



Functional Neuroimaging: *Technical, Logical, and Social Perspectives*

BY GEOFFREY K. AGUIRRE

Neuroscientists have long sought to study the dynamic activity of the human brain—what's happening in the brain, that is, while people are thinking, feeling, and acting. Ideally, an inside look at brain function would simultaneously and continuously measure the biochemical state of every cell in the central nervous system. While such a miraculous method is science fiction, a century of progress in neuroimaging technologies has made such simultaneous and continuous measurement a plausible fiction.

Noninvasive measures of neural function have been available since the 1920s, when electroencephalography (EEG) was developed to measure at the scalp the electrical signals of brain activity.¹ Other techniques, such as positron emission tomography (PET) and single photon emission computed tomography (SPECT),² followed in the 1980s and improved upon the crude spatial resolution of EEG. That is, whereas EEG provided information about activity across the entire “space” that is the brain, subsequent technologies could provide far more fine-grained information—about activity in ever-more-specific parts of, or “spaces” within, the brain. A major subsequent advance occurred in the 1990s with the advent of blood-oxygen-level-dependent (BOLD) functional magnetic resonance imaging (fMRI). This last technique has ushered in the modern revolution of neuroimaging; it provides noninvasive, rapid, high-resolution images of the brain in action, and in the past two decades in particular, these technologies have become both more precise and much more widely used. Despite

this progress, practitioners of modern neuroimaging struggle with two kinds of limitations: those that attend the particular neuroimaging methods we have today and those that would limit any method of imaging neural activity, no matter how powerful.

In this essay, I consider the liabilities and potential of techniques that measure human brain activity. There are many interventional procedures that examine the relationship between brain and behavior, but I am concerned here only with methods that measure relevant physiologic states of the central nervous system and relate those measures to particular mental states. I will consider in particular the preeminent method of functional neuroimaging: BOLD fMRI. While there are several practical limits on the biological information that current technologies can measure, these limits—as important as they are—are minor in comparison to the fundamental logical restraints on the conclusions that can be drawn from brain imaging studies. Below, I consider how these logical limitations differ for the different types of research questions that neuroimaging might be used to investigate.

The limits of neuroimaging technologies are well understood within the scientific areas in which neuroimaging is a mature technique.³ During its rapid growth, however, neuroimaging has made multiple lateral moves to new areas of intellectual investigation. A feature of this “imaging colonization” is that the method has moved more quickly than the hard-won cautionary experience would recommend. Partly because of the ubiquity and availability of the hardware and software used to collect data and produce brain images, the initial neuroimaging work in new fields has tended to lack the methodological rigor present in more established areas.

Geoffrey K. Aguirre, “Functional Neuroimaging: Technical, Logical, and Social Perspectives,” *Interpreting Neuroimages: An Introduction to the Technology and Its Limits*, special report, *Hastings Center Report* 45, no. 2 (2014): S8-S18. DOI: 10.1002/hast.294

The practical limits on the biological information the technologies can measure are minor compared to the logical restraints on the conclusions that can be drawn from the brain imaging studies.

Despite the many cautions and limitations, I will close by recognizing the astonishing power of neuroimaging techniques. Contrary to the claims of some critics, neuroimaging is not modern phrenology. As the field has been embracing new, powerful analytic techniques, there has been a shift in goals; instead of focusing on explaining how mental states arise, neuroscientists are now also trying to predict mental states—at least the mental states of those individuals who have agreed to cooperate with the study.

What Do Neuroimaging Technologies Measure?

The bioelectrical basis of human brain function has been studied since the early 1800s, although early work had to be undesirably invasive to circumvent the electrical insulation of the skull. The modern era of noninvasive measurement of human brain activity began in the 1920s with the development of the electroencephalograph.⁴ There has been a profusion of neuroimaging methods since, spawning an alphabet soup of neuro-acronyms: SPECT, PET, MEG, EEG, BOLD fMRI, and NIRS (near infrared spectroscopy).

Despite this daunting variety, the noninvasive measurement of human brain activity may be divided into several broad techniques: those that directly measure the electromagnetic fields produced by active neurons, those that measure the metabolic activity of neurons, and those that measure the local changes in blood flow that are produced by brain activity.

Directly measuring electrical activity in the brain. The fundamental means of information transfer in the nervous system is the movement of charged particles (ions) across the cell membranes of neurons. This movement produces an electrical current and changes the electrical potential (voltage) of the cells. Measurement of these voltage changes is the basis of EEG and ERP (an event-related potential recorded with EEG), while magnetoencephalography (MEG) measures disturbance of the local magnetic field produced by the neural current.⁵ These kinds of neuroimaging studies are often referred to as “electrophysiologic” techniques.

Measuring neural metabolism. Neurons consume glucose (a simple sugar) and oxygen to fuel their biological activity. Two neuroimaging technologies that measure this metabolic function are PET and SPECT. These are nuclear medicine techniques in which a radio-labeled compound

is injected into the bloodstream (or in some approaches, inhaled) and carried to the brain. In one application, glucose molecules are modified to carry a radioactive isotope of carbon. The modified glucose is taken up by nerve cells, and when the carbon atom undergoes radioactive decay, energetic particles are emitted that are recorded by a detector around the subject’s head. The amount and location of this radioactivity indicate areas of increased glucose consumption and thus neural activity. Depending on the compound to which the radioactive isotope is attached, these techniques can measure different metabolic and physiologic aspects of neural function.

