Philosophy Compass 5/2 (2010): 186–198, 10.1111/j.1747-9991.2009.00275.x

Philosophical Issues in Neuroimaging

Colin Klein*
University of Illinois

Abstract

Functional neuroimaging (NI) technologies like Positron Emission Tomography and functional Magnetic Resonance Imaging (fMRI) have revolutionized neuroscience, and provide crucial tools to link cognitive psychology and traditional neuroscientific models. A growing discipline of 'neurophilosophy' brings fMRI evidence to bear on traditional philosophical issues such as weakness of will, moral psychology, rational choice, social interaction, free will, and consciousness. NI has also attracted critical attention from psychologists and from philosophers of science. I review debates over the evidential status of fMRI, including the differences between brain scans and ordinary images, the legitimacy of forward inference and reverse inference, and deductive versus probabilistic accounts of NI evidence. I conclude with a discussion of fMRI as exploratory rather than confirmatory evidence, linking this debate to the growing literature on cognitive ontology.

1. *Introduction*

1.1. THE TECHNOLOGY OF NEUROIMAGING

Thinking depends on brain activity, but not all parts of the brain are on an equal footing. When we think about different things, different parts of our brains are active. This *functional specialization* of brain regions has profound consequences for both cognitive psychology and neuroscience. In principle, it should allow us to use psychological theories to learn something about the organization of the brain, and *vice versa*. Until relatively recently, though, it was hard for scientists to exploit functional specialization. Observation of brain-damaged patients and recordings from single neurons provided some useful evidence, but these techniques have obvious limitations.

Neuroimaging (NI) has changed that. 'Neuroimaging' refers to a collection of techniques that allow scientists to look at the whole brain as it works by detecting metabolic changes caused by increased neural activity. The two most popular NI techniques are Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI). PET is sensitive to the uptake of radioactively labeled tracers in the brain; fMRI is sensitive to the small magnetic differences between oxygenated and deoxygenated blood (see Savoy 2001 and Raichle 2009 for more detail).

By constantly measuring changes in metabolic responses as subjects perform cognitive tasks, NI can provide a fast, safe, non-invasive, high-resolution look at the working brain. This combination of features has revolutionized cognitive neuroscience. Many philosophers have become interested in NI as well, as it promises to shed light on traditional philosophical issues like moral deliberation (Greene et al. 2001, 2004; Sinnott-Armstrong 2008), rational choice (Sanfey et al. 2003, 2006; Camerer et al. 2005), ethical and social interaction (Farah 2002; Adolphs 2003; Lieberman 2007), free will (Roskies 2006), weakness of will (Hare et al. 2009), and consciousness (Lloyd 2002; Hohwy 2007).

Despite its popularity, NI is controversial. This paper will review some of the methodological issues surrounding NI. In particular, I'll be interested in the question 'What kind of evidence does NI provide, and how does it provide it?' Though NI is widely used, neither question has an obvious answer.

1.2. THE PROBLEM WITH PICTURES

Neuroimaging's evidential status may seem unproblematic. By now, most people are familiar with *neuroimages*: the striking colorized pictures that accompany fMRI studies. The fact that it's called 'neuroimaging' implies that the 'imaging' part is particularly important.

This suggests a simple answer to our question: the evidence of NI is just the pictures of the working brain that it produces. Popular accounts of imaging data reinforce this pictorial view (McCabe and Castel 2008). Some authors describe fMRI as being 'to brain science what Galileo's telescope was to astronomy' (Jacobs 2009, 121); others talk of opening up the 'black box' of the mind to direct observation (Camerer et al. 2005).

The evidence of NI cannot be like ordinary pictures, however. Neuroimages differ from ordinary images at least two important ways.

First, as Savoy notes, particular neuroimages are laden with theoretical assumptions (2005; Roskies 2007, 867). Many fMRI experiments use what is known as a 'subtractive' design. Subtractive designs look for *differences* in brain activity between two tasks, to highlight brain areas that show activation specific to one of the tasks performed.

So, e.g., if I want to find the part of the brain responsible for thinking about money, I might scan subjects as they balance their checkbook and while performing equivalently difficult (but not money-related) arithmetic problems. The resulting image would show areas that were active only in the checkbook task. I might interpret this as showing the money-deliberation part of the brain. However, this image will be inherently *theory-laden*: it cannot be interpreted without knowing the specific tasks performed and the assumptions about cognition that the experimental design embodies. Further, if you deny that balancing your checkbook requires deliberating about money, then you have no reason to think that the image shows money-deliberation brain areas. But the *image* does not tell you any of these things.

Second, neuroimages do not present raw data on brain activity. The collected signal is small, noisy, and temporally complex. Neuroimages show the results of statistical tests performed on this noisy signal. In practice, this means that neuroimages do not show brain activity *per se*: rather, they show areas where one can confidently assert that some activity happened. Similarly, the color intensities in a neuroimage do not show strength of activity, but rather a statistical measure of just how strong our confidence in the presence of activity should be. This is a very different sort of information than we are used to receiving from pictures. For example, the statistical nature of a neuroimage means that one cannot infer the *lack* of activation from a dark region, or even that a brightly lit 'active' region was more active than a dark 'inactive' region (Huettel et al. 2004, 246). Contrast this with ordinary pictures and maps, where the absence of a marker allows one to infer the absence of the property that the map depicts (Rescorla 2009).

