Neuroethics

Defining the issues in theory, practice, and policy

Edited by

Judy Illes

Director, Program in Neuroethics and Senior Research Scholar
Stanford Center for Biomedical Ethics
and
Senior Research Scholar,
Department of Radiology,
Stanford University,
Stanford, CA, USA

OXFORD
UNIVERSITY PRESS

Contents

Chapter 4

Brains, lies, and psychological explanations

Tom Buller

Introduction

As the contributions to this book attest, neuroscience is a hot topic. Advances in neuroscience, particularly in neuroimaging and psychopharmacology, now enable us to correlate psychological states with specific brain functions, and to alter psychological states by modifying neurochemistry. As Farah and Wolpe (2004) put the matter succinctly, we are now able to 'monitor and manipulate' our brain function and thereby to affect our thoughts and actions by intervening at the neurobiological level.

In broad terms, these advances have two sorts of implications for ethics. As Roskies (2002) describes, we can distinguish the 'ethics of neuroscience' from the 'neuroscience of ethics'. In the first category we can place a host of ethical questions, some of which are familiar to us from earlier parallel advances and discussions in genethics, for example those pertaining to the ethics of enhancement (Parens 1998; see also Chapter 6), or to the predisposition and prediction of behavior (Raine *et al.* 1998), or to concerns about privacy (Kulynych 2002). In the second category we can place discussions about how these advances in neuroscience are providing us with a scientific naturalized account of how we make moral judgments and, more controversially, the extent to which the neuroscientific evidence suggests a revision of the belief that we are intentional rational agents who are capable of voluntary action.

The challenge neuroscience presents for neuroethics is that there is a tension between the two categories. The types of issues raised by the 'ethics of neuroscience', for example whether cognitive enhancement diminishes respect for persons (Kass 2003), rest on the belief that our folk psychological understanding of ourselves as intentional rational agents is correct. Cognitive enhancement raises ethical questions because it is consistent with agency and psychological explanations; we want new drugs to enhance our memories, desires, and actions. However, the types of issues raised by the 'neuroscience of ethics', for example the discovery of the neural correlates for our personality, preferences, or attitudes appears to challenge the notion that our behaviors are the result of intentional rational decisions. The neuroscience of ethics, like the neuroscience of anything, is a normative enterprise that purports to achieve an explanation of this phenomenon through a naturalistic view of psychobehavioral states in terms of their underlying neurobiology. Prior to the identification of the neural correlate for a particular psychobehavioral state we would attempt to explain the behavior in terms of the person's psychological states, for example lying in terms of intentional deception; subsequent to the identification of the neural correlate, lying is explained in terms of the neurobiology of this behavior. Unless we are prepared to accept neuroscientific explanations merely in terms of token-neurobiological states, and thereby lose any explanatory gain, our naturalistic account

must explain psychobehavioral states in terms of type-neurobiological states, rather than in the context of this person's particular beliefs and intentions. The tension between these two categories could be released if we could formulate a naturalized neuroscientific account of human behavior that could accommodate the belief that we are intentional rational agents. My goal in this chapter is to argue that such an accommodation will be hard to find.

Two visions of the mind

Until comparatively recently, neuroscience did not generate much public discussion or controversy. Within the philosophical community, neuroscience has been a familiar topic to philosophers of mind; however, it has been largely ignored by those working in ethics. The explanation for this is twofold: first, discussion about finding the neural correlate for belief, emotion, consciousness, or depression was based on expectation rather than on evidence; secondly, doctrinaire philosophy maintained that descriptive neuroscience had little to do with normative ethics.

