

# Neuroimaging and Inferential Distance

Adina L. Roskies

Received: 29 November 2007 / Accepted: 16 December 2007 / Published online: 7 February 2008  
© Springer Science + Business Media B.V. 2007

**Abstract** Brain images are used both as scientific evidence and to illustrate the results of neuroimaging experiments. These images are apt to be viewed as photographs of brain activity, and in so viewing them people are prone to assume that they share the evidential characteristics of photographs. Photographs are epistemically compelling, and have a number of characteristics that underlie what I call their inferential proximity. Here I explore the aptness of the photography analogy, and argue that although neuroimaging does bear important similarities to photography, the details of the generation and analysis of neuroimages significantly complicate the relation of the image to the data. Neuroimages are not inferentially proximate, but their seeming so increases the potential for misinterpretation. This suggests caution in appealing to such images in the public domain.

**Keywords** Functional magnetic resonance imaging · Neuroimaging · Inferential distance

Functional magnetic resonance imaging (fMRI) has grown rapidly in popularity and scope since its

inception in the early 1990s. The past decade and a half has seen the publication of tens of thousands of scientific articles that employ fMRI as a research tool. A sizeable proportion of these studies have been devoted to exploring the neural basis of cognition [9, 10]; many others are technical and methodological papers aimed at improving cognitive neuroimaging techniques. While methods papers typically are concerned with improving the kind or quality of information provided by fMRI, and thus have epistemic consequences, little explicit attention has been paid to the epistemic status of neuroimages themselves, either actual or apparent. What is the relation between fMRI data and the conclusions of cognitive neuroimaging experiments? How are neuroimages perceived and interpreted? How should they be? Answers to these questions have implications for the role neuroimages play in the scientific process, and for how results are received and interpreted beyond the scientific community.

In everyday life, and in science, vision is one of our most important faculties for gaining knowledge about our world. Vision seems effortless and “direct”: we just see things as the objects they are, or see that something is the case.<sup>1</sup> Moreover, we usually think that what we see justifies us in our beliefs. That we consider vision of primary importance in establishing veridical beliefs about the world is evident in

---

A. L. Roskies (✉)  
Dartmouth College,  
Hanover, NH 03755, USA  
e-mail: adinar@mindspring.com

A. L. Roskies  
Department of Philosophy, University of Sydney,  
Sydney, Australia

---

<sup>1</sup> Although there is philosophical debate regarding whether visual perception is direct or mediated, the phenomenology of vision is of direct and unmediated access to the world.

common phrases like “seeing is believing”, and “I saw it with my own eyes”; and in the import of eyewitness testimony in court cases. Sometimes what we see is mediated by images, as in photographs or film. Through photographs we can see things with which we are not spatially or temporally collocated. For some time, photographs have been thought to be evidentially privileged: they are, in most cases, what we would see were we there. Photographic images have been typically granted an epistemic status almost as privileged as vision itself. Techniques of visualization have been extended beyond the world of objects and events that we can see “directly”, to include objects and events normally hidden from us. In this paper I will explore the extent to which the epistemic status of neuroimaging should be assimilated to that of common photographic techniques.

Neuroimages are epistemically compelling: they invite us to believe. Introspectively, the phenomenological impression of credibility is undeniable, and its effects recently confirmed by empirical studies. For example, Weisberg and colleagues (in press) explore the question of whether scientific explanations are judged to be improved when accompanied by neuroscientific data irrelevant to the point at issue. They found that both naïve subjects and those with some college-level neuroscience or psychology training rate scientific explanations as better when they include explanatorily irrelevant reference to neuroimaging data. Reassuringly, this biasing effect is not seen in experts [27]. These results support the notion that information from neuroimaging adds to the apparent epistemic weight of scientific reasoning, even when it fails to be epistemically relevant.

The reasons for this biasing effect are unclear. However, one possible explanation may be a mismatch between what I will call the *actual inferential distance* and the *apparent inferential distance* between neuroimages and that which they are meant to represent. In what follows I will explicate the notion of inferential distance further.

A central question in the philosophy of science regards how we should understand the epistemic status of scientific conclusions. Intuitively, some conclusions are more warranted than others, due to the quality of evidence, the support the evidence can provide for a particular claim, the availability of alternative explanations, and the epistemic status of assumptions that play a role in the reasoning that leads us to the claim in question. There has been little agreement about how to

unpack the notion of epistemic status. One might attempt to quantify the amount or degree of error to which a conclusion may be subject on the basis of error potentially introduced by the technical and inferential steps relied upon in the course of an experiment. Error analysis has been used in this regard. Alternatively, one might look to the range of states of affairs compatible with a conclusion based on observations, and thus the degree to which a scientific claim is equivocal. Information-theoretic approaches such as this have been championed by some [5]. Bayesian epistemology assesses the degree of confirmation or confidence we ought to have in our conclusion based on the degree of confidence we have in the steps involved in reaching it. One might also consider relevant the probability that a scientific statement will have to be jettisoned or revised if assumptions it is based on are overturned.<sup>2</sup> Each of these notions is logically independent of the others, although in fact they are often highly correlated. Regardless of which measure one prefers, or whether ultimately what matters is some combination of several of them, determining actual epistemic status will involve a characterization of the inferential steps that mediate between observations and the phenomena they purport to provide information about.

