

Chapter 1

Introduction

Sarah Richmond

Curiosity about the human brain has characterized Western science from the outset. Forbidden by the law of the Roman Empire from dissecting human corpses, the physician Galen still managed to learn an impressive amount about anatomy from his investigation of animal cadavers, and his ingenious experiments on living creatures led him to claim—correctly—that the brain controls the movements of the muscles. For centuries thereafter, other scientists' hypotheses about the human brain, however prescient they may have been, were hampered by the inability to confirm or substantiate them observationally. Not until the twentieth century, with the development of highly sophisticated imaging technology that could provide a window onto the workings of the living brain, did major advances in brain science become possible.

The potential medical gains from these imaging machines were obvious from the start. A great many diseases and disorders are known to be associated with the brain, and many of these show up, more or less directly, in its imaging. From the 1980s onwards, magnetic resonance imaging (MRI) scanners played an important role in the clinical diagnosis of a number of conditions, both in the brain and the rest of the body. The routine use of these machines made the diagnosis of brain tumours and brain damage, for example, both quicker and more accurate. A basic understanding of these techniques is now a requirement of medical training, and can be gained from any number of medical textbooks.

The focus of this book is both narrower and wider than the medical brain. It is primarily concerned with uses of brain imaging technologies that throw light on our *mental* capacities, states, and acts. (Because it is mainly neurological processes that are relevant here, the brain imaging used for these purposes is often referred to, more precisely, as 'neuroimaging'.) In this sense, our focus is on a subset of brain imaging research. However, the neuroscientific advances that relate specifically to the mind have, unsurprisingly, provoked excitement and discussion in the many disciplines to which they are relevant, extending far beyond the bounds of medicine. Commissioned to reflect this multidisciplinary work, the essays collected in this volume range widely.

Progress in the study of the neural basis of mental activity has of course influenced work in those areas of medicine concerned with the mind, including psychiatry and clinical psychology. There are hopes that brain imaging technology might become an effective means of diagnosing and understanding mental illnesses in which structural and functional changes to the brain are implicated. Outside medicine, it has also attracted the

attention of philosophers, psychologists, sociologists, criminologists, and legal scholars. Plans to capitalize on it commercially are underway, for example within market research and psychometric testing, as well as in private medicine, where several ‘neurodiagnostic’ tests are already on sale. And, although the arts are not much discussed in this volume, the influence in this domain of neuroscientific ideas and developments should also be noted, ranging from literature (especially science fiction), to film and the visual arts.

One of the most thought-provoking developments in recent neuroscience, described in some of the essays in this volume, has been the progress made with ‘mindreading’: hence our title. Using neuroimaging data, scientists are now able to ‘read’ or infer, with a significant degree of accuracy, some of the thoughts going through the mind of the person being scanned. At present this ability is limited in a number of ways. But the possibility that in the near future we will see increasingly thorough incursions of this sort into the ‘citadel of the mind’—a domain of human existence that has been regarded ever since the Stoics as a haven for private contemplation—has great power to disturb. Much has been written about the many ways in which individual privacy has been threatened by the informational revolution, but so far there has been little attention paid to the potential impact on privacy (and in particular on so-called ‘mental’ privacy) of the neuroimaging technology discussed here. This connection, between neuroscientific developments in penetrating the human mind, and the privacy of individuals, is the central concern and *raison d’être* of this volume.

As some of the chapters in this book describe, neuroscientists currently rely on various different brain imaging technologies, each of which has its own advantages and limitations. The diagram reproduced below (see Figure 1.1) offers a classification that readers may find it helpful to refer to. The resolution of magnetic resonance imaging (MRI) and functional magnetic resonance imaging (fMRI) scanners, in conjunction with the fact that they are not ‘invasive’ and do not involve the administration of radioactive substances or the use of X-rays, has made them especially central to much of the research described in the following pages.

The chapters in Part One present findings and address questions that are foundational in relation to the succeeding chapters in the volume.