Measuring neurovascular changes. Local changes in neural activity give rise to local changes in blood flow and blood oxygenation—thus the term “neurovascular coupling.” When the work of neural processing of information increases, so does the need for fuel in the form of oxygen and glucose (as does the need to clear away the byproducts of that work). Correspondingly, when neurons increase their activity, local blood vessels increase in size, allowing more blood to flow to the active area. Brain activity and blood flow is tightly coupled on a small spatial scale, with an increase in blood flow localized to three to five millimeters around an active point in the brain. Because more oxygen is delivered to active parts of the brain than is actually consumed by the neurons, there is a net increase in the concentration of oxygen in those brain areas that have an increase in neural activity.

Measurement of this local increase in tissue oxygenation following neural activity is the basis of BOLD fMRI. The way in which an MRI scanner ultimately measures this proxy of neural activity—oxygen levels in the blood—is the consequence of a chain of associations in physics and physiology. At the center of hemoglobin, which is the primary oxygen-carrying molecule in the blood, is an iron atom, which has slightly different magnetic properties depending on whether the hemoglobin is currently carrying oxygen. An MRI scanner (which is itself a very strong magnet) can detect this tiny difference in the magnetic property of blood.

In sum, an increase in the activity of neurons at a point in the brain increases blood flow to that area. This increased blood flow brings in more fresh oxygen, which alters the magnetic properties of the iron atoms in hemoglobin molecules, which can be measured by a scanner. This chain of

associations is what allows fMRI to make indirect measurements of neural activity.

These different kinds of neuroimaging techniques make different trade-offs in spatial and temporal resolution. Although the electrically based measures that I mentioned above (such as EEG) provide low spatial resolution—they are not good at detecting where in the brain the activity is occurring—they provide high temporal resolution; that is, they are good at specifying *when* a given neural activity of interest begins and ends. With the metabolic and neurovascular techniques, such as BOLD fMRI and PET, the situation is reversed. These techniques have better spatial resolution than the electrically based techniques, but they have relatively poor temporal resolution: the metabolic changes that they detect occur many seconds after the changes in brain activity that they are used to investigate.

A catalog of additional caveats and idiosyncrasies accompany all of these neuroimaging techniques. And while particular situations might favor one technique over another, BOLD fMRI has become the preeminent method for noninvasive measurement of brain activity because, in many cases, it provides a desirable balance between temporal and spatial resolution. Additionally, fMRI scanning does not expose the subject to ionizing radiation, obtains data from the entire brain, and provides a signal that is strong in relation to sources of noise. Given its near ubiquity, an in-depth understanding of how this neuroimaging tool works is helpful.

From Raw Data to Colorful Images

To produce the brain images that decorate so many scientific articles and lay press reports about neuroscience, the data collected by the scanner must be refined and processed. BOLD fMRI scanners break the brain down into tens of thousands of tiny cube-shaped spaces called “voxels,” which are the fundamental unit of measurement for fMRI studies. Voxels usually measure between 0.5 and 3 millimeters on a side and contain tens to hundreds of thousands of neurons. The scanner makes measurements of activity over the entire brain every one to three seconds, with a typical study requiring about an hour to complete for one person.

Images created from this raw data look quite different from the sharp pictures of brain activity usually associated with a clinical MRI scan. A scan of the entire brain is composed of a stack of two-dimensional slices of the brain at a given moment (see figure 1), and the entire data set is a series of such slices over time. In this image, which is coarse and fuzzy, the darker and lighter shades correspond to the relative magnetic effects of oxygenated blood across the brain. With increases in neural activity, the image becomes “brighter” at a given point as oxygen content increases with increased blood flow.

The data from a neuroimaging experiment are often analyzed by considering the profile of signals over time from each point in the brain. On the right side of figure 1, for example, the signal from a particular voxel goes up when the subject is shown a fearful face and down when the subject

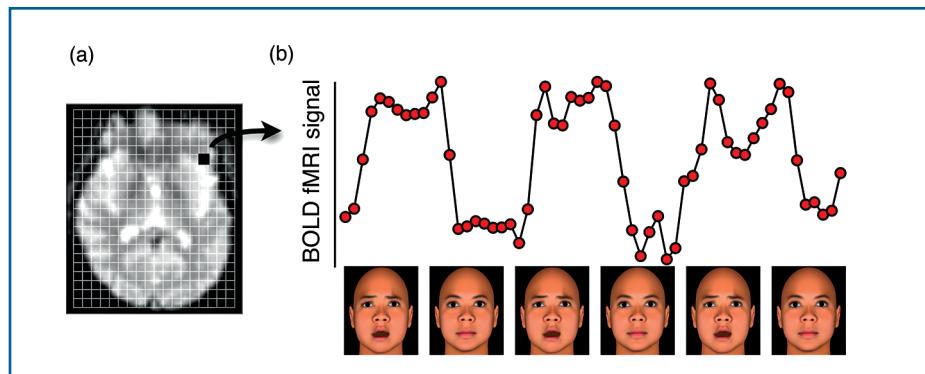


Figure 1. Functional MRI data. On the left is a two-dimensional axial (horizontal) slice through a volume of brain activity. The front of the head is toward the top of the image. The superimposed grid is a cartoon representation of the voxels from which such an image is composed, although in reality the voxels are about three times smaller. The right side of the figure shows time-series data for one voxel (indicated in black) for a hypothetical experiment in which a subject views pictures of faces. Every thirty seconds, the pictures change from a fearful

expression to a neutral expression. For the selected voxel, there might ensue an fMRI response that rises and falls in synchrony with the change in facial expression. Each point on the plot represents the fMRI signal value from the example voxel during the three seconds it takes to acquire one brain volume. The signal convincingly follows the experimental paradigm in this cartoon example, with the expected delay and smoothing of the response in time induced by the sluggish change in blood flow that is being measured.