My goals in this paper are modest. Because there is no univocal way of characterizing inferential distance, I use the term primarily metaphorically, to gesture toward whatever epistemic dimensions best characterize the relation of scientific hypotheses to observations. My aim is to describe in rough terms the relation between the conclusions of cognitive neuroimaging studies, as represented by the neuroimage, and the data upon which they are based. In so doing, I will point out some of the assumptions and theories that ground the inferential steps operative in actual scientific practice, as well as the limitations that must attend our interpretations.<sup>3</sup> In the absence of a means of quantifying inferential distance, we can view it as a comparative measure: our

<sup>2</sup> Thanks to Paul Griffiths for this suggestion.

<sup>3</sup> While ideally a discussion of actual inferential distance would provide an objective metric by which to characterize inferential distance, such a task looks to be impossible in the absence of an inductive logic. In addition, since the task of breaking down a complex inference into logical atoms is underdetermined, there is no way of uniquely identifying inferential steps in a sequence of reasoning. Thus, actual inferential distance cannot simply be determined by counting or classifying inferential steps according to a clear taxonomy of inferential kinds. However, that does not mean that inferential distance cannot be characterized at all.

conclusions cannot be more secure than the assumptions they rely upon. Future studies may provide more concrete measures of neuroimaging's epistemic status.

This paper attempts to roughly characterize the difference between the actual and apparent (or perceived) inferential distance in neuroimaging, and to account for the mismatch. "Actual inferential distance" refers to the explicit forms of inference used in scientific practice, and "apparent inferential distance" indicates a more subjective measure, intended to characterize the support people are apt to think the evidence provides for a scientific conclusion.<sup>4</sup> I distinguish actual from apparent inferential distance in order to indicate that there can be a gap between a measure of how epistemically grounded a conclusion is as based upon actual methods and reasoning, and its perceived epistemic status, for instance as apprehended by the layperson. When actual and apparent inferential distance come apart, people are prone to assign an unwarranted epistemic status to claims. Neuroimaging seems to be a case in which this mismatch is pronounced. On the one hand, the phenomena that neuroimaging studies investigate are inferentially far removed from the images produced; on the other hand, the image format is familiar and accessible. Like photographs, brain images seem to be simple and straightforward: they appear to wear their content on their sleeves. A consequence is that images are liable to be mistakenly apprehended as inferentially proximate.

The scientific content of this paper will not be new to neuroscientists who employ imaging to investigate cognitive function, though it may be enlightening to those less familiar with the scientific underpinnings of functional neuroimaging. The complexities of the technology and interpretational difficulties are points of fertile discussion and sometimes fervent debate in the field, and no attempt will be made to weigh in on them here. However, these foci of debate may be good indicators of points at which the inferential distance of steps in the reasoning process is greatest. The advances here consist in situating these issues in the context of a discussion of the epistemic role that images play, and in contrasting that role with what is plausibly and popularly taken to be an analogous technique. This way of thinking

of the epistemic issues may have several uses. First, one might investigate whether there is a gap between the actual and the apparent inferential distance in the science itself, and whether and how that might affect scientific practice. Second, the discussion may be relevant for understanding the public consumption of neuroimages, and its potential social, political, and neuroethical implications. Finally, thinking of things in this way may have particular importance for public discourse, and may help inform scientists about how to better convey their results to the public via the popular media.

### Neuroimaging as Photography

I begin with a discussion of the perceived inferential distance of neuroimaging. Superficially, neuroimages are like photographs. They are generated by machines that many think of as brain-cameras, which produce representations about events in the mind and brain. These representations are images, typically brightly colored, that appear to represent a real object (a brain or brain slice) and some of its visual characteristics; the colored contours delineate regions of neural activity.<sup>5</sup> Thus, my contention is that people tend to appreciate neuroimages as if they were photographs of brain activity [21]. Plausibly, our familiarity with photography and with the use of photographic images in science, and the automatic way in which we interpret photographs, disposes us to impart to other images that bear superficial resemblance to photographs similar evidential characteristics.

Photography is a relatively direct means of representation. Photographs of scenes or objects represent the scene or object in virtue of a causal process by which light reflected by (or emitted by) the objects in the scene produce changes on a film, resulting in an image that visually resembles its object. Our experience of looking at photographs is that we see in them the objects or scenes they are of. Although seeing a photograph as a representation of a scene involves an inference, the inference is generally automatic, and mediated by implicit background knowledge and judgments of visual similarity. In most cases, relatively little explicit

<sup>4</sup> In this context, one might view actual inferential distance as providing some sort of normative epistemic measure, and apparent inferential distance as providing a descriptive measure.

<sup>5</sup> One might here construe photography widely: perhaps we may think neuroimages are more like weather maps, but I contend that people think of weather maps as if they were photographs of atmospheric conditions as well.

theoretical apparatus needs to be deployed in order to understand the content of a photograph; in this sense they are inferentially close. Perhaps in part in virtue of this inferential proximity, photographs play important evidential roles in our society, and their status as evidence has been explored by philosophers [3, 26]. Cohen and Meskin, for instance, state: “photographs seem to have a distinctive epistemic status as compared with other sorts of pictures. Unlike non-photographic depictions, photographs typically provide evidence specially about what they depict. Most significantly, the epistemically special character of photographs is revealed by this fact: we are inclined to trust them in a way that we are not inclined to trust even the most accurate of drawings and paintings.” A number of features have been adduced to explain their importance as evidence. Here I discuss three that appear prominently in some form in the literature: mimesis, counterfactual dependence, and theory-independence. In addition, I claim that photographs are revelatory: they display their contents.

