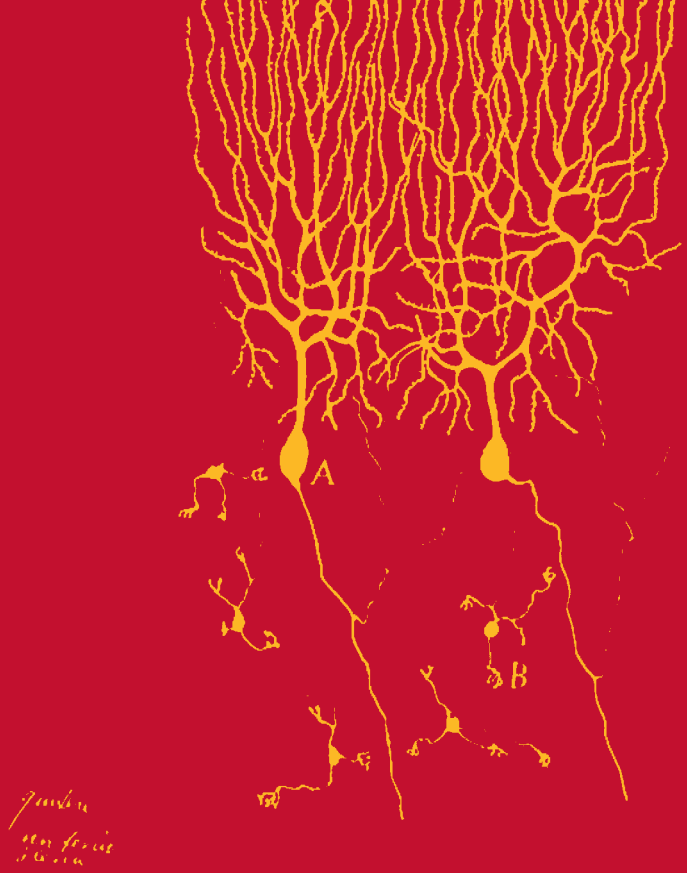


Brain Waves Module 1: Neuroscience, society and policy

January 2011

EXCELLENCE IN SCIENCE

THE ROYAL SOCIETY

Cover image: The cover image is a drawing of Purkinje cells (A) and granule cells (B) from pigeon cerebellum by Santiago Ramón y Cajal, 1899. Instituto Santiago Ramón y Cajal, Madrid, Spain.

Brain Waves Module 1: Neuroscience, society and policy

RS Policy document 01/11
Issued: January 2011 DES2015

ISBN: 978-0-85403-879-4

The Royal Society
Science Policy Centre
6–9 Carlton House Terrace
London SW1Y 5AG
Tel +44 (0)20 7451 2500
Email science.policy@royalsociety.org
Web royalsociety.org

Brain Waves Module 1: Neuroscience, society and policy

Contents

1 Introduction and summary

The *Brain Waves* project

Developments in neuroscience, the study of the brain and nervous system, are likely to provide significant benefits for society. Increasing understanding of the brain and advances in technologies to study it will enable improved treatment of neurodegenerative diseases, such as Parkinson's and Alzheimer's, and mental illnesses, including depression and schizophrenia. They will increase our insights into normal human behaviour and mental wellbeing, as well as enabling other enhancement, manipulation, and even degradation of brain function and cognition.

Neuroscience and neurotechnology reach far and wide into disparate fields. The array of 'neuro' disciplines lend themselves to applications in diverse areas of public policy such as health, education, law, and security. More broadly, progress in neuroscience is going to raise questions about personality, identity, responsibility, and liberty, as well as associated social and ethical issues. The aim of the Royal Society's *Brain Waves* project is to explore what neuroscience can offer, what are its limitations, and what are the potential benefits and the risks posed by particular applications.

This report is the first in a series of four 'modules'. Three subsequent modules will consider specific policy issues in more detail:

- Module 1: Neuroscience, society and policy;
- Module 2: Neuroscience: implications for education and lifelong learning;
- Module 3: Neuroscience, conflict and security;
- Module 4: Neuroscience, responsibility and the law.

Summary

This Module 1 report is a collection of essays that together provide a primer of current developments in neuroscience and highlight interesting issues and questions for society and policy. It is not intended as an exhaustive review of the science, nor to make specific policy recommendations. Rather, it raises key issues and questions, many of which will be explored in more depth in subsequent modules.

In Section 2 of the report we review the state of development of neuroscience and neurotechnology, discuss the translation of this knowledge into useful applications or inferences, and raise some of the implications. Section 2.1 explores neuroimaging. Here, we describe some of the existing tools used to study the brain, their limitations, and associated issues for healthcare, business, criminal justice, and privacy. In Section 2.2 we address developments in neuropsychopharmacology, their application for mental health medicines, wider issues of recreational use, as well as emerging applications for cognitive enhancement.

Neural interfaces and brain interference are the subject of Section 2.3, which

reviews those providing direct input to the brain to modulate activity and those that use outputs of brain activity to control external devices or prostheses. Applications in diverse areas of healthcare and the games industry are discussed. Sections 2.4 and 2.5 address what new developments in neuroscience may mean for our understanding of behaviour and decision-making. Section 2.4 argues that understanding of conscious and unconscious decision-making processes presents a challenge to our concepts of responsibility and free will. Section 2.5 explores how recent work in behavioural neuroscience is leading to insights about the neurobiological basis of economic decision-making.

In Section 3 we examine some of the opportunities and risks these scientific developments present, as well as the ethical questions they raise, and governance issues that need to be considered. Section 3.1 describes some of the benefits and opportunities in neuroscience, with a particular focus on new drugs to combat mental illness and cognitive decline, and the associated role of the private sector. Section 3.2 explores some of the potential risks of certain applications, such as the risks of 'medicalising' behaviour with pharmaceutical interventions, the potential dangers to human rights and privacy posed by invasive applications of neuroimaging, and the potential hazards to international security presented by the militarisation of neuroscience.

The burgeoning field of neuroethics is the subject of Section 3.3, which probes issues such as personal identity, physical and moral enhancement, responsibility, and artificial intelligence. Finally, in Section 3.4 we draw lessons from the wider governance of science and technology, suggesting approaches to marshalling neuroscience in order to maximise the benefits and minimise any risks.

Professor Colin Blakemore FRS, Chair, Steering Group, *Brain Waves* project

Acknowledgements

The *Brain Waves* project is overseen by a Steering Group, chaired by Professor Colin Blakemore FRS. These experts also serve as members of three broader Working Groups for Modules 2, 3, and 4. The Royal Society would like to thank all members of the Steering Group, most of whom have authored essays for this volume and all of whom have overseen its development and editing:

- Professor Colin Blakemore FRS (Chair), Professor of Neuroscience, University of Oxford;
- Professor Uta Frith FRS, Emeritus Professor, Institute of Cognitive Neuroscience, University College London;
- Professor John Harris, Director, Institute for Science, Ethics and Innovation (iSEI), School of Law, University of Manchester;
- Professor Nicholas Mackintosh FRS, Emeritus Professor, Department of Experimental Psychology, University of Cambridge;

- Professor Geraint Rees, Director, Institute of Cognitive Neuroscience, University College London;
- Professor Trevor Robbins FRS, Professor of Cognitive Neuroscience, Department of Experimental Psychology, University of Cambridge;
- Professor Steven Rose, Emeritus Professor, Department of Life Sciences, The Open University;
- Professor Barbara Sahakian, Professor of Clinical Neuropsychology, Department of Psychiatry, University of Cambridge;
- Professor Wolf Singer, Director, Department of Neurophysiology, Max Planck Institute for Brain Research, Germany;
- Professor Andy Stirling, Science Director, Science and Technology Policy Research (SPRU), University of Sussex;
- Professor Irene Tracey, Director, Oxford Centre for Functional MRI of the Brain (FMRIB), University of Oxford

We are very grateful for the contributions from:

- *Author of Section 2.5*: Professor Wolfram Schultz FRS, Wellcome Principal Research Fellow and Professor of Neuroscience, Department of Physiology, Development and Neuroscience, University of Cambridge; and
- *Co-author of Section 3.3*: Dr Sarah Chan, Research Fellow, School of Law, University of Manchester.

The essays in this volume were written by individual experts in their fields. There are some issues on which, necessarily, their views and conclusions differ, as well as their perceptions of benefits and risks. This range of views is not unusual, or even particular to the field of neuroscience, but demonstrates the diversity of opinion in what is still a relatively young area of science. The views expressed are those of the authors and do not necessarily represent the views of the Royal Society.

The Royal Society would also like to thank the Review Panel for reviewing and approving this report:

- Dame Jean Thomas DBE FRS (Chair), Biological Secretary and Vice-President, the Royal Society;
- Professor Tim Bliss FRS, Division of Neurophysiology, National Institute for Medical Research;
- Professor Barry Everitt FRS, Department of Experimental Psychology, University of Cambridge;
- Professor Karl Friston FRS, Scientific Director, Wellcome Trust Centre for Neuroimaging, University College London.

Science Policy Centre staff working on this project, and who have contributed to this report, are as follows:

- Dr Neil Davison, Senior Policy Adviser (Security and Diplomacy)
- Tessa Gardner, intern (September – December 2010)
- Sarah Mee, Policy Adviser

- Ian Thornton, Policy Adviser
- Dr Rochana Wickramasinghe, Policy Adviser (to September 2010)
- Rapela Zaman, Senior Policy Adviser

If you have any questions about this report please contact Dr Neil Davison in the first instance.

Science Policy Centre
The Royal Society
6-9 Carlton House Terrace
London SW1Y 5AG
tel: +44 (0)20 7451 2548
email: neil.davison@royalsociety.org

2 Contemporary neuroscience and technology

2.1 The scope and limits of neuroimaging

Professor Geraint Rees, University College London

Background

The term 'neuroimaging' refers to a group of technologies that allow some aspect of human brain structure or function to be measured without having to perform surgery. These technologies all measure the activity or structure of large populations of brain cells (neurons), in contrast with invasive techniques used in animals that measure the activity of individual neurons. Human neuroimaging instead measures aggregate signals from hundreds of thousands of neurons. However, unlike more invasive techniques, human neuroimaging acquires these signals from many tens of thousands of locations throughout the entire brain simultaneously. The techniques described below differ in their spatial resolution—how well they can distinguish between two points close together, affecting how 'sharp' the image is. They also vary in temporal resolution—the precision of the image with respect to time. There is often a trade-off between the two; however a combination of techniques can be used.

Neuroimaging techniques

Structural MRI

Structural neuroimaging acquires measurements of brain anatomy non-invasively. The surface of the brain (cortex) is a convoluted structure of folds and ridges visible to the naked eye. These are made up of a thin layer of gray matter (where neuronal cell bodies are found) underpinned by extensive white matter (or 'wiring'). Within the brain, subcortical nuclei of gray matter contain further neuronal cell bodies.

Computed tomography (CT) is a technique that uses x-rays to visualise brain anatomy in sections. It is deployed in hospitals throughout the developed world for immediate assessment of stroke and after head injury, due to its ability to detect rapidly bleeding within the skull or bone injury respectively. For research (and increasingly many clinical) purposes, it has largely been supplanted by magnetic resonance imaging (MRI).

The advantage of MRI is that the different components of brain structure, such as subcortical structures, white matter and gray matter, can be given different contrasts so that the detailed anatomical structure of the brain can be visualised. The typical spatial resolution of such structural images is between 0.5 and 1 mm^3, which represents anatomical groups of several hundred thousand neurons.

In the white matter of the brain, the diffusion of water is limited by the cell membranes of long, conducting neurons. Diffusion–weighted MRI measures the rate

and direction in which water molecules diffuse within the brain, and so can allow visualisation of the white matter. By applying mathematical 'tracking' algorithms to these images, the structure and direction of neuronal tracts that connect different brain regions can be visualised.

Functional MRI

Functional MRI (fMRI) refers to the use of MRI scanning to detect some aspect of brain function (rather than simply structure) (see Figure 1 and Figure 2). The most widely used type of functional MRI is known as Blood Oxygenation Level Dependent (BOLD) imaging. This measures

Figure 1 This is a MRI image of the visual cortex in one hemisphere of the brain. (The visual cortex is the part of the brain that is active when you look at something.) The dark grey stripes represent gray matter at the bottom of sulci and the light grey stripes represent gray matter associated with brain gyri. Overlaid on the inflated anatomical image are colored patches representing functional MRI signals associated with stimulating different regions of the visual field. (Reproduced courtesy of Geraint Rees, University College London.)

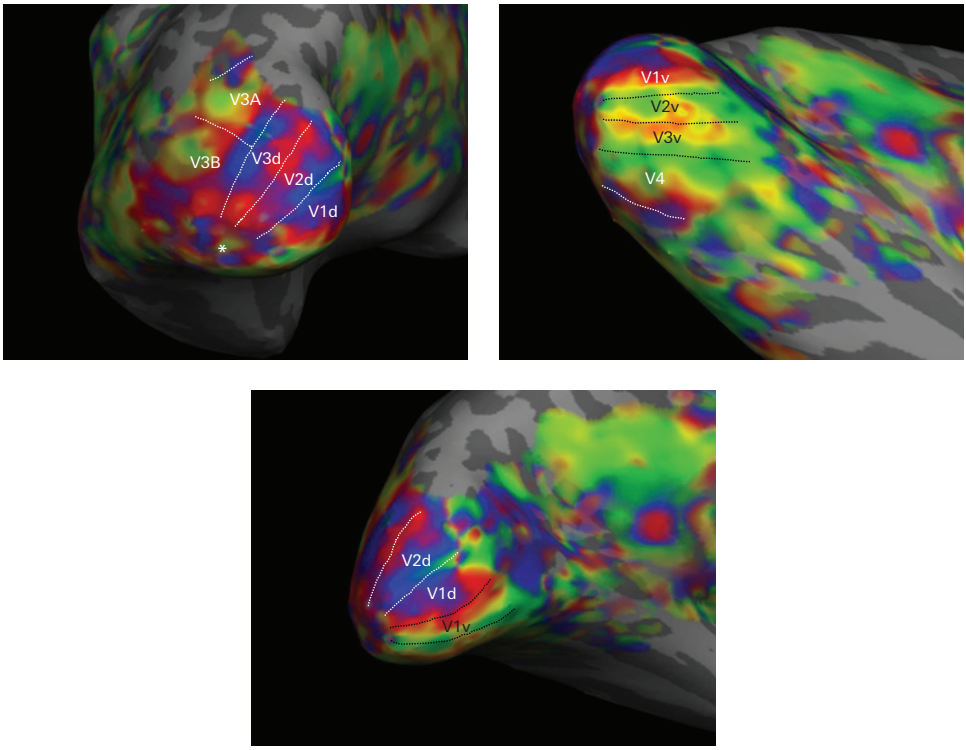
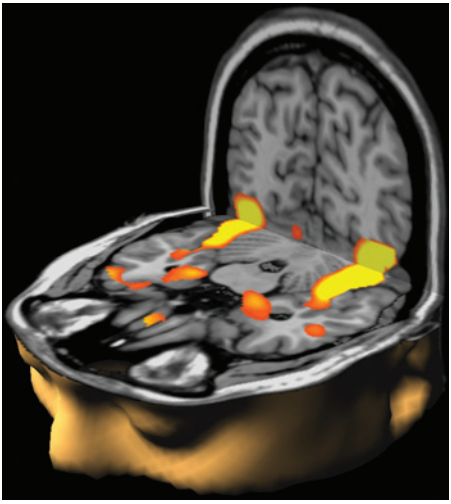


Figure 2 The visual cortex is highlighted in this brain image created using functional magnetic resonance imaging (fMRI). The surface of the brain has been expanded so the parts that are normally hidden away down the folds are 'blown out' and are shown as darker areas. (Reproduced courtesy of Mark Lythgoe and Chloe Hutton, University College London.)



changes in the level of oxygenation in the blood, which changes locally in the brain as neurons become active and consume oxygen. This leads to compensatory changes in blood flow to the active area that can be detected. BOLD contrast fMRI is therefore an indirect measure of neural activity, through the effects of changes in local blood flow in the brain. Because blood flow is regulated on a spatial scale of a few millimetres, this limits the spatial resolution of this technique; and because blood flow changes relatively slowly over a few seconds, this limits the ability of the technique to discriminate very rapid changes in neural activity associated with perception, thought and action.

Nevertheless, BOLD contrast fMRI has a number of advantages over previous ways of imaging brain function such as Positron Emission Tomography (PET) (described later), because it is entirely non-invasive and has much higher spatial resolution. Typically BOLD contrast fMRI is combined with rapid acquisition of brain imaging data. This enables the researcher to acquire a continuous series of images of the brain, one every few seconds over a period of 30 or 40 minutes, while the participant performs a particular task. This allows for inferences about the nature of the brain processes and their localisation during the task on the basis of the patterns of brain activity that are measured.

Functional MRI can also measure other aspects of brain function, such as blood flow, and it might also be possible in future to use MRI to detect the electrical currents associated with neurons directly.

EEG and MEG

Although not always considered 'imaging' techniques, electroencephalography (EEG) and magnetoencephalography (MEG) share the ability of other technologies to produce a moment-by-moment topographic 'map' of human brain activity. In contrast to the other technologies, EEG and MEG measure the tiny electrical currents or associated magnetic fields, respectively, that are associated with the summed activity of many hundreds of thousands of neurons in the human brain. These techniques measure such activity through hundreds of electrodes/sensors on the scalp and can produce an image of brain activity associated with thought processes. They provide a measure of brain activity that directly reflects the electrical activity of neurons in the brain (albeit large groups of such neurons) in contrast to the indirect measure of signals related to blood flow measured by fMRI and PET.

The principal advantage of such a measure is that it can track the activity of neurons much more rapidly over a timescale of a few milliseconds. EEG and MEG therefore have superior temporal resolution to other techniques. However their ability to resolve activity in specific brain areas is more limited, both because it is difficult to measure electrical or magnetic signals from areas deep within the brain (such as the thalamus) and because of the intrinsic mathematical uncertainty associated with the process of reconstructing how activity in specific brain areas gives rise to electrical signals in the scalp. For these reasons EEG/MEG are often used together with functional MRI as complementary approaches; one with temporal but less spatial resolution, the other with greater spatial but less temporal resolution.

Other functional neuroimaging techniques

Other neuroimaging technologies are less widespread but fill important niches by providing unique capabilities.

Positron emission tomography (PET) requires the injection of a radioactive tracer molecule. Detectors placed around the head or other body part being imaged can sense the radioactive decay of the tracer molecule in the body. This allows the reconstruction of images of the brain or other organs where the image is sensitive to the particular molecule used. Oxygen-15 labelled water can provide images of cerebral blood flow, similar to functional MRI, which has now largely supplanted PET for providing dynamic images of brain activation. When the radiotracers are molecules that bind to receptors or other sites of drug action then PET can be used to image the concentration, or changes in concentration, of neurotransmitters (the chemical messengers in the brain) at receptors. This is a particular feature of PET not shared with other neuroimaging techniques, and may be particularly important in understanding psychiatric diseases.

Near infrared spectroscopy (NIRS) is a method that measures the absorption of light at near infra-red wavelengths. In brain imaging, the resultant spectrum can be related to the concentration of oxygenated and deoxygenated haemoglobin in blood. This approach can be used non-invasively by applying a near infra-red source and array of detectors to the intact skull, which is not entirely opaque to infra-red light, particularly in babies and infants who have relatively thin skulls. NIRS can thus be used as a non-invasive neuroimaging technique measuring brain activation in babies and infants. In comparison to the other techniques, NIRS is portable and can be used at the bedside for medical applications. However it has relatively low spatial resolution due to the difficulty in localising scattered light through the intact skull, and the limited penetration of infrared light into the cortex.

In addition to these approaches, there are a number of other techniques such as single photon emission computed tomography (SPECT) that is lower in spatial resolution than PET and largely used as a clinical technique and not in neuroscientific research. (This essay has not discussed invasive optical imaging approaches used to image exposed brain tissue in animals such as 2-photon imaging.)

Uses of neuroimaging

Neuroimaging provides measurements of brain structure and activity associated with relatively large populations of neurons throughout the brain. It has been used to determine how and where in the human brain such measurements are correlated with perception, thought and action. Often this is achieved through intermediate psychological or computational theories that make predictions about how particular brain structures give rise to observed behaviour. These can then be tested by seeking evidence for signals from those brain structures consistent with theoretical predictions.

The ultimate goal of this work is to arrive at an account of how mental processes are manifested in the human brain, and how this gives rise to observable behaviour in terms of speech and motor actions or other behaviours. This approach has also been applied widely to studying disorders of the brain, whether caused during development or in adult life, and associated with neurological or psychiatric disease.

Studying brain disorders

The major use (by volume and cost) of neuroimaging technologies is in healthcare. This is principally to provide diagnostic imaging for use in clinical decision-making. In the last twenty years, neuroimaging technologies have become widely available in the developed world and this has been driven by healthcare. For example, in Europe from 1999 to 2009, the number of MRI scanners increased from an average of 0.54 to 1.38 per 100,000 people (European Community Health Indicators). Most of these scanners are in health care facilities and not all are capable of the advanced neuroimaging data acquisition and analysis discussed here. However, MRI scanning is now readily available at reasonable cost (typically around £500

per hour) both inside and outside research institutions.

The widespread dissemination and use of neuroimaging technologies coupled with the invention of functional MRI in the early 1990s, which permitted easy and non-invasive measurement of brain function, has led to an explosion in scientific studies of cognitive processes in the normal and abnormal brain. In this domain, perhaps the widest use of brain imaging technologies are to identify the processes through which information in the brain is manipulated to drive human cognition. Significant progress has been made in understanding the neural basis of cognitive processes such as attention and memory. Work in healthy volunteers (explored below in more detail) is complemented by large numbers of studies describing how such processes break down in abnormal development (eg autism) or following brain damage (eg stroke or Alzheimer's disease). These areas of research have generated thousands of research papers.

Studying the normal human brain

By understanding the relationship between brain signals and either sensory processing (the interpretation of stimuli), thought, or intention to make an action, functional brain imaging techniques hold out the prospect that they might be able to provide a non-invasive 'read out' of perception, thought and intention respectively. At present, determining the thoughts and intended actions of others requires them to voluntarily speak to us or make an action. This leads to significant problems when such abilities are impaired through neurological damage, so an ability to non-invasively 'read out' such intentions or perceptions would have potential importance (see also Section 2.3).

Recent advances in neuroimaging data analysis have enabled a limited type of 'brain reading' or non-invasive detection of particular perceptions, thoughts, or intentions to make an action. To date, all these applications of neuroimaging are restricted to decoding thought or action in quite limited experimental settings, where only a small number of possible thoughts or actions are entertained. Moreover, they generally rely on analyses performed some minutes or hours after the data have been collected. In the future, developments might mean that collection and analysis of fMRI data can take place in real time. This would raise the possibility of being able to decode an individual's thought and intended action, without requiring the participant to speak or move, in near real-time (see also Section 2.3). Nevertheless, there are important limitations to these potential applications that may be insurmountable, as discussed below.