Anyone engaged in the study of the human brain is likely to wonder why it is such an extraordinarily complex organ. How did it come to be that way? How does the human brain differ from the brains of other animals, and especially from the brains of our closest primate ‘relatives’? Advances in brain imaging have helped evolutionary theorists to develop plausible hypotheses in response to these questions. According to the Social Brain hypothesis, the unusually large size of the human brain (in relation to body size) is due to the distinctive capacities that humans have for relating to each other within a stable social group. In their chapter, the anthropologists, Susanne Shultz and R.I.M. Dunbar, defend a version of this hypothesis, focusing especially on the cognitive powers involved in the ‘Theory of Mind’ that humans are thought to deploy in order to attribute mental states to each other. The size of the human brain is not a new discovery (and does not need any sophisticated technology to establish it): the innovative contribution of

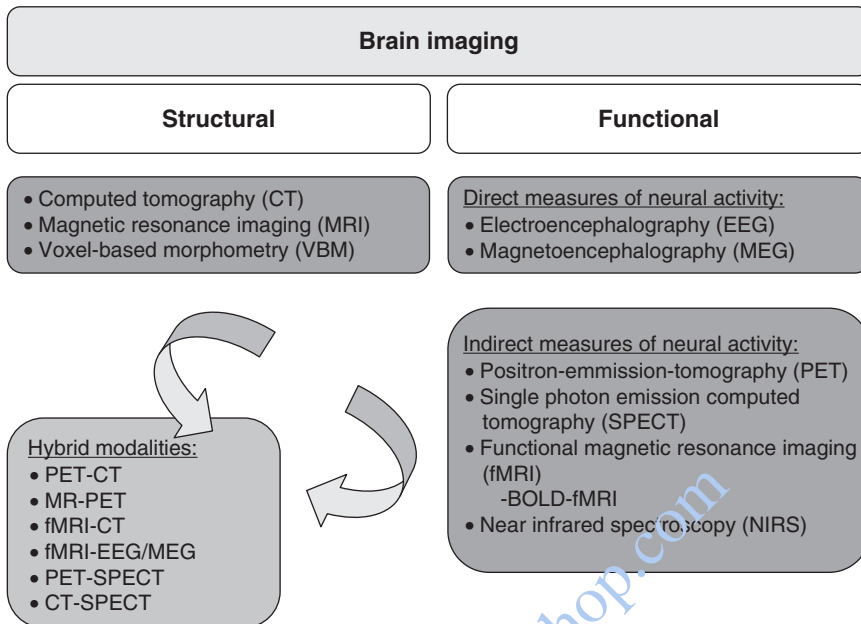


Fig. 1.1 Diagram reproduced from 'Human Functional Brain Imaging 1990–2009: Portfolio review' published by the Wellcome Trust.

neuroscience to this area of study, however, has been to enable experiments in which the brain activation of subjects doing various 'social cognition' tasks can be examined and mapped. The discovery that parts of the prefrontal cortex are consistently activated by Theory of Mind tasks provides the Social Brain hypothesis with support, and the fact that it is this area of the brain that has become larger as human brain size has increased over time also fits nicely within a theory that regards the demands of social cognition as a central 'driver' of the evolution of the human brain.

Our brains, then, are well equipped by natural selection to read other people's minds. How effectively, at the present time, can machines perform this task? In his chapter, the neuroscientist John Dylan-Haynes considers how close we are to being able to use brain imaging technology to read someone's thoughts from their brain activity. (At this point, a note on terminology is required. Although the expression 'mindreading' is often used to refer to the neuroscientists' goal, many researchers in the field prefer to talk in terms of 'brain reading'. This usage records the fact that the mental states or acts that are ascribed to a person through the use of neuroimaging data are *inferred* from that person's brain activation patterns. The brain images provide the 'text', which is 'decoded' to deliver its mentalistic content.)

Dylan-Haynes points to some exciting recent achievements: experiments have shown that it is possible to use brain imaging technology to successfully 'read out' which picture someone is looking at, or conjuring up in their mind, or which memory item they are

remembering. In addition, it can be determined which out of a range of intentions someone is entertaining, and which out of a range of concepts is figuring in their thoughts. Nonetheless, as Dylan-Haynes explains, the limitations of this research—both technical and methodological—are considerable. Technically, the resolution that can currently be achieved in brain imaging is limited, both spatially and temporally: there is a range of significant detail about brain states that cannot be captured. Brain imaging machinery is also expensive and cumbersome, and the co-operation of the subject being scanned is required to produce high-quality data. (Some of these limitations will count as a blessing in the eyes of people who are nervous about undesirable uses, including intrusive forms of surveillance and espionage, to which this technology might one day be put. These anxieties are considered in Part Four.) Methodologically, there are a host of difficulties, including the fact that there is considerable variation in the way in which the ‘same’ mental states may be represented in the brains of different people. Although the customary technical/methodological distinction is convenient, a more urgent distinction for thinking about the possibilities of future progress is between those ‘limitations’ that researchers have some chance of overcoming and those, if there are any, that are impossible to eliminate. For example, unless some ingenious way is found to accommodate it, variability between individuals may block the possibility of devising a *universal* ‘mindreading machine’, that would be capable of reading all kinds of mental content, from random minds, in detail. This problem does not, of course, rule out progress with more modest or local mindreading goals. It turns out that, where deception is concerned, the brain activation patterns across individuals are remarkably similar. If the goal is simply to determine *whether* someone is lying (regardless of the precise content of the lie), brain imaging technology can already achieve this with some degree of success and may have some advantages over the classical ‘polygraph’ equipment used in lie detection. A number of papers in this volume comment on this development.