For these reasons (and several others), Roskies concludes that neuroimages lack the kind of causal and counterfactual connection to their source that paradigmatic observational tools like photographs possess (2007 §3). Whether these differences are enough to cause problems is an open question – many of the same complaints could be made about anatomical MRIs, which provoke little skepticism (though see Joyce 2008). Nevertheless, it is clear that neuroimages have a more complicated status than pictures.

1.3. TWO PROBLEMS

In the remainder of the article, I will review two issues concerning the evidential status of neuroimages. First, I will look at the use of NI to infer brain function given a cognitive theory. Second, I will look at the (more controversial) use of fMRI data to confirm or disconfirm cognitive theories themselves. These two kinds of inference are often referred to as *forward* and *reverse* inference, respectively (Poldrack 2006, 59). This terminology should be taken with a grain of salt – the two processes are often intertwined in practice, and reverse inference is only one way to test cognitive theories.

I will argue in Section 2 that 'forward' inference is relatively unproblematic in principle (thought tricky in practice), whereas in Section 3, I will argue that 'reverse' inference is both weak and conceptually fraught. I will conclude in Section 4, by suggesting an alternative way in which NI data might be relevant to psychological theories.

2. From Psychology to Brain

2.1. WHAT SORT OF EVIDENCE IS REQUIRED?

Suppose we have a well-supported cognitive theory, and we want to know how the brain works to instantiate the processes our theory studies. Neuroimages often show facts about differential activation: that a region was more active in one condition than another. Is this the only, or the best, evidence about the organization of the brain?

Perhaps not. Consider the following case, outlined by Friston et al. A PET study showed that a particular region *R* of inferotemporal cortex was more active when subjects had to detect objects rather than shapes, but was not further activated when subjects had to additionally name shapes along with detecting them (1996, 99). One might interpret this to show that *R* is important for object detection but not object naming. A closer look at the data, however, showed a statistically significant interaction between the tasks. The amount of activity because of object recognition was greater when object recognition was required (1996, 102). It would thus be a mistake to conclude that naming is *irrelevant* to what *R* does. In technical terms, this shows a violation of the *pure insertion hypothesis*, which claims that activity attributable to a task remains unchanged when we add additional distinct cognitive tasks. The pure insertion hypothesis must be true for subtractive methods to work properly; as Friston et al. show, its truth is not guaranteed in any particular case.

This shows that simple subtractive designs might overlook important facts about functional organization. Although brain regions are functionally specialized, they also interact extensively, and rely on that interaction for proper functioning. How to tease out functional relationships in specific cases is thus an area of open research. Subtraction, interaction, and conjunction (Price and Friston 1997) have all been proposed as keys to functional organization. It is likely that complex combinations of conditions will be required to confidently assert the functional importance of an area (Henson 2005, 2006, figure 1f).

A second important debate concerns the relative importance of quantitative data. Facts about differential activity are *qualitative* facts: they tell you that a difference in activity occurred, but do not give you any indication of the magnitude of that change. Quantitative facts are important in science, however, (Tukey 1969, 86; Meehl 1978; Klein, forthcoming §5). If my theory correctly predicts that the stock market will rise next year, you should be a little impressed; far more impressive (and useful) though would be a theory that told you *how much* it will rise.

Though often reported, quantitative magnitudes in NI are difficult to interpret on their own. While there is a rough mapping between the BOLD response and the underlying vigor of neural activity, there is no mapping between the amount of metabolic activity of a brain area and the *functional importance* of the area (Buxton 2002, 423; Nair 2005, 236; Logothetis 2008, 873). Some researchers avoid this problem by looking at the functional form of responses instead. If a brain region not only responds during a cognitive task, but responds more strongly the more difficult the task is (say), then there is better evidence that the region is specially associated with performance of that task. Again, sophisticated statistical techniques have been developed to reason about these functional response patterns. (Landini et al. 2005, Chs 16–18; Sarty 2007).

However, with more advanced techniques come new controversies. In a provocative recent paper, Vul et al. argue that many studies overestimate correlation strength by a large margin (2009). The resulting debate has re-emphasized the importance of accurate quantitative estimates in NI.

2.2. WHAT IS SHOWN?

Neuroimaging shows which brain areas are associated with cognitive processes, but I have been canny about what 'associated with' really means. This is also a source of debate.

Some think that NI shows the neural correlates of mental *modules*, in the sense advocated by Fodor (1983). If true, this implies something stronger than the weak thesis of functional specialization. It would mean that many brain regions perform simple, informationally encapsulated, special-purpose computations. This picture is encouraged by NI studies that show small, intense foci of brain activation associated with simple cognitive capacities.

As modularity is a controversial thesis, many have objected to this characterization of NI results. First, the mere fact of localization does not imply modularity: as Woodward and Cowie note, the different features of modularity can also occur in non-modular systems (Woodward and Cowie 2004). Early critics of fMRI also noted that the appearance of modules was a product of data processing techniques: small foci of activation would be present whether or not the brain was actually modular (Uttal 2001, 185ff; Savoy 2001). Some took this as evidence that forward inference was simply impossible; Hardcastle and Stewart, e.g., argued that the distributed architecture of the brain precludes localization altogether, leaving nothing for NI to do.

However, more sensitive processing techniques allow fMRI to avoid the charge of bias. For example, a debate between modular and distributed views of facial processing (Ishai et al. 1999; Haxby et al. 2000; Kanwisher 2000; Yovel and Kanwisher 2004) spurred the development of alternative data analyses sensitive to distributed representations (Haxby et al. 2001; O'Toole et al. 2005). In particular cases, intense debates rage over whether activation foci represent modules or simply the most active parts of a distributed system. Rather than prejudging the question, however, NI is now an important source of evidence in these debates.