Humans are particularly skilled at judging the mental states of others, and this forms the basis of much social interaction. While there are approaches to detecting an individual's current perception or behaviour, humans are also skilled at concealing their mental states through their behaviour. For millennia there has been interest in detecting such concealed information. One type of concealment occurs during deception, and a large

number of bodily markers of deception have been proposed including the physiological signs that form the basis of polygraph tests.

More recently, there has been substantial interest in using neuroimaging to detect deception (see also Sections 3.2, 3.3, and Box 1). Much work has focused on the neural mechanisms that underpin lying versus truth telling. These show that a particular area of the brain (the prefrontal cortices) is particularly involved in inhibiting truthful responses and generating the false responses characteristic of lying. Local brain activity or patterns of brain activity can also distinguish truth-telling from lying on a trial-by-trial basis for individual cases. But there are significant limitations to these approaches, including the use of countermeasures (covert or overt measures taken by the subject in order to distort or undermine any conclusions) and the inability to generalise from studies of groups of individuals to single individuals.

Lying typically requires a deliberate conscious effort. A different approach to detecting deception focuses on unconscious processing and/or recognition of specific knowledge about an event. The so-called Guilty Knowledge Test utilises a series of multiple-choice questions, each having one relevant alternative (eg some aspects of a crime under investigation) and several neutral alternatives, all chosen to be indistinguishable by an innocent participant. If the subject's physiological (or brain imaging) responses to the relevant alternative are consistently greater than for the neutral alternatives, then knowledge of the event is inferred.

These uses of neuroimaging share two common conceptual themes. Either they are attempting to identify the presence of a particular cognitive operation (eg lying) or they are attempting to detect particular mental contents (eg having previously seen a particular object). Thus while these approaches have been examined in the context of detecting deception, they have many other potential applications. For example, there has been widespread interest in detecting mental contents as a tool to potentially communicate with seriously brain-injured individuals. Thus the ethical and policy implications of working in one area have to be understood in the context of other potential uses of the technology.

Limits of neuroimaging

Generic limits

It is worth noting that all neuroimaging techniques require participants who can (and want to) comply with the procedures and instructions entailed. The most important technical limitation of human neuroimaging is that all currently available techniques are constrained to measurements of aggregate signals from hundreds of thousands of neurons at a time. Thus any signals critical for perception, thought or actions that are encoded at a finer spatial scale may be hard to detect using neuroimaging.

The techniques outlined above differ in their relative strengths and weaknesses. Functional MRI relies on measurement of signals associated with the oxygenation level of haemoglobin in the blood. This changes at a relatively fine spatial scale

and so the spatial resolution of functional MRI is of the order of a few millimetres. But its temporal resolution is relatively poor, because blood flow changes that alter oxygenated haemoglobin concentration lag several seconds behind the electrical activity of neurons. In contrast, the electrical techniques of MEG and EEG have millisecond temporal resolution because they measure the electrical signals associated with neuronal activity directly at the scalp. However, because these signals summate over space and are altered by the scalp and tissues, the spatial resolution of these techniques is relatively poor compared to fMRI. Thus neuroimaging techniques are complementary in their limitations, which often leads to their use together to investigate a particular scientific problem.

The principal disadvantage of PET is the need for extensive radiochemistry infrastructure (both people and equipment) to develop new radiotracers. Existing radiotracers can often be made on site with relatively small needs for additional equipment. Nevertheless the technique remains comparatively expensive and inflexible compared to all other neuroimaging techniques. The temporal resolution of PET is extremely low; typically one image is acquired every thirty seconds or so. Moreover, the need to use ionising radiation places exposure limits on the numbers of scans that can be acquired (typically a maximum of 12, compared to no limits on fMRI where typically hundreds or thousands of scans are acquired for each participant).

All neuroimaging techniques measure brain activity that can only be correlated with subjective reports or observable behaviour of the individual being scanned. Such correlations do not imply a causal relationship between the brain activity and behaviour. For example, if two or more brain regions show activity patterns that are correlated with a particular behaviour, it may be that one or more of those areas are not necessary to generate the behaviour and could be damaged by illness without affecting performance. To identify a causal relationship therefore requires that neuroimaging data is combined with other evidence, such as that obtained through studying individuals with brain damage or transiently disrupting brain function through transcranial magnetic stimulation (TMS) (see Section 2.3 and Box 4).

There are also barriers to the use of currently available neuroimaging in real-world environments. All the technologies discussed here are widely available in hospital and research environments in the developed world, but there are significant obstacles to more widespread deployment in the community. MRI technologies require a large, heavy fixed installation that can only be transported in a large articulated truck. Scanning participants with MRI requires that they lie supine in an enclosed space, limiting access and interaction with the external environment, although a number of ingenious experimental approaches have been devised to circumvent this. Similarly MEG requires a cumbersome installation that requires careful electrical shielding. EEG is the most portable technology and can in principle be used in the home setting. However, local interference from other

electrical devices is not trivial and can significantly limit performance. Moreover, the electrodes that are applied to the scalp for EEG cannot be worn for extended periods or in most real-life situations. Emerging technological developments may circumvent this possibility with the creation of easily wearable and unobtrusive electrodes, but this at present remains a future development.

In addition to these generic limitations of the technologies, particular methods have specific limitations that have already been discussed above. Specific issues relating to the *application* of neuroimaging technologies are described below.

Specific issues

While neuroimaging techniques have proven impressive in their ability to provide new insights into how mental processes are realised in the human brain, these insights are typically provided by studying relatively large groups of individuals. While they can provide insights into the group average (or typically, representative) patterns of brain activity, they can provide less insight into the variability of these patterns of brain activity associated with an individual. The ability of these findings to guide therapeutic decisions or inform policy about individuals is thus typically limited. This is not an intrinsic limitation, but reflects a relative dearth of studies concerning inter-individual variability.

The use of functional neuroimaging technologies to decode specific instances of what a subject is thinking at a precise moment has been impressive. But these studies, attempting to understand specific conscious content in an individual and operate as a 'brain reading' device, suffer from a potential limitation of generalisation. Because the number of potential thoughts or perceptions an individual can have at any one time is virtually limitless, a general-purpose brain reading device would need to be able to generalise not just over all the thoughts an individual had previously experienced, but also any possible or indeed conceivable future experiences. Moreover, experimentally the process of 'brain reading' is typically studied in isolation, whereas in everyday life people typically have two or more streams of thought concurrently. It is not clear whether 'brain reading' techniques would be able to cope with such concurrency of thought.

Thus, irrespective of the possible policy issues that arise through recent developments in our ability to 'brain read', potentially serious and possibly insurmountable empirical limitations are already apparent.

The potential ability of neuroimaging to uncover covert mental states has received much recent attention. In fact, this has always been an area of intense interest to policymakers. The polygraph is a well-known approach to detecting deception (a covert mental state). Although not a neuroimaging technique, it has been repeatedly evaluated and its validity and reliability have been challenged for decades in systematic reviews and evaluations. In addition to questions about its reliability and validity, the polygraph is particularly vulnerable to

countermeasures. Such vulnerability to countermeasures is shared by all neuroimaging efforts to date that attempt to detect deception through measuring brain activity directly.

Moreover, it remains unclear whether any neuroimaging technique is in fact superior to the poor performance of the polygraph, as there have been no direct comparisons. One of the important limitations of any such evaluation is that techniques for detecting deception are typically developed and validated in populations of young individuals simulating deception. It is not clear whether such simulated deception corresponds in any way to deception carried out in the real world. Moreover it is not clear whether any indices of deception (whether brain based or otherwise) detected in young healthy adults generalise in any way to older individuals or groups with mental illness – to name but two groups over-represented in prison populations. These issues limit both the validity and the potential use of neuroimaging technologies in detecting deception (see also Box 1).

Policy questions and issues

Healthcare

The widespread availability and use of clinical neuroimaging, coupled with the computerisation of health service imaging platforms, means that there are now very large numbers of (principally structural) brain images collected within the National Health Service. These are mainly used for diagnostic purposes and clinical care, but represent a potential resource unparalleled in scope and scale for the large-scale study and characterisation of illness. At the same time, large scientific projects such as UK Biobank that tracks biomedical indicators and their relation to the health of a very large sector of the UK population are beginning to consider adding neuroimaging measures. An important emerging policy area is the structure and organisation of such collections of tens or hundreds of thousands of images, including who controls access to such resources, the role of patient consent, and the ethical issues surrounding their potential use to identify predisposition to disease, or to reveal information about groups of individuals perhaps only loosely connected to the initial reason for their undergoing clinical neuroimaging.

Many of the important applications of neuroimaging assess some form of brain plasticity (changes in connections among different parts of the brain) either in response to injury (such as stoke) or imbalance in the chemical messenger systems of the brain (neurotransmitters). This raises the possibility of developing assays of functional reorganisation and plasticity to diagnose neuropsychiatric syndromes, find associations with genetic predispositions, and help predict likely outcomes for patients and their carers.

Business

The role of neural processes in determining economic decisions has been a subject of

intense research interest worldwide in recent years. This has led to commercial interest in 'neuromarketing' and the possibility that neuroimaging measures of brain activity might help in the design or marketing of commercial products, or perhaps be used as a more effective way to evaluate marketing efforts than simply asking people (see Section 2.5).

Security and criminal justice

There has been a great deal of interest in understanding whether an ability to detect covert mental states or deception might be useful in security and criminal justice arenas. Despite considerable limitations to currently available techniques described above, there is often an unspoken assumption that these can be overcome by technical development. From a neuroscientific point of view, this is an empirical question and one whose answer is by no means certain. This uncertainty and current limitations have not stopped consideration of the use of neuroimaging techniques to assess the credibility of a witness or suspect, assess an offender's criminal liability and responsibility, and more recently to attempt to prognosticate on the dangerousness of an offender or their likelihood to reoffend (Zeki and Goodenough (eds) 2004). These issues are also discussed in other essays (see Sections 3.2 and 3.3) and will be addressed in Module 4 of the *Brain Waves* project on neuroscience, responsibility and the law.

Mental Privacy

Underpinning many of the policy areas described above is the concept of mental privacy, discussed further in other essays (see Sections 3.2 and 3.3). Our mental contents are typically not directly observable and individuals can choose whether and to what extent to reveal their mental states through communication. But the advent of non-invasive imaging technologies that might, in principle, permit the decoding of mental states raises the possibility of detecting mental states without requiring direct communication from the individual (provided the participant is able and willing to co-operate). Moreover, the possibility that functional or structural images encode additional information about an individual or their behavioural dispositions raises the possibility that such images acquired for other incidental reasons might be reanalysed after acquisition to reveal this 'hidden' information.

Conclusion

There has been a rapid development in both the deployment and the capability of neuroimaging techniques in the last twenty years, particularly with the advent of non-invasive functional imaging approaches. Nevertheless, despite some spectacular advances in our understanding of the neural basis of mental processes, the practical limitations of both the techniques and their use in 'real world' situations limits their potential use outside clinical and research arenas.

References and further reading

European Community Health Indicators, available at: http://ec.europa.eu/health/ph_information/dissemination/echi/query/

Zeki S and Goodenough O (eds) (2004) Law and the brain. *Philosophical Transactions of the Royal Society B,* Vol 359.

Box 1: Neuroimaging and lie detection

Professor Geraint Rees, University College London

Neuroimaging technologies (see Section 2.1) for lie detection show some promise in being able to explore the neural correlates of deception (a covert mental state), but none currently has anywhere near sufficient sensitivity, specificity and reliability to be deployable in the criminal justice system of England and Wales. Indeed, it remains uncertain whether such technologies will ever be sufficiently robust to be used in such 'real world' settings. Nevertheless, despite the technological barriers there remains considerable interest and pressure to consider the use of such technologies, and great public interest in situations worldwide where neuroimaging evidence has been used as evidence in criminal cases.

The polygraph is a well-known approach to detecting deception. Although not a neuroimaging technique—it relies on the measurement of skin conductance, which can be influenced by arousal during deception—it has been repeatedly evaluated and its validity and reliability have been challenged for decades in systematic reviews and evaluations. In addition to questions about its reliability and validity, the polygraph is particularly vulnerable to countermeasures—covert or overt measures taken by the subject of the polygraph in order to distort or undermine any conclusions.

Such vulnerability to countermeasures is shared by all neuroimaging efforts to date that attempt to detect deception through measuring brain activity directly. Moreover, it remains unclear whether any neuroimaging technique is in fact superior to the poor performance of the polygraph, as there have been no direct comparisons. One of the important limitations of any such evaluation is that techniques for detecting deception are typically developed and validated in populations of young individuals *simulating* deception. It is not clear whether such simulated deception corresponds in any way to deception carried out in the real world. Moreover it is not clear whether any indices of deception (whether brain-based or otherwise) detected in young healthy adults generalise in any way to groups over-represented in prison populations, such as older individuals and those with mental illness. These issues limit both the validity and the potential use of neuroimaging technologies in detecting deception.

These issues will be addressed in more detail as part of Module 4 of the Brain Waves project on neuroscience, responsibility and the law.

2.2 Neuropsychopharmacology

Professor Trevor Robbins FRS, University of Cambridge

Background

The field of neuropsychopharmacology encompasses drugs that affect cognition, behaviour, and the brain. The brain itself comprises multiple networks, each of billions of nerve cells ('neurons'), which interact via the release of chemical signals, termed 'neurotransmitters', that diffuse from one nerve cell to the other across the gaps between them (synapses).

By binding to specialised proteins in the cell membrane called receptors, neurotransmitters change the sensitivity of the nerve cell membranes to potential inputs from other cells. Receptors function in two main ways: i) through short term effects induced by neurotransmitter binding that allows the passage of ions into and out of the cell, which rapidly change its function or ii) through longer term changes induced by a biochemical cascade inside the nerve cell which eventually interacts with genetic material in the cell nucleus.

Although originally it was considered that each nerve cell released only one neurotransmitter and that there were two main chemical transmitters (one excitatory and the other inhibitory) these concepts have been overturned in the last three decades by the realization that there are literally dozens of neurotransmitters, and that most neurons can release more than one type, thus enabling a very precise and sophisticated modulation of the activity of the receiving nerve cells. A simple classification of the main neurotransmitters subdivides them according to the speed and nature of their effects on neuronal circuits.

Fast signaling neurotransmitters which link large interconnected neural networks constitute mainly the amino acids glutamate and GABA (gamma-aminobutryic acid), and slow modulatory neurotransmitters, whose action is neither exclusively excitatory nor inhibitory include the monoamines; dopamine, noradrenaline and serotonin (or 5-hydroxytryptamine, 5-HT). Finally, very slow modes of action are generally characteristic of neuropeptide neurotransmitters such as cholecystokinin and vasopressin, which may be co-released with the 'classical' neurotransmitters, thus modifying their effects (see Stahl 2008).

New developments in neuropsychopharmacology

Neuropsychopharmacology has focused mainly on understanding the molecular basis of action of drugs that alter the activity of the chemical neurotransmitter systems by one or more mechanisms. Drugs may be classified as 'agonists' that simulate the action or evoke the release of natural neurotransmitters at specific receptors, or 'antagonists' that block the actions or the re-uptake of natural neurotransmitters at specific receptors. We now understand a good deal about entire

pharmacological classes of drugs that share common mechanisms of action to exert effects in the brain.

Recent important developments are brain imaging techniques (see Section 2.1) that have enabled researchers to link the molecular actions of drugs to specific behavioural or physiological effects in humans. Positron emission tomography (PET) can be used to map the location of neurotransmitter receptors in the brain and can be utilised in conjunction with pharmacological treatments to infer the activity of central neurotransmitter systems in the human brain. This technique has been used, for example, to show that drug addicts have reduced dopamine receptors in the striatum area of the brain (Volkow *et al.* 1997), thus confirming leads from animal work that the dopamine system is implicated in addiction.

The use of functional magnetic resonance imaging (fMRI) has been important for establishing 'event related' correlations between brain activity and behaviour over time. Pharmacological fMRI is a recent development which enables the analysis of drug effects on cognition and behaviour—however, it has two main difficulties: first its lack of neurochemical specificity in relation to PET and secondly the possibility of misleading responses arising from drug actions on cardiovascular factors.

Another major area of development is pharmacogenomics (Royal Society 2005). The human genome project has enabled the identification of gene variants that define individual variation in genetic make-up. Such genetic polymorphisms not only allow differences in cognition and behaviour to be related to them, but also drug effects. Indeed, given the molecular specificity of drug action, such mapping may be done more effectively than for individual variability in behaviour itself.

Prominent examples include demonstrations that the cognitive enhancing effects of amphetamine may depend on a polymorphism in the brain that regulates the activity of dopamine in the prefrontal cortex of the brain; thus either benefit or impairment may occur in response to the same dose of the drug in different individuals (eg Mattay *et al.* 2003). A second example is the finding that depression produced by the drug ecstasy is also highly variable in users and can be related to variability in mechanisms controlling the activity of serotonin in the brain. Genetic differences may also modulate the capacity of cannabis to trigger psychosis. Such observations have enormous implications for calculating the risk associated with drug-taking and for predicting benefits of drugs in the treatment of mental illness ('personalised medicine'), as well as for 'cognitive enhancement' in healthy individuals.

Varied drug effects in the brain

In functional terms, it makes sense to subdivide the main drug actions into those which bring about detrimental or incapacitating effects by impeding or damaging neural activity, and those which bring functional effects by modulating neural activity. The latter effects may be produced by drugs: i) used as mental health medicines, such as anti-psychotic or

anti-depressant medications; ii) used for their cognitive or performance enhancing effects (eg including sexual performance or for effects on personality); and iii) used recreationally or abused, leading to addiction.

Some of these categorisations are significant; for example, several so-called cognitive enhancing drugs do not produce subjective effects that usually accompany recreational use, and also do not lead to obvious physical dependence. Nevertheless, there is a good deal of overlap among these categories; for example, some abused drugs such as methamphetamine undoubtedly have neurotoxic effects, while a drug producing cognitive enhancing effects could also be 'abused' (eg both in terms of criminal use such as cheating in competitive examinations, and in terms of a psychological dependence, leading to excessive drug intake). Moreover, virtually all drugs of abuse, from opiates to stimulants, have been used at one time or another in a medical context. Stimulants were formerly used as slimming agents; they are still used in the treatment of attention deficit hyperactivity disorder (ADHD). Heroin was originally employed in cough mixtures, and opiate drugs are still employed as powerful pain-killing (analgesic) agents. Nicotine may have cognitive enhancing effects and variants of nicotine are being pursued as possible cognitive enhancing agents. Cannabis has been suggested to be beneficial in the treatment of multiple sclerosis. Some of this plurality of action arises from the fact that many drugs have multiple sites of action in the brain, even though the same receptor may be implicated (eg euphoric and analgesic effects of opiates).

In terms of basic neuroscience research it is relevant to note that each of these drug sub-divisions can also be used to discover more about brain function.

Mental health medicines

Mental health medicines are continuously being monitored, with controversies often arising about efficacy and safety of drugs, for example certain anti-depressants. Patents for many compounds are due to expire soon and there has been a worrying failure to produce many more effective drugs that utilise new mechanisms of action, although there are some possible exceptions (eg ketamine for depression and drugs affecting glutamate transmission for schizophrenia). This has led several of the major pharmaceutical companies to close down programmes in psychiatry and neuroscience, which could result in considerable harm to both health and the economy in the UK and other countries (see also Section 3.1). On the other hand, the relative success of drugs used to treat positive symptoms in schizophrenia and severe clinical depression means that companies will continue to make large profits from the sale of mental health medicines.

One major difficulty may be increasing regulatory controls on drug development accompanied by a shift to risk aversion on the part of the drug companies in the face of adverse side effects. Another may be that the expensive clinical Phase 3 trials can be scientifically suspect due to poorly

controlled factors. Relatively minor advances have come in the form of more sophisticated means of drug delivery, which minimise the number of administrations involved in medicating certain patient groups, for example formulations for ADHD that require only once daily dosing.

Recreational drugs

Use and abuse: Many different classes of drug are used to produce subjective effects (such as changes in mood, including euphoria or sedation) in the recreational context. Some of these drugs are legal (eg alcohol, nicotine) but most are not (eg stimulants such as cocaine and methamphetamine, opiates such as heroin, and hallucinogens such as cannabis and ecstasy (MDMA)). Distinct molecular actions of these different drugs have been to a large extent identified. However, similar neurobehavioural principles may govern how they might become drugs of abuse. For example, many of these drugs have some common effects on brain dopamine systems, which may be responsible for some of their reinforcing ('rewarding') actions.

We now have a considerable understanding of the basis in the brain of the effects of recreational drugs, both through studies with experimental animals (who may self-administer them) and neuroimaging studies of drug-using humans (Robbins *et al.* 2010). Some of the drugs (eg alcohol, heroin) have obvious physical withdrawal syndromes which are associated with drug dependence. These drugs may also produce psychological dysphoric (opposite of euphoric) withdrawal states, which are also postulated to occur following stimulant withdrawal. However, the tendency towards compulsive drug seeking and 'out-of-control' drug taking, which are the hallmarks of addiction, is not an inevitable consequence of drug use; there is considerable individual variation in the degree to which individuals are vulnerable to addictive properties of drugs, probably arising from both genetic and environmental influences. It has for example, been estimated that only about 20% of persistent cocaine users actually become addicted or 'dependent' on cocaine.

Addiction and dependence: Addiction has been the subject of several recent major reviews, notably the wide-ranging, interdisciplinary Foresight project *Brain science, addiction and drugs* (published as Drugs and the Future, Nutt *et al.* 2006) and the subsequent Academy of Medical Sciences (2008) report of the same name. Several key issues were raised by these reports including: i) the relative absence of effective medications for addiction; ii) the absence of reliable and informative epidemiological data concerning addiction in the UK; iii) the lack of accurate information about adverse effects and risks entailed by using certain drugs, such as ecstasy and cannabis, which has led to controversies about how such drugs should be classified under the *Misuse of Drugs Act* by the Home Office, and considerable confusion arising from a failure to appreciate the importance of individual differences in susceptibility to drug effects; and iv) the likelihood of novel

drugs of abuse emerging periodically. These may be previously known compounds for which new properties are discovered, or new compounds claimed to offer 'legal highs' (eg mephedrone) which are not controlled and are available to buy on the web. Many 'new' drugs are often variations on a theme in terms of chemical structure, for example, being derivatives of the amphetamine-class of stimulant drugs.