Briefly, a representation is mimetic if it instantiates a represented property by means of the same property [13]. Photographs are visually mimetic in various respects, for many of the visual properties of the objects and scenes that are photographed are instantiated in the photograph, such as their colors, visual texture, patterns of shading, and their perspective-relative spatial relations and outlines shapes. It is in part because of their mimetic properties that some philosophers think we literally “see through” photographs to the objects they are of [26], or that we see in them their objects [28]. Cohen and Meskin argue that photographs are distinguished in the way they preserve visual information about their subjects, while discarding information relating the perceiver to the subject [3].

Counterfactual dependence refers to the fact that the content of photographs depends systematically on the scene photographed. If the scene were different, the photograph would be different (modulo differences not visible from the perspective of the camera). This counterfactual dependence is ensured by the mechanical methods of photography [3]. The epistemic value of photographs is related to this counterfactual dependence, and our background beliefs about photographs as a type. This isn’t the case with other types of depiction such as painting. Moreover, because of our background beliefs, the way in which the photograph depends on the scene photographed is intuitively clear,

usually enabling us easily to make inferences about the nature of what was photographed from the photograph. Theory-independence is related to this ease, as well as a characteristic that Walton has termed belief-independence [3, 26]. Photographs are belief-independent in that the beliefs of the photographer do not influence the content of the photograph by disrupting or diverting the causal process by which it carries information about its subject<sup>6</sup>; the photographer’s beliefs also need not be invoked in order to interpret the photograph. These qualities I subsume this under the rubric of theory-independence. Theory-independence is also related to a psychologically salient feature of photographs: they are revelatory—they wear their contents on their faces. The above qualities are by no means uniformly endorsed, and they may hold to greater or lesser degrees of particular photographs. In general, however, these features contribute to the evidential status of photographs.

Although not much attention has been devoted to it in philosophy, the digital age has led to a change in status of the actual inferential distance of photography. As more effective techniques of digitally manipulating photographs have been developed, photography’s evidential characteristics have been compromised. The rise of digital forensics reflects this change [6]. Insofar as this is so, photography’s evidential status should wane, although the psychological evidence reflecting such a change may lag well behind. This suggests that more attention to the contrast between actual and apparent inferential distance may be relevant in photography as well.<sup>7</sup> For the purposes of this paper, however, I will speak of photography as if it indeed were as inferentially proximate as most of the literature suggests. I will refer to the characteristics of mimesis, counterfactual dependence, and theory-independence that have been invoked to explain why photographs have historically served as a reliable and psychologically compelling form of evidence in order to explore the epistemic status of imaging. Bearing these features in mind, we now turn to a characterization of actual inferential distance in neuroimaging.

<sup>6</sup> Of course the photographer’s beliefs may affect what and how he chooses to photograph, but once the physical parameters of the photograph are established, the outcome is independent of belief or theory.

<sup>7</sup> Indeed, if people’s intuitive judgments about photography’s epistemic status should change as a result of the increased ease of photographic manipulation, those changes may also affect future apprehension of the epistemic status of neuroimaging.

## Exploring the Photograph Analogy

Inferential distance is a relational measure, so in explicating inferential distance we need to specify the relata. The inferential relata for any scientific program are interest-relative, as dictated by the scientific project. The goal of cognitive neuroscience is to map functional components of mind to neural structures. Long before imaging the brain was conceivable, William James identified the task of integrating mind and brain research:

A science of the mind must reduce ... complexities (of behavior) to their elements. A science of the brain must point out the functions of its elements. A science of the relations of mind and brain must show how the elementary ingredients of the former correspond to the elementary functions of the latter [11].

Cognitive neuroimaging studies seek to fulfill these demands by enabling us to see the neural activity underlying cognitive and/or behavioral tasks in brain images constructed from MR data. Observables in these neuroimaging experiments are thus measures of subject behavior (including, for example, button pushes, reaction times, errors), and the MR signal collected during task performance. What we'd like to infer from this data are, on the one hand, behaviorally relevant component processes (the "elements of behavior"), and on the other, the patterns of neural activity that cause the MR signal (the elementary functions of the brain). The foregoing suggests we conceive of the process of imaging as consisting in two largely separable projects, what I will call the causal stream and the functional stream. The relevant relata for specifying the inferential distance of the causal stream are thus neural activity and the MR image, and for the functional stream they are the behavioral/cognitive elements and the MR image.<sup>8</sup>

<sup>8</sup> This is actually a simplification, for in many cases the relevant relatum is not the MR image, but the MR data. To the extent that the data are preprocessed and indicate values (average intensity levels, statistical significance, etc.) I will speak of them as if they are characteristics of the image, even if they are never so displayed. It remains an empirical question to what extent the scientific conclusions of scientists are affected by their visual inspection of images, and to what extent they are based purely on numerical values. The answers will likely vary on a case to case basis.

The causal stream requires the understanding of two different relations: the relation between neural activity and raw data, and the relation between raw data and image. The inferential distance of the causal part of neuroimaging can be thought of the additive distance of these two measures. The functional stream is important for interpreting the functional relevance of neuroimaging data: in order to link neuroimaging data to cognition and behavior, an understanding of experimental design and the assumptions underlying it are important. The complexity of the neuroimaging process makes it impossible to go into full detail about either stream in a short paper. Here I can indicate roughly what inferential steps are involved.