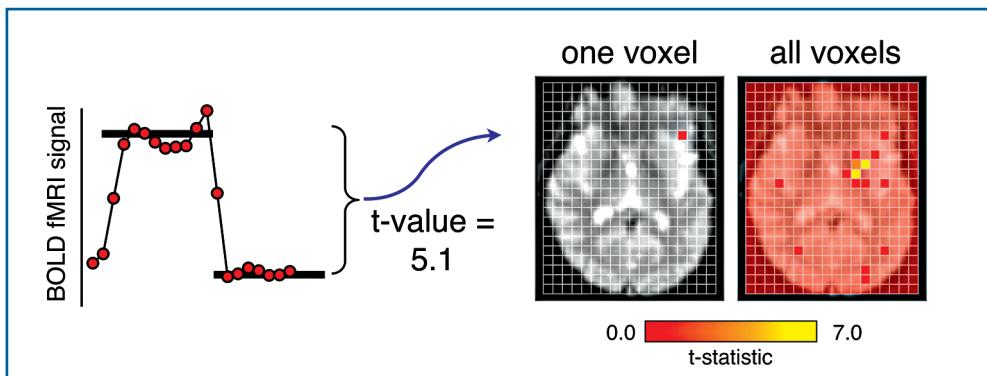


Figure 2. Calculation of a statistical map. In this example, a statistical test is used to compare the mean signal during the two experimental conditions (in reality, more nuanced models of evoked response are used). This test

is shown a neutral face, indicating that seeing a face with emotional content creates a demand for oxygen in that voxel, presumably because the neurons in that voxel become more active when the subject is looking at fearful faces. The response is a little delayed, and it persists while the fearful face is on the screen. Indeed, an important feature of the fMRI signal is that it is both delayed and dispersed with respect to the stimuli (and, presumably, neural events) that evoked the response. This is true even for brief neural events. An immediate and short-lived neural response gives rise to an fMRI signal that takes a couple seconds to start and about ten seconds to run its course. Again, this is because fMRI measures the hemodynamic events (changes in the amount of oxygen in cells) that are the downstream product of neural activity, not the neural activity itself.

Ideally, fMRI data would be acquired from motionless brains of uniform shape. Unfortunately, this is not the case, and a number of processing steps are performed on the data to correct for what researchers call “imaging artifacts.” The need for some of these steps is easily understood. For example, there is inevitably some degree of movement of the head during scanning, which can be corrected in a manner similar to how “image stabilization” technology is used in video cameras. Similarly, because the beating of the heart and the action of the lungs can affect the fMRI signal, the cardiopulmonary cycle is measured during scanning, and its physiologic artifacts removed. Other processing steps are more esoteric consequences of the properties of fMRI imaging but are no less essential.

Even after corrections for motion and other artifacts, it is usually the case that the fMRI signal produced by changes in neural activity is small compared with the random variation present in the data. This random variation in brain activity is not a result of exposure to the stimuli (in the example above, pictures of faces). Instead, there is always activity in the brain, even when we sleep. Statistical tests are

yields a statistical value that is displayed as a color code for the voxel from which the time-series was obtained. The process is repeated for every voxel in the brain volume.

required to distinguish between the changes in brain signal that result from the experiment and those random fluctuations that would have been present regardless. In most cases, the time-series data from each voxel are compared with a predicted response, and the statistical reliability of this effect is measured. Figure 2, for example, illustrates how a *t*-statistic (a measure of statistical likelihood) can be calculated by determining how much variation in the time-series data of one voxel can be explained by switching between the two different faces. This calculation yields a statistical value for the voxel—a probability that the change corresponds to the manipulation of interest. To simplify the display of these results, it is common practice to color-code the voxel from which the time-series was taken. Often, a red-to-yellow scale is used to indicate the strength of the statistical association.

This calculation can be conducted in a similar manner for every voxel in the brain, yielding a map of the statistical effect of the experiment at each point in the brain over time. This statistical map is the final result of many neuroimaging studies.

A few additional steps, however, lie between the calculation of a statistical map and the display of the result in a scientific publication (figure 3). The statistical map has a value for every point in the brain. Because there are so many voxels in a typical neuroimaging study (on the order of 100,000), it is overwhelmingly likely that random fluctuation in the fMRI signal will just happen to produce large statistical values in some voxels even if the experiment had no actual effect on those voxels. Therefore, it is common practice to adopt a threshold that corrects for the number of voxels studied and accept only those statistical values that are unlikely to have occurred by chance (figure 3). Data below this threshold are discarded, leaving a “cleaner” image representing only those changes in blood oxygenation that likely resulted from the experimental stimulus.

Next, it is common practice to digitally “smooth” fMRI data, blurring the measurement from one point in the brain with adjacent points. This may seem undesirable, as the blurring reduces the precision with which a brain activity change can be assigned to a particular point in the brain and makes it likely that very small areas of activity will be blurred to the point of being no longer visible. These limitations, however, are balanced by improvements in statistical power (by reducing the number of independent statistical tests that must be performed in the smoother data). Additionally, when analyzing data across different people, spatial smoothing helps to overcome residual differences in anatomy between subjects, differences that might otherwise render common areas of activation non-overlapping. In other words, smoothing helps reveal what is common in brain response across individuals.

A third step is to lay the data on top of a high-resolution anatomical image. As we have seen, the raw functional fMRI data are coarse, making it difficult to discern the location of specific brain structures. For this reason, the smoothed and thresholded statistical maps are usually displayed atop high-resolution anatomical images of brain structure that are acquired during the same session as the functional activation data. This step simplifies the localization of sites of neuroimaging response to brain anatomy.