Another source of controversy concerns the interpretation of 'activation'. It is common to interpret brain activation as evidence that a region performs a particular *process*: object recognition, say. This fits well with functionalism about mental states, which characterizes mental states in terms of their causal connections to one another (Armstrong 1999).

However, this is not the only possible interpretation of increased activity. Activity might signal that a region specializes in processing specific types of *information*, rather than performing specific types of processing. It might also signal regions that *represent* information relevant to a task, rather than performing specialized processing. This debate is par-

ticularly vigorous in discussions of the 'higher' functioning performed by prefrontal cortex; processing (Enger 2009), information-theoretic (Koechlin and Summerfield 2007), and representational (Wood and Grafman 2003) theories of prefrontal cortex have all been proposed. It is not obvious that these possibilities are mutually exclusive. As NI itself appears to be neutral among these interpretations,[1] however, care is required in interpreting the nature of localized brain activity.

2.3. COGNITIVE ONTOLOGY

A final issue surrounding forward inference should be of special interest to philosophers. Many NI studies are relatively non-specific, in the sense that different experiments attribute different functions to the same brain region. Price and Friston note that a region of the left posterior lateral fusiform (PLF) has been associated with a number of functions: processing visual word forms, detecting the visual attributes of animals, naming colors, decoding Braille, imagining objects, and making action decisions about familiar objects (2005, 265–7). What function, then, should we attribute to PLF? Saying that *the* function is to name colors or animals is misleading: it does more than that. Saying that it does *all* of these things is unilluminating: a good theory should unify our observations, not just list them.

Of course, these descriptions do not necessarily conflict. Activities may form a nested abstract hierarchy. Balancing a checkbook, making change, and reading a timetable are all be *a way of* doing the same thing: subtraction, which is in turn *a way of* doing mathematics, and so on. Similarly, each of the functions that PLF performs might be a way of doing a more general activity, the specific nature of which depends on the context in which it is performed. In philosophical terms, each might be a determinate performance of the same determinable function (Yablo 1992).

This is what Friston and Price conclude about PLF. They argue that all of the functions attributed to PLF are versions of integrating sensory cues with motor output—one process, but with multiple ways it can matter (2005, 268). Friston and Price conclude that there is an increasing need for a 'cognitive ontology' that makes sense of NI findings by describing them at the right level of abstraction.

This position fits well with the growing consensus among philosophers of science that the best explanations are those pitched at the right level of abstraction (Craver 2007, Ch. 6). Both Finding is the right level of abstraction for neural processes both an empirical and a philosophical task. Philosophers have spent a long-time thinking about how cognitive functions might be related to each other and to actions; these philosophical positions may have an important role to play in the integration of NI results.

3. From Brain to Psychology

3.1. ON THE VERY IDEA

The previous section focused on forward inference from psychology to brain function. This section will focus on the far more controversial use of NI: the testing of psychological theories themselves.

Consider an example. In a study on moral decision-making, Borg et al. note that:

> Imaging data suggest that moral consideration of the action-vs.-inaction distinction is mediated primarily by areas of the brain that are traditionally associated with cognition rather than with emotion (813).

This looks to be evidence that favors certain *psychological* theories about moral reasoning. Theories that claim the action-vs.-inaction distinction is reasoned about via cognitive judgments (whatever that might mean) gain support, whereas those that claim the opposite have a strike against them. This is inference from NI data back to a *psychological* theory, rather than to a theory about how the brain is organized.

Some have argued that this direction of inference is *always* illegitimate. A well-known proponent of this position is Fodor, who argues that NI cannot tell cognitive psychologists anything of use (Fodor 1999). This is just a special case of Fodor's commitment to the *autonomy of the special sciences*. Fodor thinks, on philosophical grounds, that cognitive science is methodologically independent from neuroscience (Fodor 1997). This position, known as *nonreductive physicalism*, is still popular among philosophers of mind.

Other psychologists are also skeptical of NI on principled grounds. For example, Harley argues that that brain data is irrelevant to cognitive psychology until we have a complete psychological theory – at which point NI would have nothing of interest to contribute to psychology (Harley 2004a, 11; Coltheart 2004, 22).

Early philosophical defenders of NI tried to address skepticism directly (Bechtel 2002; Landreth and Richardson 2004). However, I suspect that these debates really go far beyond fMRI. Fodor's position extends far beyond issues in NI: at stake are general issues about the relationship between cognitive psychology and neuroscience. Philosophers of science have taken on these issues in recent years, with a eye on the unique status of neuroscience (Bechtel and Mundale 1999; Bickle 2003; Craver 2007). Shallice further notes that theory-building is often a boot-strapping process that moves back and forth between abstract theories and implementation-level details and reverse inference; if so, intertheoretic scientific work can be done without completed theories on either side. (2003, S149; see also Wimsatt 1976). Finally, many of the complaints from psychologists really stem from a feeling that fMRI has shown very little *so far* (Coltheart 2006). The best response to this criticism lies in the details of particular experiments.

So suppose, at least provisionally, that NI evidence might bear on theories in cognitive psychology. The question remains: *how* does it do so? The literature offers three answers: a deductive position based on reverse inference, a deductive position based on fasificationism, and an inductive answer based on probabilistic models of theory confirmation. I will consider each in turn.

3.2. REVERSE INFERENCE

One way that theories might be tested by fMRI is by so-called 'reverse inference'. Reverse inferences contain two premises: one about the function of a brain region, and another about observed activation in an NI study. From these, a statement about cognitive psychology is deduced. In my toy inference about Borg et al. above, I reasoned this way:

(P1) If a subject judges cognitively, there is activity in region *R*.
(P2) Task T (contemplating action-vs.-inaction) cases causes activity in region *R*.