Cognitive enhancers

Considerable interest was aroused in cognitive enhancers and performance enhancing drugs (see also Box 1; and Sections 3.1, 3.2, and 3.3) by the Foresight project. This project made it clear that there may exist certain drugs with the potential ability to enhance cognitive performance, not only in patients with cognitive disorders, but also in the normal healthy population (Jones *et al.* 2007; Academy of Medical Sciences 2008).

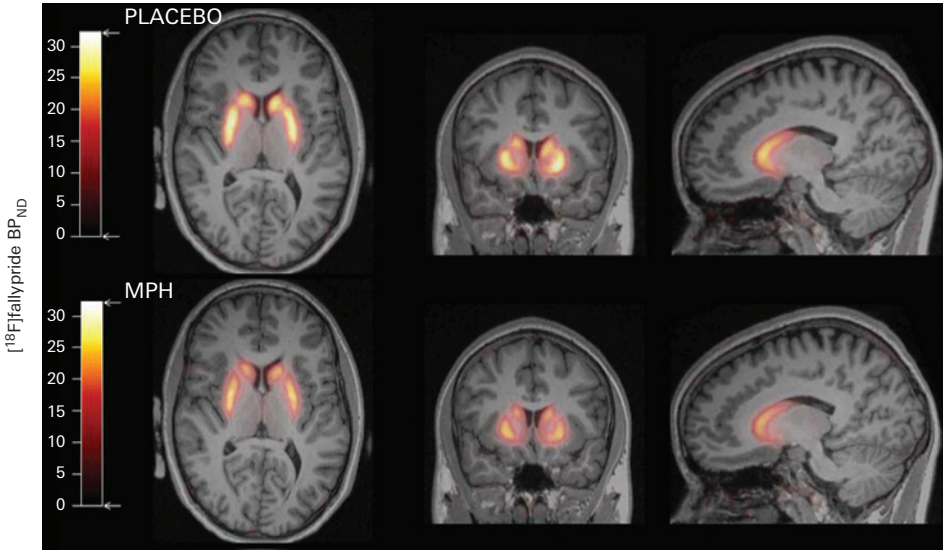
'Cognition' refers to brain processes of mental activity including: perception, attention, learning, memory, language, thinking, planning, decision-making and cognitive control. Cognition must be distinguished from overt behaviour, which is, however, largely a product of cognitive processes. Cognitive performance also depends on other important factors such as arousal level (ie the level of wakefulness) and motivation. Thus, in theory a 'cognitive enhancer' may produce its beneficial effects indirectly through effects on motivation or arousal. (This essay excludes consideration of enhancement of other aspects of behaviour such as sexual function or appetite.)

A major impetus for the emergence of cognitive performance enhancers has been the focus by pharmaceutical companies on the treatment of cognitive disorders such as dementia (including Alzheimer's and Parkinson's disease), and more recently stroke, schizophrenia and ADHD. Economic arguments underpinning this interest are compelling, given the estimated market for dementia therapy (see Jones *et al.* 2007). Such disorders are associated with different forms of pathology and so it is likely that there will be a need for different types of medication. However, it is also conceivable that a single compound may have several possible disease targets, as it may provide improvements not by removing a particular pathology but by producing compensatory improvements in neuronal systems that are intact.

There is no programme in 'big Pharma' for developing cognitive enhancers for the healthy population; regulatory and legal issues determine this lack of commercial interest. However, 'cognitive enhancing' drugs may be publicised by word of mouth, the media and the web, often leading to considerable 'off-label' use, which raises significant issues of regulation for government. A commentary in *Nature* by Sahakian and Morein-Zamir (2007) entitled *Professor's Little Helper* described the 'off-label' use of cognitive enhancing drugs such as methylphenidate (Ritalin) and modafinil by University students and academics. Many of their conclusions were endorsed by a subsequent questionnaire issued by the same journal (Maher 2008).

A compelling consideration for 'cognitive enhancers' is the diversity of candidate

Figure 3 These images show the signal acquired using [18F]fallypride (the radioactive tracer molecule) positron emission tomography following placebo (top row) and oral methylphenidate (Ritalin) (bottom row) in a healthy volunteer, superimposed on a high resolution magnetic resonance image (MRI) acquired from the same volunteer. [18F] fallypride competes with endogenous (naturally produced) dopamine for the binding to dopamine receptors. The signal reduction observed following methylphenidate relative to placebo indicates that the drug increased dopamine levels, leading to less [18F]fallypride binding due to increased competition with endogenous dopamine. (Reproduced courtesy of Natalia del Campo, University of Cambridge.)



mechanisms that are being utilised in their development. Many compounds, including well-known drugs such as methylphenidate (Ritalin) (see Figure 3), amphetamine, and nicotine exert their effects by optimising chemical signalling by monoamine neurotransmitters (eg dopamine, noradrenaline, or 5-HT) in brain regions such as the prefrontal cortex. New compounds, such as the 'AMPA-kines' work by exaggerating the effect of the fast signalling neurotransmitter glutamate in neural networks in key 'cognitive' structures of the brain such as the cerebral cortex and the hippocampus.

Drugs are also being developed as 'inverse agonists' that invert the sedative actions of related compounds to produce memory-enhancing effects. One such compound has been found to block the amnesia produced by alcohol, through acting at GABA-activated receptors. Another, modafinil, has long been licensed as a treatment for narcolepsy, but has also been shown to enhance working memory and

executive functioning in non-sleep-deprived volunteers. Intriguingly, it is far from clear how it produces its cognitive enhancing effects and this is a currently a subject of great research interest.

'Nutraceuticals' such as vitamins, supplements and other aspects of dietary modification may also produce cognitive effects, although their precise modes of action are largely obscure. Perhaps predictably, members of the public appear to find cognitive enhancement via dietary means less objectionable than by orally administered drugs, even though they may share common modes of action.

The discovery of cognitive enhancing drug effects is of particular interest in terms of boosting performance in a number of different occupations—notably in shift workers, for which controlled trials are already suggesting some benefit. Another potential application has been illustrated in a demonstration of beneficial effects of modafinil in the performance of surgeons, over conventionally-employed stimulants such as caffeine, which lead to counterproductive hand tremor (Sugden *et al.* 2010).

This area is also of significance for subsequent modules of the *Brain Waves* project. There is significant military interest in cognitive enhancement. Modafinil is thought to have been used by the French army in Iraq in the early 1990s and in 2003 the US Air Force authorised its use to combat fatigue. This use of modafinil echoes that of stimulants such as amphetamine in the Second World War, but today there is much more precise information about the benefits—and the costs—of such practice (see also Section 3.2).

Some aspects of the 'cognitive enhancement' strategy arising from recent discoveries in basic neuroscience may appear paradoxical at first sight. For example, it may eventually be possible to help patients with certain disorders such as post-traumatic stress disorder (PTSD), anxiety, or drug addiction by treating them with drugs that seemingly impair cognition—by producing selective amnesias. It is now evident from work on experimental animals that memories may be subject to experimentally-induced forgetting if evoked in the presence of certain drugs (so-called 'memory reconsolidation'). This is also true for the procedure of extinction, whereby a certain stimulus is no longer followed by its expected consequence, a process that leads to an active suppression of the normally-evoked behaviour. Extinction can be accelerated in the presence of drugs that facilitate the functioning of certain receptors.

The optimization of behaviour by pharmacological agents is a complex issue that depends on a balance of cognitive enhancement and cognitive suppression. It is now crucial that we begin to assess possible costs as well as benefits as a consequence of treatment with cognitive enhancers, especially long term or chronic consequences. Relevant considerations have been set out in the Academy of Medical Sciences (2008) report on *Brain science, addiction and drugs*.

Allied to this scientific excitement and uncertainty are similar ethical dilemmas

posed by the use of cognitive enhancing drugs by the healthy population. Whilst it may not be considered ethically dubious to aspire to enhance performance, there are obvious problems in competitive situations, eg examinations or the workplace. Other ethical issues may include the coercive use of such drugs, such as to ensure good behaviour in juveniles with ADHD (eg in certain schools in the US), or as a compulsory requirement for soldiers. The example of cognitive enhancers is a main theme of the burgeoning new subject of 'neuroethics' (see Greely *et al.* 2008) (see Section 3.3).

Incapacitating agents

Drugs acting on the central nervous system can also be used to incapacitate humans, primarily for beneficial purposes, such as the use of anaesthetics under highly controlled circumstances for medical procedures, or the sedation of violent patients. However these drugs have also drawn the attention of military and police organisations since the late 1940's for use as chemical weapons. Developers of such incapacitating chemical agents generally aim to target the central nervous system to induce unconsciousness or sedation. These weapons are distinct from 'riot control agents', such as 'tear gas', which cause local irritation to eyes, skin, and the respiratory tract, and have long been used by police forces around the world.

Since the international ban on chemical weapons was agreed in 1993 some countries have continued to develop incapacitating chemical weapons—ie drugs targeting the brain (as opposed to 'riot control agents')—for law enforcement purposes. This has led to significant controversy and concern, particularly following the use of an aerosolised spray of an opioid analgesic drug, thought to be 3-methylfentanyl (carfentanil), to break the siege of a Moscow theatre in 2002, which resulted in the death of over 120 hostages (see also Section 3.2). This issue will be addressed in more detail as part of Module 3 of the *Brain Waves* project.

As in the case of cognitive enhancing drugs, there are many ways of producing incapacitating effects including mechanisms related to anaesthesia and respiratory depression, as well as more general neurotoxicological actions. The former can be produced via a range of different mechanisms generated by our burgeoning understanding of brain mechanisms of consciousness, including the discovery of new neurotransmitter substances such as orexin. (Saper *et al.* 2005).

Conclusion

The fact that drugs can produce mind-altering effects through their chemical actions has been known for centuries, but the excitement of the field in neuroscience is that we can now explain a good many of these actions through our enhanced understanding of the chemistry and functioning of the brain. A significant realisation is that drugs can have diverse effects on different aspects of function (leading obviously to inevitable 'side-effects'); and that these effects will often encompass both beneficial (eg 'cognitive

enhancing', 'mood-enhancing', pain-killing and detrimental (physical withdrawal, neurotoxicity) actions. Thus, it becomes important to define in which precise context a drug is being used, and also to formulate the cost-benefit equations in both biological and ethical terms. (see also Sections 3.2 and 3.3).

References and further reading

Academy of Medical Sciences (2008*) Brain science, addiction and drugs*. London: Academy of Medical Sciences.

Greely H, Sahakian B, Harris J, Kessler RC, Gazzaniga M, Campbell P and Farah M (2008) Towards responsible use of cognitive-enhancing drugs by the healthy. *Nature*, Vol 456, pp 702–705.

Jones R, Morris K and Nutt D (2007) Cognition enhancers. In: Nutt D, Robbins T, Stimson G, Ince M, and Jackson A (eds) *Drugs and the Future*. Amsterdam: Elsevier. pp 241–283.

Maher B (2008) Poll results: look who's doping. *Nature*, Vol 452, pp 674–675.

Mattay V, Goldberg T, Fera F, Hariri A, Tessitore A and Egan M (2003) Catechol-O-methyl trasferase Val 158-Met genotype and individual variation in the brain response to amphetamine. *Proceedings of the National Academy of Sciences*, Vol 100, pp 6186–6191.

Nutt D, Robbins T, Stimson G, Ince M and Jackson A (eds) (2006) *Drugs and the Future: Brain science, addiction and society*. Burlington MA: Academic Press.

Robbins T, Everitt B, Nutt D (2010) *The Neurobiology of Addiction; New vistas*. Oxford: Oxford University Press.

Royal Society (2005) *Personalised medicines: hopes and realities*.

Sahakian B and Morein-Zamir S (2007) Professor's little helper. *Nature*, Vol 450, pp 1157–1159.

Saper C, Cano G, and Scammel T (2005) Homeostatic, circadian and emotional regulation of sleep. *Journal of Comparative Neurology*, Vol 493, pp 92–98.

Stahl S (2008) *Stahl's Essential Psychopharmacology: Neuroscientific basis and practical application*. 3rd edition. Cambridge: Cambridge University Press.

Sugden C, Aggarwal R, Housden C, Sahakian B and Darzi A (2010) Pharmacological enhancement of performance in doctors. *British Medical Journal*, Vol 340, p 2542.

Volkow N, Wang G, Fowler J, Logan J, Gatley S, Hitzemann R, Chen A, Dewey S and Pappas N (1997) Decreased striatal dopaminergic responsiveness in detoxified cocaine-dependent subjects. *Nature*, Vol 386, pp 830–833.

Box 2: Cognitive enhancing drugs

Professor Trevor Robbins FRS, University of Cambridge

'Cognitive enhancers' are a broad class of drugs from a number of distinct pharmacological classes with diverse molecular mechanisms of action (see Table 1). These may include disease modifying drugs (for example, arresting aspects of Alzheimer's disease pathology) or compounds that exert specific effects on particular chemical neurotransmitter systems. These serve to optimise different aspects of cognitive function, often mediated by different parts of the brain. There is growing published evidence that such actions may range, depending on the compound, from the enhancement of working memory (eg for recalling telephone numbers) or episodic memory (eg for memories of one's own life) to the augmentation of concentration and the restraint of impulsive urges. These functions may be impaired in many neuropsychiatric patient groups, including those with schizophrenia, attention deficit hyperactivity disorder (ADHD), addiction, brain damage, or Parkinson's disease, and are all targets for pharmacological remediation.

Cognitive functioning is powerfully modulated by such factors as motivation and arousal. For example, it is frequently sub-optimal even in healthy individuals as a function of circadian influences, fatigue or sleep deprivation. Thus any agent affecting those processes is bound also to also impact on cognition—often beneficially, although excesses of either motivation or arousal are likely to impair cognition. There are also probable costs as well as benefits of effects of cognitive enhancers because of the theoretical difficulty of optimizing, simultaneously, all of the brain systems that contribute to cognition. Thus what may be good for consolidating long term memory is possibly incompatible with improvements in short term memory. It is also possible that long term use of cognitive enhancers may lead to a reduction of any beneficial effect, for example, because of the phenomenon of drug tolerance, or because of the possibility of neurotoxic effects of such compounds. However, these factors have not been researched extensively.

Another complicating factor is that, as for many other drug effects, there is considerable individual variability in response and some individuals may benefit from a dose of a compound that impairs cognitive function in others. Such variability is already known to depend in part on individual genetic variation and this may enhance future programmes directed towards personalised medicine and risk assessment. The issue of cognitive enhancement in healthy individuals has gained considerable prominence on the basis of recent surveys; whilst many so-called cognitive enhancers cannot be obtained by prescription, they are often available via the internet. This raises regulatory issues, to accompany the ethical controversy that exists on the non-medical—'off-label'—use of cognitive enhancers (see also Sections 2.2, 3.1, 3.2, and 3.3).

Table 1: Candidate cognitive enhancing drugs

Mechanism of action	Examples
'Nootropic' agents (probably acting through effects on cerebral metabolism)	piracetam, aniracetam, nefiracetam, oxiracetam, pramiracetam, pipexide
Glutamatergic agents (some under clinical trial)	ampakines, memantine, D-cyclo-serine, mGLU-R5 potentiators, glyt-1 inhibitors
GABAergic (γ-aminobutyric acid) agents	inverse GABA receptor agonists eg suritozole; GABA-b receptor antagonists eg NS105
Catecholaminergic agents	methylphenidate (Ritalin), atomoxetine, propanolol, bromocriptine, L-Dopa
Cholinergic agents	nicotine and related nicotinic agonists (alpha-7; alpha4, beta 2) galantamine, rivastigmine, donepezil, choline, lecithin
Histamine R3 antagonists	ABT-39,ciproxifan
'Eugeroics' or atypical stimulants: unknown mechanism, possibly via adrenergic or hypocretin receptors, dopamine reuptake blockade, histaminergic or glutamatergic mechanisms	Modafinil
Agents acting on cerebral circulation or calcium homeostasis	vinpocetine, hydergine, phenytoin, nifedipine, nimodipine, Idebenone.
Hormones and neurohormones	dehydroepiandrosterone (DHEA) and DHEA-sulphate, vasopressin
Miscellaneous others	acetyl-L-carnitine, gingko biloba, ginseng, orotic acid, vitamin E, vitamin B6

Further reading

Jones R, Morris K and Nutt D (2007) Cognition enhancers. In: Nutt D, Robbins T, Stimson G, Ince M, and Jackson A (eds) *Drugs and the Future*. Amsterdam: Elsevier. pp 241–283.

2.3 Neural interfaces and brain interference

Professor Irene Tracey, University of Oxford

Background

A new branch of neuroscience has evolved over the past decade that is causing considerable excitement. Aimed at creating links between the human nervous system and the outside world by stimulating or recording from neural tissue, it is hoped to bring unprecedented benefit for people with sensory, motor or other disabilities of brain function. Devices that may enable restorations of these functions are known as brain-machine interfaces (BMIs), brain-computer interfaces (BCIs), neural prostheses or neural interface systems (NISs). For simplicity, here the overarching term, NIS, will be used.

NISs can be broadly separated into two types—interfaces that input to neural systems, and NISs that record electrical activity (output) and use it to *predict* cognitive intentions. 'Input NISs' constitute the majority of NISs to date and have already reached widespread clinical application. They provide direct electrical stimulation input to a specific part of the nervous system to restore or improve function by altering local neural activity (see Box 3).

'Output NISs' are another matter and it is these devices that are causing such a stir in the scientific and clinical communities, as well as among the general public. Output NISs record electrical signals from the brain and this ongoing neural activity is decoded and used to predict cognitive intentions (eg plans to perform a movement). In this way they can potentially replace a lost connection to the outside world. Also, by forming a direct connection between the brain and outside world, NISs free us from the limitations of our bodies. It is this last opportunity that has captured the public's imagination.

Devices that can decipher intentions, which to date are generally movement related, are being developed into clinically viable systems for patients who are paralysed. As is often the case with advances made in the medical arena, we are simultaneously seeing these developments penetrate the games and toy industries too. Such translation brings knowledge, acceptance and normalisation of this extraordinary concept within society.

Other techniques, sometimes known as 'brain interference' devices, and which include Transcranial Magnetic Stimulation (TMS) and Transcranial Direct Current Stimulation (tDCS), are now widely used in neuroscience research with some clinical applications (see Box 4). These devices do not encode or decode brain signals but instead inhibit or excite brain activity to produce certain behavioural outcomes. They are limited in the specificity regarding which brain region can be stimulated and

are relatively crude when compared to the precision of NISs. However, they are useful devices for allowing 'non-invasive', controlled manipulation of cortical brain regions and as such, they contribute to a more causal understanding of human brain networks, particularly when combined with other brain imaging tools such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) (see Section 2.1).

NIS research is a thriving area of neuroscience that links animal and human neurophysiology research and brings new insight into the neural basis of behaviour and perception. Advancement in NIS technology heralds new ways for understanding brain function and our understanding of neural coding and representation, plasticity, brain-behaviour relationships, as well as the neurobiology of disease. But what about extending these NISs to other forms of brain manipulation—aimed at cognitive enhancement or neural 'modification' or 'correction'? As these technologies improve, we are faced with the realisation that a new era has begun in neuroscience and with it important issues are raised that warrant discussion (see Sections 3.2 and 3.3).

Early studies that provided the impetus for NIS developments

In the 1960s and early 1970s pioneering experiments from awake, behaving nonhuman primates showed that the activity of neurons within a specific brain area (the primary motor cortex) was directly correlated to specific aspects of movement. More importantly, these movements could be predicted by the neural activity within this area. These animals demonstrated that they could wilfully control the activity in their motor cortex to alter the movement of an independent device. This provided the first evidence that primates could learn feedback control of neural activity without actually performing any bodily movements. These early observations provided the impetus for the following thirty years of research to develop viable NISs for humans (Evarts 1968; Fetz 1969; Humphrey *et al.* 1970).

The current state of the science

Two forms of neural interface are possible:

i) *Open-loop* prediction which records neural activity from multiple sites to predict behaviour. Open–loop paradigms enable us to learn what features of neural activity are critical for eliciting the measured behavioural outcome. We can then record and then decode these.

ii) *Closed-loop* control, which records neural activity to guide a device controlled by an animal/human that receives sensory feedback for learning purposes. Here, the experimental subject can use feedback (eg watching where the external device is moving) to modulate neuronal activity on an ongoing basis, so that the accuracy of the intended outcome (eg the movement of the device) can be improved. In essence, the NIS here effectively acts as a virtual mirror to real neuronal activities.

Closed-loop control is the common approach used in clinical applications, as

these largely focus on paralysed patients who require control of an artificial limb or device via intention, and where feedback to improve performance is critical. Patients suffering spinal cord injury, stroke, degenerative disorders and amputation all retain a brain mechanism to generate movement intentions. All they need is a way to deliver motor commands from the brain to an artificial aid or directly to still functioning muscles (called Functional Electrical Stimulation).

Many examples have recently been published, often using non-human primates (Truccolo *et al.* 2010; Velliste *et al.* 2008; Kipke *et al.* 2008). Here, recordings are taken and decoded from many areas of the cortex, and used for prediction of movement or control of an external robot or device.

Proof-of-concept studies in rats and non-human primates paved the way for determining the essential elements for a successful closed-loop NIS. This has now been translated to early stage human clinical trials with considerable success, and is generating much media and public interest (Hatsopoulos and Donoghue 2009; Hochberg 2006). To enable a patient with severe paralysis to regain control, communication and independence would be a tremendous achievement. Considering the number of conditions where a disconnection between a healthy brain and target muscles produces paralysis but where the patient has a normal capacity for planning and imagining movement, this is the first obvious target area of application. Patients suffering spinal cord injury, stroke, degenerative disorders and amputation all retain a brain mechanism to generate movement intentions. All they need is a way to deliver motor commands from the brain to an artificial aid.

Potential clinical applications

A pilot study of the first human long term implanted multielectrode array-based NIS, called BrainGate, is ongoing and so far deemed successful (Hochberg 2006). The NIS records signals from an implant within the part of the motor cortex responsible for the arm in patients with severe paralysis, following injury. They are now able to move a cursor on a screen and perform grip and transport actions using a robot's arm. What is astonishing about these results is that years after injury-induced paralysis, normal brain activity was still present in the motor cortex that could be wilfully modulated.

Functional imaging studies in humans show that imagined movement produces blood-flow-related changes similar to those produced during real movement, but it was assumed that the nature of the underlying neuronal activity might be subtly different and not viable for NIS. From this pilot NIS study, it is apparent that the underpinning brain patterns observed are similar whether movement is imagined or performed.

Animal studies show that the brain rapidly changes following injury (often in a maladaptive way) due to plastic mechanisms, affecting activity in primary cortical areas. However, recordings in various chronically injured patients now challenge this view. Clearly, years after sustaining dramatic injuries the neural

activity is present and capable of being harnessed to operate artificial devices. Such findings are encouraging but perhaps surprising and raise questions about how closely findings in animals are applicable to humans. These results also encourage us to reconsider attitudes towards patients who are severely disabled but have brains perfectly capable of encoding behaviour.