A wealth of new data suggests that we have moved from expectation to evidence, and even if we are skeptical that we will reach a comprehensive mature neuroscience as claimed by some, there is little doubt that neuroscience will continue to find important correlations between neurological and psychological states. Since it is not an empirical claim like the first, the second of the two explanations above appears less susceptible to revision by advances in neuroscience. Is neuroethics merely the latest attempt at a naturalized ethics? Following Hume (1739) and Moore (1902) it has been customary to insist on the distinction between fact and value and to avoid committing the naturalistic fallacy, the error of drawing normative conclusions from descriptive empirical facts (Held 1996); although a number of recent discussions that have re-examined this issue (Johnson 1993; May *et al.* 1996; Casebeer 2003). If neuroethics amounts to no more than the 'science of ethics' and the grounding of moral judgments on empirical facts, then perhaps we should be dismissive of the entire enterprise. However, as noted above, this is only half the picture. If we consider some of the issues that have provoked discussion in the neuroethical literature (e.g. cognitive enhancement, psychopharmacology, and forensic neuroimaging), it is not the case that we have simply identified the neurological basis for intelligence or depression or truth-telling and, on the basis of this, set our ethical parameters; rather, it is that these new technologies and abilities raise ethical questions.

As Flanagan (2002) argues, the challenge we face is to reconcile two 'grand images of who we are, the humanistic and the scientific'. According to the humanistic view we are, for the most part, autonomous intentional rational agents and therefore we are responsible for our actions. Contrariwise, according to the scientific view, we are animals, exhaustively described in physical terms and constrained by the ordinary laws of cause and effect. Thus on the scientific view human behavior is not caused by our intentions, beliefs, and desires. Instead, free will is an illusion because the brain is a deterministic physical organ, and it is the brain that is doing the causal work.

The reconciliation of these two visions is essential for neuroethics. Neuroethics has emerged because advances in the neurosciences present us with ethical questions; however, there would be no questions at all if we believed that ethics and neuroscience were incompatible. Consider, for example, the question as to whether neuroimaging for lie detection should be admissible in court. This is an ethical question only if there is nothing incompatible about holding (a) that we are intentional agents who sometimes attempt to deceive, and (b) that we can determine whether a person is being deceptive by directly observing brain activity. If one held that these

two claims were incompatible, then there can be no ethical question; that is, if the identification of the neural correlate of lying were taken as evidence that psychological states are just neurological states, and hence that our folk psychological concepts are, in fact, false, then it is incoherent to suppose that neuroimaging could reveal whether a person is lying. Similarly, Patricia Churchland's claim that neuroscience will be able to distinguish 'in-control' from 'out-of-control' behavior makes sense only if there is no contradiction between neuroscience and intentional voluntary human action. (Moreno 2003; P.S. Churchland 2002; see also Chapter 1). Thus neuroethics is possible, in a transcendental sense, if and only if the scientific and humanistic visions can be reconciled. In other words, neuroethics makes sense only if eliminativism and determinism are rejected (Moreno 2003; see also Chapter 3).

Consciousness and rationality

What is to say that we are intentional rational agents capable of voluntary action? Broadly, this means that we act on the basis of our conscious (and, perhaps, unconscious) beliefs and desires, that we are capable of weighing alternatives, and choosing and executing a course of action, and that the resulting actions follow directly from our thoughts and reasons for action. It is evident that consciousness and rationality are necessary conditions of agency; in order for a person to be an agent the person must be able to make conscious decisions, i.e. be *aware* of his own motivations, intentions, and beliefs, and the consequences of these actions to a meaningful extent. In this regard, agency requires self-consciousness. There are, of course, degrees of consciousness and rationality in this context, but if a person were substantially unaware of his own motivations, intentions, and beliefs, or was incapable of weighing alternatives and choosing a course of action, then it would be difficult to maintain that this person qualified as an agent. Therefore it would be difficult to hold this person responsible for action.

As Davidson (1980) has argued, an important component of rationality is the interconnectedness of psychological states and the role that individual psychological states play within a person's overall psychology. If we know that John is feeling cold and desires to get warm, and that he believes that closing the window will help, then (other things being equal) John will close the window. In a similar vein, if Sally hates cats and she has to visit her friend who has 20 of them, we can predict her opinion of the experience. This normative component of rationality plays an important role when we consider responsibility. If John knows that throwing a rock will harm the children playing and he goes ahead and throws the rock, then, other things being equal, we may conclude that he must have wanted to harm the children.