Therefore, task *T* is done via cognitive judgment.

Theories that claim that *T* is based on unconscious processes are disconfirmed by our observation, while theories that claim that *T* is based on cognitive processes are (partially) confirmed.

Unfortunately, this argument is invalid; as Poldrack notes, it commits the fallacy of affirming the consequent (2006, 60). To make it valid, we need a stronger premise:

(P1★) Region *R* is activated *if and only if* a subject is engaged in conscious reasoning.

Given this premise, the conclusion follows. P1★ is an unreasonably strong premise, however, it is equivalent to the claim that the relationship between reasoning and *R* is a *psychophysical invariant* (Cacioppo et al. 2007, 13–14) – in technical terms, that there is a one–one context-independent mapping between cognitive function and brain structure.

True psychophysical invariants are extremely rare. In Section 2.3, I reviewed evidence that the same brain region can perform a variety of functions, and in Section 2.1, I noted that activity of brain regions is often context-dependent, as it can be modulated by the performance of other cognitive tasks. Price and Friston conclude that this makes reverse inference largely impossible, because 'To infer functional specificity requires a demonstration that an area is activated only by tasks that engage its function and not others' (2005, 265). As this is rarely the case, reverse inference is a poor way to use NI to test psychological theories.

3.3. CONSISTENCY ACCOUNTS

A second way to use NI evidence is to evaluate it against the *predictions* of competing theories, with an eye to favor one or the other. On the deductive view of NI, facts about neuroimages do not function as premises in an argument. Instead, cognitive theories are taken to imply conclusions about brain functioning. These conclusions are then tested against observed brain function revealed by NI results. A mismatch indicates that the original theory was wrong; consistency with observation gives some support to the original theory.

I believe that many neuroimagers are committed a consistency account. Some are explicit about this commitment (Henson 2005, 197ff), and many more show a tacit commitment by arguing that NI results are consistent with this or that cognitive theory.[2]

The consistency account works best when cognitive theories make straightforward predictions about brain function. Debates between single-process and dual-process cognitive theories might be a case where this is so. Henson, e.g., reviews evidence that there are distinct brain regions active when subjects make different types of memory judgments about previously seen items (Henson 2006). This evidence would seem to support dual-process cognitive theories, which claim that there are two distinct types of memory (recollection and familiarity memory) rather than one single process.

Unfortunately, few cognitive theories make straightforward predictions about brain function (Mole and Klein, forthcoming). Further, it is not obvious how to *get* predictions out of many theories: the principles that would link psychological theories with claims about the brain are obscure at best (Coltheart 2006). When this is the case, it is not obvious how NI can help confirm or disconfirm a theory: *any* evidence about neural activity is consistent with theories that make no prediction about brain function.

This not an idle concern. Consider, e.g., recent debates over the role of ventromedial prefrontal cortex (vmPFC) in moral reasoning. Some argue that vmPFC generates emotional judgments in response to personal moral dilemmas, and that these judgments conflict with other, more 'utilitarian' judgment processes (Greene et al. 2001; Greene 2007; Koenigs et al. 2007; Young and Koenigs 2007). Others argue that vmPFC is responsible for generating 'prosocial moral sentiments' that are integrated in moral reasoning and an

essential part of correct moral decision-making (Moll et al. 2005, 2007). Both sides claim – convincingly, in my opinion – that the imaging data is consistent with their theory (Greene 2007; Moll and de Oliveira-Souza 2007a, b). These are clearly distinct theories. As far as the NI data goes, therefore, a consistency account must say that we are at an impasse.

This impasse is not a coincidence. The consistency account of NI is a *falsificationist* view of theory-testing (Popper 1959). On a falsificationist account, theories consistent with the data get to remain in play, while theories that are inconsistent with the data must be abandoned. Falsificationism does not allow one to decide between unfalsified theories, however. And as philosophers have long noted, most theories can be made consistent with apparently recalcitrant evidence simply by adopting appropriate auxiliary hypotheses (Klee 1996, Ch. 4; Harley 2004b, 50). Finally, recall that neuroimages are theory-laden: the very same theories that are being tested are also the source of the contrasts that generate the neuroimages themselves. The effect of this dependency is poorly understood. Consistency accounts, while widely held, thus do not seem to provide a strong basis for inference about cognitive functioning.

3.4. PROBABILISTIC ACCOUNTS

A final approach to NI data is a probabilistic one. In a probabilistic, Bayesian framework, NI data D supports a hypothesis H just in case it *raises the likelihood of that hypothesis being true*. That is, H is supported just in case it would be more likely that we would see some NI data D if H was true than it would be if H was false (Mole et al. 2007).

The probabilistic account has a number of clear advantages over the deductive account. Intuitively, we want NI data to be *informative*: that is, we want it to weigh in favor of some hypotheses rather than others. Evidence might be consistent with two hypotheses, but make it much more likely that one hypothesis is true than another. The paper says that the Cubs lost. This is *consistent* with the hypothesis that the Cubs won and the hypothesis that they lost (there might have been a misprint, after all, or the reporter might have made a mistake, or…). However, this report surely *favors* the hypothesis that the Cubs won: as the paper is generally reliable, it is much more likely to read that they won if they actually won. Similarly, we should want imaging data that increases the likelihood of a theory relative to its competitor theories (Mole et al. 2007) Keeping this in mind reminds us that theories are rarely tested in a vacuum: we are usually interested in how they fare against other, competing theories.