The research programme outlined by Dylan-Haynes is further discussed in the chapter by the philosopher Tim Bayne, who examines some of the most important experimental findings in close detail, considers potential objections, and clarifies their significance. (One of the studies discussed by Bayne was led by Sean Spence, a psychiatrist at the University of Sheffield. Professor Spence was on the original list of contributors to this volume but, sadly, died before his essay could be submitted. The editors are sorry to report this missing contribution from the book, but glad to point out that Spence’s influence is, nonetheless, recorded in its pages.) Bayne draws attention to the fundamental circumstantial difference between the experimental setting—a laboratory—and the outside world, and the difficulty in successfully extrapolating findings from one domain to the other. In relation to this transfer, lie detection, for example, raises an unsurprising challenge: while scientists can, for the most part, count on the truthfulness and co-operation of the subjects being tested in the lab, it is unlikely that the population on whom lie detectors will most often be trained in the ‘real world’ will be similarly motivated to help. Another problem that needs to be noted is the highly constrained range of outcomes within which current experimental work is conducted: where it is a matter of deciding which out of just two possible intentions a subject is entertaining,

scientists have succeeded in using brain imagery to identify the right one; outside the laboratory, however, it is rarely the case that the possibilities are so limited. Bayne's chapter also introduces the deep philosophical issue of the authority of brain-based attributions of mental states, in comparison with the methods we use in everyday life. Faced with someone who sincerely denies that they feel any pain, we will surely not feel entitled to insist, on the basis of neuroimaging evidence, that they are mistaken and that they actually *are* in pain: the more plausible response will be to defer to the speaker. Later chapters in the book return to this issue of 'first person authority'.

Where mental activity is concerned, fMRI technology is often used to map ongoing brain activity. In conjunction with pattern-recognition software, this technology has driven recent advances in the imaging of transient mental states and acts. But examination of some *structural* aspects of the brain (that can also be captured with MRI scanners) also throws light on people's more enduring mental characteristics ('traits'). The chapter by the cognitive neuroscientists Geraint Rees and Ryota Kanai explores and discusses some of these findings. The surprisingly robust correlations that have been established between brain structure and a range of human traits allow us to envisage a scenario in which scrutiny of someone's brain structure will deliver information about their behavioural dispositions, and even their political commitments, that might not otherwise be available. We can imagine that such information might be of interest to prospective employers; moreover, as brain structure is not something that a person can readily influence, this method of finding out about someone might be more appealing than some of the current methods (interviews, observation, psychometric tests) in which there is sometimes considerable room for manipulation. (However, as some of the other chapters discuss, manipulation can present a problem in functional imaging.)

The essays in Part Two focus on various medical applications of brain imaging to investigate patients' minds. Two papers discuss the use of brain imaging technology in relation to patients afflicted with disorders of consciousness that make it difficult to know whether they retain awareness. Some patients who survive acute brain injury enter a condition known as the 'vegetative state' (VS). The criteria for diagnosis of this condition include 'wakefulness without awareness': obviously, in order to establish whether these criteria are fulfilled, a physician has to establish that the patient does indeed lack awareness. But as the patient may not be able to communicate by the normal behavioural means (such as by speaking or motor signals), it is difficult to rule awareness out with confidence. The possibility that, after a certain period of time, the option of withdrawing life-support (nutrition and hydration) from a patient thought to be in VS may be considered—in addition to the possibility that the patient may be in pain—makes medical uncertainty on this score especially harrowing. The neuroscientist Adrian Owen describes some remarkable research that has allowed a small number of these patients, handicapped in the ways described, to apparently communicate residual awareness by action of a purely 'mental' type, in which no bodily movements are required. The patient is instructed, prior to being scanned, to *imagine* performing some activity, such as walking around their house or playing tennis. (The choice of a paradigmatic motor activity is not accidental, because the experiment exploits the well-validated fact that the brain