Finally, the combined functional and anatomical volume must be displayed. Because the result is three-dimensional, some display options are cumbersome. For example, some neuroimaging studies display the results as a series of two-dimensional slices through the data, starting (for example) at the bottom of the brain and moving up, slice by slice. An alternative popular approach is to produce a three-dimensional rendering of the cortical surface and then display the color-coded, above-threshold statistical results on top of it. In some applications, the cortical surface is digitally separated from the underlying brain structure and flattened

into a sheet, allowing the areas of statistical response from within the curves and folds of the brain to be seen easily.

Each of these display steps is designed to facilitate the communication of results while still accurately reporting the data. It is important to realize, however, how far we have traveled from the initial, fuzzy functional activation data in figure 1a to the final, polished result in figure 3. While these steps are well justified from a scientific standpoint, in aggregate they have come to constitute an aesthetic: there is a certain way that neuroimaging results are “expected” to look. I will consider the consequences of these aesthetic choices below.

Practical Limits of fMRI

While there has been remarkable progress over the past century, there remain considerable limits on the quantity and quality of information that may be derived from neuroimaging techniques in general and from fMRI in particular.

Many of these limits relate to the possibility that the imaging method may fail to measure a relevant aspect of nervous system function. While we have so far considered “brain activity” in the aggregate, there is actually a diverse array of cell types in the brain (neurons and supporting glial cells) all communicating with a variety of neurotransmitters and stimulating different types of receptors. Beyond the ongoing electrical activity of neurons, brain function includes activity-driven changes in DNA transcription, regulation of the expression of channels in the cell membrane, synaptic pruning, and even the modulation of neurogenesis in some areas of cortex. Functional MRI reduces this symphony of biological function to a single note of neural “activity.” The fMRI signal is driven primarily by the release of neurotransmitters at the connection points between neurons, which (under most physiologic circum-

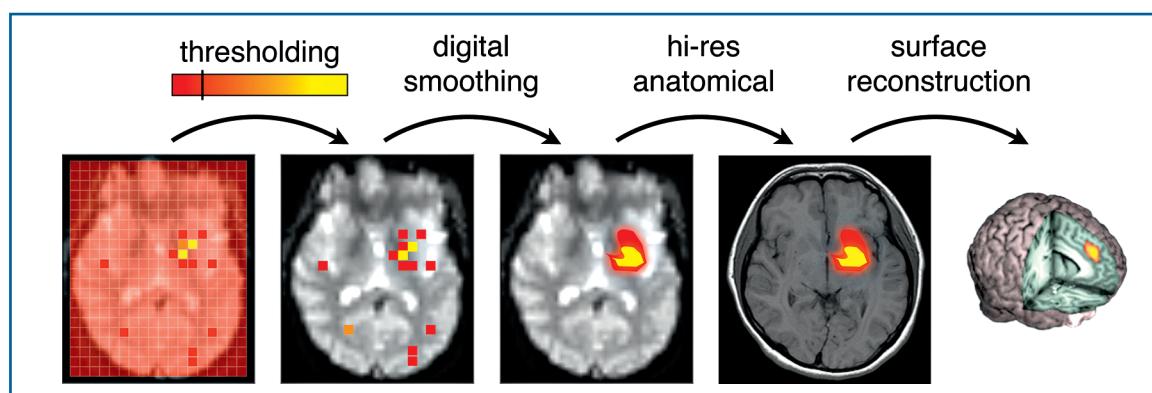


Figure 3. Display of a statistical map. Starting from the initial, volumetric statistical map, several manipulations of the data are undertaken to report the result. These include 1) thresholding the map to include only those voxels with

results that were unlikely to have arisen by chance, 2) smoothing the data in space, 3) displaying the areas of response on an anatomical image, and 4) creating a digital reconstruction of the cortical surface.

*The steps taken to produce neuroimages constitute an aesthetic:
there is a certain way that neuroimaging results
are “expected” to look.*

stances) is proportional to the electrical activity of neurons. The measure does not indicate which neurotransmitter was released, if it had an excitatory or an inhibitory effect on the receiving neuron, or even if the activity is within a local network of neurons or the result of activity from distant cell bodies. The practical consequence of what this jargon-laden summary describes is that the same fMRI signal may be created by very different brain cell operations. There are substantial limits to our knowledge of the relationship between changes in the imaging signal and neural function. The fMRI signal is probably insensitive to many critical aspects of neural activity or computation—it just cannot detect all relevant neural changes that are associated with a mental state of interest.

Further, the imaging signal is integrated over several seconds and several hundreds of thousands of neurons. Consequently, some neural activity cannot be detected using fMRI. For example, a neural signal that rises and falls every second would appear as an unchanging, average signal with fMRI; the blood vessels simply cannot change rapidly enough to keep up with that speed of alternation of neural activity. Similarly, two different mental states might have identical aggregate levels of BOLD signal in a particular brain region but differ critically in how this activity is distributed over a population of neurons. Functional MRI may not pick up these differences.

Additional limitations are caused by generalizing neuroimaging results from individuals to populations. Different people have differently shaped brains. These differences cannot be resolved through simple scaling, and they extend to the complex pattern of folds that comprise the cortical surface. While there have been enormous advances in computational techniques to match brain anatomy between individuals, irreducible differences between people in the mapping of structure-to-function provide one boundary on the translation of group results to individuals. In addition, different people have different lifestyles and medication regimens, which can alter the fMRI signal in surprising ways. The technique relies critically on the steady response of blood flow to changes in neuron activity, but medical states and drug effects can alter this neurovascular coupling, confounding measurements of brain activity. For example, drinking coffee before an fMRI scan can markedly increase the signal, while taking ibuprofen can abolish it.