Can these conceptions of consciousness and rationality that underlie agency be accommodated by neuroscience? The last 30 years have witnessed a vast number of articles and books on the topic of consciousness, and there is a spectrum of positions on this topic. These range from eliminativists and reductionists on the one side (P.S. Churchland 1989, 2002; P.M. Churchland 1992) to non-reductionists (Chalmers 1996; Davidson 1980), dual-aspect theorists (Nagel 1974, 1986), and 'mysterians' (McGinn 1993). According to neuroscientific reductionism, empirical discoveries will identify elements in the reducing theory (neuroscience) that correspond to elements in the reduced theory (folk psychology) in conjunction with 'bridge laws' connecting the elements of both theories (P.S. Churchland 1989, 2002; Bickle 2003). It is customary to think of reductionism as making both ontological claims about the identity of properties between the reduced and the reducing theories, and explanatory claims about the appropriate way to describe the world. Ideally, neuroscience will provide us with a more accurate and successful way of explaining human behavior than folk psychology, for we will be able to

explain in greater and more accurate detail why a person is feeling depressed or paranoid or has a constant craving for food. It is for these reasons, one supposes, that Ramachandran (2003) claims that psychiatry will be reduced to neurology. As Ramachandran discusses in the case of hysteria and other 'psychological disturbances', this disturbance can be usefully explained in terms of what is happening in the person's brain.

The goal of a reductive neuroscience is to provide us with greater explanatory success. In principle, this is achieved by identifying correlations between psychological states and neur biological states, and the added predictive and explanatory value of neuroscience; for instance, the identification of the function and role of neurotransimitters such as dopamine and serotonin provide us with us with a better way of explaining and understanding depression. As part of this reduction, it is predicted that neuroscience will reveal that some parts of our folk psychology are mistaken. This might occur, for example, if we were to find that our folk psychological notions of consciousness and free will do not correlate with the emerging neuroscientific evidence (P.S. Churchland 1989; Ramachandran 2003; Greene and Cohen 2004). As one example we can consider the research by Libet (1985) that appears to undermine the notion of free will through the identification of neurological activity that precedes a conscious decision to act. For the sake of argument, let us accept that Libet's studies do show that my conscious decision to raise my arm was preceded by brain activity—the 'readiness potential'. What impact should these results have on the conceptions of rationality and consciousness that underlie agency and responsibility? Should we revise these in order to match the neuroscientific facts? While revision may be the appropriate course of action to take, why should we choose this course of action rather than draw conclusions about the limitations of neuroscience? What we are faced with is not simply a choice between the scientific and humanistic visions but a philosophical dispute about the meaning of terms such as 'free will', 'rationality', and 'consciousness.' According to the reductionist neuroscientific approach, empirical evidence will either show that these terms refer to particular neurobi logical states or that they are dispensable terms belonging to an outdated mysticism. According to the non-reductionist humanistic approach these terms refer to psychobehavioral states, and therefore the fact that they have no neurobiological correlate is, frankly, unimportant. To put the matter differently, if we think that free will can be confirmed or denied by empirical evidence, then we should be concerned about Libet's findings; however, if we do not believe that free will is an empirical notion, then these findings are irrelevant. To repeat, this is not to deny these neurobiological findings; it is only to say that their importance depends upon a prior commitment to the ontological and explanatory success of neuroscientific reductionism.

As Morse (2004) argues, emerging neuroscientific information is relevant if it provides an explanation of why a person is capable of acting rationally; however, without further argument, there is little justification as to why this new information should be used as a determination that the person is incapable of acting rationally, independent of any social or pragmatic reasons (see also Chapter 3). Suppose that neuroscience will further reveal how a person's decision-making capacities are adversely affected by stress or depresion. What relevance does this information have for our notion of rationality? Perhaps new neurobi logical information will ead us to consider revising our moral and legal notions of rationality. However, if we do decide to do so, it would be for social or pragmatic reasons. Independent of such reasons, it is difficult to see how this new information could be used without committing the naturalistic fallacy. Certainly, there are good parsimonious reasons why one might wish to argue for the ontological reduction of psychological states to

neurobiological states but, particularly in the context of neuroethics, one needs further arguments to show that such a reduction will have an appropriate degree of explanatory success.

