for example, the shorter the duration of scanning needed to obtain robust results. Randomness describes the predictability of the order of the experimental conditions. For certain classes of study (e.g. tests of memory), it is important that the subject be unable to anticipate the upcoming trial type. In general, increasing the randomness of the design will tend to decrease the detection power. Finally, estimation efficiency is the ability of the design to measure the precise shape of the hemodynamic response function. It is mentioned here for completeness, but for most clinical studies, estimation efficiency will be a relatively unimportant consideration.

A blocked fMRI design is one that maximizes detection power at the expense of randomness. In these designs, two or more conditions alternate in a fixed order over the course of a scan. For most hypotheses of interest, these blocks of time will not be utterly homogeneous but will consist of several trials of some kind presented together. Blocked designs have the obvious difficulty that the subject can anticipate trial types, which may be undesirable in some settings. On the other hand, blocked designs have superior statistical power compared with all other experimental designs. This is because the fundamental frequency of the boxcar can be positioned at an optimal location with respect to the filtering properties of the HRF and the low-frequency noise. For typical shapes of the HRF and distributions of temporal noise, this ideal balancing point occurs with epochs of about 20–30 s in duration.

Event-related designs model signal changes associated with individual trials, as opposed to blocks of trials. This makes it possible to ascribe changes in signal to particular events, allowing one to randomize stimuli, assess relationships between behavior and neural responses, and engage in retrospective assignment of trials. These designs have reduced power compared with blocked designs. Conceptually, the simplest type of event-related design to consider is one that uses only a single stimulus type, and uses sufficient

temporal spacing of trials to permit the complete rise and fall of the hemodynamic response to each trial; a briefly presented picture of a face once every 16 s for example. This is frequently termed a sparse event-related design. Importantly, while this prototypical experiment has only one stimulus, it has two experimental conditions (the stimulus and the intertrial interval). If one is willing to abandon the fixed ordering and spacing of these conditions, more complex designs become possible. For example, randomly ordered picture presentations and rest periods could be presented as rapidly as once a second. The ability to present rapid alternations between conditions initially seems counterintuitive, given the temporal smoothing effects of the HRF. While BOLD fMRI is insensitive to the particular high-frequency alternation between one trial and the next, it is still sensitive to the low-frequency ‘envelope’ of the design. In effect, with closely spaced, randomly ordered trials, one is detecting the low-frequency consequences of the random assortment of trial types. These rapid event-related designs are fairly sensitive to the accurate specification of the HRF for their success.

There is a large range of ‘hybrid’ designs that seek to balance detection power and randomness. For example, ‘stochastic variation’ designs¹⁶ can admit some (incomplete) degree of unpredictability to the ordering of the stimuli while still maintaining relatively high detection power. The recent work of Liu¹⁷ provides an exhaustive consideration of the trade-offs between detection power and randomness.

It should also be noted that there are many other, ‘specialist’ fMRI experimental designs that do not easily fit the categories discussed so far. Within-trial discrimination designs¹⁸ are used, for example, to discriminate periods of neural activity within a behavioral trial. The benefit of this approach is that closely spaced neural events can be discerned, even if their order cannot be randomized (for example, the delay period that falls between seeing a stimulus to be remembered and making a response based upon that stimulus). Neural-

onset asynchrony designs^{19,20} are used to detect differences in the timing of neural activity evoked by different stimuli. Here, a sparse event-related design is used, along with exquisite coupling of the timing of stimulus presentation to image acquisition. A difference in the time of onset of the smooth, BOLD hemodynamic response evoked by two different stimuli within a cortical region is sought. Traveling-wave stimuli are used to define topographic maps of cortical responses, the most familiar being the retinotopic organization of early visual areas.⁸ These designs use stimuli that vary continuously across some sensory space (e.g. retinal eccentricity), and identify, for any point within a cortical area, what was the optimal position of the stimulus within the sensory space for the evocation of neural activity. These designs are often combined with cortical flat-map techniques for the display of results.²¹

STATISTICAL THRESHOLDS IN CLINICAL FMRI

The preprocessing and statistical analysis of BOLD fMRI data is a sizable and complicated topic that cannot be given a comprehensive review here (for a good overview, see Ashburner et al²²). One topic, however, that is worth discussing in the context of clinical fMRI is the balance of statistical control of false-positive and false-negative results.