## The Causal Stream

### Neural Activity to Raw Data

Neuroimaging is so often characterized as a tool for measuring brain activity that it frequently comes as a surprise to people that the magnetic resonance signal is insensitive to the electrical signals generated by neurons. The BOLD (blood oxygenation level dependent) signal, which is by far the most widely used paradigm for functional neuroimaging, is not a direct measure of neural firing, or any immediate causal consequence of neuronal activity.<sup>9</sup> Instead, the MR signal is a measure of a quantity that changes as a downstream consequence of neural activity, and bears to it a complex relation. Sometimes the BOLD signal is described as measuring changes in blood flow that correlate with changes in neural activity. Even this is inaccurate, for BOLD imaging is also relatively insensitive to blood flow. As its name suggests, the BOLD MR signal is dependent on the blood oxygen-

<sup>9</sup> For the purposes of this paper I focus on BOLD echo-planar imaging, which dominates the cognitive neuroimaging literature. Those familiar with MR imaging will recognize that I have said nothing about this or other pulse sequences. Pulse sequences can be tailored to make the MR image more or less sensitive to different factors. This imparts MR imaging with a great degree of flexibility. Details about inferential distance will depend upon what sorts of imaging parameters are chosen. A discussion of this is beyond the scope of this paper, but relevant details can be found in Buxton [2].



ation level. It is a direct measurement of the dephasing of spins of water molecules in blood, caused by local changes in magnetic susceptibility. The magnetic susceptibility of water (in this case, water in the blood) is affected by the local concentration of deoxygenated hemoglobin. Increased levels of deoxyhemoglobin reduces the BOLD signal; reduced concentrations increase it. Neural activity is linked to susceptibility changes in the following way: it increases local metabolic demands, and these are compensated for by an increase in local blood flow and capillary volume. Although active neurons consume oxygen and thus increase the amount of deoxyhemoglobin in the blood, the increased supply of oxygenated blood outstrips the oxygen consumed, resulting in a net increase in the concentration of oxygenated hemoglobin, and a concomitant increase in the MR signal. Thus, a constellation of factors, some with opposing effects, are altered in the wake of neural activity.

The correlation between neural activity and increased MR signal is empirically well-confirmed [14, 15, 16], but the quantitative relationship between changes in magnetic susceptibility and neural activity is not well understood. The underlying physiological mechanisms and their quantitative relationships remain phenomena of scientific inquiry. Our confidence in our ability to relate changes in the fMRI signal to neural activity is dependent upon empirical generalization and corroboration by other methods that more directly measure metabolic demands and blood flow [2, 17, 23].

When considering the steps that allow us to draw conclusions about neural activity from imaging data, it becomes evident that a large number of neural states could conceivably give rise to the same signal, and furthermore, we currently lack means of ruling out many of those possibilities as improbable.<sup>10</sup> For example, with so many factors contributing to susceptibility effects, it is likely that the same net magnetic susceptibility changes can result from different combinations of changes in blood flow, volume, and oxygen extraction. If these correspond to different macroscopic activity patterns, then the signal is univocal with respect to this class of activity

patterns. Moreover, the spatial and temporal resolution of the technique is such that the information it provides is orders of magnitude coarser than the neural firings we suppose are at the appropriate grain for understanding many aspects of cognition and behavior.

In addition, when making claims for the functional relevance of changes in activity it is important to bear in mind that evidence suggests that neuroimaging data reflects local synaptic activity, irrespective of whether it is excitatory or inhibitory. Thus, although the functional implications of these two types of synaptic transmission are quite different, the MR signal does not allow us to distinguish them. Moreover, the MR signal may reflect subthreshold neural activity and modulatory inputs from distant areas in addition to neural firing. It is not clear whether it could distinguish other physiologically relevant parameters, such as large changes in the firing rate of a few neurons, small changes in the firing rates of many neurons, or changes in neuronal synchrony in the absence of changes in mean firing rates. Finally, because the MR signal is not a direct measure of neural activity, the signal may be temporally and even spatially out of register with the activity changes that are ultimately the phenomenon of interest. For example, the vascular response to neural activity occurs with a delay of several seconds, and has a timecourse significantly broader than the timecourse of neural activity that it reflects. While this might appear to be a problem, studies have found that the hemodynamic response is relatively stereotypical, so the temporal factors that distinguish the blood flow from the neural signals can be well accounted for during statistical modeling [7, 12]. In addition, both the geometry of the vasculature and the particular pulse sequences used in generating the MR data can sometimes lead to artifacts in which signal increases appear to come from regions at some remove from the active neural tissue. All these possible sources of the MR signal illustrate the inferential distance between the signal and the scientific conclusions we draw from it; in order to bridge the gap we have to make a variety of assumptions about the temporal and spatial relationships between blood flow changes and neural activity. To the extent that we can expect these issues to manifest equally across conditions, their influence will be effectively negated with direct contrast analyses. While these assumptions are corroborated

<sup>10</sup> This highlights the fact that inferential distance is a measure that can change given our state of knowledge. The more refined or well-confirmed our theories upon which inferences are based, the less distance there is.

by experimental findings in some paradigms, we must assume that the relations that have been established hold in diverse brain areas and independent of task.

Thus, functional MRI doesn't directly measure neural activity, but rather reflects the net effect of other physiological factors that are causally related in a rather complex way to downstream consequences of neural activity. From this signal, neural activity is inferred to have increased or decreased. Inferences are grounded by a rough theoretical conception of the physiological changes that are influenced by activity and that themselves influence the MR signal; the rough theory is anchored at points by empirical results from other types of studies that more directly measure the relationship of blood flow, glucose and oxygen metabolism, and neural activity [20]. While the general relationship is well-established, the inferences that are justified by our current knowledge are rather coarse. Warrant for making fine-grained inferences about local neural activity is much more limited.