Minds, brains, and agents

In the first chapter of his book *Descartes' Error*, Damasio (1994) recounts the now well-known story of Phineas Gage. Gage was a hardworking, even-tempered, and respected construction foreman who worked laying track for the Rutland and Burlington Railroad. On 13 August 1848, Gage was helping to set detonations used to clear rock from the future course of the rail track. Momentarily distracted, Gage forgot to put sand into the hole in the rock in which the explosive powder had already been placed. He then inserted the tamping iron, a 3ft 7in iron bar weighing 13 pounds (Macmillan 2004) into the hole to pack down the sand but the iron caused a spark which ignited the explosive powder. The explosion sent the tamping iron up through the base of Gage's skull, through the front of his brain and out through the top of his head, eventually landing more than 100 ft away (Damasio 1994). Remarkably, Gage was not killed by the accident; however, after he recovered it became clear that 'Gage was no longer Gage.' His personality and behavior had changed, and he became profane, impatient, irreverent, and unreliable.

The tragic accident that befell Phineas Gage has helped us to understand better how injury to a particular part of the brain can lead to specific types of behavioral effect. As Damasio states, '[T]here is no question that Gage's personality change was caused by a circumscribed brain lesion in a specific site' (Damasio 1994; see also Chapter 2). Recent advances in neuroimaging have greatly increased our ability to locate brain abnormalities and have helped to identify these abnormalities as explanations of psychological and behavioral changes (P.S. Churchland 2002; Ramachandran 2003). Furthermore, through the identification and study of abnormal cases, we are able to gain a clearer understanding of the relationship between brain function and behavior. In this regard, one can see an obvious parallel between the growth of knowledge in genetics and in neuroscience; abnormal states of affairs (diseases, disorders) have been linked to damage to a particular gene or part of the brain which, in turn, has enabled us to identify the function of the gene or part of the brain in the normal case.

As Patricia Churchland (2002) has argued, the distinction between being in control and being out of control is not a sharp one but a matter of varying degrees. A person may have greater ability for freedom of choice with regard to some types of decisions than others, and this ability may be affected by a variety of factors, including damage to particular parts of the brain, and the levels of neuromodulators and hormones. It is plausible to contend, as Churchland does, that we will be able to identify the major neurobiological factors that underlie a person's ability to be in control and, conversely, those that undermine it. On the basis of the above neuroscientific evidence we may be able to identify the factors that impede a person's ability to be in control, and hence, in the absence of such factors, we will by default have an account of the neurobiological factors that are consistent with being in control; nevertheless, I want to argue that this does not warrant the claim underlying Churchland's hypothesis.

To begin, the neuroscientific evidence cannot on its own be the basis for the distinction between being in control or out of control. The evidence is not sufficient for this task because being in control or having freedom of choice are normative notions distinct from the

neuroscientific facts. Independent of social standards and expectations, it is difficult to understand what it would mean to say that one type of neurological functioning is any more or less in control than any other, or that a behavior is any more or less rational. We are able to gain some idea of the normal level of the neuromodulators because, *ex hypothesi*, we have discovered in an important set of cases that individuals who were not in control had abnormal levels. In the absence of these behaviors and society's judgment of them, one neuromodulator level is no more normal than any other.

Furthermore, what counts as being in control or being out of control will vary according to time and place; what is viewed in Richmond as refined self-possession may in Reyjkjavik be regarded as wanton excess. Less figuratively, different cultures may have different ideas as to what constitutes depression or hyperactivity. If this is correct, then it is either the case that the notion of being in control cannot be grounded in neuroscientific terms, or we have to be generous in these terms and admit a range of possible neurological states as underlying this notion. The first of these options suggests that we should abandon or revise our hopes for a successful ontological reduction of psychobehavioral states to neurobiological states; and the second questions the possibility of successful explanatory reduction. If the neuroscientific evidence does not match up with our social or ethical notions of what constitutes decision-making capacity, then it is difficult to see how neuroscience can successfully explain this capacity in neuroscientific terms.