A decoder must simultaneously collect information on both neural activity and behaviour in order to map one to the other. This is not possible in patients who have already lost the behavioural component (eg movement). However, passive visual observation of a task produces neuronal responses near identical to those observed when the task is performed. This knowledge provides us with the concept of 'mirror-neurons' (neurons that behave the same when a task is observed as when that task is performed) that can be used in the development of NISs bespoke to patients while contributing to a better understanding of the role of such mirror-like neurons in humans.

As NIS technologies improve, broader applications will become available. For example, better software in the NIS can generate smoother movements of external devices, as well as allowing more motor control and movement options, which might suit severely disabled individuals. One context where this ability is potentially applicable is for patients in vegetative states and comas. There has been a lot of recent publicity surrounding data from neuroimaging studies showing that patients who are in a persistent vegetative state may have wilfully controlled thoughts (Owen *et al.* 2006). The media coverage surrounding these findings indicates that the public wants to engage and debate the implications of this work (Cruse and Owen 2010).

Brain-Body Uncoupling

One potentially very interesting extrapolation of NISs in the future is in 'brain-body uncoupling'. As described above, output NISs work because the same group of neurons which normally move a limb can instead be co-opted to move an artificial device. Fascinatingly, in non-human primates it has been observed that these neurons quickly adapt such that no or minimal movement of their own limbs occur when moving the device in the same manner. The capacity of the brain to disconnect from its body parts and interact with the world via artificial means has wide-ranging and significant implications.

NISs as tools for neuroscience research

NISs have a considerable future role in many areas of basic neuroscience research. NIS and other neuroimaging studies have taught us that large networks of neurons beyond specialised motor regions are involved in even the simplest of movements, as many brain areas can encode movement to some extent. These findings argue against focussing solely on functionally specialised brain regions in order to understand systems behaviour.

Further, NISs offer an opportunity to understand how the activity of individual neurons and the surrounding groups of brain cells (measured through field

potentials) inter-relate in coding information. These two types of electrical signals are often studied separately—but NISs take in both types of information.

The closed-loop NIS provides an unprecedented opportunity to explore learning in the human brain in the context of short-term and long-term improvements. Because the NIS forms a direct, causal link between the recorded brain area and behaviour, any behavioural changes can be attributed to changes in neural activity from the recorded brain area and not downstream areas (eg spinal cord, muscles)—as these have been removed from the system. This is not to say that there aren't changes occurring upstream in other brain areas that in turn provide altered inputs to the recorded brain area, but this can now be studied using distributed NISs.

Recurrent NIS and 'cognitive prostheses'

There is considerable interest in creating a recurrent NIS system where neural activity is recorded and processed in real-time to control electrical stimulation of particular areas of the brain or muscles through other implanted electrodes. The brain could then learn to incorporate this activity into normal function. Initial devices were developed using a 'neurochip' that interacts continuously with the brain of a monkey, allowing continuous operation during free behaviour and sleep. Future applications might include using NISs to directly control a patient's paralysed muscles (instead of an external device) where the normal pathway has been damaged.

More interesting and ethically challenging applications include those producing long-term modification of the strength of connections between brain cells. Given that the strength of connections between cells underpins learning and behaviour, NISs could potentially 'force' the brain to react in a certain way to a certain stimulus or 'learn' something. However, given our ignorance of the way that memory is structured this seems a distant possibility.

Technological developments make it possible to implement recurrent NISs in higher-order cognitive areas of the brain—eg hippocampus, producing what could be called a 'cognitive prosthesis'. This could cause significant changes in how the brain operates and functions.

Such recurrent NIS induced plasticity opens opportunities for experimentation and clinical translation (eg to facilitate recovery from stroke), but also manipulation and exploitation (see Section 3.2).

Technical challenges and opportunities

The primary requirements of an NIS for the broad range of neuroscience applications are recording and/or stimulating from a number of discrete parts of the brain at requisite spatial resolutions for specific periods of time, safety, usability, reliability, patient acceptance, and cost. A number of electrode technologies are in various stages of development with wire bundles and arrays being the simplest technology and most widespread type of implantable electrodes. Microfabricated electrode arrays (neural probes) are more complex

but can be customised to meet specific experiment requirements to have a larger 'design space'.

Technical developments are thus progressing that will allow for more complicated NISs to be stably implanted that will record robustly and reliably for many years from different brain areas at the same time. Wireless transmission of information is also a real possibility. In addition, developments in better understanding of 'background' brain activity may allow for more detailed control of the behavioural output. Rather than controlling a massless, virtual object or physical device via a controller, it is important to develop NISs that can control a system that has physically realistic dynamics – early results for this are promising.

Finally, NISs to date have largely relied upon visual feedback. However, incorporating other forms of sensory feedback will be important, such as tactile and proprioceptive (sense of the body's position and movement) sensations. For example, an NIS might be able to (artificially) stimulate muscles directly (via functional electronic stimulation) *and* receive feedback from the body's network of sensors that provide proprioception. Or, during movement, an NIS could stimulate the areas of the brain that elicit tactile and proprioceptive experiences in individuals where these are lacking.

Less / non-invasive NISs and extrapolation to the games and toy industry

Non-invasive NISs use brain activity recorded from the scalp rather than via direct implantation, and support reasonably high brain-based control after extensive user training. The games and toy industries have leapt into this area realising how such devices have captured the public imagination. Early examples are simple games where participants wear a headset, focus on a small ball in a cylinder, and control the motion of the ball through thought. More complex games have players move a tiny foam ball through a mini–obstacle course using fans that cause the balls to rise. Both toys employ EEG (a method of measuring brain activity through the scalp, see Section 2.1) and utilise a wireless headset equipped with sensors that read alpha and beta waves, relaying signals to the toys.

Using more sophisticated developments in neurotechnology, a personal interface for human computer interaction has been developed by Emotiv. It is described by the company as 'a high resolution, neuro-signal acquisition and processing wireless neuroheadset that uses sensors to collect data to detect player thoughts, feelings and expressions and connects wirelessly to a computer'. Emotiv claims: 'you can use your thoughts, feeling, and emotion to dynamically create color, music, and art; it brings life changing applications for disabled patients, such as controlling an electric wheelchair, mind-keyboard, or playing a hands-free game; you can experience the fantasy of controlling and influencing the virtual environment with your mind'. The headset can be linked to software applications by converting detected events from 'thoughts' into any combination of keystrokes: For example, smile detection can be linked to characters such as ':)', so that chat applications know when you smile.

Less invasive NISs can record from the surface of the brain. This requires less penetration of the brain than intracortical recordings and has a higher spatial resolution than scalp based NISs, and other benefits including potential longer-term stability than intracortical recordings.

Conclusion

NISs exist and are being developed at a rapid pace by both the industrial and academic sectors with wide-reaching application to basic science, clinical problems and the games and toy industry. Extrapolation to other fields, for example intelligence and defense agencies, will occur in the coming years (see also Section 3.2). At this point no legislation exists to control their use, as in general, devices are subject to considerably less stringent approval systems than pharmaceuticals. Nevertheless, the proven track-record of many types of input NISs and early indications with the more advanced output NISs suggest that potential benefits exceed the potential for ill-use and harm.

References and further reading

Cruse D and Owen A M (2010) Consciousness revealed: new insights into the vegetative and minimally conscious states. *Current Opinion in Neurology*, pp 656–660.

Evarts E (1968) Relation of pyramidal tract activity to force exerted during voluntary movement. *Journal of Neurophysiology*, Vol 31, No 1, pp 14–27.

Fetz E (1969) Operant conditioning of cortical unit activity. *Science*, Vol 163, No 870, pp 955–958.

Hatsopoulos N and Donoghue J (2009) The science of neural interface systems. *Annual Review of Neuroscience*, Vol 32, pp 249–266.

Hochberg L, Serruya M, Friehs G, Mukand J, Saleh M, Caplan A, Branner A, Chen D, Penn R and Donoghue J (2006) Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, Vol 442, No 7099, pp 164–171.

Humphrey D, Schmidt E and Thompson W (1970) Predicting measures of motor performance from multiple cortical spike trains. *Science*, Vol 170, No 959, pp 758–762.

Kipke D, Shain W, Buzski G, Fetz E, Henderson J, Hetke J and Schalk G (2008) Advanced neurotechnologies for chronic neural interfaces: new horizons and clinical opportunities. *Journal of Neuroscience*, Vol 28, No 46, pp 11830–11838.

Owen A, Coleman M, Boly M, Davis M, Laureys S and Pickard J (2006) Detecting awareness in the vegetative state. *Science*, Vol 313, No 5792, p 1402.

Truccolo W, Hochberg L and Donoghue J (2010) Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. *Nature Neuroscience*, Vol 13, No 1, pp 105–111.

Velliste M, Perel S, Spalding M, Whitford A and Schwartz A (2008) Cortical control of a prosthetic arm for self-feeding. *Nature*, Vol 453, pp 1098–1101.

Box 3: Neural interface systems

Professor Irene Tracey, University of Oxford

'Input Neural Interface Systems (NIS)' are now widely utilised in medicine. If we consider these input 'neuroprostheses' as falling into categories of sensory (eg auditory, visual), motor (eg bladder function), pain relieving (eg spinal cord stimulation) and cognitive (eg Alzheimer's), it is clear there is widespread need and potential.

Cochlear implants, auditory brainstem implants and auditory midbrain implants are the three main types of auditory prostheses, with the first cochlear implant dating back to 1957 and being the most successful of all three types. Unlike hearing aids that simply amplify sound to send it through the external ear, a cochlear implant has a microphone that receives sound from the external environment and sends it to a processor that digitises the sound, filters it into separate frequency bands and then sends these signals to the appropriate region in the cochlear corresponding to these frequencies. Estimates from 2006 suggest that over 100,000 are in use worldwide.

Likewise, a visual prosthesis creates a sense of an image by electrically stimulating nerve cells in the visual system. A small camera wirelessly transmits to an implant, which maps the image across an array of electrodes—stimulating over a thousand locations in the retina to create an image.

For pain relief, spinal cord stimulators are used to stimulate sensory cells in the spinal cord to produce a tingling sensation in the area of a patient's, which provides pain relief. Again, they are increasingly implanted for intractable chronic pain.

Motor prosthetics are devices that support function of the autonomous nervous system, including an implant for bladder control whereby a device delivers intermittent stimulation which improves bladder emptying. Directly stimulating muscles (functional electrical stimulation) is another means whereby motor prosthetics are used to produce movement of a limb in patients with motor disabilities, and of course cardiac pacemakers would be another related example. Cognitive prostheses, which are discussed in more detail in Section 2.3, aim to restore cognitive function to patients with brain tissue loss due to injury, disease, or stroke by performing the function of the damaged tissue – this can be via 'input' NIS or 'output' NIS. An example of a common 'input' NIS in this domain would be deep brain stimulation for alleviation of Parkinson's disease related symptoms.

In short, as these many devices become safer, and our understanding of how the brain works increases, our capacity to develop the devices we will see their use increase for the betterment of patients' quality of life and well-being.

Box 4: Transcranial magnetic stimulation

Professor Geraint Rees, University College London

Transcranial magnetic stimulation (TMS) is a brain stimulation technique that uses electromagnetic induction to induce weak electrical currents in the brain using a rapidly changing magnetic field. A coil of wire enclosed in plastic is held close to the scalp over the brain area to be stimulated. When a current is passed through the coil, a magnetic field is produced oriented at right angles to the plane of the coil. Magnetic fields pass through the scalp and skull and activate nerve cells in the part of the brain underlying the coil. The effect of this activation is to transiently disrupt the stimulated brain areas, and to produce activity in distant brain areas connected to the stimulated areas by neural connections (axons). TMS can both stimulate the underlying brain region to produce a muscle twitch and nullify brain activity by introducing 'noise'. The effects of stimulation are temporary and there are no long-term side effects, though there is a very small risk of inducing an epileptic seizure during stimulation. This possibility is minimised through adherence to internationally agreed safety guidelines.

TMS belongs to a larger family of non-invasive brain stimulation technologies, including transcranial direct current stimulation (tDCS) where weak electrical currents are passed through the skull through attached electrodes to modulate the activity of neurons in the brain. All of these techniques are commonly used to investigate whether activity of neurons in a particular brain area is necessary for a particular function. If transient stimulation impairs a particular mental process or function, then it can be concluded that the stimulated brain area is causally necessary. TMS can be used to investigate the timing of mental processes, by showing that only stimulation at one particular *time* during a process is effective at causing disruption. This ability to investigate causality has made it a powerful tool complementary to neuroimaging techniques that are purely observational, and so cannot be used on their own to determine whether a brain area is *necessary* for behaviour (see Section 2.1).

This ability to change brain activity raises the question of whether TMS might also be capable of deliberately manipulating brain activity and therefore changing thought or behaviour (see Section 3.2). However, there are very significant barriers to this that have not been overcome. While the fictional film *Eternal Sunshine of the Spotless Mind* envisages a future in which a TMS-like device can selectively erase unwanted memories, in reality such memories (as with all thoughts, perceptions and actions) are encoded in the brain at a very fine spatial scale and interwoven with other neuronal representations. Because TMS and other techniques necessarily operate at a much coarser spatial scale they cannot effect such selective disruption, even if the desired pattern of neuronal representations was known. Moreover, the effects of these

approaches are transient, lasting a few hundred milliseconds at most, and not permanent. Thus, while approaches to 'brain reading' show some ability to decode individual thoughts, perceptions and actions, there is no commensurate non-invasive technique available, either now or in the immediately foreseeable future, that will disrupt or change those patterns of activity selectively.

2.4 A determinist view of brain, mind and consciousness

Professor Wolf Singer, Max Planck Institute for Brain Research, Frankfurt

Background

Progress in brain research allows scientists to explore the way that the connections between neurones in the brain (the 'functional architecture') guides higher brain functions such as perception, reasoning, decision-making, planning and consciousness. These studies indicate that mental functions are based on neuronal processes in very much the same way as the many other behavioural functions that are controlled by the brain. This view is in conflict with our intuitions and requires reconsideration of a number of ethical issues. This essay provides a determinist view of behaviour and decision making that is not shared by all (see Section 3.3 and Box 5).

Implicit assumptions

In their efforts to understand the organisation and function of the brain, neuroscientists make a number of assumptions that are taken for granted. For instance, they consider the nervous system to be a specialised organ whose structural and functional features depend on genetic and other factors in the same way as for other organs. The general layout of the various centres of the brain and of the connections among them is specified by genetic instructions. However, to a much greater extent than any other organ, the properties of the brain can be modified according to a person's experiences. During the initial development of the brain, which in humans lasts until about the age of 20, the wiring of the brain undergoes major modifications and these depend on environmental influences. At birth most of the neurons are already present and have migrated to their final position but many of them are not connected yet. Many new connections are formed only after birth and this outgrowth of connections continues until adulthood. These newly formed connections are maintained or reduced, guided by the activity of neurons, and by any external factors that affect neuronal activity. Therefore, all interactions with the physical and socio-cultural environment influence the final layout of neuronal connections in the brain. These structural modifications are complemented by lifelong learning processes (which are discussed further in Module 2 of the *Brain Waves* project, Neuroscience: implications for education and lifelong learning). It is likely that this is a result of existing connections being strengthened or weakened depending on use. In this way, there is ongoing brain modification after the age of 20.

The structure of the mature brain is therefore determined by three processes:

i) Genetic instructions that specify the general lay out of the brain;

ii) Epigenetic shaping of connections that adapts the brain to its environment during development;

iii) Lifelong adaptation in response to experience.

Neurobiologists assume that all functions of the brain are determined by its structure and the connections between neurons (known as the 'functional architecture' of the brain). In contrast to the situation for computers, in which different hardware components are reserved for different operations such as storage of data and computations, all functions of the brain are realised by widely distributed networks of neurons and uniquely determined by the functional architecture of these connections. This defines the flow of signals and the structure of the complex activity patterns that are the basis of brain functions. These include not only basic functions such as the ability to perceive, remember, and act but also higher functions such as the ability to decide, control attention, generate emotions and finally to understand and generate speech, to consciously deliberate and to be aware of oneself as an independent, autonomous and intentional agent. It follows from this view that mental phenomena are the consequence and not the cause of neuronal interactions. A thought or a decision is the result of preceding computations and therefore cannot *per se* influence the functioning of neuronal networks. The future dynamics of neuronal networks are influenced by the neuronal activity patterns that underlie the thoughts and decisions.

The evidence

Observations suggest that the development of complex nervous systems is the result of a continuous, self-organizing process. Throughout development, close correlations exist between the maturation of distinct brain structures and the emergence of particular brain functions. The behaviour of simple organisms can be fully accounted for by the functions of their neuronal networks; and the same is likely to be true for more complex organisms.

The close relationship between the function of brain structures and mental phenomena has been demonstrated in clinical studies by the loss of specific functions after structural damage. More recently non-invasive imaging technologies (see Section 2.1) have provided overwhelming evidence that even our most 'private' thoughts, decisions and emotions are preceded by the activation of defined networks of neurons (Frith and Frith 2010; Frith and Rees 2004).

It follows that both subconscious and conscious processes are the result of neuronal interactions. Our perceptions, emotions, decisions, plans, thoughts, arguments, and value assignments are shaped by sequences of neuronal states that are causally linked.

Conscious versus unconscious processing

These causally linked brain processes determine which of our many mental contents actually reach consciousness. Evidence indicates that we are aware of

only a tiny fraction of the neuronal activities that guide and control our behaviour. Some signals are always excluded from conscious processing, such as those involved in controlling blood glucose levels and kidney function. The same is true for implicit knowledge that determines how we perceive, decide and react, which we use without being aware of the fact that we have it. It is contained in the layout of the brain's functional architecture, which in turn is mostly determined through genetics and shaped during early development without conscious control.

In contrast to this implicit knowledge, knowledge acquired by learning has access to consciousness and can therefore be subject to conscious deliberation. However, the number of items that can be held simultaneously in consciousness is limited and it is not possible to retrieve them 'at will'. Everybody is familiar with the problem of not being able to recall names or situations that have previous been well remembered.

Wrong intuitions

The evidence described above contradicts our intuition that we can always freely decide what we are going to do next and which factors we are going to consider when we plan future acts. If mental processes are the consequence of neuronal processes then decisions are the result of self-organizing neuronal processes that converge towards the most probable stable state in the given conditions. These conditions usually comprise a very large number of variables that influence neuronal activity. One decisive condition is the specific functional architecture of the brain, which varies from individual to individual because of differences in genetic dispositions, developmental imprinting, and experience. The other relevant condition is the activation state preceding the moment of decision making. Activation patterns that have access to consciousness are subjectively experienced as causes or arguments and can therefore be quoted as reasons for a particular decision. We say 'we have decided in this way because . . . 'and then we give the reasons that we are consciously aware of. However, much of the activity that actually prepared and determined the decision process escapes conscious recollection.

Imaging studies indicate that there is sometimes a clear dissociation between the onset of neuronal activity patterns that finally result in the decision to move a finger and the moment when subjects become aware of their intention (Moser *et al.* 2010; Soon *et al.* 2008). After having been identified with functional magnetic resonance imaging (fMRI) the neuronal activation patterns associated with the voluntary act to press a key with either the right or the left hand, subjects are asked to perform freely paced key presses while their brain activity is measured. In this case the onset of activation patterns predicting a right or a left hand press can precede by up to ten seconds the subjects' awareness of having decided to press.

In this experiment, the instruction is stored in working memory, and the motor act (the key press) is prepared before the subjects

become aware of what they are going to do next. Here, the 'intention' is concordant with the action. However, there are also conditions where the real causes of an action dissociate from the reported intentions. For instance, subjects can be instructed to perform a particular action without being aware of having been instructed. When subjects that comply are asked why they performed the action, they reply as if they intended to do so (eg 'I decided to do this because I wanted to ….'). In this case the reported reasons for a particular action do not match the real reason, but are experienced as if they had been at the origin of the performed action. These examples illustrate impressively that only a fraction of the neuronal processes that prepare decisions and guide behaviour have access to conscious recollection and that neurobiological processes precede the awareness of having reached a decision.

Differences between unconscious and conscious decisions

Intuition tells us that there is a difference between being able consciously to weigh arguments and then decide what to do, and acting spontaneously without reflecting consciously different arguments and possible consequences. This is not necessarily contradicted by what we know about subconscious processing as described in the section above. The rules governing subconscious and conscious decisions differ (Engel and Singer 2008; Dehaene 2008). The neuronal mechanisms underlying decision making are the subject of intense research and likely to differ depending on the task being done (see also Section 2.5). One line of evidence suggests that evidence is accumulated until a particular threshold is reached, and a decision is then made (Shadlen *et al.* 2008). This correlates to an increase of activity in neural circuits representing the different stimuli. The circuit with the fastest rise in activity will reach the threshold soonest. A different, related model suggests that decisions are the result of competitive interactions between neuronal networks representing different options. This model rests primarily on behavioural experiments and requires further neurobiological examination (Keysers *et al.* 2008). It is likely that most of these decision processes do not require conscious deliberation, or conscious recollection of the variables involved).

Conscious decision making appears to follow different rules. As consciousness has a limited capacity, only a small number of variables can be dealt with. This requires the contents of declarative memory (memory that can be recalled) to be scrutinised sequentially, which takes time. Moreover, this rational strategy can result in well adapted solutions only if the variables involved are sufficiently reliable. For most decisions reached in this latter way subjects can correctly report the relevant arguments.

Decision processes occurring without conscious deliberations seem to be less constrained. They can exploit the rich database of subconscious heuristics (problem solving) and therefore can process many more variables in parallel and cope better with unreliable and 'noisy' variables. Often, these subconscious

processes lead to more adapted responses than the conscious deliberations, in particular if there is little time for the preparation of a decision, if multiple interdependent variables have to be considered simultaneously and if the reliability of these variables is low.

Usually, subconscious decisions result in immediate action. However, they may also be inhibited by simultaneously occurring conscious deliberations. In this case the solutions presented by the subconscious and conscious decision mechanisms need not be congruent. We experience this dissociation when we say that we have decided after having carefully scrutinised all arguments but that the outcome somehow does not feel right. Vice versa certain decisions are felt to be absolutely right even though they are considered entirely irrational. Thus, both the subconscious and the conscious decision mechanisms are equally relevant in determining future behaviour. Both are based on neuronal processes that influence in a causal way future states of the brain but they follow somewhat different rules.

Why is intuition in conflict with neurobiological evidence?

