3-14-2014

Functional Neuroimaging: Technical, Logical, and Social Perspectives

Geoffrey K. Aguirre

University of Pennsylvania, aguirreg@mail.med.upenn.edu

Recommended Citation

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/neuroethics_pubs/134
For more information, please contact repository@pobox.upenn.edu.
Functional Neuroimaging: Technical, Logical, and Social Perspectives

Abstract
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. 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. 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.

Disciplines
Bioethics and Medical Ethics | Neuroscience and Neurobiology | Neurosciences
Neuroscientists have long sought to study the dynamic activity of the human brain while people think, feel, and act. 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 it a plausible fiction.

Non-invasive 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. Other techniques, such as positron emission tomography (PET) and single photon emission computed tomography (SPECT), followed in the 1980s and improved upon the crude spatial resolution of EEG. 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, as it provides non-invasive, 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 limitations: those that attend the particular neuroimaging methods we have today, and those that would limit even the imaginary technique of omniscient imaging.

I consider in this paper 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 with the fundamental, logical restraints upon the conclusions that can be drawn from brain imaging studies. Below, I consider how these logical limitations differ for the different types of questions that might be asked using neuroimaging.

The limits of neuroimaging technologies are well understood within the scientific areas in which neuroimaging is a mature technique. 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. The ubiquity and availability of the hardware and software to collect data and produce brain images has contributed to a tendency for the initial neuroimaging work in new fields to lack the methodological rigor present in more established areas.
Despite many cautions and limitations, I will close this paper by recognizing the astonishing power of neuroimaging techniques. Contrary to the claims of some critics, neuroimaging is not simply a modern phrenology. As the field embraces new, powerful analytic techniques, there has been a shift from explanation to prediction. This transition brings neuroimaging to the brink of science fiction, providing a method to access the private mental states and traits of individuals who choose to cooperate with a brain scan.

What do Neuroimaging Technologies Measure?

The bio-electrical basis of human brain function has been studied since the early 1800s, although initially with undesirable invasiveness to circumvent the electrical insulation of the skull. The modern era of non-invasive measurement of human brain activity began in the 1920s with Hans Berger’s development of the electroencephalograph. There has been a profusion of neuroimaging methods since, spawning an alphabet soup of neuro-initialisms: e.g. SPECT, PET, MEG, EEG, BOLD fMRI, ASL, and NIRS (see glossary and annotated bibliography at the end).

Despite this daunting variety, the non-invasive measurement of human brain activity may be divided into several broad techniques: (a) those that directly measure the electro-magnetic fields produced by active neurons, and (b) those that measure the metabolic activity of neurons, and those (c) that measure the local changes in blood flow that are produced by brain activity.

(a) 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, while MEG measures disturbance of the local magnetic field produced by the neural current. These kinds of neuroimaging studies are often referred to as “electrophysiologic” techniques.

(b) 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 then travels 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 pinpoints areas of increased glucose consumption and thus neural activity. Depending upon the compound to which the radioactive isotope is attached, these techniques can measure different metabolic and physiologic aspects of neural function.
(c) Measuring neuro-vascular changes

A fortuitous property of brain physiology is that local changes in neural activity give rise to local changes in blood flow and blood oxygenation. This "neuro-vascular" coupling is sensible from a teleological perspective. Increases in the computational activities of neurons requires an increased supply of oxygen and glucose fuel, and an increased need to clear away the byproducts of metabolism. 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 spatial scale, with an increase in blood flow localized to only 3-5mm around an active point in the brain. More oxygen is delivered to active parts of the brain than is actually consumed by the neurons. Therefore, there is a paradoxical increase in the concentration of oxygen in brain areas with increase activity.

Measurement of this local increase in tissue oxygenation that follows neural activity is the fundamental basis of the preeminent neuroimaging technique: Blood Oxygen Level Dependent (BOLD) functional Magnetic Resonance Imaging (fMRI). The manner by which an MRI scanner ultimately measures this proxy of neural activity is the consequence of a chain of associations in physics and physiology.

Hemoglobin is the primary oxygen carrying molecule in the blood. At its center, each hemoglobin molecule contains an iron atom. Iron, in turn, has slightly different magnetic properties depending upon whether the hemoglobin is currently carrying oxygen or is not. An MRI scanner (which is itself a very strong magnet) can be made sensitive to this tiny difference in the magnetic property of blood when it has more or less oxygen present.