activation patterns generated by merely *imagining* performing some motor activity are similar to those generated by *actually* performing that activity.) If the neuroimaging data consequent upon this instruction shows activation in the relevant ‘activity’ areas of the patient’s brain, there is extremely good reason to believe, Owen suggests, that a significant degree of awareness is retained: the patient has demonstrated, by means of this ‘neural proxy’ for behaviour, that he or she retains the ability to follow instructions, to carry out mental tasks in accordance with them, and to exercise their will. Owen argues compellingly that, given the uncertainty attendant on the traditional, clinical ‘bedside’ tests for awareness, the use of fMRI data can provide a valuable additional tool in the assessment of behaviourally non-responsive patients.

The clinical neuroscientists Athena Demertzi and Steven Laureys take up another urgent question in relation to patients with VS or MCS (minimally conscious state): are these patients in pain? As with the variant question about awareness, the absence of behavioural indicators in these patients makes it difficult to reach a confident verdict. But, the authors point out, the question is far from academic: decisions about whether or not life-support should be continued, as well as about whether or not analgesics should be administered, are likely to be influenced by a belief that the patient is in pain. Demertzi and Laureys outline some research, using neuroimaging, that aims to understand the neural realization of pain and argue that advances in this area will contribute to the development of evidence-based guidelines on the treatment of patients with disorders of consciousness. They acknowledge, along the way, the fascinating and intransigent philosophical difficulty presented by the ‘gap’ between knowledge of a person’s brain states and knowledge of what that patient is experiencing.

The other three essays in this section of the book are concerned with mental health. In this area, neuroscientific research has aimed to provide new information about patients’ brains that could provide a basis for diagnosis, prediction, and even treatment. Anticipating the transition, in the near future, from the research stage of brain imaging technology to clinical applications of it, Emily Borgelt, Daniel Buchman, and Judy Illes—a team of neuroethicists based in Canada—sought out the views of the clinicians whose work would be affected. These mental health professionals would have the task of presenting and explaining test findings to their patients, and integrating them within the process of securing informed consent in relation to diagnosis and treatment. The survey of 61 practitioners revealed a range of thoughtful responses. Alongside the hope that test images might clarify patients’ understanding of their condition, worries were expressed about the illusory authority—the ‘pseudo-precision’—that might be associated with hi-tech brain images. There was also a fear that the presence of brain scan results in the medical files of an already stigmatized group of patients might have an additional stigmatizing effect. This chapter also considers the fascinating iterative problem, raised elsewhere in this volume, in relation to ‘informed consent’. If, as seems possible, neuroimaging technology might one day be used to determine a patient’s decision-making capacity—which is a requirement for informed consent—does it make sense to seek that patient’s informed consent to that very test?

The clinical psychiatrist Brendan D. Kelly considers the clinical usefulness within psychiatry of brain imaging techniques. Adding his voice to the call for caution, Kelly reminds us of the imaginative power that the very idea of brain imaging exerts on those suffering from mental illness and their families, people whose situation makes them 'eager to believe that brain imaging can offer a precise ... description of a demonstrable aberration ... which will then be corrected'. Sensationalist coverage of neuroscientific studies in the media has unfortunately helped to fuel further unrealistic expectations of the technology. Kelly shows that so far brain imaging techniques have primarily been of use within psychiatry for ruling out other possible causes of symptoms, for example brain tumours and dementia. His discussion demonstrates the importance, alongside the continuing research into the neurobiology of psychiatric illness, of specific research into its clinical usefulness.