We tend to feel that there is an agent in our brain that is at any time free to make a decision that overrides that of the deterministic neuronal machinery. There are at least two reasons for this. First, we are only aware of the results of the neuronal processes in our brain, and not of the mechanisms of these processes. Second, we tend to assume that these processes are essentially linear. As linear processes cannot account for the processes that we ascribe to ourselves and others, we postulate the existence of an intentional agent that is not fully determined by the neuronal machinery. Linear deterministic systems behave like reliable clocks. They cannot self-organise, cannot be creative and their future behaviour is fully determined by initial conditions. Their trajectories can only be changed by external forces. In contrast, we experience ourselves and others as creative, and able to take new courses without perceivable external influences. However, if one accepts that the brain is a complex self-organizing system with non-linear dynamics, an independent intentional agent is dispensable. All characteristics attributed to such an agent are emergent properties of non-linear, self-organizing systems, such as the brain.

Potential consequences for our legal systems

The neurobiological evidence reviewed above is incompatible with the view that a person, at the moment of having reached a decision, could have decided otherwise. However, this assumption is said to be at the base of our legal systems and the justification for attributing responsibility to a person and sanctioning deviant behaviour. Consequently, it could be argued that if this premise is false, persons are not responsible for their actions and therefore exempt from sanctions. The original assumption implies a separation between an intentional agent (the person) and the mechanics of the nervous system

that are required for the execution of the agent's orders. What would be the consequences if one abandoned this traditional dualist approach and adhered to the interpretations suggested to us by modern neuroscience? Would this require a change of legal practices or only a revision of the interpretation of deviant behaviour? These issues, and others discussed below, will be considered as part of Module 4 of the *Brain Waves* project on neuroscience, responsibility and the law.

The view that a person is responsible for what she or he does (meaning that they are the causal agent) is not invalidated by neurobiological evidence, because all authorship remains with the deciding and acting person. How about the measures taken to prevent deviant behaviour or to protect society from harm? Should they be modified in view of the evidence that the delinquent, in the moment of his or her action was unable to decide otherwise?

Punishment and confinement

Punishment is an evidence-based means of prevention because of its educative and deterrent effect. Constraining freedom is likely to remain as a means of protecting others. Punishment also seems to fulfil a second function, in satisfying the human need for justice. Legal systems do not seem to only sanction the amplitude of deviance but also the severity of the consequences of deviant behaviour. Crossing a red light without causing an accident may lead to a temporary loss of the driver's license. However, if the consequence is a severe accident with casualties, the punishment tends to be much more severe, even though the deviant behaviour, the 'subjective guilt', was exactly the same.

Attenuating conditions

What light does neurobiological evidence shed on the attribution of mitigating neurological factors? It is unlikely that this would change anything if the purpose is to find out how likely it is that anyone else would have done the same thing in the same circumstances. To determine this, it would be necessary to examine the options available at the moment of the decision, and the extent to which the subject was able to fully exploit the brain mechanisms required for reaching the decision. This comprises evaluation of dependencies from external pressure, from internal constraints such as addictions or compulsory drives, the time available for a decision and the ability to rely not only on subconscious heuristics but also on conscious deliberations. Conscious deliberation is attributed particular significance because most of the social imperatives that have been acquired through education are amenable to conscious processing.

Viewed from a neurobiological perspective, this does not assume anything about 'free will' because it only examines how far the behavioural dispositions or character traits of a person exhibiting deviant behaviour differ from the normal distribution. If most of the human beings raised in a particular cultural context would have acted in the same way given the conditions, sanctions are moderate because attenuating factors will prevail. In contrast, if the conclusion is

that under the given circumstances the deviant behaviour has to be considered as extremely far from the norm, sanctions are usually drastic.

Brain research posits that behavioural dispositions, character traits and the decision mechanisms available to a person are determined by the functional architecture of the person's brain. Is it then that legal systems in practice evaluate the extent to which the functional architecture of a law breaker's brain deviates from the normal distribution and adjusts sanctions accordingly? If this were a commonly accepted stance, would it interfere much with the common practice of sanctioning deviant behaviour or would it only lead to a more empathic attitude towards persons who happen to end up at the negative tail of the normal distribution?

The following simple thought experiment is meant to illustrate this point. If a person has committed what is considered to be cold blooded murder in order to obtain some benefit and a tumour is discovered in his frontal lobe, extenuating conditions may be granted. One might argue that this tumour disrupted the pathways that link the storage site of moral values with the inhibitory centres that would normally prevent the fatal action. From a neurobiological point of view, however, one might argue that any person capable of committing such a crime must always have some abnormalities in the functional architecture of his or her brain, even if this abnormality is not detectable with current technologies. Genetic dispositions could have limited the storage capacity of the networks in which moral values and imperatives get stored, or they may have led to abnormally weak control mechanisms for the inhibition of actions. The same abnormalities may have been caused by developmental mishaps, insufficient installation of moral imperatives through education, or deficient inhibitory mechanisms due to lack of training during brain maturation. If all these features of the functional architecture are actually in the normal range, then one would have to assume temporary abnormalities in the system's dynamics, for instance caused by metabolic disturbances, or by some highly unlikely but still possible deviations of the brain's dynamics.

Even though none of these abnormalities are detectable with currently available methods this does not detract from the conclusion that there must have been a neuronal cause for the deviant behaviour whatever its exact nature. Thus, members of our society who had the misfortune to possess a brain which ended up at the negative end of a normal distribution should have our empathy. But does this exempt our society from its duty to protect all its members and to define what is tolerable and what is not?

References and further reading

Engel C and Singer W (Eds) (2008) *Better than Conscious? Decision Making, the Human Mind, and Implications for Institutions*. Strüngmann Forum Reports. Cambridge, MA: MIT Press.

Frith U and Frith C (2010) The social brain: allowing humans to boldly go where no other species has been. *Philosophical*

Transactions of the Royal Society B, Vol 365, No 1537, pp 165–176.

Frith C and Rees G (2004) *Brain Imaging: The Neural Correlates of Consciousness*. Oxford: Oxford University Press.

Dehaene S (2008) Conscious and nonconscious processes: Distinct forms of evidence accumulation? *In*: Engel C and Singer W (Eds) *Better than Conscious? Decision Making, the Human Mind, and Implications for Institutions*. Strüngmann Forum Reports. Cambridge, MA: MIT Press, pp 21–49.

Keysers C, Boyd R, Cohen J, Donald M, Güth W, Johnson E, Kurzban R, Schooler L, Schooler J, Spelke E and Trommershäuser J (2008) Explicit and implicit strategies in decision making. *In*: Engel C and Singer W (Eds) *Better than Conscious? Decision Making, the Human Mind, and Implications for Institutions*. Strüngmann Forum Reports. Cambridge, MA: MIT Press, pp 225–258.

Moser E, Corbetta M, Desimone R, Frégnac Y, Fries P, Graybiel A, Haynes J, Itti L, Melloni L, Monyer H, Singer W, von der Malsburg C and Wilson M (2010) Coordination in brain systems. In: Von der Malsburg C, Phillips W and Singer W (Eds). *Dynamic Coordination in the Brain. From Neurons to Mind*. Strüngmann Forum Reports. Cambridge, MA: MIT Press, pp 193–214.

Shadlen M, Kiani R, Hanks T and Churchland A (2008) Neurobiology of decision making: An intentional framework. *In*: Engel C and Singer W (Eds) *Better than Conscious? Decision Making, the Human Mind, and Implications for Institutions*. Strüngmann Forum Reports. Cambridge, MA: MIT Press, pp 71–101.

Soon C, Brass M, Heinze H and Haynes J (2008) Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, Vol 11, pp 543–545.

Box 5: The biological becomes personal: philosophical problems in neuroscience

Dr Sarah Chan and Professor John Harris, University of Manchester

Early historical beliefs notwithstanding, we recognise the brain as fundamental to shaping who we are, our personality and personal identity. Consider the still-hypothetical prospect of whole-brain transplantation: most people would probably regard this as whole-body transplantation (and not as a brain transplant). Nevertheless, and despite modern scientific understanding of the brain, the philosophical relationship between brain, body, mind and identity remains elusive.

The inherently problematic nature of this can be explored through two related but conceptually distinct questions: 'Am I my mind?', and 'Is my mind my brain?' Clearly, 'we' are not just our brains or our minds: our sense of identity is closely associated with our physical bodies; our experience of the world, though expressed in one form as brain activity, necessarily includes the phenomenon of embodiment. Equally, however, the brain-transplant thought experiment illustrates that selfhood is not solely attached to the body. This is reflected in social and clinical attitudes towards 'brain death' and current philosophical thinking on 'personhood' (Radin 1982; Harris 1985). But are we then embodied minds—a mind within a body—or is embodiment itself an essential element of mind? Moreover the relationship between the physical and mental properties of the human organism is philosophically problematic. Neurobiology attempts to explain mental processes and the workings of the mind in terms of physical neuronal processes in the brain. But is mind (mental) *merely* a property of brain (physical); can all mental states be reduced to physical phenomena? Against such reductionist approaches, we might argue that mental states are not *simply* physical phenomena, any more than a poem is simply words on a page (see also Section 2.4).

The mind-body problem has been the subject of extensive philosophical inquiry since long before the brain was recognised as the organ responsible for thought. Yet although we can now relate states of mind to physical 'states of brain', that does not resolve the fundamental philosophical questions (Ryle 1969).

Despite the perhaps inextricable entanglement of our physical, mental and psychological natures, it is evident that much of what we see as our 'selfness' is brain-dependent. This raises, however, further questions that assume new significance in light of neuroscience. For example, am I still the same person when my brain changes—through injury or disease, the influence of drugs or surgery, or simply experiences and the passing of time? Again, we may now be able to observe, understand, and even affect the physiological basis of these changes more directly.

However, this provides no simple answers to such questions but throws them into sharper relief.

The idea that mind and self have biological foundations has profound implications for our understanding of free will and responsibility. If all our desires and impulses to act can be reduced to neurochemistry, how can 'we' be responsible for our actions or choices? Studies showing that decisions manifest physically in the brain even before one is consciously aware of the decision being made, for example, throw into question our usual assumption that it is the conscious mind that exercises free will. And if we do not have free will, if our brains are deciding for us, then can we truly be responsible for our actions? Responsibility implies volition and intention, not just causation; there is a difference between doing something intentionally and unintentionally, yet both involve causation and consequences and both have correlated brain states. The origin and explanation of free will and of intent, however, remain open to debate (see also Section 2.4).

This applies not only to our everyday understanding of moral responsibility but also to the concept of legal responsibility. When we hear the defence 'My brain made me do it', how should we respond? Neuroscience may radically change our concept of criminal and legal responsibility, and the way we react to wrongdoing (Greely 2006). This topic will be explored in Module 4 of the *Brain Waves* project on responsibility and the law.

References and further reading

Dennett D (1969) *Content and Consciousness.* London: Routledge.

Greely H, The Social Effects of Advances in Neuroscience: legal problems, legal perspectives. *In*: Illes J (ed), *Neuroethics: Defining the Issues in Theory, Practice and Policy*. Oxford: Oxford University Press.

Harris J (1985) *The Value of Life*. London: Routledge.

Radin M (1982) Property and Personhood. *Stanford Law Review,* Vol 34, p 957.

Ryle G (1969) *The Concept of Mind.* London: Hutchinson.

2.5 Reward, decision-making and neuroeconomics

Professor Wolfram Schultz FRS, University of Cambridge

Background

Recent work in behavioural neuroscience has examined how humans and animals make economic decisions about gains and losses. This research is based conceptually on animal learning theory and economic choice theory. It investigates how individual neurons in animals and brain regions in humans process information about gains (rewards) and losses (punishments) and use this information to make economic choices. As animal work forms an important part of this research, investigations are more frequently directed at reward than punishment. The field also incorporates data from brain lesions and pharmacological manipulations of neurotransmitters. The term 'neuroeconomics' refers to the neurobiological basis of economic decision making. It combines economic choice theory with neuroscience in order to understand the neuronal basis for economic decisions. For review articles see Glimcher *et al*. 2008 and Schultz 2008.

The functions of rewarding and punishing outcomes are not defined by specific sensory receptors but are inferred from their influence on behaviour. Rewards are objects or events that induce approach learning (positive reinforcement in animal learning theory), serve as arguments for economic (value-related) decisions, engage positive emotions such as joy, and produce hedonic feelings. Rewards support elementary processes such as eating, drinking and sex, and engage individuals in such diverse behaviours as novelty seeking, foraging, trading on stock markets and social interactions. By contrast, punishments have largely opposite effects to rewards, inducing avoidance learning, negative emotions such as fear, and feelings of displeasure. Obtaining rewards and avoiding punishments is crucial for individual and gene survival.

Aims of neuroeconomics: perspective of neuroscience

The neuroscientist wants to know how the brain works, notably how it organises the behaviour that is crucial for the survival of the organism and its genes. What does the neuroscientist expect from neuroeconomics?

Theoretical concepts

Economic choice theory provides important definitions of decision variables and valid decision models. A decision process can be broken down into its different components, which is essential for designing well controlled behavioural tasks. The impact of rewards and

punishments is quantified by gains and losses in subjective value, called utility, which can be assessed by revealed choice preferences. In general, outcomes are variable and uncertain. Outcomes are therefore adequately described by probability distributions, which are characterised by their expected value and risk. 'Risk' refers to the uncertainty of the outcome, rather than simply the chance of losing. The term 'ambiguity' denotes uncertainty when probability distributions are incompletely known. Risk and ambiguity influence decision making by impacting on the subjective value of risky outcomes.

Behavioural tools

Economics has developed behavioural and analytic tools to assess crucial components of choice behaviour. Examples are the following:

- the Pest procedure (Parameter estimation by sequential testing) assesses values of unknown relative to known outcomes;
- the definition of subjective value by Expected Utility Theory;
- the impact of risk on outcome valuation and choices;
- the weighting of outcome probabilities, loss aversion and reference dependent coding by Prospect Theory;
- the role of past experience in updating predictions of outcomes; and
- the identification of 'irrational' choices of suboptimal outcomes.

Aims of neuroeconomics: perspective of economics

The behavioural economist wants to know how humans make economic decisions in order to survive best in a competitive world with limited resources. The recent financial crisis has shown that many aspects of everyday life and world economics depend on the risk-taking behaviour of a few individuals. What does the economist expect from neuroeconomics?

Biological plausibility

Behavioural economics has established theoretical concepts for individual decision making. Identification of neuronal signals for particular theoretical decision variables would demonstrate its biological plausibility, inform decision models, allow deconstruction of whole decision processes and identify crucial steps that might possibly go wrong.

Identifying brain states related to suboptimal decision making

Controlled, quantitative assessments of anomalies in decision making may help to characterise patterns of inconsistent, suboptimal, 'irrational' choices. Such choices may constitute normal biological phenomena that may not be easily changed without explicit policies. Suboptimal choices would become more understandable if physical, neurobiological, correlates were found.

Reward systems show subjective rather than objective value coding. Nevertheless,

humans tend to make irrational choices, which do not maximise the expected monetary utility of the outcome. Such irrational behaviour is observed often when emotions are involved. For instance, in the ultimatum game, one player proposes how to divide a given amount of money. The second player either accepts the offer and both players get their share, or rejects it and both players receive nothing. Rejection of any non-zero offer in this game is economically irrational; it is attributed to the sense of or outrage. Receiving fair offers activates the insular and prefrontal cortex, suggesting a neurobiological correlate for this emotion. Future economic theory might attribute an economic value to this emotion and generate new value functions with which choices involving emotions can become optimal and thus 'rational'. In this way, neuroeconomics may lead to new normative theories that provide valid explanations of consistent choices across a wider range of human and animal choice behaviour than current utility and prospect theory.

We assume that brains evolved to assure the survival of genes and their carriers in *most* situations. Assuring survival in *all* possible, however unlikely, situations would require extra brain matter and function that would make brains less efficient and their carriers less competitive. The existence of visual illusions demonstrates the boundaries imposed by such efficiency principles. Even with new normative theories of economic decision making, exceptional circumstances may remain in which economic choices will be inadequate, in particular in situations for which our brains have not (yet) evolved.

Current state of neuroeconomic research: how the factors that influence decisions are represented in the brain

The neuroeconomic investigation of reward and punishment is based on the understanding that outcomes are defined by their value and by the risk with which they occur (see Schultz 2006). The brain's reward (or gain) system includes the dopamine system, orbitofrontal cortex, striatum, amygdala and other cortical and subcortical structures in the brain. Punishment (or loss) is processed in similar structures, with the exception of the dopamine system which is concerned more with reward than punishment (see also Section 2.2).

Value

The value of a particular outcome depends on the type of outcome, its magnitude and its probability. Reward neurons code reward value by graded firing, increasing or decreasing activity. Later rewards, which are subjectively valued less (temporal discounting), induce lower value responses, suggesting subjective rather than objective value coding. Neurons code value relative to explicit references in all major reward structures. This neuronal property may underlie reference-dependent reward valuation which is a central tenet of prospect theory. Prefrontal value responses show distorted responses to reward probability in line with prospect theory.

Risk and ambiguity

Reward neurons encode risk separately from value (see Schultz *et al.* 2008). Prefrontal risk signals differ between risk avoiders and risk takers. Risk reduces value signals in risk avoiders and increases value signals in risk takers. Thus, individual risk attitudes may reflect variations of prefrontal function. Some neuronal responses are stronger for ambiguity (uncertainty of an outcome when the probability of distribution of its occurrence is not known) than for risk.

Predictive learning

Predictions contain information about the value and risk of future outcomes and are necessary for making informed decisions. Neurons in all reward respond to stimuli that predict reward value, and some neurons respond to stimuli that predict risk. Predictions are acquired and updated through experience when the actual outcome differs from the predicted outcome (prediction error). Dopamine neurons signal prediction errors for reward value during learning. The insular cortex signals prediction errors for risk.

Current state of neuroeconomic research: how one possible action is selected from multiple options

We decide what to do in a particular situation by assessing the relative values of different responses. Current work in neuroscience aims to determine the way in which the different variables and choice options are represented in the brain. These studies are based on a number of theories (see also Section 2.4), for example:

Reinforcement learning theory

The ultimate behavioural consequence of a decision is an action. Each possible action leads to a particular outcome which has specific value, and the action leading to the highest value is selected. Neurons track the outcome values of the individual actions ('action value'). Rational decision makers select the action producing the highest action value. This action is the result of a competition between neurons carrying different action values.

Theories of evidence accumulation

Neuronal activity in parietal and frontal cortex develops gradually as evidence about the values of the available options accumulates. So-called diffusion-race models conceptualise how the first option reaching a threshold is chosen (see Gold and Shadlen 2007).

Current state of neuroeconomic research: social decisions

One of the great promises of neuroeconomics is the use of formal tests for studying neuronal processes underlying social interactions, including the role of emotions in decision making.

Apes and monkeys

Behavioural work on nonhuman primates identifies 'rational' maximisation without

regard for others in the ultimatum game. Neurophysiological work on monkeys uses economic games against computer opponents, showing activity related to subjective preferences and past actions and rewards. However, the extent to which monkeys perceive such games as a social activity is unclear. Ongoing neurophysiological work on interacting monkeys investigates basic processes such as video game competition and reward observation.

Outcome observation in humans

Human social economic neuroimaging studies employ straightforward situations such as observation of reward and punishment received by others. For example:

- The effect of reward on the striatum is relative on the reward received by others;
- Seeing another person making errors in predicting a reward whilst learning the same task activates the striatum;
- Observing another person receiving painful stimuli activates a neuronal correlate for empathy in the cingulate and insular cortex;
- Following the behaviour of others in the valuation of outcomes activates the striatum.

Games in humans

Human studies often use formal economic games to standardise and measure emotions. The most popular tools are:

- Dictator game: The player splits a received endowment with another player. The game tests fairness and aversion to inequity;
- Ultimatum game (described above): The game tests fairness, inequity aversion and outrage;
- Trust game: The amount of money one player gives to another player is multiplied by the experimenter. The other player gives back an amount which is multiplied again, and so forth. The game tests trust, reciprocation, cooperation and fairness;
- Prisoner's dilemma: Rewards are assigned to two players depending on their cooperation with each other. If the first player cooperates but the second player 'betrays' his/her, the first player receives the worst possible outcome and the second player the best, thus tempting the second player. If both cooperate or both betray each other, they receive intermediate outcomes, which are higher for cooperation than betrayal. The game tests cooperation against defection and self-interest:
- Public goods game. Players can put goods into a pot to be split between the group. The game tests prosocial attitude, cooperation, altruism and defection.

Several brain activations are observed in such studies (see Frith and Singer 2008). Many socially and emotionally positive situations activate the brain's reward system, including fairness in the ultimatum game; cooperation in prisoner's dilemma;

reciprocation and intention to trust in the trust game; and altruistic punishment of players who were uncooperative in the trust or prisoner's dilemma game. Donations to charities activate the striatum, despite one's own money loss. Activations in the amygdala correlate with the trustworthiness of faces (see Figure 4); correspondingly, lesions of the amygdala reduce the recognition of trustworthiness of faces. Many aversive situations activate the brain's punishment system in the anterior cingulate and insula, including receiving unfair and likely unacceptable offers in the ultimatum game. In line with its cognitive control functions in behaviour, the prefrontal cortex is activated when unfair offers in the ultimatum game are rejected.

Policy questions

Neuroeconomics is starting to uncover the neurobiological basis for the properties of economic valuations and decisions. In some cases, these valuations lead to daily choices that could be disadvantageous to individuals and society. By increasing our understanding of these processes we may be able to find means to avoid their detrimental consequences.

We know that certain reward values are coded 'inaccurately' in the brain. For example, we know that the reward processes in the striatum tend to discount the values of future rewards (temporal value discounting). This may be a factor that leads us to invest less in provisions for the future (such as education, healthcare or pensions) than we 'should' do. Similarly, events of low probability (such as an appalling murder or terrorism) can be overvalued in the brain (the rewarding nature of novelty), a factor that might contribute to resources being diverted from issues that affect the majority of the population.

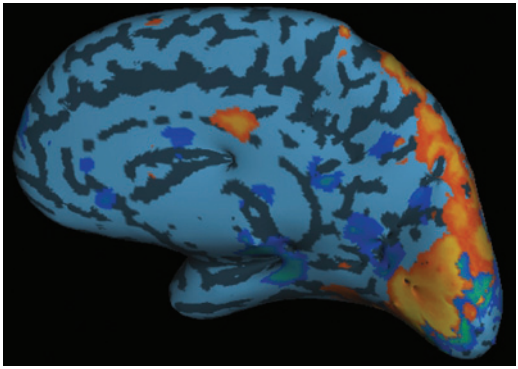
The encouraging news is that social cooperation activates the reward centres of the brain that are involved in plasticity and learning. As we understand more about the processing of reward in the brain, it may be possible to influence our own brains and those of future generations to reward the things that are actually rational for us in the longer term.

Individual variations in how our brains process reward

Variations in brain function across individuals appear inevitable given the complex nature of the brain. As our understanding of individual neurobiological variations in reward processing (both natural and pathological) increases, we can unpack the economic consequences of these variations. Early indications highlight possible roles for this variation in reward processing in very diverse areas:

- risk avoidance and anxiety disorders;
- risk seeking (an area where the behaviour of a few individuals may impact economically on the majority);
- obesity;
- drug addiction (both pathological influences of drugs on reward signals and natural variation); and
- gambling.