Moreover, it is not at all clear what we gain in explanatory terms by being able to distinguish in-control from out-of-control behavior in neuroscientific terms. For the sake of argument, let us grant that for each and every psychological state or activity there is a correlative neurophysical state. Accordingly, we should not be surprised that the distinction between being in control and being out of control correlates with neurophysical states. But correlation is not the same as explanation. It does not follow from the fact that psychobehavioral differences correlate with neurophysical differences that differences at the neurophysical level explain the differences at the psychobehavioral level. We might discover that there are neurophysical differences that correspond to the distinction between introverts and extroverts. Does this mean that the neurophysical differences explain the difference between introverts and extroverts? Not necessarily, because the neurological differences may have no relation to the psychological aspects that are relevant to a person's being an introvert or an extrovert.

Neuroscience can inform our notion of intentional rational action by revealing how abnormal brain function can explain a person's behavior. In this way a judge might decide that Smith is not responsible for his actions if Smith's actions or volitions are beyond his control. In such a case we think it appropriate to explain Smith's behavior in these terms not only because we believe that the cause of his actions and/or volitions was not intentional, but also because we believe that this enables us to better predict what he will do. However, in the normal course of events we do not attempt to explain behavior in neurophysical terms because we simply take it for granted that neurophysical explanations are irrelevant. If Jones throws a brick through a shop window it is no excuse to say that his brain did it. Since all our psychobehavioral states correlate in some fashion with neurophysical states, we should either jettison completely the belief that we are intentional rational agents or deny neuroscience any broad explanatory role in the normal case. Neuroscientific explanations become relevant only in those cases where we have independent reasons for claiming that the person's behavior is not intentional or the person is not capable of acting rationally. In such cases, neuroscience can provide an explanation of why this is so.

Neuroimaging and 'brain-reading'

Neuroscientific reductionism can be appropriately categorized as an internalist theory since the theory claims that the content of a person's psychological state can be individuated according to the state of that person's brain (see also Chapter 2). For example, to say that Sally believes that it is snowing is to make a remark about the current state of he brain; if the content of her psychological states were to change, then, according to our best neuroscientific theory, we could infer that the state of her brain had changed.

This internalist perspective underlies an issue that has already provoked considerable discussion in the neuroethics literature, namely the use of neuroimaging to 'read' a person's mind. Discussion of this issue has divided into two topics: firstly, the use of neuroimaging to detect whether a person is telling the truth (Farah 2002; Langleben *et al.* 2002; Farah and Wolpe 2004) or to identify a person's attitudes or preferences (Hart *et al.* 2000; Phelps *et al.* 2000); secondly, the threat that neuroimaging poses to our privacy (Kulynych 2002; Farah and Wolpe 2004; Wolpe *et al.* 2005; Illes and Racine 2005).

Two potential examples of lie detection by neuroimaging have been discussed in the literature. First, Langleben and his colleagues performed a study that involved subjects bluffing versus telling the truth about symbols on playing cards (Langleben *et al.* 2002). As a result of the experiments, they identified the anterior cingulate cortex and superior frontal gyrus as 'components of the basic neural circuitry of deception', and they drew the conclusion that 'cognitive differences between deception and truth have neural correlates detectable by fMRI' (Langleben *et al.* 2002). Secondly, Farwell has described a brain fingerprinting device that, according to him, can detect whether a person recognizes a word or phrase shown on a computer screen according to changes in the P300 event-related potential (Farewell 1991; Knight 2004). Farwell claims that, using this device, one can determine whether a suspect was familiar with the details of a crime scene, and hence the device could be used to determine whether the person was telling the truth. With regard to attitudes and preferences, two studies have reported finding neurological evidence that our brains respond differently to faces from our own or a different racial group (Hart *et al.* 2000; Phelps *et al.* 2000).