David Linden's chapter takes this debate further. Pointing out that the absence of 'objective' diagnostic tests makes psychiatry the 'oddball' of medical disciplines, Linden explains the difference that the availability of such tests might make. As things stand, psychiatric diagnoses rely heavily on patients' self-report. In cases where it is safe to assume the sincerity and co-operation of the self-reporting patient this is not usually a problem (although questions may arise, even where a patient's intentions cannot be faulted, about the accuracy of his or her account). In other cases, however, there may be reason to doubt the sincerity of the patient: someone might with good reason want to avoid a diagnosis of psychopathy; it is conceivable, also, that someone might attempt to deceive a doctor by 'faking' psychiatric symptoms, in order to gain some advantage from a diagnosis (such as leave from work). In these cases the availability of a corroborative 'objective' diagnosis would be helpful. A significant amount of research has been undertaken that aims to map, on the basis of neuroimaging data, psychiatric symptoms and conditions. However, as Linden shows, there are important limitations to the results. One problem (also noted elsewhere in this volume), is known as 'reverse inference': the presence of some symptom (such as auditory hallucinations) can only legitimately be inferred from a particular pattern of activation in a brain area if the same area is not also activated by other processes. Alongside this type of difficulty, there is the intriguing 'inherent circularity' of the attempt to establish objective diagnostic tests whose reliability can only be assessed by reference to the findings of traditional, clinical diagnoses—based in large part on self-report. The possibility of someone's manipulating, in the course of a scan, their own brain activation patterns also threatens to undermine its use in outwitting deceitful patients. Still, neuroimaging data may have a useful role to play in predicting the onset of mental illness in patients who have not yet developed symptoms and so are not yet in a position to report them. This information might be useful for people with a high illness risk, and would allow for treatment (including preventative measures) to be introduced at an early stage.

The essays in Part Three discuss the possible impact of neuroscientific technology within criminal law and defence. In both these domains, it might be claimed, this technology might usefully further the state's legitimate interests in protecting the public.

(Against the objection that public protection is not the only, or even the primary goal, of the judicial system it can be pointed out that in the case of some offenders it is at least one area of concern.) In relation to criminal justice, it is thought that neurological data may in the future play an evidential role, especially in establishing facts about a defendant's state of mind—facts that might make a difference to the ascription, for example, of full responsibility. Although some attempts have already been made to use neuroimaging data in criminal trials, it is not clear so far that any decision has turned on it.

The forensic psychiatrists Colin Campbell and Nigel Eastman consider, in their essay, the 'evidential' potential of this sort of data. They discuss a range of scientific studies that have used imaging technology to examine the brain structures and functions of so-called 'anti-social' populations, and point to the extreme difficulty in extrapolating, from this research, any facts that would be relevant to the mental state of an individual 'anti-social' defendant, especially in relation to an offence allegedly committed at a specific time and in specific circumstances. More broadly, Campbell and Eastman emphasize the conceptual and methodological difficulties involved in any attempt to 'translate' neuroscientific findings into terms that are fit for purpose by the law, and warn against the risk that efforts in this domain will serve illiberal socio-political agenda. Reminding us of the erstwhile prestige of phrenology, before its scientific demise, Campbell and Eastman caution against exaggerated expectations of the potential of neuroscience to explain human behaviour.

In his discussion of the prospects for bringing neuroscientific evidence to bear in relation to a claim of diminished capacity in the US, the lawyer and psychologist Stephen Morse also sounds a sceptical note. Morse offers a critical description of the present state of US law and legal thinking, indicating the points at which the concept of diminished capacity might come into play. With Campbell and Eastman, Morse believes that as things stand neuroscientific data cannot be of help in establishing a defendant's mental state at the time of the crime of which he or she is accused. In relation to broader diagnoses, such as the capacity of the defendant for rational action more generally, Morse argues compellingly that the criteria of folk psychology (in terms of which legal reasoning about actions is conducted) will always 'trump' any putative conflicting evidence from neuroscientific data. Morse's discussion also illuminates the ways in which the rights guaranteed by the US Constitution might apply to the concerns about privacy raised by the use of neuroimaging examinations within the judicial system.

Privacy is a good that, recent experience has shown, governments are often ready to 'trade' against security. The chapter by bioethicists Jonathan Moreno and Sonya Parashar discusses uses of brain imaging technology to enhance national security that have been of interest to defence agencies in the US, and the predictable repercussions on privacy. As Moreno and Parashar point out, the ordinary person's 'reasonable expectations' of privacy that the law seeks to protect in normal peacetime conditions are unlikely to be heeded once national security is threatened. Members of the armed forces, moreover, automatically cede virtually all their privacy rights and, if the interests of defence are thought to require military personnel to use unapproved devices, their 'informed consent' may be dispensed with. Research in this area has focused on the development of

accurate lie-detection equipment, based on brain imaging, which, we can imagine, might play a useful role in future intelligence activity. In addition, US defence officials have been interested in the contribution that brain imaging might make to personality assessment and other psychological measures. Information from this source about relevant aspects of potential soldiers' capacities, for example, might in the future guide military recruitment.

Concerns about the broader potential social and cultural impact of brain imaging, some of which are touched on in the chapters in the first three parts of the volume, are explored and discussed in greater depth in the essays in Part Four.