A probabilistic account also avoids other problems that plague deductive accounts. NI is sensitive to *any* change in brain activity, whether or not that change is functionally important. If cognitive state *A* (say, reasoning about moral dilemmas that involve one's family members) reliably gives rise to a functionally unimportant cognitive state *B* (say, anxiety), an NI experiment will likely show brain areas that are associated with *both A* and *B*. But *B*-related activity is a *functional byproduct* of *A*: something we are likely to see no matter which theory of *A* is true. (Compare: the throbbing noise made by your heart is a functional byproduct of the heart's true function of pumping blood. As such, the observation that the heart does not favor any particular functional theory about hearts, even if those theories predict throbbing.) Functional byproducts are consistent with every theory, but *favor* none of them: a probabilistic account correctly rules them out as evidence.

Finally, a probabilistic account makes possible a weaker version of reverse inference: that is, given some brain activation, we can infer the *likelihood* that a particular cognitive

process occurred. This would get around some of the problems noted with a strictly deductive version of reverse inference.

A probabilistic account thus allows NI data to test psychological theories, and does so in a philosophically defensible way. However, it is not clear how close it comes to answering the practical skeptics about fMRI. In many cases it is difficult to assign relative probabilities, and when one can, the evidence presented by NI often seems relatively weak. Probabilistic reverse inference is similarly shaky. For example, Poldrack used existing studies to estimate the Bayes factor of reverse inferences about language processing. He found that reverse inference could provide a 'positive but relatively weak increase in confidence' in cognitive hypotheses (2006, p. 62).

Probabilistic inference often provides only a suggestive, inconclusive test of many cognitive hypotheses. Probabilistic accounts are most favorable to studies of simple sensory processing, where the relevant alternatives are clearly delineated and evidence from other disciplines gives some rough estimate of the conditional probabilities in question (Mole and Klein, forthcoming). This is not the case for more complex cognitive tasks.

4. Confirmation vs Exploration

The discussion so far has assumed that the job of NI is to confirm or disconfirm hypotheses. An alternative view of NI is that it provides *exploratory,* rather than confirmatory, data (Rapoport 1991, A142; McKeown et al. 2006). Confirmatory data analysis is meant to raise or lower our confidence in a particular hypothesis. Exploratory data analysis, by contrast, simply looks for patterns in the data and suggests interpretations of them (Tukey 1977; Good 1983). In the imaging literature, exploratory data analysis is sometimes called *data-driven* analysis (as opposed to *hypothesis-driven* confirmatory analysis).

The hypotheses that result from data-driven analyses are not (necessarily) *supported* by the data, and should not be taken as such. Exploratory analysis accepts a much greater risk of false positives than confirmatory analysis, and makes no attempt to rule out all plausible competing hypotheses. Exploratory data analysis is useful, however, when the basic organizing principles of a domain aren't well-known. Cognitive neuroscience is surely in this position: it is not obvious that our cognitive theories carve the mental world in the correct way, and it is not obvious that the brain will mirror any such carving we come up with. This is another way to emphasize the need for a proper 'cognitive ontology' in Price and Friston's sense. The debate over PLF in Section 2.3 is less a debate about what PLF does than a debate about w*hat mental processes there are* (Poldrack 2008, 224). The use of fMRI to build these cognitive ontologies is an exploratory rather than confirmatory use of the data.

Building a data-driven cognitive ontology is partly an empirical task, of course, and interesting advances have already been made. A meta-analysis by Kober et al. showed six functional groupings of brain areas engaged in emotional processing, none of which correspond in a straightforward way to ordinary concepts of emotion (2008). Michael Anderson has used a data-driven fMRI meta-analysis to support his 'massive redeployment hypothesis' of cognitive architecture, and so to suggest a new relationship between perception and action (2007a, b). On a smaller scale, data-driven analyses of fMRI data have revealed novel information processing strategies in the brain (O'Toole et al. 2005).

Interpreting these results can also involve a philosophical component. Consider again the debate between Greene et al. and Moll et al. on moral decision-making. This debate is as much a problem of philosophical interpretation as an empirical dispute. Must emotion and reason necessarily conflict in judgment? Or are the 'moral sentiments' crucial to

our ordinary moral understanding? Greene seems to believe that the emotional responses are necessarily indifferent to the agent's considered reasons. Moll, in contrast, believes that emotions can track reasons. This is a philosophical debate as much as an empirical one (Arpaly 2000), and philosophical positions can influence the interpretations of the resulting NI data.

A focus on confirmatory data analysis might suggest that the impact of philosophers is destined to be relatively limited: aside from sorting out big-picture methodological debates, it looks like philosophers should simply sit back and let their favorite hypotheses be tested. Emphasis on exploratory analysis, on the other hand, suggests the possibility of dynamic interdisciplinary collaboration between neuroimagers and philosophers. As Harrison notes, 'it is clear that we have conceptual work to do before we fire up the scanner' (2008, 321). That conceptual work ensures that philosophers can help the nascent science of NI as much as they can learn from it.

Short Biography

Colin Klein's research focuses on philosophy of mind and philosophy of science, particularly where they intersect in philosophy of psychology. He is interested in the role of idealized models in psychology, and the challenges they raise for interpreting theories. He is interested in general questions about theory testing and intertheoretic explanation in psychology (and lately, in the fraught issues surrounding statistical significance testing). In addition to his work on neuroimaging, he has published on pain perception, multiple realizability, scientific reduction, enactivism, and computationalism. His work appears in *Journal of Philosophy*, *Philosophical Psychology, Philosophical Studies, Philosophical Quarterly, Synthese,* and the forthcoming volume *Foundational Issues of Human Brain Mapping*. He earned a dual BA from Franklin and Marshall college, and a PhD in philosophy from Princeton University.