How does the actual inferential distance so characterized compare to the perceived inferential distance? Although brain images may be thought to be mechanically derived direct measures of a state of affairs just as photographs are, we can realize here relevant departures from the features thought to ground the inferential proximity of photographs. Consider the relation of counterfactual dependence. While neuroimaging is counterfactually dependent upon the blood oxygenation characteristics of the brain, accounting for its suitability as a scientific tool, the above discussion elaborated some ways in which we lack a clear understanding of precisely what the dependence relation is between signal and neural activity. Because it is likely that many different kinds or patterns of activity can result in the same fMRI signal, our ability to infer relevant characteristics of the activity is limited. This is much less true for photography, where the counterfactual relation is relatively transparent and in general the possibilities are intuitive and well-delineated.<sup>11</sup> In addition, the foregoing discussion emphasizes the extent to which the interpretation of the functional relevance of the MR signal is theory-dependent. Certain theoretical

relations must hold between the MR signal and neural activity in order for us to be warranted in interpreting the signal as reflecting activity. The limits of our interpretation are governed by the nature of these relations.

### Raw Data to Image

So far, I have discussed the inferential links that allow us to draw conclusions about brain activity from the MR signal. However, it is not the raw signal that is typically analyzed in order to make claims about neural activity, but rather pooled and processed data from multiple trials, and often, multiple subjects. This pooling is necessary, for the signal to noise ratio for neuroimaging is quite low; data from multiple scans is averaged in an attempt to minimize effects of random noise. Finally, the data is often represented as an image, and in constructing such a representation certain conventional choices are made that can dramatically affect the visual appearance of the image.

Clues to the mismatch between the actual and apparent inferential proximity of neuroimages may be gleaned from consideration of the sorts of transformations in format that data undergoes in neuroimaging. In photography, negatives and prints are both isomorphic to a two-dimensional projection of their subjects. In contrast, the format of data from functional neuroimaging experiments is not originally in the format of an image, but rather in terms of data structures that encode numerical values of phase and frequency-dependent signal intensity collected in an abstract framework called "k-space". Visual representations of data in k-space bear no visual resemblance to images of brains. These data are transformed to spatial values of signal intensity with a Fourier transform, resulting in an image that looks roughly brain-like. This analytical transformation doesn't alter the information encoded in the data, and thus may be taken not to introduce any inferential distance, in that it introduces no error or range of possible causes. However, the radical transformations in format are indication of the indirectness of the technique relative to photography. The fact that these transformations are completely invisible to the consumer of the image implies that neuroimages are not revelatory, and illustrates the inaptness of the imposition of an epistemic framework paradigmatic of photography.

Technical debates about proper statistical and data analysis abound in the literature, ranging from

<sup>11</sup> Again, these distinctions are quantitative; in some cases even traditional photography can mislead (for example, photographs of the Ames Room). The arguments here are meant to reflect more of a difference in degree than in kind.

questions about how to correct for multiple comparisons to whether analysis should be hypothesis-driven or whether brute-force statistics suffice. This paper will not take on these technical questions, but notes their prevalence in order to underscore the point that to the extent that there is disagreement about the proper way to analyze the data, we can presume that different background assumptions are called into play, and the different analytical approaches may affect the actual inferential distance of an imaging study. This supposition is confirmed by the fact that choosing different ways to process the data may affect results. For example, which voxels are identified as statistically significant can be affected by experimenter's choices, such as whether the raw data is spatially filtered or not. To the extent that we know that the same raw data can produce different results depending upon reasonable choices made about processing, we know there is not a one-to-one correspondence between data and image. This is true of any form of data reduction, but it emphasizes the need to know what analysis procedures are used in the determination of the results of a study, and the importance of understanding their effects. This illustrates again the differences between neuroimaging and photography with respect to the nature of the counterfactual dependence relation. It also suggests that images are not revelatory in that the choices that are made are not visible in or recoverable from the image itself, and thus the proper interpretation cannot be discerned on the basis of the image alone.

Averaging data across different subjects introduces further complexities. There is considerable variation in individual brain structure as well as in structure/function relationships. Variation in structure is addressed by warping the individual data in order to register the brains in a common space, but we lack independent means for assessing the adequacy of the warping algorithms [4, 8, 12]. While warping may address complexities introduced because of anatomical variability, functional/anatomical variability remains an issue. Dispersion in the location of functionally active areas can reduce the statistical significance of the data and/or produce resultant images that show activated regions that don't correspond to any actual functionally-delineated brain region. Regardless of how variability is dealt with, it is interesting to note that although what results from this averaging procedure is a brain image, it is not an image of a

brain. Brain images are best thought of as generalizations, not particulars. In this respect, brain images are more akin to scientific diagrams or schematics than to photographs [18]. While generality is not a feature unique to scientific images, it is a feature foreign to photography. Photographs are always of particulars. They thus support existential inferences: from a photograph of a scene with a dog in it, one can conclude that a dog was there. This does not hold true for neuroimages, for there may be no particular that corresponds with the generalization. In this way, too, they are unlike photographs.