As Illes and Kulynych have argued (Illes and Raffin 2002; Kulynych 2002; Illes 2003; Illes *et al.* 2003), neuroimaging raises important concerns about a person's right to privacy and confidentiality. These concerns apply not simply to the accidental revelation of medical information (e.g. a fMRI scan inadvertently revealing an abnormality) but also to a person's preferences or knowledge. If neuroimaging can reveal a person's psychological states (thoughts, attitudes, or preferences), then it is clear that privacy concerns are raised. (One need only consider the potential use of such machines by security services, law enforcement, or marketing firms).

As described, the ethical questions raised by neuroimaging depend upon an internalist conception of the mind, for these questions are predicated on the understanding that, in principle or in fact, we can determine what a person is thinking by examining that person's brain. Clearly, there are privacy concerns associated with neuroimaging, but is the notion of 'brain-reading' coherent and a legitimate cause for concern? I think that there are a number of reasons why one might challenge the internalist conception of the mind that underlies some of the ethical concerns associated with neuroimaging. In a recent paper discussing the 'promises and perils' for emerging lie-detection technologies, Wolpe *et al.* describe various neuroimaging technologies and research, including that of Langleben. As the authors thoroughly discuss, there are good reasons to be cautious about

the conclusions that one can draw about lie detection and neuroimaging outside the parameters of the study design. Nevertheless they are prepared to conclude that 'there are fundamental neurological differences between deception and truth at the neurological level…' (P. Wolpe *et al.* 2005).

I want to raise a number of reasons why one might resist this internalist picture and the conclusion drawn above. First, one might object to this conclusion on the grounds that the tests reveal not truth-telling or lying but our beliefs about how society views lying. For example, Langleben *et al.* (2002) report that subjects did not report being anxious during the test. However, the state of anxiety is distinct from the belief (albeit perhaps false) that one ought not to lie. Perhaps the neurological feature that the tests reveal is not lying, but the belief or sentiment that one ought not to do something that is generally condemned. In a society of liars, would the neurological features associated with deception be associated instead with truth-telling?

More importantly, perhaps, there are many different types of lies: prevarication, tact, 'white lies', self-deception, and lies told as part of a game or a play (Illes 2004). If there were a neurological feature that correlated with the distinction between truth-telling and deception, then we would expect all these types of lying to be correlated with this neurological feature. It is, of course, an empirical question whether we will find such a correlation, but given earlier remarks about how our normative notions are culturally variable, I believe that the class of psychobehavioral states classified as lying will have no useful explanatory correlation at the neurophysical level. And if we do not find the appropriate correlations between neuroscience and the psychobehavioral states, then what implications does this have for our folk psychological notion of lying? Furthermore, problems are raised by mistaken cases of lying. Imagine that Jones believes that John Kerry won the last US Presidential election but lies to his wife and tells her that George W. Bush won. Technically, we might suppose that neuroimaging would reveal that Jones was lying, but even if he is shown to be doing so by this technology, the question of whether, in fact, he is lying is still open. A related objection is that we might find that one type of lie may be realized by different neurophysical states—the problem of multiple realizability. In this case it would still be true that each type of lying corresponds to a neural correlate but we gain little or no explanatory advantage by identifying lying in this way.

Finally, there is also a perhaps more fundamental concern about neuroimaging and 'brain-reading' that pertains directly to the internalist perspective. In their book *Philosophical Foundations of Neuroscience*, Bennett and Hacker (2003) criticize neuroscientific reductionism on the grounds that it commits a 'mereological fallacy', i.e. the fallacy of ascribing properties to parts that should be ascribed to the whole. According to their Wittgensteinian perspective, terms such as 'believes', 'reasons', 'understands', or, for that matter, 'lies' should be properly thought of as descriptions not simply of mind but of behavior. To say that Smith understands Spanish or believes that it is going to rain is to say something about how Smith will behave, what other beliefs Smith will have, and what statements Smith will agree to. All of this is to say that we should understand particular psychological states in terms of the role that they play within a person's overall psychology.