BOLD fMRI data are typically analyzed in a ‘massively univariate’ approach. One begins with a statistical model that contains covariates (predictors) of the expected pattern of BOLD fMRI responses, and this model is then evaluated at each of the (upwards of 40 000) voxels that comprise the entire brain dataset. The product is a statistical map, in which every voxel in the brain contains a corresponding statistical value for the contrast of the covariates of interest. The final step of the analysis involves assigning a level of statistical significance to those values. If the dataset were composed of a single voxel, then this would be a straightforward enterprise: a *t*-value of greater than 1.96 would be significant

at a $p = 0.05$ level (presuming many degrees of freedom and a two-tailed test). Because there are many voxels, however, we must correct for the likelihood that noise alone might render one t -value significant if many are tested. Solutions to perform this correction in the face of spatial smoothness (which renders adjacent voxels non-independent) exist within gaussian random field theory.²²

Performing the appropriate, mapwise correction to control the false-positive rate can frequently yield a rather stringent statistical value necessary to label any result significant. This, in turn, raises concerns about ‘false-negative’ results, in which true experimental effects might be missed because the experiment is underpowered. While this is of some general concern in the normative studies of cognition to which fMRI is frequently applied, it is of particular concern for clinical applications of fMRI. Within the clinical context, the desired balance between false positives and false negatives is altered. For example, if one is using an fMRI study to define areas of ‘eloquent’ language cortex to be spared during a tumor resection (discussed further below), the cost of a false-negative result is quite high: cortex that may be important for language processing is not identified and is improperly removed. Therefore, an assessment of statistical power, while infrequently performed in cognitive neuroscience studies, should play an essential role in well-constructed clinical applications.

It may further be the case that the level on which statistical control of the false-positive rate is sought may differ between normative and clinical applications. Cognitive neuroscience imaging studies typically control the mapwise or regionwise false-positive rate, meaning that if 20 statistical maps were produced under null-hypothesis conditions (i.e. in the absence of any actual experimental treatment), only one would on average be expected to contain even a single false-positive voxel. In clinical applications, it may be sufficient to control the false-positive rate of voxels within a statistical map, as opposed to across statistical maps. Such a measure is

provided by the false-discovery rate (FDR) approach.²³ Instead of controlling the false-positive rate at a mapwise level, the FDR method controls the proportion of false-positive voxels present within a single map. For example, an FDR threshold of 5% implies that, of the voxels identified as significant within a statistical map, 5% are on average expected to be false positives. The FDR threshold is adaptive, in that it becomes more stringent in the face of reduced signal, and in the limit is equivalent to traditional mapwise thresholds in datasets that contain no experimentally induced signal change. This is neither better nor worse than traditional mapwise control of the statistical significance, but is instead a different stance with regard to inference. FDR methods will likely be of considerable use in clinical applications. For example, it may be desirable to express the confidence of results of functional mapping for surgical planning in terms of the specificity of the population of voxels identified.

MODES OF CLINICAL INFERENCE WITH fMRI

Introduction

With a general understanding of the properties of fMRI and the types of paradigm designs that might be used, there now follows a survey of the types of clinically relevant information that fMRI can obtain. Different ‘modes of inference’, or ways of applying fMRI methods to answer particular clinical questions will be discussed. Each mode of inference requires certain assumptions and provides for different logically supported conclusions. Particular attention will be paid to those sorts of conclusions that can be deduced logically from the results of an fMRI study and those that, while not logically required, may be found to hold empirically. These categories are not meant to be exclusive or exhaustive, but hopefully will provide a guide to thinking about the properties of different clinical functional neuroimaging studies. While reference will be made to fMRI in particular, these notions apply in

general to any correlative neuroimaging method (e.g. positron emission tomography (PET) and event-related potentials). Further, the discussion will be restricted to functional imaging approaches that have as their central measure a change in neural activity. This is to distinguish these applications from other types of ‘functional’ imaging, such as receptor binding assays, measurements of resting cerebral blood flow, and other measures of metabolic function that are not directly related to alterations in regional neural activity.