The final point to make regards the conventions of image construction. Brain images typically indicate activation levels or statistical significance by overlaying colored areas on a greyscale structural image. The representation of the analyzed data relies upon a number of conventional choices that can affect both the final informational content of the image, and its visual appearance. Thresholding, for example, reduces the informational content of the image, and displays only information reaching or exceeding the threshold chosen by the experimenter. Other methods discard activated areas that fail to contain a minimal number of statistically significant contiguous voxels. These methods produce images that typically look cleaner than ones that merely compute statistical significance, and the impression of cleanliness may affect people's apprehension of the image and their appraisal of its epistemic status. Color manipulations may also have psychological impact. Some experimenters use color gradations to indicate relative levels of activity, while others use color gradations to indicate relative levels of statistical significance. The very same visual image can thus represent quite different, albeit related, contents. Again, the nature of these contents is not discernable from the image alone. Finally, even the choice of color scheme, which is purely conventional and does not affect the information content of the image, can affect the way an image looks to an observer. Blues are cool colors, while reds are hot, and thus often blues are used to represent deactivations, and reds to represent activations. It is an empirical question whether purely arbitrary color assignments or other conventional decisions have any psychological effects on the way such images are apprehended or interpreted.

The fact that some conventional decisions affect the visual appearance of the image without changing



its content implies that images are not mimetic in the way photographs are. That neuroimaging is not mimetic is underscored when we recognize that the properties which they are meant to represent, location and level of neural activity, are not visual properties, though they can be visually represented. Understanding what is represented requires a knowledge of the choices made in analysis and image construction, and is therefore theory-dependent and non-revelatory. In considering the causal stream, then, we have seen ways in which imaging differs from photography in terms of mimesis, counterfactual dependence, theory-independence, and revelation. We now turn to consider that other part of functional neuroimaging which allows us to link the image data to cognitive function.

### The Functional Stream

Despite the popularity of phrenological cartoons showing brain areas for loyalty, humor, or remembering where one put one's car keys, it has long been recognized that complex cognitive functions are not atomistically located in the brain. For instance, relatively early in the development of functional neuroimaging techniques, Petersen and Fiez wrote:

Elementary operations, defined on the basis of information processing analyses of task performance, are localized in different regions of the brain. Because many such elementary operations are involved in any cognitive task, a set of distributed functional areas must be orchestrated in the performance of even simple cognitive tasks. A functional area of the brain is not a task area: there is no "tennis forehand area" to be discovered. Likewise, no area of the brain is devoted to a very complex function; "attention" or "language" is not localized in a particular Brodmann area or lobe. Any task or "function" utilizes a complex and distributed set of brain areas. [19]

The goal in neuroimaging is to assign function or processes to brain regions. One might suppose that a neuroimaging experiment starts with a subject getting into a scanner, but that is far from the case. Many critical factors in a neuroimaging experiment depend upon task analysis and appropriate experimental design.

Raw MR images look quite unlike the dark images with colorful splotches that grace the pages of popular magazines and scientific journals. MR provides information about the global levels of magnetic susceptibility in blood and tissue, so images corresponding to unanalyzed data show activation throughout the brain. The bulk of metabolic activity and blood flow is devoted to supporting ongoing, non-task-related brain function. Little can be discerned from visual inspection of these images resulting from scans during performance of different tasks. In fact, only a small fraction of what is recorded by neuroimaging corresponds to task-related changes [20]. If so, how do neuroimages provide information about the functional architecture of cognition? A number of methods have been developed to help identify regions involved in different tasks. These involve comparing MR data across different task conditions. Whether one uses standard subtractive techniques or more recent event-related designs, having a functional decomposition of a task is essential to good experimental design as well as interpretation.

A functional decomposition is a working theory of what sorts of processes are involved in task performance. In subtraction, one typically tries to identify task components by subtracting from the data collected during performance of a task of interest the data from another task involving many of the same components, ideally save one. Thus, one might believe of a memory task that it involves seeing a stimulus, holding it in working memory, seeing a second stimulus, judging whether the two are the same, and pressing one button if they are, and another if they are not. In order to try to isolate the region(s) involved in working memory one would then want to design a similar task in which the subject sees another stimulus but is not prompted to remember it, sees a second stimulus, makes a judgment, and presses one of two buttons dependent on the judgment. If all worked according to plan, subtracting the scan from the non-memory condition from the one generated by the task of interest, one would isolate the brain region(s) involved in holding something in working memory. This is an ideal characterization: despite widespread misunderstanding (e.g. [24, 25]), it is not necessary to isolate a single task component in order to learn something about cognition. Nonetheless, the interpretability of

the data will depend upon having an understanding of the way in which complex tasks functionally coincide and ways in which they diverge in their demands.

A consequence of relying on these methods is that experimental design and outcome is heavily dependent on one's theoretical commitments to functional architecture. If, for example one was mistaken in one's theory of the task components, one would misidentify the function of the region(s) resulting from a subtraction. For instance, with respect to the preceding example, if there were more task components in the first scan unmatched by corresponding ones in the contrast, more regions would result from the subtraction than the ones actually subserving memory; if more were involved in the contrast task, deactivations would result; if the task analysis were largely mistaken, function could not properly be identified (though one could still conclude that the differences in performance of the two tasks involved the regions resulting from the subtraction).

Similarly, when analyzing results of event-related designs, one must identify the types of events and the temporal profile of their occurrence during a sequence of task executions. This identification of event-types and their time-courses is essential to the workings of the general linear model (GLM) that tries to maximize the amount of signal fluctuations accounted for by fitting the data with different event-type time courses. Here too, then, misidentification of event-type or the appropriate occurrences can lead to errors in functional attribution.