Figure 4 When you recognize a familiar face the parts of the brain highlighted in orange on this functional magnetic resonance image (fMRI) light up. In this case 'lighting up' means there is increased blood flow to the areas that are working hardest. Functional MRI allows these regions to be visualized. They can then be superimposed onto a 3D reconstruction of the brain to get a precise picture of the location of those regions. (Reproduced courtesy of Mark Lythgoe and Chloe Hutton, University College London.)



Neuromarketing

A practical application of neuroeconomics concerns consumer behaviour. Demand for goods can be influenced by several factors, many of which impact on the reward system of the brain. Neurons in all reward structures respond to stimuli that predict rewards, and these can be conditioned in a Pavlovian manner without the active participation of the subject. For example, reward predictions stimulate dopamine neurons, and the stimulation of dopamine neurons in rats induces approach behaviour and learning. Understanding more about these reward predictive stimuli could help us appreciate the biological foundations of individual consumer behaviour and perhaps even intervene when this behaviour becomes detrimental.

In addition, a recent study on wine tasting shows that the same wine elicits a stronger reward value signal when its price information is inflated. These data suggest a neurobiological basis for the influence of external influences such as pricing on subjective valuation. Furthermore, studies show that the brain values rewards relative to external references, such as the current economic situation and the rewards of social partners. This may explain in part why the wealthy are not necessarily any happier than those less well off.

References and further reading

Frith C and Singer T (2008) The role of social cognition in decision making. *Philosophical Transactions of the Royal Society B*, Vol 363, pp 3875–3886.

Glimcher P, Caterer C, Fehr E and Poldrack R (2008) *Neuroeconomics: Decision making and the brain*. London: Academic Press.

Gold J and Shadlen M (2007) The neural basis of decision making. *Annual Review of Neuroscience*, Vol 30, pp 535–574.

Schultz W (2006) Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, Vol 57, pp 87–115.

Schultz W (2008) Neuroeconomics. *Philosophical Transactions of the Royal Society B*, Vol 363, pp 3765–3886.

Schultz W, Preuschoff K, Camerer C, Hsu M, Fiorillo C, Tobler P and Bossaerts P (2008) Explicit neural signals reflecting reward uncertainty. *Philosophical Transactions of the Royal Society B*, Vol 363, pp 3801–3811.

3 Neuroscience and society

3.1 Benefits and opportunities

Professor Barbara J Sahakian, University of Cambridge

Background

Neuroscience is of growing importance in the 21st century given the rapid technological advances in this area and their impact on society. For example, neuroscience is critical to the understanding of the brain in health and disease and in developing more accurate diagnosis and new treatments across the lifespan. To illustrate, cognitive training treatments are under development for disorders such as attention deficit hyperactivity disorder (ADHD) (see Figure 5) and substance abuse, and drugs that protect the nervous system from degradation are being developed for Alzheimer's disease. Recent innovative proof of concept studies for drugs such as ketamine and scopolamine suggest that it might be possible to treat patients for depression effectively and very rapidly.

For neuropsychiatric disorders and also for brain injury, cognitive enhancing drugs are being used and further developed to improve functional outcome, quality of life and wellbeing (see also Section 2.2 and Box 2). This area of research in neuroscience raises important neuroethical issues, such as the increasing use of so-called 'lifestyle drugs' including 'smart drugs' by healthy people. The neuroethics of this and other areas is discussed in more detail in Section 3.3 (see also Section 3.2).

During the next decade, there will be marked advances in the use of neuroscience in education in helping children to learn in schools. This will lead to evidence-based programmes of individualised or personalised learning. This area, termed 'educational neuroscience', is addressed in depth in Module 2 of the *Brain Waves* project, Neuroscience: implications for education and lifelong learning.

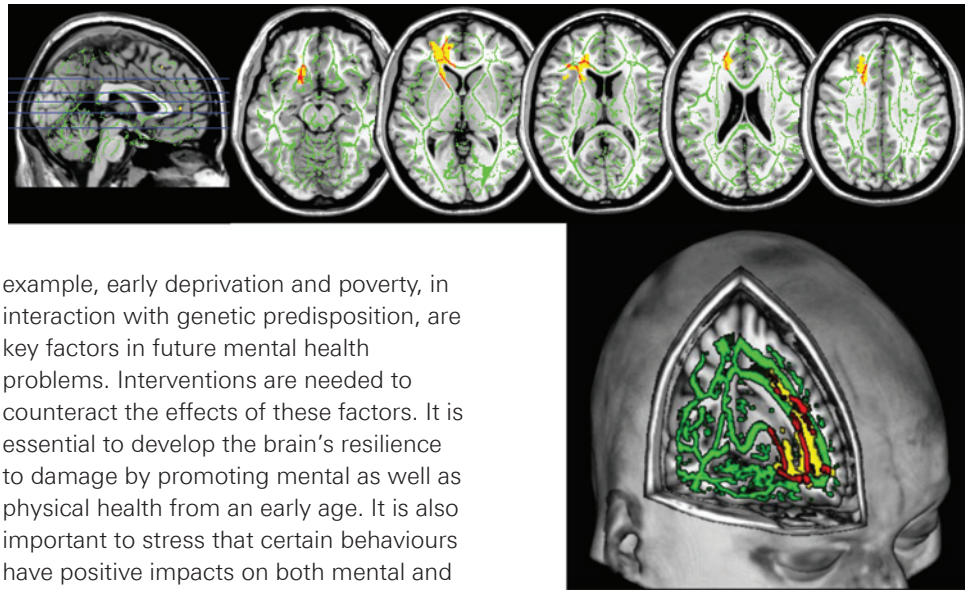
Other areas of neuroscience too will flourish, including those that intersect with social sciences and genetics, such as studies on whether, and under what conditions, we choose prosocial over antisocial or non-cooperative behaviour. The intersection of neuroscience with information technology will expand and further develop, including neural interface systems, neuroprosthetics and functional Magnetic Resonance Imaging (fMRI) feedback (see Sections 2.1, 2.3 and Box 3).

Deep brain stimulation is another area which continues to develop rapidly for treatment for neuropsychiatric disorders, including depression, obsessive compulsive disorder and Parkinson's disease. This involves surgically implanting a medical device that stimulates specific parts of the brain. Similarly, stem cell research applied to the treatment of neuropsychiatric disorders continues to progress (see Section 2.3 and Box 4).

Application to policy

Certain areas of neuroscience will be important for public health policy. For

Figure 5 This image shows a) in green: major white matter tracts generated from a population of 32 subjects-16 healthy volunteers and 16 adult patients with attention deficit/ hyperactivity disorder (ADHD) b) in red: regions where these tracts were found to be abnormal in ADHD patients compared to controls and c) in yellow: regions of reduced white matter density in ADHD patients compared to controls. A, b and c are overlaid on a high resolution brain template acquired with magnetic resonance imaging (MRI). (Reproduced courtesy of Natalia del Campo, University of Cambridge.)



example, early deprivation and poverty, in interaction with genetic predisposition, are key factors in future mental health problems. Interventions are needed to counteract the effects of these factors. It is essential to develop the brain's resilience to damage by promoting mental as well as physical health from an early age. It is also important to stress that certain behaviours have positive impacts on both mental and physical health, and these behaviours should be promoted accordingly. For instance, exercise stimulates the development of nerve tissue in some regions of the brain (Olson *et al.* 2006).

Another example is the importance of early detection and treatment, as well as the need for new treatments, for neuropsychiatric disorders, such as depression and Alzheimer's disease. For example, genetic, cerebrospinal fluid (CSF) and blood, cognitive and neuroimaging biomarkers can play an important part in early identification of these disorders. This would enable some disorders to be prevented, while others could be treated effectively before they develop a chronic, relapsing or progressive course. Furthermore, new insights into underlying mechanisms, coupled with the use of more selective cohorts in clinical trials, is essential for the development of effective drugs in Alzheimer's disease.

Strong links should be fostered between academia and industry to allow valuable collaborations to facilitate drug development and evaluation. As we move further into the 21st century, it is important

to develop novel treatments, drug-based and otherwise, based on symptoms rather than the specific diagnosis. For example, it would be more productive to focus on symptoms, such as impulsivity, across diagnostic categories (such as mania, ADHD and substance abuse) or specific cognitive functions (eg impaired episodic memory) irrespective of disease diagnosis (Sahakian *et al.* 2010; MRC 2010). To illustrate this, a treatment that reduces impulsive behaviour may do so whether a person has a diagnosis of ADHD or substance abuse. Similarly, a treatment for certain memory problems may be useful for improving cognition and functional outcome in both mild Alzheimer's disease and first episode schizophrenia. Symptoms, such as impulsivity, reflect genetics and underlying neurobiology and are therefore more likely to be tractable targets for treatment, compared with a heterogeneous category from diagnostic manuals.

Industrial partnerships could promote the development of biomarkers and neurocognitive training research, as well as novel drug development. Proof of concept studies with ketamine and scopolamine indicate novel areas for the pharmaceutical industry to pursue which would result in great benefits for patients with depression (Berman *et al.* 2000; Zarate *et al.* 2006; Drevets and Furey 2010). Unlike the currently used selective serotonin reuptake inhibitors (SSRIs), with which depressed patients take weeks to show improvements, new drugs with different mechanisms work very rapidly. In addition, small molecule drug design is of growing importance as it holds potential to deliver better and more effective treatments. Translational medicine—the process of turning biological discoveries into drugs and medical devices that will help patients—is a UK strength. In addition, these partnerships could promote the development of pharmacogenomics, the discipline behind how genes influence the body's response to drugs. This might lead to a 'personalised medicine' approach to mental wellbeing, with individuals being given treatments tailored according to their genotype (see also Section 2.2).

Other novel approaches to understanding and treating brain diseases will include synapse proteomics. Irregularities in the sets of synapse proteins (chemicals in the junctions between nerve cells) cause particular clinical symptoms shared between diseases. This provides a novel route for the pharmaceutical industry, since 'blockbuster' drugs that have large markets and are useful in many individuals can be developed. Sets of synapse proteins can be targeted by drugs that should be useful in the treatment of multiple diseases.

The discovery of hundreds of new synapse proteins also provides a new set of potential drug targets. For example, to date, by far the greatest effort of pharmaceutical companies has been on the neurotransmitters (chemicals that send a signal from one nerve cell to the next) and their receptors and uptake systems. These comprise less than 10% of all synapse proteins, which leaves the other 90% to investigate. This provides an investment opportunity. The study of molecular functions at synapses and their

importance in behaviour is rapidly developing. Synapse proteomics will be of increasing importance in genomic diagnostics of brain diseases. For instance, from data on genetic disorders that affect the nervous system, it was found that over 130 brain diseases are caused by mutations in synapse proteins (Bayes and Grant 2009; Fernandez *et al.* 2009). It is already clear that autism, schizophrenia and bipolar disorder involve dozens of synapse proteins. This further emphasises the important point that future diagnostics will be likely to involve careful clinical phenotyping (classification of observable characteristics) using, for example, cognitive testing and brain imaging, but also genetic diagnosis.

Other novel neuroscientific approaches to understanding and treating neuropsychiatric disorders, such as Parkinson's disease, exist. For example, there are highly innovative techniques using optics to control neural activity (Gradinaru *et al.* 2010). These techniques allow for the control of cellular activity by exposure to light. Specific cells can be excited or inhibited by different wavelengths of light, a technique that is both spatially and temporally precise. These optogenetic methods (an emerging field that combines optics and genetics to probe neural circuits) have been demonstrated to be effective in many species and are likely to have wide applications in the future (Boyden *et al.* 2005). Optogenetics has been used to investigate synaptic connections within neuronal networks (Petreanu *et al.* 2007) and synaptic plasticity (Zhang *et al.* 2008). Used alongside Magnetic Resonance Imaging (MRI), it may prove especially valuable (Wells *et al.* 2010). Moreover, the technique has proved beneficial in restoring visual function in mice (Lagali *et al.* 2008) and so has applicability to treating forms of human blindness. Optogenetic therapy also holds promising potential for the treatment of neurological diseases. The technique has now been used in the study of Parkinson's disease where it can be used to investigate how symptoms are produced by different pathways in the brain (Kravitz *et al.* 2010). A recent study in *Nature* (Kravitz *et al.* 2010) indicated in a mouse model of Parkinson's disease that regulation of motor behaviours by optogenetic control was possible and that a modulation of circuitry may represent an effective therapeutic strategy for ameliorating motor deficits in patients with Parkinson's disease. While optogenetic techniques hold great promise for translational studies to aid understanding of neural mechanisms in neuropsychiatric disease, it remains to be determined whether these techniques can be used in humans.

Opportunities

The prospect of new technologies and new treatments for neuropsychiatric disorders and brain injury for the benefit of patients, society and the economy is extremely exciting. Many opportunities for their commercialisation exist, including those described above. These may be developed through public-private partnerships in some instances.

Important innovation in the area of neurocognitive activation (cognitive

training) could be exploited commercially for the benefit of society relatively rapidly. This is an important new technique which could easily be exploited through the UK Games Industry. It could be used in entertainment games for training impulse or cognitive control in children and adolescents with ADHD or substance abuse problems. It could also be used for training episodic memory in elderly people with amnestic mild cognitive impairment (aMCI), the onset stage of Alzheimer's disease. Klingberg (2010) has demonstrated the power of the technique by showing that training on a working memory task is associated with changes in brain activity in specific regions of the brain, as well as changes in the densities of certain types of receptors in healthy people. Therefore this technique could be useful for young and old healthy people, in addition to those with neuropsychiatric disorders and brain injury. The challenge for the games industry or other markets is to transform this training from a chore into a fun and enjoyable activity.

Other areas provide financial opportunities for neuroscientific development, although the gain to society is unclear, such as neuromarketing (see Section 2.5). Studies in this area have used decision making and purchasing paradigms together with fMRI technology to indicate, for example, that a preference for the drink Coke may be influenced by brand image rather than by the taste itself.

Challenges and solutions

Neuroscience is coming of age and can be successfully applied to important problems in health and disease to ensure the UK is economically competitive and that our society flourishes. Therefore it is unfortunate that there has been a recent withdrawal of some drug companies, including UK-based ones, from the development of new drugs for the treatment of psychiatric disorders (Miller 2010) (see also Sections 2.2 and 3.2).

Two actions which might stimulate central nervous system (CNS) research and development enterprise are: i) to extend the patent life of a new drug for psychiatry to ensure that the enormous development costs are taken into account; and ii) to make new mental health treatments a priority by speeding access and development of new drugs/products for mental health, as is done for HIV/AIDS.

A more difficult problem to solve is the sensitivity of 'accepted outcome measures' to change and early stage disease, such as those used by the US Food and Drug Administration (FDA) and European Medicines Agency (EMA). This is particularly true where early detection is important and where neuroprotective drugs can be given. Treatments, including pharmacological ones, need to be given early—before the patient has marked impairments in occupational and social functioning and reduced quality of life and wellbeing. (The standard diagnostic manuals require a decline and impairment in social and occupational functioning for a diagnosis of dementia). Therefore this challenge may lead to an opportunity to innovate the drug development process, for instance targets for treatment may become closely related to genetics and neurobiology (eg impulsivity, episodic

memory) rather than diagnostic categories (eg schizophrenia, ADHD) (see Sahakian *et al.* 2010). Another challenge is the true translation of Proof of Concept studies to Phase 3 clinical trials, as it is difficult to replicate exactly the methodology used in these two stages. Finally, other challenges for research studies include extensive bureaucracy (such as paperwork for research and development funding, ethical reviews, and intellectual property rights protection) and the reluctance of research sponsors to tolerate risk.

Other European opportunities in response to this need to 'repair' the CNS research and development enterprise may include: funding via organisations such as the Innovative Medicines Initiative (IMI),a partnership between the European Community and the European Federation of Pharmaceutical Industries and Associates (EFPIA); grant funding from the EU via the national research councils; development of a capability cluster in mental health; and increased integration between disciplines, particularly psychiatry and neurology.

Conclusion

In summary, there are extensive opportunities for neuroscience to contribute evidence-based advice to policymakers, health professionals, the private sector and the public. Public engagement in neuroscience is a rapidly growing area. 'Smart drugs' or cognitive enhancing drugs (eg modafinil) and their benefits for healthy people have been of keen interest. If these drugs are shown by the pharmaceutical industry to be safe and effective in long-term studies, policy change might allow these to be marketed through the usual routes. This would reduce potential harms of current internet purchase of these 'smart drugs' and may be of particular aid to certain groups in society such as elderly people (see also Sections 2.2, 3.2 and Box 2).

Furthermore, there is a global market for adaptive learning technologies. Marketing e-books and TV programmes are thus far an unexploited area. These types of initiatives can have importance for lifelong learning. Education and learning are known to enhance cognitive reserve, and better cognitive function is associated with better wellbeing (Beddington *et al.* 2008). Both science and technology and higher levels of cognitive abilities and education are linked to increased prosperity (eg increased gross domestic product) (see, eg, Royal Society 2010; Rindermann 2008). Furthermore, investment in mental health has provided substantial economic benefit in the past and should continue to do so in the future (Health Economics Research Group, Office of Health Economics, RAND Europe 2008).

Neuroscience can provide us with the tools to make the most of our minds and also the necessary platform for an economically competitive and flourishing society. Now it is up to us as neuroscientists, the government and society to transform this potential into a reality.

References and further reading

Bayes A and Grant S (2009) Neuroproteomics: understanding the

molecular organization and complexity of the brain. *Nature Reviews Neuroscience*, Vol 10, pp 635–646.

Beddington J, Cooper C, Field J, Goswani U, Huppert F, Jenkins R, Jones H, Kirkwood T, Sahakian B and Thomas S (2008) The mental wealth of nations. *Nature*, Vol 455, pp 1057–1060.

Berman R, Cappiello A, Anand A, Oren D, Heninger G, Charney D and Krystal J (2000) Antidepressant effects of ketamine in depressed patients. *Biological Psychiatry*, Vol 47, No 4, pp 351–354.

Boyden E, Zhang F, Bamberg E, Nagel G and Deisseroth K (2005) Millisecond-timescale, genetically-targeted optical control of neural activity. *Nature Neuroscience*, Vol 8, No 9, pp 1263–1268.

Drevets W and Furey M (2010) Replication of scopolamine's antidepressant efficacy in major depressive disorder: A randomized, placebo-controlled clinical trial. *Biological Psychiatry*, Vol 67, No 5, pp 432–438.

Fernandez E, Collins M, Uren R, Kopanitsa M, Komiyama N, Croning M, Zografos L, Armstrong J, Choudhary J and Grant S (2009). Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Molecular Systems Biology*, Vol 5, p 269.

Gradinaru V, Zhang F, Ramakrishnan C, Mattis J, Prakash R, Diester I, Goshen I, Thompson K and Deisseroth K (2010) Molecular and cellular approaches for diversifying and extending optogenetics. *Cell*, Vol 141, pp 1–12.

Health Economics Research Group, Office of Health Economics, RAND Europe (2008) *Medical Research: What's it worth? Estimating the economic benefits from medical research in the UK*. London: UK Evaluation Forum.

Klingberg T (2010) Training and plasticity of working memory. *Trends in Cognitive Sciences*, Vol 14, pp 317–324.

Kravitz A, Freeze B, Parker P, Kay K, Thwin M, Deisseroth K and Kreitzer A (2010) Regulation of Parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature*, Vol 466, pp 622–626.

Lagali P, Balya D, Awatramani G, Münch T, Kim D, Busskamp V, Cepko C and Roska B (2008) Light-activated channels targeted to ON bipolar cells restore visual function in retinal degeneration. *Nature Neuroscience*, Vol 11, pp 667–675.

MRC (2010) MRC Review of Mental Health Research - Report of the Strategic Review Group. London: Medical Research Council.

Miller G (2010) Is Pharma running out of brainy ideas. *Science*, Vol 329, pp 502–504.

Olsen A, Eadie B, Ernst C and Christie B (2006) Environmental Enrichment and Voluntary Exercise Massively Increase Neurogenesis in the Adult Hippocampus via Dissociable Pathways. *Hippocampus*, Vol 16, pp 250–260.

Petreanu L, Mao Y, Sternson S, Svoboda K (2009) The subcellular organization of neocortical excitatory connections. *Nature*, Vol 457, pp 1142–1145.

Rindermann H (2008) Relevance of education and intelligence at the national level for the economic welfare of people. *Intelligence*, Vol 36, pp 127–142.

Royal Society (2010) *The Scientific Century: securing our future prosperity.* RS Policy document 02/10.

Sahakian B, Malloch G and Kennard C (2010) A UK strategy for mental health and wellbeing. *The Lancet*, Vol 375, pp 1854–1855.

Wells J, Walker-Samuel S, Marina N, Figueiredo M, Teschemacher A, Spyer M, Gourine A, Kasparov A, and Lythgoe M (2010) In-vivo optogenetic activation of cortical neurons and astrocytes with MRI at .9.4T. (*abstract*).

Zarate C, Singh J, Carlson P, Brutsche N, Ameli R, Luckenbaugh D, Charney D and Manji H (2006) A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. *Archives of General Psychiatry*, Vol 63, No 8, pp 856–864.

Zhang F, Prigge M, Beyrière F, Tsunoda SP, Mattis J,Yizhar O, Hegemann P and Deisseroth K (2008) Red-shifted optogenetic excitation: a tool for fast neural control derived from Volvox carteri. *Nature Neuroscience*, Vol 11, pp 631–633.

3.2 Risks

Professor Steven Rose, The Open University

Background

The rapid advances in neuroscience of the past decades, and those that can be anticipated in the near and mid-future, bring with them not merely new knowledge of how the brain works, with profound implications for humanity's understanding of itself, but also actual and potential technologies with associated benefits and risks. Both have been widely discussed in the media. This section is concerned to balance the risks against the perceived benefits discussed more extensively in other sections (see eg Section 3.1).

Is increased knowledge of the brain itself hazardous?

Some argue that increasing knowledge of brain mechanisms and their relationship to such deeply personal and private characteristics as memory, cognition, emotion and even consciousness, risks diminishing our sense of having independent agency and free will. We will be nothing other than neuronal machines, driven by the complex firing patterns of the cells in our brains, themselves the inevitable product of the interplay between genes and environment during our development.