Noting people's tendency to react with alarm to the idea that brain imaging technology might one day allow others to read our minds, the philosopher Sarah Richmond explores an imagined possible scenario in which, as a result of this technological development, our minds have in fact become transparent to others: the Transparency Scenario. In response to the suggestion that this situation would realise our worst 'Big Brother' fantasies, she points out that the right future to think about (at least as things stand) is not a 1984-style police state in which the latest means of surveillance are oppressively used against the citizens, but a society in which reasonable decisions have been made on the basis of hard thinking about how the technology should be regulated to allow people maximally to benefit from it, and minimally be harmed. Richmond's paper examines some possible effects of 'transparency' on our interpersonal transactions and relationships, pointing out that alongside some clear costs of the incursion into mental privacy, there could also be gains. An unsettling question remains, that philosophy cannot answer, about how far, in the envisaged circumstances, people—and their expectations of each other—might change.

Another philosopher, Annabelle Lever, approaches the utopian hopes and dystopian fears raised by neuroscientific developments from a different perspective. Pointing out that there are many potential actions that we would greatly fear if we had not internalized moral and legal norms that prohibit them, she argues that the most important task, in responding to the challenges posed by neuroscience, is to advance and clarify our understanding of the nature and value of privacy. This is by no means straightforward: privacy is a notoriously fuzzy concept, whose diversity of instances has led some writers to deny that it has any unified meaning. Moreover, as Lever observes, disagreement about the value of privacy is common: is it, as some communitarians suggest, often over-valued, at the expense of other important goods? Lever shows that these issues can be productively approached by considering the various ways in which the protection of privacy can advance our interests as participants in a democratic society. An excellent example is the secret ballot, the protection of which plays an important role in enabling us freely to express our choice of political representatives. As with other technological advances, Lever claims, the case of neuroscience highlights the importance of the humanities and social sciences in understanding how we might best respond to them.

Roger Brownsword, a lawyer whose research has focused especially on the law in relation to technological developments, considers the challenge set for regulators and policy-makers by brain imaging technology. Helpfully focusing on a narrow sense of the

term ‘privacy’, appropriate to the brain imaging context, Brownsword considers various conceptions that regulators could adopt to protect it. While the frequent appeal, in law, to someone’s ‘reasonable expectation of privacy’ might appear to provide an attractive benchmark, Brownsword warns that in a world in which new ways of finding things out about people are increasingly used, expectations are *also* likely to change, with the effect that the boundaries of the private are repeatedly driven back. This chapter also considers the ingenious resources that technology can now provide regulators with: the range of so-called PETs (privacy enhancing technologies—including filters, settings, passwords) that can be used to block the ‘wrong’ people from accessing protected information. But reliance on these measures, Brownsword thought-provokingly points out, runs a paradoxical risk: by taking the element of choice away from agents, these methods threaten the moral sensibility, and even the dignity, of the community. Retention of our dignity as responsible moral agents may require us to choose—rather than to be forced to—obey a norm.

In her chapter, the bioethicist Sarah J.L. Edwards looks in detail at the ways in which privacy is currently protected by the laws and regulations applicable to brain scanning procedures, and in particular at current practice in relation to informed consent. Edwards points out that it is likely that this protection will be insufficient in the future, given the ways in which we can reasonably expect neuroimaging technology and its uses to develop. One development already under way (and discussed elsewhere in this volume) is the use of brain scans for commercial purposes. Unless additional controls are introduced, future patients may find themselves consenting in fairly broad terms to a brain scan, ordered for medical purposes in a hospital environment, that generates data that is subsequently put to a commercial use that they had not intended to support. Edwards emphasizes the high risk of ‘incidental findings’ posed by brain imaging in comparison with other familiar investigative tests, which further complicates the possibility of securing a person’s genuinely ‘informed’ consent in advance. As neuroscientific research extends our ability to ‘read’ information of a personal nature from brain images, the ‘sensitivity’ of the information these images contain will be heightened. Brain images can already reveal information that the owners of the brains in question might prefer not to know (for example, that there are early signs of an unpleasant disease). We can easily imagine a great many other facts that are currently concealed, especially about our mental lives, rising to the surface in the content of images of our brains. As Edwards reminds us, there is an urgent need for public debate about the measures that may be required in future to protect the sort of world in which we would want to live. The essays collected in this volume represent an early contribution to this debate.