There is much to be said about the issues surrounding functional decomposition. Many worry that our hypothesized functional decompositions reflect a naïve and inaccurate understanding of the cognitive processes underlying the performance of complex tasks, and worry that such inaccuracies would invalidate neuroimaging [25]. If indeed our understanding of tasks was massively in error and remained so, imaging would prove to be a problematic technique for investigating cognitive function. However, ontology is rarely if ever handed to us on a silver platter in any science, yet in general we maintain some confidence that science can provide us better and more accurate ways of looking at the world. Certainly neuroimaging should not be seen as working in vacuum. It should and does draw upon various other disciplines for theoretical insight and empirical corroboration. Data and theories from cognitive

psychology provide a theoretical background with which to work, and results from anatomy, comparative neurobiology and neuropsychology can also help to constrain theories of brain function. Functional triangulation and corroboration from other brain imaging studies are also essential to the development and refinement of our psychological theories about functional decomposition [1, 22]. While some are apt to see the glass half-empty, those more optimistic about the scientific endeavor are inclined to point to the practical success of scientific bootstrapping in order to see the glass as half-full. Any way you look at the glass, however, it is shot through with theoretical commitments. One cannot discern its contents purely by visual inspection.

What is the relevance of this discussion for the issue of inferential distance? It seems that the inferential distance for functional attributions is going to be a function of how well-established various functional components of tasks are. In task analysis, sensory and motor components are relatively trivial to identify, but identifying more cognitive components is more speculative. Their plausibility is tested by experimental manipulations designed to target a specific functional component, for example by changing task load or creating interference. Understanding the functional or computational components of various cognitive tasks is a primary focus of cognitive psychology, and has also been addressed by other fields such as computer science and linguistics. Just how well established a particular component is will vary, and judgments about its nature will depend upon the evidence in its favor, the availability of alternative accounts, and its ability to unify different task accounts. In the case of neuroimaging, the direction of explanation need not be entirely from the side of cognitive psychology. We can also rely upon convergent evidence from the neurosciences to help inform our functional decompositions. Neurophysiology in nonhuman animals and lesion studies and other noninvasive techniques in humans can provide independent confirmation for our hypotheses, increasing our confidence in our functional specifications. However in general, the farther things get from the sensory and motor peripheries, the more speculative the inferences that support functional attribution to brain areas.

The above concerns actual inferential distance. Other features of the functional stream also affect

apparent inferential distance. For example, in neuroimaging the data generated are always dependent upon the chosen tasks, and subtle differences in task can matter to the resulting image<sup>12</sup>. With subtractive designs, the data displayed in the image are contrastive, and it is impossible to read off the image itself what the tasks are that contribute to its content. In particular, the baseline or contrast task is effectively invisible, contributing to what isn't seen or what is seen as deactivation. The image that results from an experiment, therefore, is heavily dependent upon both the task of interest and upon what contrast task is chosen, yet the contrast tasks are often overlooked by naïve consumers of functional images as irrelevant. Because the images don't explicitly carry the information about what the relevant tasks are for interpreting the image, they are not revelatory. Moreover, the choice of tasks and interpretation depend upon the experimenter's hypothesized functional decomposition. Thus, the content of a neuroimage is highly dependent upon the tasks used in its generation, and an understanding of those tasks is theory-laden. Again, when comparing the functional stream to our characterization of photography, we see that neuroimaging fails to share many features of photographs that make them inferentially proximate. In contrast to photography, interpreting the functional significance of neuroimages is highly theory-dependent.

Cognitive function itself is not represented in images, so trivially images are not mimetic with respect to their functional significance. All imaging experiments rely upon the consumer to correlate functional descriptions with imaging results. In general, scientists tend to be more aware of the importance of other information presented alongside neuroimages, such as figure legends and long discursive accounts of experimental design and methods that are part of every scientific paper. However, rarely is adequate information presented in the popular venues in which neuroimages appear, and their visually arresting qualities and seeming revelatory nature makes detailed information seem redundant or unnecessary. Often, it seems, a short caption will do. The foregoing discussion is meant to underline how misleading this can be.

<sup>12</sup> Indeed, even differences as subtle as the word frequency of lexical stimuli have been shown to have noticeable effects on imaging outcomes.

Of the two streams, the functional stream is more problematic. The causal stream involves assumptions for which our evidence is at many places incomplete or inconclusive, but this can forseably be remedied with further work. The assumptions that are invoked in the functional stream are more global and more difficult to independently verify, and at some level involve the entire conceptual framework upon which neuroimaging is based. As mentioned before, this is a problem faced by all the natural sciences—nature does not come with a manual delineating her joints. While this by no means invalidates the technique or undermines the inferences about functional anatomy which it enables, a proper appreciation of the epistemic status of neuroimages requires acknowledging these dependencies.

### The Illusion of Inferential Proximity

Visualization is an important tool for gaining and conveying scientific understanding. Images provide ways of representing a manifold of information at once, and we apprehend some relationships better when represented visually than when verbally described. However, the familiarity we have with the visual images we are best acquainted with—photographic images—and the epistemic status we accord to them does not always easily translate to images in science.