This interpretation is itself vigorously contested among philosophers and neuroscientists—compare for instance, the views of Singer on the one hand (see Section 2.4) with Chan and Harris on the other in this volume (see Section 3.3). Some distinguished neuroscientists celebrate such a reduction of mind to brain (eg Crick 1994) whilst some philosophers view it with foreboding (eg Habermas 2003). I take neither view. Instead I would argue that any genuine increase in knowledge of brain processes and their pre- and post-natal development can only enrich our understanding of ourselves. Nor can such increased knowledge replace or diminish the insights into what it is to be human that come from philosophy, the social sciences or the humanities—although these disciplines will need to take the findings of neuroscience into account, just as neuroscience will need to respect these other perspectives and understandings. Here, therefore, there should only be benefits, providing one can pick one's way through the 'over-hyping' of apparent neuroscientific claims and the attendant prophesies of doom that so attract media attention.

Neurotechnoscience

Technosciences involve research enterprises in which the old distinctions between science and technology have broken down. It is not just that one cannot say where the science stops and the technology begins, but that the two are inseparable, each both driving forward and being driven by the other. And it is as a technoscience that the risks and benefits

of advances in the study of the brain may be seen as more finely balanced. To judge these requires reflecting not merely on the potential of developments discussed in the previous sections, but also on the intentions and goals of those who fund them—primarily the State, pharmaceutical companies, medical charities, supranational organisations such as the European Commission, and the military. In awarding funds each organisation will have specific ends in view, and these must be taken into account as we consider their implications (see also Section 3.4).

Policy issues

Neurogenetics and the pharmaceutical industry

Many neurological and psychiatric disorders run in families, and in some cases genes have been identified that increase the probability of a person having a particular disorder (see also Section 3.1). Huntington's disease (HD), which results in increasing motor and mental incapacity in middle life, is a single gene disorder where both the gene and its biochemical and cellular consequences are well known and predictable. The same is true for rare forms of Alzheimer's disease (AD), which strike in mid-life rather than, as in most cases, older age. There are also known genetic risk factors for the more common forms of the disease, but these are probabilistic rather than predictive.

The situation is less clear for the common psychiatric disorders such as depression or schizophrenia, whose mode of transmission is obscure and for which, despite decades of research, no unequivocal and replicable genetic markers have been identified. Rather, the modern techniques of genome wide association studies (GWAS) suggest that there may be tens or even hundreds of genes which, individually or in combination and varying from individual to individual, may affect the risk for the disorder. Under these circumstances, the benefits of genetic testing of an individual for the predisposition may well be outweighed by the risks of false diagnosis (either positive or negative) with its consequent implications for how a person plans his/her life. Even when a relatively certain genetic diagnosis can be made, as in the case of HD, experience has shown that many of those known to be at risk of the disease decline the test, preferring uncertainty, especially where there is no effective treatment, as is still the case for HD.

Some commercial companies offer tests for one of the genetic risk factors for AD, the gene *ApoE4*, but this practice worries both genetic counsellors and ethicists, as a positive result is indicative only of a potential risk factor and provides little guidance as to how one should live one's life, whilst adding to a person's burden of anxiety such that every small slip in memory may be taken as a warning of the impending onset of the disease. Nonetheless in an unregulated or only lightly regulated biotechnological economy, the availability of such tests is likely to increase substantially in the coming years (see eg Collins 2010).

One of the major goals for neuropharmacology must be to develop

drugs to treat or alleviate these conditions. Both the record and the prospects are mixed. In the case of AD, even though the triggering events for the disease are unknown, the biochemical cascade that leads to the accumulation of plaques and tangles and the death of neurons is well understood, providing many potential targets for drug action. Whilst the present generation of drugs is not very effective, there are several promising new developments. Most such drugs are aimed at slowing the cognitive decline characteristic of AD. In this sense the drugs are cognitive enhancers (see also Section 2.2, 3.1 and Box 2). The benefits of such drugs—at the least in enabling AD sufferers to maintain a longer period of independent living—are clear. However, as they do not prevent the inexorable neurodegeneration that the disease entails, some have questioned whether living with the knowledge of that fate rather than benign neglect is always beneficial.

There is some evidence that drugs like the statins, used to control cholesterol levels, and folic acid may be marginally neuroprotective, as are 'brain exercises' on the 'use it or lose it' principle (see also Section 3.1). There have also been very recent claims to have identified a 'marker' molecule, present in cerebrospinal fluid that can predict the onset of the disease before any behavioural indications of its onset. The benefits that would accrue from developing a truly protective agent against neurodegeneration would be enormous, even when set against the risks of the widespread prophylactic drug taking over many years.

For psychiatric diagnoses such as schizophrenia and depression, the situation is less optimistic. Despite earlier hopes and claims, the newer generation of selective serotonin reuptake inhibitor drugs (SSRIs) to treat depression have turned out, when widely prescribed, to be little more effective that the earlier ones in treating what the World Health Organization has categorised as a worldwide epidemic of depression. However, people's responses to the drugs are very variable, and one hope, though not yet realised, is that GWAS might make it possible to identify subsets of the population who could benefit most from any of the different classes of drugs that are available, Nonetheless, it is perhaps in recognition of these difficulties, and the sheer seeming intractability of the problem that, as mentioned by Robbins and Sahakian (see Sections 2.2 and 3.1), several of the major pharmaceutical companies have abandoned their drug discovery programmes for these disorders, whilst retaining those for frank neurological conditions such as AD or Parkinson's Disease (Miller 2010).

Drugs for social control?

The US Diagnostic and Statistical Manual (DSMIV), regarded as the psychiatrists' bible, is currently undergoing revision, but a category of disorders that is likely to remain in one form or another is that which relates not to relatively clear-cut psychiatric conditions but those concerned with an individual's behaviour—conditions such as (in the current classifications), conduct disorder, oppositional defiance

disorder and, attention deficit hyperactivity disorder (ADHD).

The frequency with which such diagnoses are being made has increased dramatically over the past two decades. ADHD (then called minimal brain dysfunction), considered to affect no more than one in several hundred children in the UK in the 1980s, is now estimated to be present in from 1–5% of children—mainly boys—between 6 and teenage. The diagnosis is based primarily around a child's unruly, disobedient or inattentive behaviour at school and home. There are no unequivocal neurological or neurochemical markers to correlate with the diagnosis. There are, however, pharmaceutical approaches to correcting or modifying these behaviours, most notably the amphetamine-like drug methylphenidate (Ritalin).

Ritalin prescriptions in the UK have increased from around 2000 a year in the early 1990s to approaching 600,000 a year today, though with very marked regional variations. Ritalin and related drugs certainly make a child calmer in class, less troublesome to teachers and parents. However, the long-term effects of the drug, on a growing child's brain and behaviour, have not been fully assessed.

An unanswered question is why a disorder considered rare several decades ago should now be diagnosed with such frequency. Did it exist unrecognised earlier, with children being called naughty or delinquent rather than as having a brain disorder? (The increase in autism diagnoses over the decades is a comparable example). Have society's criteria for what is or is not acceptable changed? Or might the problem lie less in the brain of the child and more in the parenting practices or social environment?

There is a more general issue at stake here, and that is the use of drugs as a means of social control. Whilst there are undoubtedly children who could benefit from the drugs, such medicalisation of behaviours regarded as outside the norm, to use Chan and Harris' term (See Section 3.3), may turn out to be an example of what Stirling categorises as missed opportunity or forced tramlines (see Section 3.4). By focussing on the individual and positing that the source of his or her distress lies in some molecular disorder in the brain, we may miss the broader social public health context (from economic insecurity to poor schooling or inadequate parenting) that affects the brain and makes the distress manifest. In focussing on 'a pill for every ill' do we risk moving towards what a neurophysiologist once referred to (approvingly) as a 'psychocivilised society'—and what, less approvingly, Aldous Huxley wrote about in *Brave New World*?

More immediately, in the US, where the use of Ritalin has been more widespread for far longer than in the UK, the FDA has called attention to the extent to which it is being widely traded amongst schoolchildren on the grounds that its attention-enhancing effects are an aid to study and preparing for exams. As discussed by Robbins (see Section 2.2), a number of respondents—predominantly from the US—to a survey conducted by *Nature* amongst its readers as to whether they use cognitive enhancers also reported

that they had or currently still used Ritalin in this way.

Ritalin is of course not the only drug to be considered as a cognitive enhancer (see Section 2.2 and Box 2). A wide variety of agents with very different mechanisms of action have been proposed, mainly on the basis of animal experiments and often without good evidence as to their effects in humans, to act as cognitive enhancers, giving rise to much ethical debate (see Section 3.3) and media speculation. Concern has been expressed over whether the use of such drugs outside a medical context confers an unfair advantage on their users in competitive examinations. Should such 'steroids for the brain' be regulated as in sport? Even if it were possible—and the wide availability of Ritalin for purchase via the web suggests that it would be difficult—would it be appropriate? How different, ethically, is the deliberate and voluntary taking of a cognitive enhancer to employing a tutor or enjoying the educational advantages that, in a profoundly unequal society such as Britain, come with class and income (Rose, 2002)?

Imaging and neural interfaces

As discussed by Rees (see Section 2.1), once they had moved beyond the experimental province of physicists, virtually all modern neuroimaging techniques were developed with diagnostic aims in mind: Computed Axial Tomography (CAT or CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI) and functional MRI (fMRI) are examples. Perhaps the exception is magnetoencephalography (MEG). Barring the usual problems with all scanning techniques—of both false positives and false negatives—the clinical benefits to neurosurgery and neurology scarcely need stressing. The only risk seems to be the general one: that modern medicine and clinical practice sometimes seem to substitute diagnosis for treatment—as for instance in the use of MRI scans to detect spinal disc injury as a substitute for, or addition to, clinical judgement even when surgical intervention is not contemplated.

Similarly, at first glance the neural interfaces discussed by Tracey (see Section 2.3 and Box 3) would seem only beneficial, in their potential to compensate for loss of sensory or motor abilities—though their military applications are more disturbing. However, the science fiction possibilities of such prostheses—as described for example by Gibson (2000)—open the prospect of a dystopic world of cyberpeople: part human, part engineered, reflecting in fiction Habermas's concerns (see also Section 3.3).

There are, however, two main areas in which developments in imaging techniques raise immediate social concerns. Both relate to the consequences of living in an increasingly security- and surveillance-obsessed society. Could brain imaging be a further step down the path pioneered by the ubiquitous CCTV cameras and national DNA databases—that is, as a further restriction on an individual's privacy? Might brain imaging provide an internal surveillance of our thoughts, emotions and intentions to add

to the external surveillance of the cameras? Not *Brave New World* but *1984* updated?

Claims that brain imaging ('Brain Fingerprinting') could serve as an updated form of lie detection have been treated with some scepticism by the imaging community (See Box 1), but a number of commercial companies now offer devices claiming to do just this, mainly based on the measurement of evoked response potentials—an application of the electroencephalography (EEG) technique discussed by Rees (see Section 2.1). The company websites suggest their use to detect whether someone 'is a terrorist' or has visited 'terrorist training camps' as well as in the courts to help determine guilt or innocence. The use of such techniques has been admitted into trials in India and the US, though not so far in the UK (Rose 2006). These issues will be addressed in more detail in Module 4 of the *Brain Waves* project on neuroscience, responsibility and the law.

Of at least as great concern are the suggestions that brain imaging can provide a prospective diagnosis of a person's potential for criminal activity or psychopathic violence. There have been claims that MRI or fMRI could detect characteristic differences in brain structures or neural activity between incarcerated men convicted of murder or other violent crimes and either non-violent criminals or 'normal' people. However a major problem with such studies is that they are post-hoc. The violent murderers studied have been in prison, often for long periods of time, have a history of the use of both legal and illegal drugs, and many other potentially relevant experiences. Although there have been claims that it is possible to detect propensity to psychopathy, based on psychological and even biological measures in young children, whether useful predictive brain differences could be found prior to crime, conviction and imprisonment is simply not known. Furthermore, even if such predictions could reliably be made, it is a cardinal principle of law that intentions or predispositions in the absence of acts are not crimes—unless we are truly to move towards a 1984-type category of 'thought crime.' The argument that individuals 'at risk' should be scanned and subjected to an appropriate remedial or control regime, and the individual could choose whether to be subject to it rather than coerced, fails on the grounds both that any such brain findings are likely to include many false positives, and a lack of knowledge about just what would constitute an appropriate remedial regime. It is doubtful whether there would be any added value provided by a brain scan above that given by standard psychiatric evaluation.

More directly interventive are the expanding applications of transcranial magnetic stimulation (TMS) (see Box 4). TMS has been proposed as a potential therapeutic approach to mitigate the effects of neurological conditions such as Parkinson's, to alleviate depression, or as an alternative or adjunct to cognitive behaviour therapy in obsessive/compulsive disorders. In these senses it would seem to carry a similar balance of benefits and possible risks as do pharmacological interventions. However,

the declared interest of the US military (see below) in exploring the uses of TMS as a method of thought and behaviour control, though still in the realm of basic research and even science fiction, raises the same issues of privacy versus social control and manipulation discussed above.

The military and neurotechnoscience

Military interest in neurotechnoscience is directed towards two general goals: improving the efficiency of one's own forces, and diminishing that of an enemy. Most available information comes from the US, where the military and its research organisations such as the Defence Advanced Research Projects Agency (DARPA) have a long record of funding work in this area, and where information is much more freely available than in other more secretive states. Current concerns centre on two areas: psychoactive chemicals; and behaviour modification via brain stimulation. These issues will be addressed as part of Module 3 of the *Brain Waves* project on neuroscience, conflict and security.

Psychoactive chemicals: Pharmaceuticals to improve the performance or motivation of one's own military have a long history (from hashish to marijuana and amphetamine) and widespread use. For example during recent conflicts it is reported that US pilots were supplied with Modafinil to improve alertness and concentration during long flights (see also Section 2.2).

The use of any toxic chemical, including neuroactive chemicals such as nerve gasses, as weapons of warfare is prohibited in international law by the Chemical Weapons Convention. However, the Convention allows for the use of long-standing 'riot control agents'—'tear gasses' such CN and CS that cause local irritation of skin and mucous membranes—for 'law enforcement including domestic riot control'. It is into this grey area between 'police' and 'military' deployment that some countries have sought to introduce incapacitating chemical weapons with central affects on the brain to induce unconsciousness or sedation. The first use of incapacitating chemical weapons was in Moscow in 2002 when Russian Special Forces stormed a theatre to release hostages taken by Chechen militants. A derivative of the opioid fentanyl was pumped into the theatre with the intention of incapacitating the militants before the troops stormed the building—but the drug killed over 120 of the hostages (Davison 2009). Concerns have been raised of the risk such weapons' development poses to the international ban on chemical weapons (see also Section 2.2).

Physical methods: Other basic research efforts aim to investigate the potential for affecting the brain and central nervous system using different powers, frequencies, and pulses of electromagnetic radiation. These would join emerging directed energy weapons, including lasers and the so called 'Active Denial System', a new weapon that projects a millimetre wave beam of radiation to heat the skin with the aim of causing a burning

sensation without permanent damage (Davison 2009). This weapon has recently been installed on trial in a Los Angeles prison. Whether one judges the availability of such techniques as a benefit or a risk depends on whether one sees them as preventing unwanted social disorder or increasing the power of a militarised State to control citizens.

Of more direct military significance has been the research conducted over several decades under DARPA contracts to develop 'distance' methods to control or manipulate brain and thought processes, in the past by high intensity microwave beams and more recently by magnetic pulses through TMS (Rose 2006). Despite the claims by some groups in the US that they have been unwittingly subject to such experiments by the military, the evidence is that at present the available technology requires that the individual be placed directly into the appropriate equipment. Thought control at a distance remains in the realm of science fiction.

Conclusion

Many of both the benefits and the risks of advances in neuroscience still lie in the future. Some we can anticipate, others will be unintended or unforeseen. Much neuroscience is still 'upstream' of application. However the time to consider and make choices about the directions in which neuroscience could or should be pursued is now, whilst policy makers and society at large still have time to consider the potential developments and how, and to what extent, they may be directed or controlled (see also Section 3.4). The further 'downstream' the technologies have moved, the harder such choices will be.

References and further reading

Collins F (2010) *The Language of Life: DNA and the revolution in personalised medicine*. HarperCollins.

Crick F (1994) *The Astonishing Hypothesis: the scientific search for the soul*. London: Simon and Schuster.

Davison N (2009) *'Non-lethal' Weapons*. Basingstoke: Palgrave Macmillan.

Gibson W (2000) *Neuromancer.* Voyager.

Habermas J (2003) *The Future of Human Nature.* Cambridge: Polity Press.

Miller G (2010) Is Pharma running out of brainy ideas? *Science*, Vol 329, pp 502–504.

Rose S (2002) Smart drugs: do they work? Are they ethical? Will they be legal? *Nature Reviews Neuroscience*, Vol 3, pp 975–978.

Rose S (2006) *The 21st Century Brain: Explaining, mending and manipulating the mind.* London: Vintage.

3.3 Neuroethics

Dr Sarah Chan and Professor John Harris, University of Manchester

Background

'Neuroethics' is a term that has gained prominence over the last decade to describe ethical issues arising in relation to various advances in neuroscience (Dana Centre 2002; Illes and Raffin 2002; Farah 2005). The kinds of technologies that provoke neuroethical concerns include neuroimaging techniques such as functional magnetic resonance imaging (fMRI) (see Section 2.1), the use of pharmacological agents to alter brain function (see Section 2.2), neurosurgery and other physical interventions (see Section 2.3). In addition, basic research in neuroscience continues to expand our knowledge of the biological basis for the brain's functioning and for the mental, psychological and behavioural correlates of neurobiological phenomena (see also Sections 2.4 and 2.5). This raises further ethical and philosophical challenges as to the implications of these findings and how they should be interpreted and used.

As with many new areas of technology and the sub-fields of ethics that have developed in association with them, the emergence of neuroethics as a specific area poses the question: are the associated ethical issues novel and distinct from general ethical concerns raised in the context of medicine, health care and biotechnology?

Some of the issues related to neuroethics are common across many areas of biomedical technology: for example, questions relating to participation in neuroscience research, or concerns regarding risk in relation to experimental technologies. Aside from these, however, neuroscience does challenge us to contemplate the brain and its relationship to the mind in new ways. Our increasing understanding of brain function as a biological phenomenon leads us to reconsider profound philosophical questions about free will, responsibility, identity, and the nature of consciousness. Although the questions themselves may be age-old, neuroscience both encourages us to revisit them and casts them in a new light—though not necessarily one that illuminates the answers any further (see Section 2.4 and Box 5).

Additionally, neuroscience is likely to have far-reaching social implications. The way in which we constitute ourselves and other persons as 'neurological subjects' (Cunningham-Burley 2010) inevitably affects our understandings of ourselves and our relationships with others. For example, clinical and social attitudes towards mental illness have evolved considerably over the last century, influenced by our changing understanding of the neuroscientific aspects of psychiatric conditions and the consequent perception of mental disorders as rooted in physical pathology (see also Section 3.2).

There are also difficult and in some cases urgent questions about how we ought to make use of the knowledge and the

technologies emerging from this field, in contexts ranging from health care to the legal process and even political or social control (see Sections 2.4, 3.1, and 3.2).

The ethical concerns and social changes engendered by neuroscience and all it entails will require extensive consideration with respect to policy and regulation: how are we to manage and respond to these concerns and changes? How can and should we guide the uses of neuroscience in the public interest?

This essay presents a brief overview of the issues at the forefront of neuroethics as it has developed to date and highlights some that require further attention and others that may emerge in the near future. Lastly this essay will identify some current priorities for policy consideration and suggest an approach to ethical policy-making in neuroethics.

Neuroimaging

Neuroimaging technologies such as positron emission tomography (PET) and fMRI enable us, in a manner of speaking, to 'see' mental processes in terms of the architecture and activity of the brain. (See Section 2.1) This has certainly increased scientific knowledge of the brain as a biological organ but does not fully answer philosophical and human or personal questions about conscious thought and the nature of mind. For example, we now have a scientific understanding of the neurological processes involved in moral reasoning and can observe some of the biological correlates of the mental phenomenon of consciousness—but will this produce a deeper philosophical or legal understanding of what it is to be conscious, or a moral agent? (See also Section 2.4.)

Whether or not this is the case, neuroimaging does have some ethical consequences in its application, including ramifications for diverse areas of medical ethics. For example, in relation to end of life decisions: recent research shows that patients previously thought to be in a permanent vegetative state actually demonstrate some level of brain function sufficient to express (and therefore presumably to have) preferences. This result provoked intense debate over whether treatment protocols for such patients should be revised to take account of their newly-discovered capacity for some level of thought or at least reaction.

Privacy concerns

Foremost amongst the concerns raised about neuroimaging technologies is the fear that the use of neuroimaging to 'read minds' may lead to unacceptable infringements of personal privacy and civil liberty. At present, given the primitive state of the technology, such concerns are probably overblown. Neuroimaging, although it may be able to detect general mental states such as emotion and even, at a rudimentary level, more specific conceptual/thought patterns, currently it cannot be used to determine precisely the content of thoughts. Insofar as it can currently be interpreted to form general conclusions about individuals or tendencies within a population, it is little different to other physical indicators of

mood or mental state (see also Sections 2.1 and 3.1).

Forensic uses

Of more concern is the danger that the possibilities currently offered by neuroimaging may be misinterpreted and thus misapplied, overstretching the current capabilities of the technology and leading to false assumptions being made about the extent to which brain states may constitute 'windows on the mind'. This is particularly worrying in the forensic context, for example neuroimaging lie detectors (see Section 2.1 and Box 1). An attitude of scientific reductionism to forensic uses of neuroscience might lead to a one-dimensional legal approach which would be less likely to serve the requirements of justice or due process. Overly scientistic approaches to subjective fact-finding disregard the many complex factors that feed into legal and judicial decision-making. A rigorous and contextualised understanding of neuroscience and its place within social structures such as the law is required to avoid such pitfalls. This topic will be addressed as part of Module 4 of the *Brain Waves* project on neuroscience, responsibility and the law.

This is one area for the cautionary shaping of policy, both regarding direct uses and to help develop understanding on the part of publics and policy-makers. Scientists too have an important role to play in this process, through communication but without exaggeration and with a realistic attitude towards what is possible currently and in the near future.

Genetic neuroscience

Those seeking to map and explore the novel territory of neuroethics have often made reference to the genomics era and the congruent evolution of genethics to address ethical and social implications of genetic science as a paradigm for the development of neuroscience and neuroethics. As well as clear parallels between the two, there are degrees of overlap: the intersection of genetics with neuroscience, for example in behavioural genetics, multiplies the possible ethical dilemmas.

Already in a number of instances, behavioural genetics has raised novel issues: the correlation of particular sets of chromosomes with aggressive personality types has led to at least one legal attempt at a 'genetic defence' (Farahany and Burnet 2006) for criminal behaviour, while the discovery of specific genes thought to be associated with risk-taking and aggression, in connection with racial genotyping, may reinforce unwarranted and unwanted social attitudes towards certain racial groups (Wensley and King 2008) (See also Section 3.2).