Finally, a limitation of brain imaging technology that will be welcomed by those concerned about invasions of

mental privacy is the requirement for a cooperative subject. The measurement of subtle aspects of neural response often requires an hour or more of data collection, during which the subject must hold still and attend to the instructions and stimuli presented. Those who conduct neuroimaging studies in children, patient populations, the elderly, or almost any group other than eager college students quickly discover how difficult it is to obtain these data, even from willing participants. It would be easy for a coerced subject to deny a neuroimager useful data through any one of several countermeasures, ranging from averting one's eyes from the stimuli to making small head movements, to actively directing one's thoughts to other subjects. And for those concerned about clandestine neuroimaging, it is reassuring to observe that there is nothing subtle about a thirty-five-ton machine that clangs at one hundred decibels while in operation.

Each of these practical barriers, however, and many more not considered here, are the target of a sustained research assault. Already, clever experimental design can provide access to some neural information below the ostensible temporal and spatial resolution of the technique.⁶ Advances in between-person brain alignment, based on the pattern of cortical surface topology, appear able to overcome much of the anatomical difference among individuals, providing surprisingly good registration of functional mapping across individuals. Subject cooperation has also become less critical with the advent of “resting-state” neuroimaging methods. In these studies, discussed in greater detail below, the pattern of correlated activity across brain regions is obtained while subjects daydream without explicit instructions or stimuli. The observed patterns have been found to correlate with subject traits and states and to support classification of individuals into diagnostic categories. (For reflections on the difficulties associated with finding “neural signatures” for diagnostic categories, see Helen Mayberg’s essay in this collection.) It is as important to recognize that progress is being made in overcoming many practical barriers as it is to recognize that some of the limits to progress are in-principle—they are what I will call “theoretical limits.”

The Theoretical Limits of Neuroscientific Inference

There is diversity not only in neuroimaging methods but also in the kinds of studies in which these techniques

are used. Despite the variety, we can break these studies into two overarching types. The first falls in the domain of neuroscience and asks “why” and “how” questions about the relationship between brain and behavior. These studies examine the “neural basis” of a particular emotional or mental state and seek to understand complex cognitive operations by looking at more elemental neural and psychological units. The second, broad category of study adopts an atheoretical stance and asks whether a particular neuroimaging finding can predict (with some accuracy) behavior now or in the future. I consider first these neuroscientific applications of brain imaging that seek to understand the relationship between neural mechanisms and mental states.

Neuroscientific applications of brain imaging generally pose questions about the relationship of the brain and behavior in one of two directions: forward or reverse inference. Forward inference studies examine the anatomical, neural correlates of a given mental operation and are often used to investigate localization questions—that is, to work out which areas of the brain are active during a particular experimental condition. Generally, the subject is presented with instructions and stimuli that are designed to selectively evoke a particular mental or emotional state of interest and the neuroimaging method identifies if and where changes in neural activity accompany that cognitive process.⁷

For example, does perception of a human face activate a particular area of the brain different from that activated by perception of other stimuli (such as utensils or plants)? The key assumptions for this type of study design are that a given cognitive process exists and that the task that the research subject performs in the scanner isolates only that cognitive process. Various manipulations of stimuli and instructions are used in an attempt to isolate the mental operation of interest from other processes that invariably are present (for example, the neural processes required for the subject to push a button or prepare a response or understand what the instructor is saying, and so on). Often, an “experimental” condition is contrasted with a “control” condition, which is designed to trigger all of the cognitive processes present in the experimental condition *except* the cognitive process of interest. This approach is sometimes referred to as “cognitive subtraction,” since the aim is to “subtract” the undesired mental states, leaving behind the mental state to be studied. The assumption that a mental state or cognitive operation can be purely isolated through a behavioral manipulation lies at the heart of every forward inference study and is a key inferential weakness. Even the cooperative subject might, for example, engage other, confounding, mental operations unintentionally, rendering this assumption invalid.

Neuroscientific studies also examine the relationship between brain and behavior in the “reverse” direction. Reverse inference studies leverage knowledge about the neural cor-

relates of particular mental states to learn something about an imperfectly understood behavior. One begins by assuming that neural activity in a particular area of the brain is a marker of the presence of a particular mental state and no other. For example, neural activity of a certain magnitude at a certain spot in the fusiform gyrus may be assumed to indicate that the subject is in the behavioral state of seeing a human face. The subject then performs a task that may or may not activate the cognitive process of interest. For example, ambiguous stimuli are presented that may be perceived as a face or a vase. If the specified neural activity is seen, the conclusion is drawn that the subject saw a face at that moment in time. Reverse inference studies have been the basis of the rapidly growing fields of emotional, social, and economic neuroscience, in which activity in certain brain locations is taken as evidence of a particular emotional or cognitive state. For example, neuroimaging has been used to claim that witnessing the pain of others produces a sensation of fear that is equivalent to that produced by experiencing pain oneself.⁸

What provides the evidence that a particular region of the brain is uniquely activated by a specific cognitive process? Logically, only an exhaustive, enumerative induction in which neuroimaging is used to examine every possible cognitive process, under every possible circumstance. Since this is obviously practically impossible, 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. Naturally, neuroimaging studies differ in the quality of the evidence to support the reverse inference. In some cases, it can be readily shown that different mental operations are, in fact, capable of activating the same brain region—that is, that the reverse inference study is invalid. For example, as early as 1998 it was recognized that neural activity in the pre-frontal cortex did not necessarily imply the mental state of remembering something briefly, a popular assumption up to that point.⁹

From Neuroscientific Cause to Behavioral Prediction

I have considered neuroimaging studies that seek to understand cause—specifically, how does neural activity cause behavior? That is, what is going on in the brain when a person behaves one way instead of another? This worthy enterprise faces the inferential challenges common to any scientific endeavor that seeks a mechanistic or causal relationship between the phenomena being studied. Namely, imperfect control over the thing the experiment manipulates (in this case, behavior and mental states) leads to uncertainty when concluding how the manipulation relates to the measurement.