Localization of necessity

One of the first clinical applications of fMRI was to presurgical mapping (see Chapter 10). The desired inference in this setting is to identify cortical tissue that is necessary for a given mental operation so that it is not removed along with pathological tissue during a subsequent surgical procedure. This mode of inference might be called ‘localization’. For example, one might wish to identify those cortical areas around a glioma that are necessary for language, in order to minimize the risk of producing aphasia following tumor resection.²⁴ In such studies, the subject is presented with a task designed to selectively evoke a particular cognitive state of interest. The key assumption here is that the behavioral paradigm can isolate the mental operation of interest. Various techniques might be used (e.g. cognitive subtraction or parametric manipulation, discussed earlier) to isolate the mental operation of interest from the other processes that invariably are present (e.g. button pushing, preparing responses, etc.).

A critical aspect of clinical fMRI for localization is that, in a strict sense, the desired conclusion cannot be logically supported by the study! We wish to identify cortical areas necessary for a mental operation, in the sense that surgical removal of the area would impair the patient’s ability to perform the task. The converse inference is also important: that we can identify areas that are not necessary for the mental process. Does finding activation of a cortical region in a functional neuroimag-

ing study imply that the region is necessary for the cognitive process? In short, the answer is no. The primary cause of this state of affairs is the observational, correlative nature of neuroimaging. Although we make inferences regarding cognitive processes, these processes are not themselves directly subject to experimental manipulations. Instead, the investigator controls the presentation of stimuli and instructions, with the hope that these circumstances will provoke the subject to enter a certain cognitive state and no other. Although cooperative, the subject may unwittingly engage in confounding cognitive processes in addition to that intended by the experimenter, or alternatively, may fail to differentially engage the process. For example, a subject might constantly engage in the process of declarative memory formation, even during periods of time when he is ‘supposed’ to be performing some other, control behavior. It is therefore not possible to know if observed changes in neural activity in a brain region are the result of the evocation of the cognitive process of interest or an unintended, confounding process. Negative results (even in the face of arbitrarily high statistical power) are also not conclusive, not only because of the failure of perfect control of evocation of cognitive processes, but also because of the possibility that the neuroimaging method employed is not sensitive to the critical change in metabolic activity (e.g. the pattern of neuronal firing as opposed to the bulk, integrated dendritic activity).

Despite these caveats, it is still possible that clinical fMRI studies of the localization of mental operations can be successful. The reason is that, while it is not logically required that a localization study be able to define necessary regions, it may be the case that empirically the necessity of a cortical region for a mental operation correlates very highly with the results of an imaging scan. For example, there has been interest in using BOLD fMRI to replace the ‘Wada’ or intracarotid amobarbital test (see Chapter 6). Performed to guide surgical resection of epileptic foci, each internal carotid artery is in

turn catheterized and instilled with anesthetic to determine which hemisphere is dominant for language. The hope is that BOLD fMRI can be used to determine which hemisphere responds to language tasks and replace this invasive procedure. While it is not logically required that the Wada test and the BOLD fMRI results be in accord, practically this has found to be the case. Indeed, the careful work of Binder and his colleagues has been focused upon finding just the right behavioral paradigm that provides this high degree of correlation. In effect, the ability of a clinical fMRI study to localize necessary cortical regions for a given mental operation must be demonstrated empirically by reference to invasive methods (e.g. surgical resection, Wada testing, or transcranial magnetic stimulation), and cannot be assumed based upon the findings of imaging studies alone.