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. The scanner, then, measures this alteration of magnetic strength. This chain of associations is what allows fMRI to make indirect measurements of neural activity.

These different kinds of neuroimaging techniques make different sacrifices in spatial and temporal resolution. The electrically-based measures provide high temporal resolution, with the ability to examine the evolution of neural activity across the brain on a millisecond basis. However, the distortion of electrical and magnetic fields produced by the skull and scalp leave these techniques with coarse spatial resolution. In contrast, the metabolic and neuro-vascular techniques, such as BOLD fMRI and PET, have better spatial resolution, but relatively poor temporal resolution as changes in these metabolic measures lag behind the changes in brain activity.

A catalog of additional caveats and idiosyncracies accompany all of these neuroimaging technique. And while particular situations might favor one technique over another, BOLD fMRI has become the preeminent method of non-invasive 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.

Because of the near-ubiquity of fMRI, I will focus upon this neuroimaging tool, and next consider how the data collected by the scanner is refined and processed to produce the colorful brain images that now decorate so many scientific articles and lay-press reports of neuroscience.

**BOLD fMRI: From Raw Data to Colorful Images**

BOLD fMRI scanners break the brain down into tens of thousands of tiny cubes called “voxels,” which are the fundamental unit of measurement for fMRI studies. Voxels are usually between 0.5-3 mm per side and they contain tens to hundreds of thousands of neurons. The scanner makes measurements of activity over the entire brain every 1-3 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. Figure 1a shows a single, horizontal slice through the brain at one point in time. A scan of the entire brain is composed of a stack of these 2-dimensional slices, and the entire data set is a series of these 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 lighted 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 signal over time from each point in the brain. Figure 1b shows the time-series response for a single voxel in a hypothetical experiment. You can see that the signal from this particular voxel goes up when the fearful face is shown and down when the neutral face is shown, 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. You may also notice that the response is a little delayed and that it lasts for a while (it isn’t a quick spike). 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—the person in the scanner sees the fearful face and experiences an immediate and short-lived response to it, but the scanner detects a slightly delayed response that stretches out over several seconds. This is because fMRI measures the physiologic, hemodynamic events (changes in the amount of oxygen in cells) that are the downstream product of neural activity and not the neural activity itself.

Another way of stating this point is to say that the spatial and temporal resolution of fMRI is limited by the neuro-vascular coupling that is the source of contrast (it is limited by the nature of what is being measured). While MRI images can be obtained as rapidly as every 100 milliseconds, and with spatial resolution in the tenths of a millimeter, this fine resolution has little practical advantage for fMRI. Changes in neural activity, which are often very quick and can be very localized, give rise to a change in blood flow and oxygenation (and thus a change in the fMRI signal) that evolves over seconds and that spreads over several millimeters, making it impossible to identify precisely which neurons were active. As a result of these inherent limits, fMRI images are seldom acquired more frequently than once a second and are typically composed of voxels no smaller than 1 mm per side.

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 “image stabilization” technology in video cameras. Similarly, because the beating of the heart and the respiratory cycle can affect the fMRI signal, the cardio-pulmonary cycle is measured during scanning and its physiologic artifacts are removed. Other processing steps are more esoteric consequences of the properties of fMRI imaging, but are no less essential to conduct.

Even after correcting for motion and other artifacts, it is usually the case that the fMRI signal that is 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

---

**Figure 2**

BOLD fMRI signal

\[ t\text{-value} = 5.1 \]

one voxel all voxels

0.0 7.0

t-statistic
exposure to the stimuli (in this example, face pictures). Instead, there is always activity in the brain, even when we sleep. Statistical tests are required to distinguish between the changes in brain signal that result from the experiment, and those random fluctuations which 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 examining the amount of variation in one voxel’s time-series data that can be explained by switching between the two different faces. This calculation yields a statistical value for the time-series data for one 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.