As the field of neuroimaging has been embracing new, powerful analytic techniques, there has been a shift in goals, with neuroscientists now trying to predict mental states.

There is, however, another type of neuroimaging study that asks rather different questions and consequently avoids these inferential challenges. In this approach, the goal is not to understand how or why behavior and brain states are related, but simply to provide accurate *predictions* about behavior from brain state measurements. Eschewing a neuroscientific account of *how*, this prediction approach asks only *if* a given pattern of brain imaging data predicts a given behavior now or in the future. As an illustration, consider the inferential differences between the following claims that might be made of a neuroimaging study:

- Activity in the nucleus accumbens (a part of the brain associated with “reward” signals) is increased when a person addicted to cocaine views pictures of drug paraphernalia. Therefore, drug-addicted people experience a sense of craving when viewing pictures of drug paraphernalia.
- The amount of activity in the nucleus accumbens when a person is shown pictures of drug paraphernalia predicts the likelihood that someone will have a positive urine test for cocaine in the week following the brain imaging study.

The first claim concludes that a particular mental state has been evoked in people by viewing activity within their brain. This conclusion depends on knowing that activity in the specified brain region (the nucleus accumbens) is uniquely associated with a sense of craving. Supporting this assumption is fundamentally difficult. While the study might be able to convincingly show that people who use drugs and those who do not have different brain responses when shown certain pictures, attributing those differences to a particular mental state (craving) is challenging. Indeed, the claim that the brain activity difference between the groups is *caused by being drug addicted* is challenging to establish. Perhaps drug users are simply more alert and pay closer attention to the pictures in the experiment, resulting in a larger brain response for this group compared to the nonaddicted control subjects.

In contrast, the second claim concerns matters that are more easily measured and verified. The study has confirmed a prediction: if there is a large brain response to the pictures shown to the subject, then they are more likely to use cocaine in the following week. Apart from challenging the accuracy of the urine testing, the conclusion is

fairly unassailable. This kind of claim makes no mention of mental states or mechanism. The predictive accuracy of responses from this part of the brain for drug use *could be* because this brain region is related to craving. But the prediction could also succeed if the increased activity in the drug addicted is caused by some other mental state, or even by some confounding medical or physiologic state that just happens itself to predict drug use. The shift in stance from providing a causal account of the mechanisms of cognition to simply trying to make accurate predictions of behavior frees the study from many inferential shackles—from needing to make a claim about causes that might be unwarranted.

Almost any neuroimaging study, with a variety of experimental designs and data analysis approaches, may be used in the service of either of these two types of claim. In the last decade, however, two neuroimaging approaches in particular have become increasingly closely associated with the second, predictive application of neuroimaging, indicating a shift among some researchers from studying cause to predicting an effect.

MVPA and Resting-State Studies

Multivoxel pattern analysis (MVPA) is the name given to neuroimaging studies that generally avoid localizing mental operations and consider instead the distributed and complex pattern of neural response that certain stimuli or experimental conditions might trigger across the entire brain. Typically, these studies have an initial “training” phase, during which the distributed neural response to a set of stimuli or conditions is measured. The goal is to repeatedly evoke a particular mental state or behavior in the participant under a variety of circumstances. The commonalities in brain response across these presentations are identified by computer software (generally, “machine-learning algorithms”). At the end of the training phase, the computer has identified that a particular pattern of brain response is triggered by a particular stimulus or behavior. This pattern is often more subtle or complex than a simple “blob” of focal brain activity and is instead distributed within and potentially across brain regions.

Next, there is a “test” phase, during which the subject is shown other stimuli or asked to engage in certain behaviors and the software attempts to classify the resulting brain states according to the patterns learned during the training

phase. In effect, the software is trying to determine whether the current pattern of brain activity of the subject can be matched to one of the patterns learned during the training phase. In this manner, MVPA studies have been used to “read out” from the pattern of brain activity (for example) which one of many pictures a person is currently viewing or whether the subject is currently mentally adding or subtracting a pair of numbers.¹⁰

The information used for the training and testing phases of these studies can be of different kinds and in principle can be conducted within or across subjects.¹¹ For example, a computer classifier can learn the consistent pattern of brain activity brought about by different uncomfortable stimuli (such as a hot wire on the skin or a sharp pin) and then in the test phase try to predict the subjective pain report a given person will provide to yet another, different sort of painful stimulus. Or consider this example: People who smoke could undergo brain scanning while they watch different videos of smoking-cessation advertisements. The machine-learning algorithm could be trained to associate a particular pattern of brain response with the chance that the person will quit cigarettes a month later. In the test phase, the experiment might ask if the pattern of brain activity observed in response to new advertisements can predict the effectiveness of their use in a smoking-cessation campaign.

The key challenge that these predictive studies face is demonstrating that the ability of the brain measurement to predict behavior can extend beyond the particular type of stimuli or experimental states studied. For example, can the measurement of subjective sensation of the pain associated with the hot wire or pin prick generalize to other types of painful stimuli? Chronic pain? Emotional pain? A strength of prediction studies, however, is that these questions may be addressed in a fairly direct empirical manner, without a claim of understanding the neural mechanisms that underlie the generalization. Indeed, if one cares only about making accurate predictions, the details about which aspects of brain function are providing the predictive information are irrelevant. If a brain scan can accurately predict which smoking cessation videos will effectively modify behavior, in some sense it doesn’t matter what aspect of neural activity makes the prediction possible. In practice, however, there is necessarily a rich interplay between the search for neuroscientific explanations and the search for accurate predictions. For example, the prediction software can become more effective when it is designed using knowledge of the principles of neural response and representation in different brain regions. Moreover, the details about how the software maps brain states to behavioral states can serve as the basis for deep neuroscientific insights, perhaps explored in subsequent “hypothesis-driven” studies.