Detection of dynamic pathology

Essentially the complement of the previous application, fMRI may be used to identify cortex with pathological neural activity. The properties of this mode of inference derive from the use of fMRI to detect spontaneous patterns of neural activity that are unlike neural activity evoked by normal mental operations. The prototypical use is the detection of the cortical origin of seizure activity (see Chapter 6). Unlike many of the other applications of fMRI discussed here, the localization of pathological neural activity does not rely upon a behavioral or stimulus paradigm to create a particular pattern of neural activity, but instead is designed to detect endogenous, pathological neural patterns. Of course, the study might create circumstances favorable to the induction of seizure activity (e.g. sleep deprivation or photic stimulation). In the case of seizure mapping, the goal might be simply to diagnose the presence of seizures, or to identify cortex that, if removed, would reduce or eliminate the seizure activity. There are other neurological disorders that are marked by the presence of pathological patterns of dynamic neural activity. For

example, migraine is marked by spreading neural depression that might be detected using fMRI.²⁵

In some cases, the clinician may know the timing of the neural events to be detected through symptom occurrence or from other forms of monitoring (e.g. simultaneously acquired scalp electroencephalography (EEG)). In this case, identification of the location of the pathological activity is relatively straightforward, as the shape of the HRF can be used to predict the pattern of BOLD fMRI signal that would result from a region that had neural activity time-locked with the symptoms or neurophysiological recording. Under other circumstances, the clinician may wish to identify brain areas that demonstrate pathological patterns of neural activity, even when it cannot be specified when those events took place. To do so, it is necessary to specify signal parameters that can distinguish between normal and abnormal neural patterns. For example, one might find that a propensity for seizure activity in the mesial temporal lobes is marked by an oscillation in the fMRI signal of a particular frequency from this location. Actual applications would require far more sophisticated methods of signal processing²⁶. Distinguishing such pathological signals from background noise and imaging artifacts presents the primary challenge for this type of application of fMRI.

Diagnostic and therapeutic classification

There is a broad range of clinical applications of fMRI that might fall under the title diagnostic and therapeutic classification. The goal here is to use the measured neural response to some behavioral or stimulus paradigm to place patients within a diagnostic category or to predict their response to therapy. fMRI might be used in this way to classify patients with psychiatric disorders or to predict which patients are likely to recover from brain injury and benefit from rehabilitation. In fact, the simple example study presented at the outset of this chapter, in which the response of the visual cortex to a

flash of light might classify patients as having optic nerve disease, is an example of this mode of inference.

Unlike applications that attempt to discern the necessity of a cortical area for a functional process, fMRI studies performed for diagnostic and therapeutic classification need only to establish correlation for their inference. Suppose that a particular magnitude of neural response within the frontal cortex during an attention paradigm strongly predicts if a child will benefit from drug therapy for attention deficit disorder. In a clinical sense, it is irrelevant if that frontal region is actually necessary for performance of the task, or if the pharmacological therapy actually acts at that cortical site – what is important is that the fMRI study is very good at predicting response to therapy, regardless of the underlying mechanism.

The critical requirement for a useful classification study is that the neuroimaging data provide information that could not be obtained by a more readily available physiological or behavioral measure. One might, for example, attempt to develop an fMRI test for Alzheimer's disease (see Chapter 3). Across an elderly population, activation in the hippocampus during performance of an episodic memory task might be found to correlate rather well with subsequent pathology obtained at autopsy. Such a test would only be useful, however, if the correlation were better than that provided by the behavioral performance on the test! This is a crucial point, as neural activity as measured by fMRI can be found to very accurately reflect simple behavioral performance measures.²⁷

Under what circumstances might there be a divergence of performance and neural activity, such that the fMRI study might be of superior predictive value than behavioral measures? A ready analogy is provided by the application of the electromyogram/nerve conduction study (EMG/NCS) to prediction of recovery from peripheral nerve injury. Immediately following a compressive lesion to the peroneal nerve in the leg, for example, a patient might no longer be able to flex his

foot at the ankle. On examination, the patient is unable to generate any measurable force in the affected muscles, but the clinician knows that some patients with this examination go on to improve whereas others will remain permanently weak. An EMG/NCS of the affected muscles and affected nerve can provide this prognostic information by determining if the nerve still has the ability to conduct electrical impulses, in which case the nerve is in continuity and there is a good chance for recovery over the next few months. The discrepancy between the predictive value of the clinical examination and that of the electrophysiological test is the result of a 'floor effect' in the measurement of weakness. No matter whether or not the nerve was in continuity, the effect of the trauma was to produce complete weakness now.

fMRI has great promise to become the EMG/NCS of cortical rehabilitation. It might be used to demonstrate that a cognitive pathway (e.g. for language) is 'in continuity' despite the presence of a severe deficit immediately following a lesion. Not just the degree of predicted recovery, but the assignment to different rehabilitative programs, could be informed by these methods.