There are a few additional steps, however, that 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 only accept statistical values that are unlikely to have occurred by chance (Figure 3). Data below this threshold are discarded, leaving a “cleaner” image representing only changes in blood oxygenation that are likely to have 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 initially seem like an undesirable thing to do, 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 need to be performed in the smoother data). Additionally, when analyzing data across different people, spatial smoothing helps to overcome residual differences in anatomy between subjects that might otherwise render common areas of activation non-overlapping (i.e. smoothing helps reveal what is in 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 2-dimensional slices through the data, starting (for example) at the bottom of the brain and moving up, slice-by-slice. An alternative, and increasingly popular, display approach is to produce a three-dimensional, digital rendering of the cortical surface and then display the color-coded, above-threshold statistical results upon 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 travelled from the initial, fuzzy functional activation data (Figure 1a) to the final, polished result (Figure 3, right). 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 considered so far “brain activity” in 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 on-going 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 neuro-transmitter at the connection points between neurons, which (under most physiologic circumstances) 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 upon 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 this jargon-laden summary is that the same fMRI signal may be obtained for a diverse set of underlying brain cell operations. We do not know what “more” of the imaging signal means in terms of neural function. Also, the fMRI signal is likely insensitive to many critical aspects of neural computation—it just cannot detect all relevant neural changes in response to a mental state of interest.

Further, the imaging signal is integrated over several seconds of time 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.

There are additional limitations that follow from generalizing neuroimaging results from individuals to populations. Different people have different shaped brains. These differences cannot be resolved through simple scaling, and they extend to the complex pattern of gyral and sulcal 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 upon the steady response of blood flow to changes in neuron activity. Medical states and drug effects can alter this neuro-vascular coupling, confounding measurements of brain activity. For example, drinking coffee before a fMRI scan can markedly increase the signal, while taking an ibuprofen can abolish it.

Finally, a perhaps welcome limitation of brain imaging technology 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 beside eager college students, quickly discover how difficult it is to obtain these data even from willing participants. It would be trivially easy for a coerced subject to deny a neuroimager useful data through any one of several countermeasures, ranging from averting one’s eyes from presented to stimuli, to making small head movements, to actively directing one’s thoughts to other subjects. And for those who would fear clandestine neuroimaging, cataloging our mental states and traits without our knowledge, it is
reassuring to observe that there is nothing subtle about a 35-ton machine that clangs at 100 decibels while in operation.

It should be noted that each of these practical barriers, 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. Advances in between-person brain alignment, based upon the pattern of cortical surface topology, appear able to overcome much of the anatomical difference between 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. Further, incremental improvement of neuroimaging approaches will wear down these limits, and bring the technology ever closer to the theoretical limit of what might be learned.

Theoretical Limits of Neuroscientific Inference

There is diversity not only in neuroimaging methods, but 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 a-theoretical stance, and asks whether a particular neuroimaging finding can predict (with some accuracy) behavior now or in the future. We consider first these neuroscientific applications of brain imaging that seek to understand the neural mechanisms that link brain function and mental states.

These neuroscience 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 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.

For example, does perception of a human face activate a particular area of the brain different from that evoked by perception of other stimuli? The key assumption for this type of study design is that a given cognitive process exists and that the task isolates only that cognitive process. Various manipulations of stimuli and instructions are used in an attempt to isolate the mental operation of interest from the other processes that invariably are present (e.g., neural process required to push a button or prepare a response or understand what the instructor is saying, etc.). Often, an "experimental"
condition is contrasted with a "control" condition that is designed to evoke all of the cognitive processes present in the experimental period except the cognitive process of interest. This approach is sometimes referred to as “cognitive subtraction”, as the aim is to “subtract” away the undesired mental states, leaving behind the mental state to be studied. This 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.

Neuroscience studies also examine the relationship between brain and behavior in the “reverse” direction. Reverse inference studies leverage knowledge about the neural correlates 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 which may or may not evoke 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.

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. This is obviously practically impossible, 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. Naturally, neuroimaging studies differ in the quality of the evidence to support the reverse inference. In some cases, it may 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.

From neuroscientific cause to behavioral prediction

I have considered neuroimaging studies that seek to understand cause, specifically how does neural activity cause behavior (or, what is going on in the brain when a person behaves one way verses another?). This worthy enterprise faces the inferential challenges common to any scientific endeavor that seeks a mechanistic or causal 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.

There is, however, another type of neuroimaging study that asks rather different questions, and consequently avoids these inferential limitations. 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:

(a) 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.