Notes

⋆ Correspondence: 1420 University Hall MC 267, 601 S Morgan St. Chicago, Illinois, United States, 60613. Email: cvklein@uic.edu.

[1] Though recent work suggests that fMRI might be more sensitive to relative synaptic activity than action potentials, which might have some bearing on the question (Viswanathan and Freeman; Logothetis 873ff). For an example of a study where this makes a difference to localization, see Rilling et al. 2004.

[2] This reading is not uncontroversial; see Harman, forthcoming, for an alternative perspective.

Works Cited

Adolphs, Ralph. 'Cognitive Neuroscience of Human Social Behaviour.' *Nature reviews. Neuroscience* 4.3 (2003): 165–178.

Anderson, Michael L. 'Massive Redeployment, Exaptation, and the Functional Integration of Cognitive Operations.' *Synthese* 159.3 (2007a): 329–345.

——. 'The Massive Redeployment Hypothesis and the Functional Topography of the Brain.' *Philosophical Psychology* 20.2 (2007b): 143–174.

Armstrong, David. *The Mind-Body Problem: An Opinionated Introduction*. New York: Westview Press, 1999.

Arpaly, Nomi. 'On Acting Rationally Against One's own Best Judgment.' *Ethics*. 110.3 (2000): 488–513.

Bechtel, William. 'Decomposing the Mind-Brain: A Long-Term Pursuit.' *Brain and Mind* 3 (2002): 229–242.

—— and Jennifer Mundale. 'Multiple Realizability Revisited: Linking Cognitive and Neural States.' *Philosophy of Science* 66 (1999): 175–207.

Bickle, John. *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Boston: Kluwer Academic Publishers, 2003.

Borg, J.S., Hynes, C., Van Horn, J., Grafton, S. and Sinnott-Armstrong, W. 'Consequences, Action, and Intention as Factors in Moral Judgments: An FMRI Investigation.' *Journal of Cognitive Neuroscience* 18.5 (2006): 803–817.

Buxton, Richard B. *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*. New York: Cambridge University Press, 2002.

Cacioppo, John T., Louis G. Tassinary and Gary G. Berntson. 'Psychophysiological Science: Interdisciplinary Approaches to Classic Questions About the Mind.' *Handbook of Psychophysiology*. 3rd edn. Eds John T. Cacioppo, Louis G. Tassinary, Gary G. Berntson. New York: Cambridge University Press, 2007. 1–18.

Camerer, C., Loewenstein, G. and Prelec, D. 'Neuroeconomics: How Neuroscience can Inform Economics.' *Journal of Economic Literature* 43.1 (2005): 9–64.

Coltheart, M. 'Brain Imaging, Connectionism, and Cognitive Neuropsychology.' *Cognitive Neuropsychology* 21.1 (2004): 21–25.

Coltheart, Max. 'What has Functional Neuroimaging Told us About the Mind (so far)?' *Cortex* 42.3 (2006): 323–331.

Craver, C. F. *Explaining the Brain*. USA: Oxford University Press, 2007.

Enger, T. 'Prefrontal Cortex and Cognitive Control: Motivating Functional Hierarchies.' *Nature Neuroscience* 12.7 (2009): 821–822.

Farah, Martha J. 'Emerging Ethical Issues in Neuroscience.' *Nature Neuroscience* 5 (2002): 1123–1129.

Fodor, Jerry. *The Modularity of Mind*. Cambridge: MIT press, 1983.

——. 'Special Sciences: Still Autonomous After All These Years.' *Philosophical perspectives: Mind, Causation, and World* 11 (1997): 149–163.

——. 'Diary.' *London Review of Books* 21.19 (1999): 68–69.

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. and Dolan, R. J. 'The Trouble With Cognitive Subtraction.' *Neuroimage* 4.2 (1996): 97–104.

Good, I. J. 'The Philosophy of Exploratory Data Analysis.' *Philosophy of Science* 50.2 (1983): 283–295.

Greene, J. D. 'Why are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains.' *Trends in Cognitive Sciences* 11.8 (2007): 322–323.

——, Sommerville, R. B., Nystrom, L. E., Darley, J. M. and Cohen, J. D. 'An FMRI Investigation of Emotional Engagement in Moral Judgment.' *Science* 293.5537 (2001): 2105–2108.

——, Nystrom, L. E., Engell, A. D., Darley, J. M. and Cohen, J. D. 'The Neural Bases of Cognitive Conflict and Control in Moral Judgment.' *Neuron* 44.2 (2004): 389–400.

Hardcastle, Valerie Gray and C. Matthew Stewart. 'What Do Brain Data Really Show?.' *Philosophy of Science* 69.s3 (2002): S72–S82.

Hare, Todd A., Colin F. Camerer and Antonio Rangel. 'Self-Control in Decision-Making Involves Modulation of the VmPFC Valuation System.' *Science* 324.5927 (2009): 646–648.

Harley, T. A. 'Does Cognitive Neuropsychology Have a Future?' *Cognitive Neuropsychology* 21.1 (2004a): 3–16.

—— 'Promises, Promises.' *Cognitive Neuropsychology* 21.1 (2004b): 51–56.