Surrogate measure of behavioral state

Some diseases produce symptoms that are subjective, and can only be imperfectly measured by observation or patient report. The experience of pain is an example of this kind (see Chapter 9). In other cases, patients experience symptoms that they under-report, such as those with drug addiction who minimize their degree of drug craving. Some patients may have important internal cognitive states that are not evident to the clinician, such as patients who are 'locked in' from pontine lesions or those who are in a minimally conscious state following more extensive cortical damage.²⁸ Finally, there are patients who feign neurological deficit either due to psychopathology or in hope of secondary gain. In each of these cases, fMRI might be used to measure an internal mental

state of a subject that is not easily obtained through simple behavioral observation.

This application of fMRI reverses the usual direction of inference that is employed in neuroimaging studies. One begins by assuming that a particular cortical region is activated by a particular mental operation. The fMRI data are then examined to determine if increased neural activity was present within the specified region during the task – and, if so, the conclusion is drawn that the subject experienced the particular mental state of interest. To support this mode of inference, it is necessary to first demonstrate that neural activity in the monitored location is actually indicative of the mental state to be measured. What can provide this kind of evidence? Logically, only an exhaustive neuroimaging examination of every possible cognitive process, under every possible circumstance, could provide the necessary evidence. This is obviously impossible in practice, so a series of neuroimaging experiments that demonstrate activation of a particular region during a given cognitive process and no other usually suffices to support the assumption (a logical inference termed enumerative induction). For example, over a series of experiments, one might demonstrate that the degree of activation within an area of the insula is proportional to the degree of craving for rewarding stimuli in control subjects (e.g. food or nicotine²⁹). In a population of patients undergoing behavioral therapy for drug addiction, measurement of the degree of activation within the insula in response to pictures of drug paraphernalia might be taken as a surrogate measurement of the effectiveness of reducing drug craving.

The general challenges of using surrogate markers in clinical trials and other therapeutic settings have been well addressed.³⁰ In the particular context of clinical fMRI, it should be noted that the value of the surrogate measure application is dependent upon the soundness of the assumption that activation in a region is a unique identifier of a given mental state. For early cortical areas, a unique mapping of neural activity to a particular

sensory impression may be well justified, but as more abstract behavioral states are assessed by observation of association cortical areas, the assumption becomes more tenuous.

CONCLUSIONS

These categories should provide a useful guide for considering different clinical applications of fMRI techniques. As mentioned, this is not an exhaustive list, and one can conceive of clinical applications that do not fit well within these categories. For example, this chapter has not addressed the use of fMRI in patients to study redundant cortical systems for mental operations. This application is primarily used to better understand normative systems for cognition,³¹ but may see clinical application as well.

Finally, it should be noted that these techniques of clinical fMRI have been described here in terms of the measurement of regional, bulk neural activity. More subtle measures are also possible. For example, one can assess the degree of effective connectivity between different cortical regions – the extent that one cortical region influences neural activity in another region.³² Such a metric could be used with any of these described clinical applications. One might find, for example, that the signature of attention deficit disorder is decreased effective connectivity between prefrontal and parietal regions during attention-demanding tasks, independent of the average level of neural activity in each region.

REFERENCES

1. Belliveau JW, Kennedy DN Jr, McKinstry RC, et al. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science* 1991; 254: 716–9.
2. Ogawa S, Menon RS, Tank DW, et al. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophys J* 1993; 64: 803–12.
3. Duong TQ, Kim DS, Ugurbil K, Kim SG. Spatiotemporal dynamics of the BOLD fMRI signals: toward mapping submillimeter cortical columns