(b) The amount of activity in the nucleus accumbens in response to 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 is dependent upon knowing that activity in the brain region (the nucleus accumbens) is uniquely associated with a sense of craving. This is a fundamentally difficult assumption to support. While the study might be able to convincingly claim that there is a difference in brain response to the pictures between people who use drugs and those who don’t, attributing those differences to a particular mental state (craving) is challenging. Indeed, the claim that the brain activity difference between the groups is due to being drug addicted *per se* is challenging. Perhaps (for example) drug users are more alert, and simply pay closer attention to the pictures in the experiment, resulting in a larger brain response for this group compared to the non-addicted 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, 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.

Importantly, this second kind of claim makes no mention of mental states or mechanism. The predictive accuracy of responses from this point in 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 from some other mental state, or even from some confounding medical or physiologic state that just happens itself to predict drug use. The shift in stance from providing an account of the mechanisms of cognition to simply trying to make accurate predictions of behavior frees the study from many inferential shackles.

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 inferential claim. In the last decade, however, two particular neuroimaging approaches have become increasingly 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**

Multi-voxel pattern analysis (MVPA) is the name given to neuroimaging studies that generally avoid localizing mental operations, but instead consider the distributed and complex pattern of neural response that stimuli or experimental conditions might evoke across the entire brain. Typically, there is 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 evoked by a particular stimulus or behavior.

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 evoked brain states according to the patterns learned during the training phase. In effect, can the pattern of brain activity the subject currently has 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.

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. For example, a computer classifier can learn the consistent pattern of brain activity evoked by different uncomfortable stimuli (e.g., a hot wire on the skin or a sharp pin), and then in the test phase be used to predict the subjective pain report a given person will provide to some novel painful stimulus. As another example, people who smoke could undergo brain scanning while they watch different videos of smoking cessation advertisements. The machine learning algorithm would then be trained to associate a particular pattern of brain response to the chance that the person will have 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 pain 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 overt concern for the neuroscientific basis of the success or failure of the generalization. Indeed, if one cares only about making accurate predictions, the details of what aspects of brain function are providing the predictive information are irrelevant. If a brain scan can predict with high accuracy which smoking cessation videos will effectively modify behavior, it in some sense doesn’t matter what aspect of neural activity enables this prediction. Yet in practice, there is a rich and necessary interplay between the neuroscientific and predictive stances. 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 of 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, another kind of neuroimaging approach is also 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, and 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, functionally connected regions. These connectivity maps may then be compared across populations (clinical or otherwise) and between different behavioral states. A frequent initial critique of these studies is to challenge the very 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, this fact is actually the entire point. When given no explicit instructions regarding 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 an MVPA studies, a radical inferential stance adopts an agnostic view towards the meaning of particular neural patterns in such studies, and is content to assert that given neural patterns predict with a specified accuracy certain behaviors or group memberships. As with MVPA, 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 non-invasive recordings of human brain activity has a nearly 100 year history, and the 20th anniversary of the development of functional MRI has recently passed. During this time, enormous effort has been expended to refine, validate, and expand 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 clear-eyed understanding of the assumptions upon which any claims are based. Despite this, critics of neuroimaging can point accurately to many studies that make breathless—and arguably baseless—claims regarding 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? Further, why is it that neuroimaging studies seem to have such an outsize influence upon scientific discourse?

I believe that the marked variability of research quality in neuroimaging may 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 are nearly ubiquitous in modern hospitals, and while some equipment upgrades improve the quality of the data, even a standard, clinical MRI machine may be used for neuroimaging studies. This produced a rapid and revolutionary democratization of neuroimaging research.
The easy collection of data, however, raised the challenge: How would it be analyzed? Functional MRI data sets are large and, as described above, must be subjected to fairly involved pre-processing 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 this data is quite specialized, and requires in addition 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 how to produce a 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, made freely available by the FIL (Functional Imaging Laboratory of Wellcome Trust Center), provided for push-button analysis of neuroimaging data. This software placed complicated data analysis and statistical methods behind an easy to use graphical interface. This removed a barrier 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 the authors of SPM, investigators were empowered to analyze fMRI data with minimal understanding of the many statistical processes and assumptions which lay behind the software buttons and the brain images that they ultimately produced. A playful analogy to this circumstance considers the difference between long and arduous martial arts training that produces both power and responsibility, and the provision of a gun.