Bennett and Hacker's conclusion bears directly on neuroimaging studies that claim to detect a neurological basis for certain preferences or attitudes (Hart *et al.* 2000; Phelps *et al.* 2000). In their study, Hart and colleagues report neurophysical differences between how individuals respond to images of members of their own racial group and how they respond to images of members of a different group. As a result of this neuroscientific evidence, would we be justified in claiming that a person's racial preferences have a neurobiological basis?

Imagine that a neuroscientific study revealed that there was a neurological difference between how people respond to accents that are similar to or different from their own. The results of this study do not warrant any general conclusions about people's preferences about accents because it is not clear that there is any obvious connection between how a person's brain might respond to a certain auditory event and that person's moral or general preferences. There is nothing incoherent about the possibility that someone might respond differently (more favorably?) in neuroscientific terms to English accents than to Welsh accents, yet believe that Welsh accents are much more pleasant.

The objection here pertains to the interconnectedness of our psychological states, preferences, and attitudes. If we know that Davis prefers Welsh accents, then we can plausibly infer that, other things being equal, he would prefer to live in Wales or to be surrounded by his Welsh friends—this is all part and parcel of what it means to have a preference. But this interconnectedness of psychological states has no echo at the neurological level. Detecting neurologically that Davis responds differently (more favorably?) to Welsh accents than to English ones has no implications for any other neurophysical states, including those that correlate with the other psychological states that are the substance of his preference.

A modest accommodation

In this chapter I have attempted to argue that there are challenges to a successful accommodation between a naturalized neuroscientific account of human behavior and the belief that we are intentional rational agents. In particular, I have tried to show, first, that the internalist perspective that underlies reductionist neuroscience and some of the current neuroethical issues is at odds with our folk psychological notions. Secondly, I have argued that the relevance of neuroscience to our notion of agency is derivative upon prior normative considerations, and hence that the impact of neuroscientific advances on matters of agency and responsibility is limited.

As advances in neuroscience improve our ability to monitor and manipulate brain function, so the discussion of the ethics of such actions will continue. These advances are also ensuring that the discussion continues on fundamental philosophical questions regarding the status of folk psychology and the reach of science. It is these two sets of discussions that make neuroethics such an exciting field and, as long as we can continue to distinguish normative from metaphysical questions, neuroethics will continue to grow.

References

Bennett MR, Hacker PMS (2003). *Philosophical Foundations of Neuroscience.* Oxford: Blackwell.

Bickle J (2003). *Philosophy and Neuroscience: A Ruthlessly Reductive Account.* New York: Springer.

Casebeer WD (2003). *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition.* Cambridge, MA: MIT Press.

Chalmers DJ (1996). *The Conscious Mind: In Search of A Fundamental Theory.* New York: Oxford University Press.

Churchland PM (1992). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* Cambridge, MA: MIT Press.

Churchland PS (1989). *Neurophilosophy: Toward a Unified Science of the Mind–Brain.* Cambridge, MA: MIT Press.

Churchland PS (2002). *Brain-Wise: Studies in Neurophilosophy.* MIT Press, Cambridge.

Churchland PS (2004)

Damasio AR (1994). *Descartes' Error: Emotion, Reason and the Human Brain.* London: Picador.

Davidson D (1980). *Mental Events: Essays on Actions and Events.* Oxford: Clarendon Press.

Farah MJ (2002). Emerging ethical issues in neuroscience. *Nature Neuroscience* **5**, 1123–9.

Farah MJ, Wolpe PR (2004). Monitoring and manipulating brain function: new neuroscience technologies and their ethical implications. *Hastings Center Report* **34**, 34–45.

Farwell and Donchin (1991). The truth will out: Interrogative Polygraphy ('lie-detection') with event-related potentials, *Psychophysiology* **28**, 531–547.