Harman, Gil. 'Words and Pictures in Reports of FMRI Research.' *Foundational Issues of Human Brain Mapping*. Ed. Stephen José Hanson, Martin Bunzl. Cambridge: MIT Press, forthcoming 2009.

Harrison, G. W. 'Neuroeconomics: A Critical Reconsideration.' *Economics and Philosophy* 24.03 (2008): 303–344.

Haxby, J. V., Hoffman, E. A. and Gobbini, M. I. 'The Distributed Human Neural System for Face Perception.' *Trends in Cognitive Sciences* 4.6 (2000): 223–233.

——, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L. and Pietrini, P. 'Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex.' *Science* 293.5539 (2001): 2425–2430.

Henson, Richard. 'What can Functional Neuroimaging Tell the Experimental Psychologist?' *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology* 58.2 (2005): 193–233.

——. 'Forward Inference Using Functional Neuroimaging: Dissociations Versus Associations.' *Trends in Cognitive Sciences* 10.2 (2006): 64–69.

Hohwy, J. 'The Search for Neural Correlates of Consciousness.' *Philosophy Compass* 2.3 (2007): 461–474.

Huettel, Scott A., Allen W. Song and Gregory McCarthy. *Functional Magnetic Resonance Imaging*. Sunderland, Massachusetts: Sinauer Associates, Inc, 2004.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L. and Haxby, J. V. 'Distributed Representation of Objects in the Human Ventral Visual Pathway.' *Proceedings of the National Academy of Sciences of the United States of America* 96.16 (1999): 9379–9384.

Jacobs, A. J. 'Do I Love My Wife? An Investigative Report.' *Esquire* (2009): 120–125.

Joyce, Kelly A. *Magnetic Appeal: MRI and the Myth of Transparency*. Ithaca: Cornell University Press, 2008.

Kanwisher, Nancy. 'Domain Specificity in Face Perception.' *Nature Neuroscience* 3.8 (2000): 759–763.

Klee, Robert. *Introduction to the Philosophy of Science: Cutting Nature at Its Seams*. New York: Oxford University Press, 1996.

Klein, Colin. 'Images are not the Evidence of Neuroimaging.' *British Journal for the Philosophy of Science*, forthcoming.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K. and Wager, Tor D. 'Functional Grouping and Cortical-Subcortical Interactions in Emotion: A Meta-Analysis of Neuroimaging Studies.' *Neuroimage* 42.2 (2008): 998–1031.

Koechlin, E and C. Summerfield. 'An Information Theoretical Approach to Prefrontal Executive Function.' *Trends in Cognitive Sciences* 11.6 (2007): 229–235.

——, Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M and Damasio, A. 'Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements.' *Nature* 446.7138 (2007): 908–911.

Landini, Luigi, Vincenzo Positano and Maria Filomena Santarelli. *Advanced Image Processing in Magnetic Resonance Imaging*. Boca Raton, CRC Press, 2005.

Landreth, Anthony and Robert C. Richardson. 'Localization and the new Phrenology: A Review Essay on William Uttal's *The New Phrenology*.' *Philosophical Psychology* 17.1 (2004): 108–123.

Lieberman, Matthew D. 'Social Cognitive Neuroscience: A Review of Core Processes.' *Annual Review of Psychology* 58 (2007): 259–289.

Lloyd, D. 'Functional MRI and the Study of Human Consciousness.' *Journal of Cognitive Neuroscience* 14.6 (2002): 818–831.

Logothetis, Nikos K. 'What we can do and What we Cannot do With FMRI.' *Nature* 453 (2008): 869–878.

McCabe, David P. and Alan D. Castel. 'Seeing is Believing: The Effect of Brain Images on Judgments of Scientific Reasoning.' *Cognition* 107 (2008): 343–352.

McKeown, M. J., Wang, Z. J., Abugharbieh, R. and Handy, T. C. 'Increasing the Effect Size in Event-Related FMRI Studies.' *IEEE Eng Med Biol Mag* 25.2 (2006): 91–101.

Meehl, Paul E. 'Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology.' *Journal of Consulting and Clinical Psychology* 46 (1978): 806–834.

Mole, Chris and Colin Klein. 'Confirmation, Refutation and The Evidence of FMRI.' *Foundational Issues of Human Brain Mapping*. Ed. Stephen José Hanson, Martin Bunzl. Cambridge: MIT Press, Forthcoming 2009.

Mole, C., Kubatzky, C., Plate, J., Waller, R., Dobbs, M. and Nardone, M. 'Faces and Brains: The Limitations of Brain Scanning in Cognitive Science.' *Philosophical Psychology* 20.2 (2007): 197–207.

Moll, Jorge and Ricardo de Oliveira-Souza. 'Moral Judgments, Emotions and the Utilitarian Brain.' *Trends in Cognitive Sciences* 11.8 (2007a): 319–321.

—— and ——. 'Response to Greene: Moral Sentiments and Reason: Friends or Foes?' *Trends in Cognitive Sciences* 11.8 (2007b): 323–324.

——, Zahn, R., de Oliveira-Souza, R., Krueger, F. and Grafman, J. 'The Neural Basis of Human Moral Cognition.' *Nature Reviews Neuroscience* 6.10 (2005): 799–809.

——, de Oliveira-Souza, R., Garrido, G. J., Bramati, I. E., Caparelli-Daquer, E. M. A., Paiva, M. L. M. F., Zahn, R. and Grafman, J. 'The Self as a Moral Agent: Linking the Neural Bases of Social Agency and Moral Sensitivity.' *Social Neuroscience* 2.3 (2007): 336–352.