The ready availability of neuroimaging hardware and software has meant that the technique can be readily adopted in new areas of intellectual inquiry, without having to bring along the hard-won cautionary experience that result from years of training and experience. While best practices may be understood in areas of inquiry for which neuroimaging is a mature technique, this is often not the case in the first, exciting rush of novel work in a new field. Essentially, the phylogeny of 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 upon data subsets that were themselves selected by that statistical test, and to test for statistical interactions when interaction results are claimed. Each of these statistical errors are well understood outside of (and within subsets of) neuroimaging research, but must be reinforced anew. A deeper and more worrisome issue is that many choices are available regarding the analysis parameters of neuroimaging data (e.g., how 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 pre-determined analysis pathway. This is not a statistical error that is by any means unique to neuroimaging data, but it is one that is magnified by the ease with which the dials of the analysis may be turned on a computer screen.

The power of the brain images themselves must also be acknowledged. As I discussed earlier, the transition from the raw, functional imaging data to the final, published brain image involves many individual processing steps, each of which can be well justified in scientific terms. In aggregate, however, these steps serve to 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 (e.g., the cortical folds and subcortical nuclei). Indeed, the colored blobs of functional activity look like they could have grown there, revealed by the scanner in the same way a multiple sclerosis plaque or benign brain tumor are revealed by a structural scan (Figure 4). The appearance of functional neuroimaging data facilitates an interpretation of the results as representative of a static, innate property of the brain, as opposed to the outcome of particular experimental circumstances. Recent advances of multi-voxel analysis notwithstanding, this display aesthetic has also encouraged now outdated interpretations of brain function and mental operations as highly localized (i.e., confined to one area of the brain).

Figure 4

Room for astonishment

In seeking to understand the influence of neuroimaging in general, and fMRI in particular, it must finally be acknowledged that this is an astonishing scientific tool. Despite the blurring effect of the hemodynamic response, fMRI is routinely used to measure the activity of as few as 23,000 neurons integrated over 3 seconds of time.¹ 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

¹ For details of this calculation, see: https://cfn.upenn.edu/aguirre/wiki/public:neurons_in_a_voxel
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 non-invasively, 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, which are ushering in a second age of neuroimaging research, with increased rigor, improved applicability to clinical populations, and a focus upon empirical prediction. Like any other scientific enterprise, there are fits and starts, with keen-eyed critics playing their part to sharpen the approach.

**Figure Legends**

**Figure 1.** BOLD “echoplanar” functional data. (a) Shown is a 2-dimensional, axial (horizontal) slice through a volume of brain activity data. The front of the head is towards the top of the image. The super-imposed grid is a cartoon representation of the voxels from which such an image is composed, although in reality the voxels are about 3 times smaller. (b) The time series data for one voxel (indicated in black) for a hypothetical experiment. Consider a study in which a subject viewed pictures of faces. Every 30 seconds these pictures changed from having a fearful expression to a neutral expression. For the selected voxel, there might ensue a fMRI response that rose and fell in synchrony with the change in facial expression. Each point on the plot represents the fMRI signal value from the example voxel during the 3 seconds it took 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 yields a statistical value which is displayed as a color code for the voxel from which the time-series was obtained. This process is repeated for every voxel in the brain volume.

**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: (i) thresholding the map to only include those voxels with results that were unlikely to have arisen by chance; (ii) smoothing the data in space; (iii) displaying the areas of response on an anatomical image; and (iv) creating a digital reconstruction of the cortical surface.

**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.

**Glossary and structured bibliography**
<table>
<thead>
<tr>
<th>Initialism</th>
<th>Name</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPECT</td>
<td>Single-photon emission Computed Tomography</td>
<td></td>
</tr>
</tbody>
</table>

**Inference (forward and reverse)**


**Reviews of MVPA and resting state studies**
Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nature reviews Neuroscience, 8(9), 700–711. doi:10.1038/nrn2201


**Statistical errors in neuroimaging**

Vul, E, Harris, C, Winkielman, P & Pashler, H (2009) "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(3) 274-290.

Circular analysis in systems neuroscience – the dangers of double dipping


Simmons J., Nelson L., Simonsohn U. (in press) "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allow Presenting Anything as Significant", Psychological Science