Here I have compared the actual inferential distance in neuroimaging with the perceived distance, on the hypothesis that people tend to view neuroimages as if they were photographs of brain activity. Superficially, they share many qualities with other types of image that we interpret unproblematically. However, I have argued that unlike photographs, neuroimages are not mimetic, for visual characteristics of brain activity are not faithfully replicated in the image. Since neural activity has no relevant visual properties, the conventions of the brain image are representational translations of certain nonvisual aspects of properties related to neural activity, namely magnetic properties of the water of the brain. Moreover, the raw data generated by the scanner does not look like a brain at all: it bears no relation to a negative in the photographic process, for its visual characteristics are not isomorphic to the image generated. Both photography and imaging are useful evidential techniques because of their counterfactual

dependence on aspects of their objects of investigation. Neuroimaging is especially scientifically valuable because it provides evidence about brain activity in awake behaving humans. However, the counterfactual dependencies in neuroimaging are often mediated by provisional or incomplete theories, and many details are still poorly understood, whereas in photography the relations are largely unproblematic. The generation and interpretation of neuroimages is highly theory-dependent, again distinguishing imaging from photography. Interpretation of neuroimages takes place within a framework largely established by means of the technique itself. Finally, the ways in which neuroimaging differs from photography are not generally apparent in the images themselves—the images are not revelatory with respect to the considerable degree of background information necessary in order to understand how to interpret neuroimages. All of these factors combine to create an illusion of inferential proximity: Neuroimages are inferentially distant from brain activity, yet they appear not to be. For this reason, caution is needed in interpreting them; moreover, the inferential distance between image and neural activity makes their use in public discourse especially fraught and risky.

**Acknowledgements** This work was supported in part by a grant from the Australian Research Council. Thanks to Mark Colyvan, Richard Joyce, Neil Levy and C.C. Wood, for helpful comments on earlier drafts.

## References

- Bechtel, W. 2002. Aligning multiple research techniques in cognitive neuroscience: Why is it important? *Philosophy of Science* 69: S48–S58.
- Buxton, R.B. 2002. *Introduction to functional magnetic resonance imaging: Principles and techniques*. Cambridge: Cambridge University Press.
- Cohen, J., and A. Meskin. 2004. On the epistemic value of photographs. *Journal of Aesthetics and Art Criticism* 62(2): 197–210.
- Crivello, F., T. Schormann, N. Tzourio-Mazoyer, P.E. Roland, K. Zilles, and B.M. Mazoyer. 2002. Comparison of spatial normalization procedures and their impact on functional maps. *Human Brain Mapping* 16(4): 228–250.
- Dretske, F. 1981. *Knowledge and the flow of information*. Cambridge: MIT Press.
- Farid, H. 2006. Digital doctoring: How to tell the real from the fake. *Significance* 3(4): 162–166.
- Friston, K.J., C.D. Frith, R. Turner, and R.S.J. Frackowiak. 1995. Characterizing Evoked Hemodynamics with fMRI. *NeuroImage* 2(2, Part 1): 157–165.
- Gispert, J.D., J. Pascau, S. Reig, R. Martinez-Lazaro, V. Molina, P. Garcia-Barreno, et al. 2003. Influence of the normalization template on the outcome of statistical parametric mapping of PET scans. *NeuroImage* 19(3): 601–612.
- Illes, J., M. Kirschen, and J.D.E. Gabrieli. 2003. From neuroimaging to neuroethics. *Nature Neuroscience* 6: 205.
- Illes, J., and E. Racine. 2005. Imaging or imagining?: A neuroethics challenge informed by genetics. *The American Journal of Bioethics* 5(2): 5–18.
- James, W. 1890. *Principles of Psychology*, vol. 1. New York: Henry-Holt and Co.
- Kruggel, F., and D. Yves von Cramon. 1999. Alignment of magnetic-resonance brain datasets with the stereotactical coordinate system. *Medical Image Analysis* 3(2): 175–185.
- Kulvicki, J. 2006. *On images: Their structure and content*. New York: Oxford University Press.
- Logothetis, N.K. 2003. The underpinnings of the BOLD functional magnetic resonance imaging signal. *Journal of Neuroscience* 23(10): 3963–3971.
- Logothetis, N.K., and J. Pfeuffer. 2004. On the nature of the BOLD fMRI contrast mechanism. *Magnetic Resonance Imaging* 22(10): 1517–1531.
- Logothetis, N.K., and B.A. Wandell. 2004. Interpreting the BOLD signal. *Annual Review of Physiology* 66: 735–769.
- Mukamel, R., H. Gelbard, A. Arieli, U. Hasson, I. Fried, and R. Malach. 2005. Coupling Between Neuronal Firing, Field Potentials, and fMRI in Human Auditory Cortex. *Science* 309(5736): 951–954.
- Perini, L. 2005. Explanation in two dimensions: Diagrams and biological explanation. *Biology and Philosophy* 20(2): 257–269.
- Petersen, S.E., and J.A. Fiez. 1993. The processing of single words studied with positron emission tomography. *Annual Review of Neuroscience* 16: 509–530.
- Raichle, M.E., and M.A. Mintun. 2006. Brain work and brain imaging. *Annual Review of Neuroscience* 29: 449–476.
- Roskies, A.L. (2008). Is neuroimaging like photography? *Philosophy of Science* (in press).
- Roskies, A.L., and S.E. Petersen. 2001. Visualizing human brain function. In *Frontiers of Life: The Intelligent Systems*, eds. E. Bizzi, P. Calissano, and V. Volterra, III: 87–109. New York: Academic.
- Shmuel, A., M. Augath, A. Oeltermann, and N.K. Logothetis. 2006. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nature Neuroscience* 9(4): 569–577.
- Uttal, W. 2001. *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge: MIT Press.
- Van Orden, G.C., and K.R. Paap. 1997. Functional neuro-images fail to discover pieces of mind in parts of the brain. *Philosophy of Science* 64(Supplement): S85–S94.
- Walton, K. 1984. Transparent pictures: on the nature of photographic realism. *Critical Inquiry* 11: 246–276.
- Weisberg, D.S., F.C. Keil, J. Goodstein, E. Rawson, and J.R. Gray. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience* (in press).
- Wollheim, R. 1987. *Painting as an Art*. Princeton: Princeton University Press.