Flanagan O (2002). *The Problem of the Soul: Two Visions of the Mind and How to Reconcile Them*. New York: Basic Books.

Greene J, Cohen J (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London, Series B* **359**, 1775–1885.

Hart AJ, Whalen PJ, Shin LM, McInerney SC, Fischer H, Rauch SL (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neuroreport* **11**, 2351–5.

Held V (1996). Whose Agenda? Ethics versus Cognitive Science. In L. May, M. Friedman and A. Clark eds. Mind and Morals: Essays on Ethics and Cognitive Science, MIT Press, Cambridge, 6–87.

Hume D (1739). *A Treatise of Human Nature* (ed. Selby-Bigge LA). Oxford: Clarendon Press, 1967.

Illes J (2003). Neuroethics in a new era of neuroimaging. *American Journal of Neuroradiology* **24**, 1739–41.

Illes J (2004). A fish story? Brain maps, lie detection and personhood. *Cerebrum* **6**, 73–80.

Illes J, Racine E (2005). Imaging or imagining? A neuroethics challenge informed by genetics. *American Journal of Bioethics* **5**, 1–14.

Illes J, Raffin TA (2002). Neuroethics: an emerging new discipline in the study of brain and cognition. *Brain and Cognition* **50**, 341–4.

Illes J, Kirschen M, Gabrieli JD (2003). From neuroimaging to neuroethics. *Nature Neuroscience* **6**, 205.

Johnson M (1993). *Moral Imagination: Implications of Cognitive Science for Ethics*. Chicago, IL: University of Chicago Press.

Kass L (2003). *Beyond Therapy: Biotechnology and the Pursuit of Happiness*. New York: Harper Collins.

Knight J (2004). The truth about lying. *Nature* **428**, 692–694.

Kulynych J (2002). Legal and ethical issues in neuroimaging research: human subjects protection, medical privacy, and the public communication of research results. *Brain and Cognition* **50**, 345–57.

Langleben DD, Schroeder L, Maldjian JA, *et al.* (2002). Brain activity during simulated deception: an event-related functional magnetic resonance study. *NeuroImage*, **15**, 727–732.

Libet B (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* **8**, 529–66.

McGinn C (1993). *The Problem of Consciousness*. Oxford: Blackwell.

Macmillan M (2004). The Phineas Gage homepage. http://www.deakin.edu.au/hbs/GAGEPAGE/ (accessed 22 April 2005).

May L, Friedman M, Clark A (eds) (1996). *Mind and Morals: Essays on Ethics and Cognitive Science*. Cambridge, MA: MIT Press.

Moore GE (1902). *Principia Ethica*. London: Prometheus Books, 1988.

Moreno JD (2003). Neuroethics: An agenda for neuroscience and society. *Nature Reviews Neuroscience* **4**, 149–53.

Morse SJ (2004). *Neuroscience and the Law: Brain, Mind, and the Scales of Justice*. New York: Dana Press.

Nagel T (1974). What's is it like to be a bat? *Philosophical Review* **83**, 435–50.

Nagel T (1986). *The View From Nowhere*. New York: Oxford University Press.

Parens E (1998). *Enhancing Human Traits: Ethical and Social Implications*. Washington, DC: Georgetown University Press.

Phelps EA, O'Connor KJ, Cunningham WA, *et al.* (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience* **12**, 729–38.

Raine A, Meloy JR, Bihrle S, *et al.* (1998). Reduced prefrontal and increased subcortical brain functioning assessed using positron emission tomography in predatory and affectivemurderers. *Behavioral Science and Law* **16**, 319–32.

Ramachandran VS (2003). *The Emerging Mind*. London: Profile Books.

Roskies A (2002). Neuroethics for the new millenium. *Neuron* **35**, 21–3.

Wolpe RP, Langleben DD, Foster, K (2005). Emerging neurotechnologies for lie detection: promises and perils. *American Journal of Bioethics* **5**, 15–26.