- using the early negative response. *Magn Reson Med* 2000; 44: 231–42.
4. Logothetis NK, Pauls J, Augath M, et al. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 2001; 412: 150–7.
 5. Malonek D, Grinvald A. Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: implications for functional brain mapping. *Science* 1996; 272: 551–4.
 6. Jezzard P, Song A. Technical foundations and pitfalls of clinical fMRI. *NeuroImage* 1996; 4: S63–75.
 7. Moonen CTW, Bandettini PA (eds). *Functional MRI*. Berlin: Springer-Verlag, 1999.
 8. Engel SA, Glover GH, Wandell BA. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex* 1997; 7: 181–92.
 9. Detre JA, Alsop DC. Perfusion fMRI with arterial spin labeling. In: Bandettini PA, Moonen C (eds). *Functional MRI*. Berlin: Springer-Verlag, 1999: 47–62.
 10. Aguirre GK, Zarahn E, D’Esposito M. The variability of human BOLD hemodynamic responses. *NeuroImage* 1998; 8: 360–9.
 11. Zarahn E, Aguirre GK, D’Esposito M. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage* 1997; 5: 179–97.
 12. Aguirre GK, Detre JA, Alsop DC. Experimental design and the relative sensitivity of BOLD and perfusion fMRI. *NeuroImage* 2002; 15: 488–500.
 13. D’Esposito M, Zarahn E, Aguirre GK, Rypma B. The effect of normal aging on the coupling of neural activity to the bold hemodynamic response. *NeuroImage* 1999; 10: 6–14.
 14. D’Esposito M, Deouell LY, Gazzaley A. Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. *Nat Rev Neurosci* 2003; 4: 863–72.
 15. Price CJ, Friston KJ. Cognitive conjunctions: a new experimental design for fMRI. *NeuroImage* 1997; 5: 261–70.
 16. Friston KJ, Zarahn E, Josephs O, et al. Stochastic designs in event-related fMRI. *NeuroImage* 1999; 10: 607–19.
 17. Liu TT. Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part II: Design of experiments. *NeuroImage* 2004; 21: 401–13.
 18. Zarahn E, Aguirre G, D’Esposito M. A trial-based experimental design for fMRI. *NeuroImage* 1997; 6: 122–38.
 19. Menon RS, Luknowsky DC, Gati JC. Mental chronometry using latency-resolved functional MRI. *Proc Natl Acad Sci USA* 1998; 95: 10902–7.
 20. Henson RNA, Price CJ, Rugg MD, et al. Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. *NeuroImage* 2002; 15: 83–97.
 21. Sereno MI, Dale AM, Reppas JB, et al. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 1995; 268: 889–93.
 22. Ashburner J, Friston K, Penny W. *Human Brain Function*. Amsterdam: Elsevier, 2003.
 23. Nichols T, Hayasaka K. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statist Methods Med Res* 2003; 12: 419–46.
 24. Atlas SW, Howard RS 2nd, Maldjian J, et al. Functional magnetic resonance imaging of regional brain activity in patients with intracerebral gliomas: findings and implications for clinical management. *Neurosurgery* 1996; 38: 329–38.
 25. Huang J, Cooper TG, Satana B, et al. Visual distortion provoked by a stimulus in migraine associated with hyperneuronal activity. *Headache* 2003; 43: 664–71.
 26. Esteller R, Echaz J, D’Alessandro M, et al. Continuous energy variation during the seizure cycle: towards an on-line accumulated energy. *Clin Neurophysiol* 2005; 116: 517–26.
 27. Kirschen MP, Chen SH, Schraedley-Desmond P, Desmond JE. Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: an fMRI study. *NeuroImage* 2005; 24: 462–72.
 28. Schiff ND, Rodriguez-Moreno D, Kamal A, et al. fMRI reveals large-scale network activation in minimally conscious patients. *Neurology* 2005; 64: 514–23.
 29. Pelchat ML, Johnson A, Chan R, et al. Images of desire: food-craving activation during fMRI. *NeuroImage* 2004; 23: 1486–93.
 30. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989; 8: 431–40.
 31. Price CJ, Friston KJ. Scanning patients with tasks they can perform. *Hum Brain Mapp* 1999; 8: 102–8.
 32. Buchel C, Friston KJ. Assessing interactions among neuronal systems using functional neuroimaging. *Neural Networks* 2000; 13: 871–82.