In addition to MVPA studies, *resting-state connectivity* studies are well suited to predictive inferences. Resting-

state connectivity studies collect data from different brain regions while the subject is in a putative “rest” state in the scanner, not exposed to stimuli or instructions. Data showing the patterns of resting-state signals across the brain allow researchers to divide the brain into different but functionally connected regions. These connectivity maps may then be compared across populations (clinical or otherwise) and between different behavioral states. Critics of these studies often challenge the notion that there is a “rest” state and that it would look the same in different people or groups. Rather than a limitation of this approach, however, the difficulty of identifying a particular “rest” state is actually the entire point. When given no explicit instructions about how to manipulate their internal mental operations, people engage in spontaneous mental activity that is revealing about their behavioral state or predictive of their behavioral traits. As with MVPA studies, one can be agnostic about the meaning of particular neural patterns and be content to assert that given neural patterns predict, with a specified accuracy, certain behaviors or group memberships. As with MVPA, however, the actual conduct of this line of scientific inquiry is more often nuanced, relying on an interplay between prediction and neuroscientific hypothesis testing.

Neuroimaging and the Sociology of Science

The ability to obtain noninvasive recordings of human brain activity has a nearly one hundred-year history, and the twentieth anniversary of the development of fMRI has recently passed. During this time, enormous effort has gone into refining, validating, and expanding the analytical and inferential tools of neuroimaging. It is now possible to conduct an fMRI study with great confidence in the statistical validity of the results and with a clear-eyed understanding of the assumptions on which any claims are based. Despite this, critics of neuroimaging can point accurately to many studies that make breathless—and arguably baseless—claims about the brain and behavior. What factors might contribute to the discrepancy between the available rigor and promise of neuroimaging techniques and the proliferation of studies of questionable quality? And why do neuroimaging studies seem to have such an outsize influence on scientific discourse?

I believe that the marked variability of research quality in neuroimaging can be attributed to the manner in which neuroimaging techniques have spread within and between academic disciplines. Before the advent of fMRI, PET scanning was the primary means of obtaining images of brain activity. PET scanners are specialized medical devices, requiring a cyclotron and the injection of radioisotopes, which greatly limited the availability of the technology. The development of fMRI, however, radically altered this situation. MRI scanners can be found in almost every modern

The ready availability of neuroimaging hardware and software has meant that fMRI can be readily adopted in new areas of intellectual inquiry where hard-won cautionary experience is lacking.

hospital, and while some equipment upgrades improve the quality of the data, even a standard, clinical MRI machine may be used for neuroimaging studies. This development produced a rapid proliferation of neuroimaging research.

The easy collection of data, however, raised this challenge: How would it be analyzed? Functional MRI data sets are large and, as described above, must be subjected to fairly involved preprocessing and statistical analyses to account for the many idiosyncratic properties of the data in space and time. The quantitative and statistical knowledge necessary to analyze these data is quite specialized and requires some skill with computer programming. While it might be easy to place a subject in the scanner and collect a gigabyte of neuroimaging data, the challenge of producing a meaningful result in the form of a brain image remains.

As it happens, the ubiquity of MRI scanners intersected with the availability of free, open-source software capable of performing these analyses. Statistical parametric mapping (SPM) software, freely available from the Functional Imaging Laboratory of Wellcome Trust Center, provided for push-button analysis of neuroimaging data. SPM software placed complicated data analysis and statistical methods behind an easy-to-use graphical interface, removing one of the barriers to new users of neuroimaging methods. Indeed, the democratization of scanner and software was perfectly timed with the rapid rise of the Internet in the early 1990s, making the software available for remote download.

Consequently, little more than research access to a hospital and an Internet connection was needed to perform

neuroimaging experiments and produce pictures of the brain in operation. Inevitably, despite the best intentions and most strenuous of educational efforts on the part of SPM authors, investigators were empowered to analyze fMRI data with minimal understanding of the many statistical processes and assumptions that lay behind the software buttons and the brain images that they ultimately produced. The difference between the understanding that the SPM authors brought to the data analysis and what many investigators who used their software possessed is somewhat like the difference between training in martial arts—a long and arduous process that produces both power and responsibility—and buying a gun.

The ready availability of neuroimaging hardware and software has meant that fMRI can be readily adopted in new areas of intellectual inquiry without the need to bring along the hard-won cautionary experience that results from years of training and experience. While best practices may be well established in areas of inquiry in which neuroimaging is a mature technique, they are often unknown in the first, exciting rush of novel work in a new field. Essentially, the evolution of neuroimaging technique is recapitulated anew each time fMRI seeds new ground.

This history partially accounts for the steady production of papers that aim to illuminate and correct statistical errors that appear in the neuroimaging literature. Recent high-profile publications have cautioned investigators to appropriately correct for multiple statistical comparisons across the brain volume, to avoid performing statistical tests on data subsets that were themselves selected by that

statistical test, and to test for statistical interactions when interaction results are claimed.¹² (For more, see Martha Farah's paper in this collection.) Each of these statistical errors is well understood outside of (and within subsets of) neuroimaging research, but understanding must be reinforced anew. A deeper and more worrisome issue is that many choices are available regarding the analysis parameters of neuroimaging data (for example, how



Figure 4. "Organic" activation. A typical neuroimaging result is shown between (left) a FLAIR image of a white matter lesion and (right) a gadolinium-enhanced meningioma on a T1-weighted sequence.

much to smooth, which regions of the brain to examine, whether to remove certain effects of no interest). Many degrees of freedom are therefore available to the investigator to explore the data in search of a desired result and to offer that result as if it were the inevitable consequence of a predetermined analysis pathway. This statistical error is not by any means unique to neuroimaging data, but it is magnified by the ease with which the dials of analysis can be turned on a computer screen.