——, Ricardo De Oliveira-Souza and Roland Zahn. 'The Neural Basis of Moral Cognition: Sentiments, Concepts, and Values.' *Annals of the New York Academy of Sciences* 1124 (2008): 161–180.

Nair, Dinesh G. 'About Being BOLD.' *Brain Research Reviews* 50 (2005): 229–243.

O'Toole Alice, J., Jiang, Fang, Abdi, Herve and Haxby James, V. 'Partially Distributed Representations of Objects and Faces in Ventral Temporal Cortex.' *Journal of Cognitive Neuroscience* 17.4 (2005): 580–590.

Poldrack, Russell A. 'Can Cognitive Processes be Inferred From Neuroimaging Data?' *Trends in Cognitive Sciences* 10.2 (2006): 59–63.

——. 'The Role of FMRI in Cognitive Neuroscience: Where do we Stand?' *Current Opinion in Neurobiology* 18.2 (2008): 223–227.

Popper, Karl. *The Logic of Scientific Discovery*. London, Hutchinson & Co., 1959.

Price, C. J. and K. J. Friston. 'Cognitive Conjunction: A new Approach to Brain Activation Experiments.' *Neuroimage* 5.4 Pt 1 (1997): 261–270.

—— and K. J., Friston. 'Functional Ontologies for Cognition: The Systematic Definition of Structure and Function.' *Cognitive Neuropsychology* 22.3 (2005): 262–275.

Raichle, Marcus E. 'A Brief History of Human Brain Mapping.' *Trends in Neurosciences* 32.2 (2009): 118–126.

Rapoport, S. I. 'Discussion of PET Workshop Reports, Including Recommendations of PET Data Analysis Working Group.' *Journal of cerebral blood flow and metabolism* 11.2 (1991): A140.

Rescorla, Michael. 'Cognitive Maps and the Language of Thought.' *British Journal for the Philosophy of Science* 60.2 (2009): 377–407.

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E. and Cohen, J. D. 'Opposing BOLD Responses to Reciprocated and Unreciprocated Altruism in Putative Reward Pathway.' *Neuroreport* 15.16 (2004): 2539–2543.

Roskies, Adina. 'Neuroscientific Challenges to Free Will and Responsibility.' *Trends in Cognitive Sciences* 10.9 (2006): 419–423.

——. 'Are Neuroimages Like Photographs of the Brain?' *Philosophy of Science* 74 (2007): 860–872.

Sanfey, Alan G. and James K. Rilling et al. 'The Neural Basis of Economic Decision-Making in the Ultimatum Game.' *Science* 300.5626 (2003): 1755–1758.

——, George Loewenstein, F., McClure, S. M. and Cohen, J. D. 'Neuroeconomics: Cross-Currents in Research on Decision-Making.' *Trends in Cognitive Sciences* 10.3 (2006): 108–116.
Sarty, Gordon E. *Computing Brain Activity Maps From FMRI Time-Series Images*. Cambridge: Cambridge University Press, 2007.
Savoy, Robert L. 'History and Future Directions of Human Brain Mapping and Functional Imaging.' *Acta Psychologica* 107 (2001): 9–42.
——. 'Experimental Design in Brain Activation MRI: Cautionary Tales.' *Brain Research Bulletin* 67.5 (2005): 361–367.
——. 'Functional Imaging and Neuropsychology Findings: How can They be Linked?' *Neuroimage* 20 Suppl 1 (2003): S146–54.
Tukey, John W. 'Analyzing Data: Sanctification or Detective Work?' *American Psychologist* 24 (1969): 83–91.
——. *Exploratory Data Analysis*. Reading: Addison-Wesley, 1977.
Uttal, William R. *The New Phrenology*. Cambridge: MIT Press, 2001.
Viswanathan, Ahalya and Ralph D. Freeman. 'Neurometabolic Coupling in Cerebral Cortex Reflects Synaptic More Than Spiking Activity.' *Nature Neuroscience* 10.10 (2007): 1308–1312.
Vul, E., Harris, C., Winkielman, P. and Pashler, H. 'Puzzlingly High Correlations in FMRI Studies of Emotion, Personality, and Social Cognition.' *Perspectives On Psychological Science* 4.3 (2009): 274–290.
Sinnott-Armstrong, Walter. Ed. *Moral Psychology. Volume 3: The Neuroscience of Morality*. Cambridge: MIT University Press, 2008.
Wimsatt, William C. 'Reductionism, Levels of Organization, and the Mind-Body Problem.' *Consciousness and the Brain: A Philosophical Investigation*. Globus, G., Savodnik, I. and Maxwell, G. New York: Plenum Press, 1976. 205–267.
Wood, J. N. and J., Grafman. 'Human Prefrontal Cortex: Processing and Representational Perspectives.' *Nature Reviews Neuroscience* 4.2 (2003): 139–147.
—— and ——. 'The Mind is Not (Just) a System of Modules Shaped (Just) by Natural Selection.' *Contemporary Debates in Philosophy of Science*. Ed. C. Hitchcock. Oxford: Wiley, 2004: 293–312.
Yablo, Stephen. 'Mental Causation.' *The Philosophical Review* 101.2 (1992): 245–280.
Young, Liane and Michael Koenigs. 'Investigating Emotion in Moral Cognition: A Review of Evidence From Functional Neuroimaging and Neuropsychology.' *British Medical Bulletin* 84 (2007): 69–79.
Yovel, Galit and Nancy Kanwisher. 'Face Perception: Domain Specific, not Process Specific.' *Neuron* 44.5 (2004): 889–898.