The power of the brain images themselves must also be acknowledged. (Again, see Farah's paper for more on this issue.) Although there is a scientific justification for each step in the transition from raw, functional imaging data to final, published brain image, in aggregate these steps produce brain images that have a distinctive "organic" aesthetic. The favored presentation of neuroimaging data creates the appearance that the activation is a natural property of the brain, with a scale and smoothness that matches the anatomical structures (the cortical folds, subcortical nuclei, and so on). Indeed, the colored blobs of functional activity look as if they could have grown there and are simply *revealed* by the scanner in the same way a multiple sclerosis plaque or benign brain tumor is revealed by a structural scan (see figure 4). The appearance of functional neuroimaging data encourages a viewer to interpret the results as representing a static, innate property of the brain rather than a particular set of experimental circumstances. Recent advances of multivoxel analysis notwithstanding, this display aesthetic has also encouraged now-outdated interpretations of brain function and mental operations as highly localized (that is, as confined to one area of the brain).

Room for Astonishment

As we seek to understand the influence of neuroimaging in general and fMRI in particular, it must finally be acknowledged that this is an astonishing scientific tool. Pushed to its limits, fMRI can resolve the separate contribution to cognition of as few as 23,000 cortical neurons (out of 16 billion) on a second-to-second basis.¹³ This is a monumental accomplishment. Functional MRI provides a series of volumetric images of the entire human brain, from motor cortex to cerebellum, each image composed of tiny resolution elements smaller than the tip of your pinkie, assembled into a movie of brain activity that updates every three seconds. All this is done noninvasively, without radiation, while the subject lies comfortably in a scanner.

Despite the limitations imposed by physiology and scientific fallibility, the aggregate of insights produced by neuroimaging is impressive, and growing rapidly. MVPA and resting-state approaches are powerful new analytic

tools that are ushering in a second age of neuroimaging research, with increased rigor, improved applicability to clinical populations, and a focus on empirical prediction. Like any other scientific enterprise, there are fits and starts, with keen-eyed critics playing their part to sharpen the approach.

References

1. M. G. H. Coles and M. D. Rugg, *Event-Related Brain Potentials: An Introduction. Electrophysiology of Mind* (Oxford: Oxford Scholarship Online Monographs, 1996), pp. 1-27.
2. M. N. Wernick and J. N. Aarsvold, *Emission Tomography: The Fundamentals of PET and SPECT* (Amsterdam, Netherlands: Elsevier, 2004).
3. N. K. Logothetis, "What We Can Do and What We Cannot Do with fMRI," *Nature* 453 (2008): 869-78.
4. D. Millet, "The Origins of EEG," 7th Annual Meeting of the International Society for the History of the Neurosciences, Los Angeles, CA, June 3, 2002, at <http://www.bri.ucla.edu/nha/ishn/ab24-2002.htm>.
5. C. C. Gallen, E. C. Hirschkoff, and D. S. Buchanan, "Magnetoencephalography and Magnetic Source Imaging: Capabilities and Limitations," *Neuroimaging Clinics of North America* 5, no. 2 (1995): 227-49.
6. See G. K. Aguirre, J. A. Detre, and J. Wang, "Perfusion fMRI for Functional Neuroimaging," *International Review of Neurobiology* 66 (2005): 213-36.
7. G. K. Aguirre, "Functional Imaging in Behavioral Neurology and Cognitive Neuropsychology," in *Behavioral Neurology and Cognitive Neuropsychology*, ed. T. E. Feinberg and M. J. Farah (New York: McGraw Hill, 2003); R. A. Poldrack, "Can Cognitive Processes Be Inferred from Neuroimaging Data?" *Trends in Cognitive Sciences* 10 (2006): 59-63.
8. A. Olsson, K.I. Nearing, and E. A. Phelps, "Learning Fears by Observing Others: The Neural Systems of Social Fear Transmission," *SCAN* 2 (2007): 3-11.
9. M. D'Esposito et al., "Human Prefrontal Cortex Is Not Specific for Working Memory," *NeuroImage* 8 (1998): 274-82.
10. K. A. Norman et al., "Beyond Mind-Reading: Multi-voxel Pattern Analysis of fMRI Data," *Trends in Cognitive Science* 10, no. 9 (2006): 424-30.
11. M. D. Fox and M. E. Raichle, "Spontaneous Fluctuations in Brain Activity Observed with Functional Magnetic Resonance Imaging," *Nature Reviews Neuroscience* 8, no. 9 (2007): 700-11.
12. E. Vul et al., "Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition (The Paper Formerly Known as 'Voodoo Correlations in Social Neuroscience')," *Perspectives on Psychological Science* 4, no. 3 (2009): 274-90; N. Kriegeskorte et al., "Circular Analysis in Systems Neuroscience—The Dangers of Double Dipping," *Nature Neuroscience* 12, no. 5 (2009): 535-40; S. Nieuwenhuis, B. U. Forstmann, and E.-J. Wagenmakers, "Erroneous Analyses of Interactions in Neuroscience: A Problem of Significance," *Nature Neuroscience* 14 (2011): 1105-07; J. Simmons, L. Nelson, and U. Simonsohn, "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allow Presenting Anything as Significant," *Psychological Science*, forthcoming.
13. For details of this calculation, see G. K. Aguirre, "Number of Neurons in a Voxel," accessed September 18, 2013, https://cfn.upenn.edu/aguirre/wiki/public:neurons_in_a_voxel.