New medical applications

Probably the biggest area of medical application in neuroscience has been the development of psychopharmacological agents to alter various aspects of brain function (see Section 2.2 and 3.1). Psychoactive drugs such as antidepressants are becoming increasingly widely used; many drugs provide effective relief for conditions which previously represented serious impairments.

Electrochemical or physical interventions may also have therapeutic effects: for example, deep brain stimulation to treat diseases such as Parkinson's has been highly successful in a number of cases (see Box 4). Neurosurgery may alter brain function as a side-effect of removing harmful growths such as tumours, or may aim to achieve alterations in function through physical intervention.

Other possibilities include neural interfaces to enable humans (or indeed other animals) to control electronic or robotic devices—something that has immense possibility in terms of prosthetic applications (see Section 2.3 and Box 3). Yet these developments also lead us to question the biological nature of the human body and, perhaps, to redefine what we consider to be human. When machines are as much a part of 'us' as our own bodies the boundary between the biological and mechanical becomes blurred. And when we consider the extent to which we already depend on machines for our everyday existence—computers, electronic devices, modes of transport—it becomes apparent that perhaps the modern human is already part-machine.

Harmful uses

These promised new technologies also have potential downsides (see also Section 3.2). Some have speculated that psychopharmacology may be used in a harmful way to achieve mind control or manipulation of minds to suit some ulterior motive. (Drugs are not the only possibility—it might argued that advertising and the psychological research that enables commercial operators to sell products more effectively is also a form of mind manipulation.) Another worry is that knowledge of neuroscience may be used to create new incapacitating chemical agents—a form of neurochemical warfare (see Sections 2.2 and 3.2). These issues will be addressed in Module 3 of the *Brain Waves* project on neuroscience, conflict and security.

The dilemma in relation to many of these neuroscience applications, as with many new scientific advances, is that the selfsame technologies or knowledge used for the beneficial medical treatments described might also be used for harmful purposes. How we should deal with the so-called 'dual use' problem is one of the serious issues facing not just neuroethics but science ethics as a whole (see also Section 3.4).

Medicalising the mind

There is one concern over dual-use technologies that we would argue is misdirected: not just their alternative or 'dual' use to cause harm, but their use to cause *too much* good—in other words, to enhance as well as to cure. Neuro-enhancement carries its own set of cautions to be considered, as will be discussed—however, simply doing more good should not in itself be a problem.

The perception that enhancement is wrong while treatment is right relates to another, more subtle but more significant consequence of neuroscience: the medicalisation of the mind (see also Section 3.2). The scientific definition of a

biomedical model of the 'healthy' brain and by extension the 'healthy' mind is constructed in opposition to mental illness and psychiatric disorder, where health and normality are the binary opposite of disease and dysfunction.

This characterisation of certain conditions as abnormal or pathological affects our perception of self and of others. It may be stigmatising to be classed as mentally ill; the translation of neuroscience into diagnostic psychiatry criteria, and the effect this may have on people must take account of the possibility of harmful effects on those who are diagnosed. On the other hand, the understanding of mental illness in terms of physical pathology might reduce feelings of personal responsibility and attribution of blame for what can then be interpreted as an unfortunate accident of biology rather than purely mental and a flaw intrinsic to the self. Receiving a neuroscientific diagnosis may be something of a relief, not only because it affords what appears to be concrete and hence a feeling of control, but because it allows separation of the ill/diseased state of the body part 'at fault', and the negative values associated with this state, from the self.

Medicine and its conception of health provides a tremendous normative force that is often seen as stipulating the proper and improper uses of technologies (Fukuyama 2002; Habermas 2003). However, there is the danger that ascribing excessive normative weight to a medically-determined concept of health risks removing the power of individuals to evaluate their own condition and make decisions about it for themselves. This is perhaps of particular concern with respect to mental health and neuroscience. The construction of medical norms for mental and psychological attributes may be seen to exclude and devalue the abnormal, thereby dictating what constitutes an 'acceptable' way to think and to be.

There is also a further concern related to the development of new therapies. The normativity of medicine provides legitimation for those interventions designated as 'medical therapies', aimed at curing 'disease'. But this phenomenon may be capitalised upon in reverse: by redefining the state of disease, an apparent moral need, not to mention a market, is created for new therapies. Some have speculated that the rise in availability and range of psychoactive drugs is due at best to overdiagnosis or a broadening of diagnostic categories, and at worst to the creation of new diseases to encourage uptake of new marketable products (see Section 3.2).

Neuro-enhancement

Perhaps the question we ought to ask about all neurotechnologies is not whether they constitute a medical therapy for a recognised dysfunction or disease, but whether *in and of themselves* they have some beneficial effect, regardless of how we classify them. It is certainly true that many of the new treatments available have huge potential to improve function—and yet they have often been criticised for precisely that: for being 'enhancements' rather than 'treatments'.

Cognitive enhancement

Cognitive enhancement is one of the oldest forms of human enhancement and one of the most valued. Improving cognitive powers and capacities is one of the quintessential human activities: tool-using, social organization, speech, written language, and more recently universal formal education and the use of computers, are all dramatic enhancers of cognition and knowledge acquisition. Chemical cognitive enhancers (see also Section 2.2 and Box 2) may now provide us with another means by which to seek the same end (Chan and Harris 2006; Greely *et al.* 2008; Foresight Mental Capital and Wellbeing Project 2008). While many are rightly concerned that our powers might outstrip our ability or willingness to control their harmful effects, the paradox will remain that cognitive enhancement is also the most promising option we have to prevent or remedy potential harms caused (Rees 2004; Perrson and Savulescu 2008; Harris 2011).

Mood enhancement

Many of the psychoactive drugs available today in the medical context and the recreational context—or sometimes both—are what might be termed 'mood enhancers' (see Section 2.2). This could mean that they enhance mood either in the sense of heightening whatever emotions the user is experiencing, or in the sense of improving mood towards some 'better' state however this may be defined: by personal preference, by medical normativity or by moral philosophy, among others.

Along with these drugs come fears that we will become a population of lotus-eaters constantly bathed in a sea of drug-induced euphoria and contentment, and concerns about authenticity of emotion and experience, once we can control our emotions and the way in which we react to experience through the application of suitable chemical modifiers (see also Section 3.2). It might even be argued that there is something inherently wrong in seeking to alleviate distress and negative emotions through artificial means, that we as humans have a need for suffering This argument, though, is in many ways analogous to arguments about hard work as a virtue in other areas of enhancement (Kass 2003), and is susceptible to the same criticisms (Harris 2007).

Moral enhancement?

One topic that has received recent attention is the possibility of moral enhancement: that alongside the other mental and psychological capacities we might improve upon, it may also be possible to 'make us better people' in another sense, through enhancing our ability to act in a moral way. This idea seems plausible and perhaps appealing at first glance: morality in the sense of 'goodness' is perceived as a desirable quality and something for which we should strive. As science begins to examine the biological foundations of morality, for example through identifying regions of the brain that appear to be involved in moral decision-making or exploring the evolutionary origins of moral behaviour, it is tempting to think that

morality is something that can be easily identified, isolated and improved. Along these lines, some have argued not only that enhancement is a moral imperative but that the greatest imperative is toward moral enhancement (Perrson and Savulescu 2008).

Upon closer examination, however, this apparently simple proposition is highly problematic. Ethical expertise[1] is not being better at being good, rather it is being better at knowing the good and understanding what is likely to conduce to the good. Those with the insight, sympathy, empathy and knowledge to have formed clear ideas of what might conduce to the good are not necessarily better at making the world a better place, for a number of familiar reasons.

Some of these are to do with the problem of 'akrasia' or weakness of will, one form of which was brilliantly summarised by George Bernard Shaw when he defined virtue as 'insufficient temptation'. A bigger problem is that the sorts of traits or dispositions that seem to lead to bad conduct are also the very same ones required not only for virtue but for any sort of moral life at all.

This problem was effectively articulated by Peter Strawson in a famous essay entitled 'Freedom and Resentment' (Strawson 1960). Strawson was concerned not with moral enhancement but with the problem of free will; in the course of combating some absurd forms of determinism he points out that certain strong emotions, including aversions, are an essential and even desirable part of certain valuable motives or attitudes to others. Could we in short have the sorts of feelings that are appropriate and possibly necessary to morality if we did not feel strong aversion, for example to someone who deliberately and unjustifiably injured those we love?

The space between knowing the good and doing the good is a region entirely inhabited by freedom. Without the freedom, good cannot be a choice and if freedom disappears along with it goes virtue.

Machine minds?

One group of technologies that are not always considered within the scope of neuroethics, but perhaps should be, is computer science and especially artificial intelligence (AI). The modern idea of artificial, machine-based intelligence has been around for over 50 years, and as machines become ever more capable, the possibility of creating a computer that can 'think for itself' seems increasingly plausible.

Efforts to create AI have produced so-called AI computers that can solve intricate mathematical problems, perform complex search operations and, famously, play chess. There is as yet no evidence that any of these *think* in the way ordinary humans do, but it is undeniable that AI is becoming increasingly sophisticated and perhaps one day they will.

Another related area of research is computer-based brain simulation. Simple

1 In this section we follow lines taken by John Harris in the introduction to his *Enhancing Evolution* Princeton University Press, Princeton and Oxford Paperback Edition 2010.

modelling of neurobiological systems is already possible and plans to create a simulation of the entire brain are in progress. The question is whether a computer that simulates the whole human brain to a sufficiently realistic degree would become, in some sense, a human mind or indeed any other sort of mind. We might be reluctant to give a 'human' label to a non-biological entity lacking any of the physical attributes by which we normally identify humans; yet if it were able to think like us, ought we not to recognise it as one of us? This harks back also to the philosophy of mind problems outlined above: would a computer that simulated the physical workings of a brain also be thus a mind, though not housed in a human body?

When would we say that a computer had become conscious? How would we know, and what would or should we do about it? The issues we face with AI reflect those raised by neuroethics with respect to our own brains and minds: how do we quantify consciousness, and what is its moral significance?

Policy concerns in neuroethics

There are a number of pressing policy concerns in neuroethics today. Perhaps the foremost relates to the increasing commercialisation, not just of neuroscience but of science in general. In relation to neuroscience and to other technologies, we need to be aware of the effect that proprietary intellectual property rights has on the entire process of scientific discovery as well as access to the fruits of this process (The Manchester Manifesto 2009). Not only has it been shown that current mechanisms for proprietising innovation may have adverse consequences for the progress of science through restricting the openness with which scientists feel they can communicate about their research (Royal Society 2003), but commercialisation, through altering the balance of incentives for innovation, may actually change the course of science itself—not necessarily for the better (Chan, Sulston and Harris forthcoming).

Another issue is how we ought in general to regulate the use of emerging technologies (see Section 3.4). While risk and safety are quite rightly the foremost concerns in relation to introducing new biomedical technologies, it does not follow that the rational approach to managing risk is to limit their use only to circumstances where the likely benefit is so vast as to justify unequivocally the fuzzy, unquantified risk that we fear. If benefits of whatever sort are possible, they perhaps should—and undoubtedly will—be sought after. To take again the example of chemical cognitive enhancers, if it proves that 'healthy' adults can benefit from these, we should be pro-active about assessing risk in these circumstances rather than hiding behind over-restrictive regulation (see also Section 2.2 and Box 2).

As discussed, the issues encompassed within neuroethics range from largely theoretical philosophical questions to immediate practical concerns regarding possible improper uses of present technology. While the latter self-evidently require addressing, it is important to

remember that the role of ethics particularly relating to policy and regulation is not simply reactive and restrictive but should be forward-thinking and facilitative where needed. The aim of science ethics should be to ensure that science is done as best it can be and also does the best it can, to improve lives and increase the welfare of persons. This will be achieved not merely through preventing misuses, but by guiding the direction and application of science to bring about beneficial outcomes as well as avert harmful ones.

References and further reading

Chan S and Harris J (2006) Cognitive regeneration or enhancement: the ethical issues. *Regenerative Medicine*. Vol 1, No 3, pp 361–366.

Chan S, Sulston J, and Harris J (forthcoming) *Ethical Incentives for Innovation*. In: Stiglitz, J *et al*. (eds) *Intellectual Property Rights: Legal and Economic Challenges for Development*. Oxford: Oxford University Press.

Cunningham-Burley S (2010) *Engaging with neuroscience: examining neurological subjectivity and what this means for debates about enhancement technologies* (conference paper at 'Beyond The Body: Perspectives on Enhancement', 9–11 April 2010, Manchester).

Dana Centre (2002) Neuroethics: Mapping the Field. Brainwork, Special Issue.

Farah M (2005) Neuroethics: the practical and the philosophical. *Trends in Cognitive Science*, Vol 9, No 1, pp 34–30.

Farahany N and Bernet W (2006) Behavioural Genetics in Criminal Cases: Past, Present and Future. *Genomics, Society and Policy*, Vol 2, No 1, pp 72–79. Regarding the case of Stephen Mobley.

Foresight Mental Capital and Wellbeing Project (2008) *Final Project Report*. London: Government Office for Science.

Fukuyama F (2002) *Our Posthuman Future.* London: Profile Books.

Greely H, Sahakian B, Harris J, Kessler R, Gazzaniga M, Campbell P, and Farah M (2008) Towards responsible use of cognitive enhancing drugs by the healthy. *Nature*, Vol 456, pp 702–705.

Habermas J (2003) *The Future of Human Nature.* Cambridge: Polity Press.

Harris J (2007) *Enhancing Evolution.* Princeton NJ: Princeton University Press, Chapter 7.

Harris J (2011) Moral Enhancement and Freedom, *Bioethics*, Vol 25, No 2, pp 102–111.

Illes J and Raffin T (2002) Neuroethics: an emerging new discipline in the study of brain and cognition. *Brain and Cognition,* Vol 50, No 3, pp 341–344.

Kass, L (2003) Ageless Bodies, Happy Souls. *The New Atlantis*. Spring.

Perrson I and Savulescu J (2008) The perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity. *Journal of Applied Philosophy*, Vol 25, No 3, pp 162–177.

Rees M (2004) *Our Final Century.* London: Arrow Books.

Royal Society (2003) *Keeping Science Open: the effects of intellectual property policy on the conduct of science*. London: Royal Society.

Strawson P (1960) Freedom and Resentment. *Proceedings of the British Academy*, Vol 48, pp 1–25.

The Manchester Manifesto (2009) Available at: http://www.isei.manchester.ac.uk/TheManchesterManifesto

Wensley D and King M (2008) Scientific responsibility for the dissemination and interpretation of genetic research: lessons from the 'warrior gene' controversy. *Journal of Medical Ethics*, Vol 34, No 6, pp 507–509.

3.4 Governance of neuroscience: challenges and responses

Professor Andy Stirling, University of Sussex

Over past decades, global debates over the nature, pace and direction of research and innovation have yielded important lessons concerning the governance of new technologies. We can identify three broad lessons and seven more specific 'syndromes'. From these, we can extract three principles for the 'social appraisal' of neuroscience and neurotechnology. The lessons of past technologies are all the more clear and emphatic for being hard-won in protracted high-stakes controversies, like those over nuclear power, hazardous chemicals, and GM foods. Yet received wisdoms, prevailing values and incumbent interests can obscure many key features of these lessons. The challenge is compounded because some details may be ambiguous, with different contexts and interpretations suggesting contrasting policy implications. As a result, there are dangers that some familiar but readily-avoided mistakes may be repeated in the field of neuroscience and technology. If society is to maximise prospects for realising the many positive potentials in this area—and minimising risks—then it is crucial that these lessons be heeded (EEA 2001; Voß *et al.* 2006).

Three lessons

Lesson one

One lesson is taught by cumulative experiences in areas like agriculture, energy, medical and material science: *Levels of knowledge that are sufficient for a technology to meet initial narrow practical goals, are rarely sufficient to predict the full range of eventual indirect impacts.*

Initial visions for many technologies entirely failed to appreciate wider consequences, both beneficial and adverse. On the positive side, for example, are wireless broadcasting, semiconductors, lasers or the World Wide Web. More negatively, there are examples like asbestos, various chemicals (eg DDT, PCBs, TBT, and CFCs), thalidomide and retail use of ionising radiation. In all these cases, narrow initially-prioritised benefits are now generally held to be significantly outweighed by unforeseen and unintended negative impacts. A vast array of other examples occupy intermediate positions in this spectrum like the urban automobile, nuclear power, industrial agrochemicals, and genetically modified crops. In each of these areas there is active debate over their complex, pervasive, and intertwined benefits and impacts (ESTO 1999; EEA 2001; Stirling 2009).

For good or ill, several of what turned out to be the most significant side effects of many new technologies lay well beyond the bounds of initial quantification (or even imagination). Indeed, conventional regulatory efforts at anticipation and prior

appraisal were in many cases so circumscribed and rudimentary that a realistic range of implications was typically finally appreciated only by learning through actual pursuit of the technologies themselves in the real world. A resulting dilemma is that awareness of our ignorance may actually increase with growing knowledge and experience. Here, it is a particular feature of neuroscience that—irrespective of its possible applications—an increase in such knowledge may in itself be hazardous (see Section 3.2).

Neuroscience challenges many deeply held convictions around the nature of 'free will' (see Sections 2.4, 3.3 and Box 5). The more we think we understand human agency, the more difficult it becomes to sustain the necessary fictions around its exercise. As science is translated into technology, an expanding array of economic, institutional and political commitments compounds exposures to the consequences of ignorance. Incentives grow to interpret 'absence of evidence' as 'evidence of absence' of harm—leaving us systematically more vulnerable. This places a high premium on provision for early and thorough 'social appraisal' of the potential applications and implications of neuroscience. Rather than seeking to avoid political debate, we should find ways to make this as rigorous, comprehensive and critical as possible (Wynne 1992; EGSG 2007; Leach *et al.* 2010).

Lesson two

A second lesson is that: *Enthusiastic and highly visible championing of some specific new family of technologies by the most influential scientific and powerful industry or government bodies, provides no guarantee that these technologies will actually come to be established in the envisaged forms*.

A classic example of this is provided by the case of nuclear power. This is not a partisan point. Despite current renewed interest, there remain serious questions over the relative scale of the future promise of nuclear power when compared with alternative low-carbon energy options. Yet in the thirty years after the Second World War, nuclear power was almost universally expected in mainstream scientific, industry and government circles to present a complete and ubiquitous energy source from the 1990s onwards. From famous early prognoses that nuclear electricity would be 'too cheap to meter', confident predictions of exponential global growth rates for this technology were persistently made by supposedly neutral and authoritative scientific bodies right up until the 1980s—long after the actual pace of development was stalled and moving into reverse. In reality, construction rates for new global nuclear capacity dipped almost to zero in the 1990s and presently remain significantly smaller than that of wind power—with the discrepancy tending to increase (Wiliams and Edge 1996; Leach *et al.* 2010).

Experience with nuclear power is recounted here, because the passage of time has allowed some sense of perspective. In fields like neuroscience, there has accumulated insufficient experience to determine the validity of some of the more ambitious aspirations

and claims. Already, however, there are grounds for caution concerning the 'over-hyping' of certain possible neuroscience applications. It has become clear, for instance, that benefits of genetic testing must be qualified by recognising the significance of associated risks (no matter how small) of false diagnosis. Likewise, hopes for a transformative new generation of drugs to treat depression have thus far turned out to be unfulfilled (see also Section 3.2). In the same way that early expected trends in uptake of GM foods in Europe can now be seen to have been destined for frustration, so these tentative experiences suggest grounds for caution over exaggerated claims and aspirations.

Elsewhere in the life sciences—for example with monoclonal antibodies and genome sequencing—we have become quite familiar with the dynamics of 'hype cycles' generated by privileged parochial interests. This points to the ever-present possibility that even the most authoritative expectations for any given application of neuroscience will fail to be fully realised—at least on initially stated timescales. Again, this presents an imperative for deeper, more comprehensive and critical social appraisal and political debate (Voß *et al.* 2006; SPRU 2009).

Lesson three

A third—even more crucial—lesson is that: *The particular paths followed by scientific and technological developments in any given area are not pre-determined by nature*.

This insight does not deny that physical reality and technical feasibility impose highly demanding constraints on what is possible in the real world. Nor does it diminish the essential role of science in opening up and driving forward many new possibilities. The point is that the particular directions taken by innovation in a field like neuroscience are powerfully shaped by social, as well as material, factors. This is true even of long-established consumer products like the QWERTY keyboard or VHS videos. Even in highly competitive markets, these technologies came to global domination, despite being widely seen as inferior to possible (even existing) alternatives. The same is true of infrastructures like narrow (rather than broad) gauge railways, AC (rather than DC) or centralised (rather than distributed) electricity supply, submarine-derived civilian nuclear power reactors, early petrol-driven internal combustion engines (compared to contemporary steam and electric technologies, which might feasibly have been advanced to similar effect) and current general urban dependency on the automobile (Williams and Edge 1996; EGSG 2007; SPRU 2009).

Each of these cases demonstrate how economic and political power structures, property rights, income distributions, dominant values, and incumbent market interests can help 'lock in' certain innovation pathways and 'crowd out' others. Resource constraints, path dependencies, learning effects and scale economies in a finite world, all mean that our global society cannot equally realise the full potential of all feasible and viable innovation trajectories. Whether deliberately or blindly, societies *choose* which paths to follow or not. Looking forward, we see examples in choices

between particular low carbon energy or sustainable agriculture pathways. Likewise within neuroscience—and between neuroscience and alternative applications—those innovation pathways that we end up pursuing will typically represent only a small subset of those that were initially potentially viable. Which direction we go in depends as much on political as on technical choices. This is why social appraisal of neuroscience research and the regulation of associated technologies should pay as much attention to driving intentions and goals and associated power structures as to claimed benefits and supposed risks. Are these for military or civilian ends; northern or southern markets; public or private applications; IP-intensive or open-source development? (SPRU 2009; Leach *et al.* 2010).

Seven syndromes

From these three lessons, there emerges a series of seven more specific 'syndromes' to look out for in the governance of neuroscience and technology. These interact, but each may be distinguished in its specific causes, symptoms or remedies. For ease of recall, each is labelled below with a simple phrase. An appreciation of these interweaving dynamics in knowledge, technology and encompassing society is essential to any robust design for associated governance intervention (Leach *et al.* 2010).

Foregone Futures

Global society as a whole may fail to fulfil the benefits of some particular set of possibilities for innovation, by pursuing instead alternative trajectories that later leave these practically unrealisable. For instance, this is widely seen as true of locked-in global water supply infrastructures—which militate against more efficient and environmentally beneficial institutions and practices. Likewise, neuroscientific interventions in areas like attention deficit hyperactivity disorder (ADHD), cognitive enhancement, or brain interference may foreclose possibilities of alternative institutional or behavioural provision (see also Section 3.2). Here as elsewhere, it could emerge that potentially preferable 'open source' policies or practices are eclipsed by greater enthusiasm of existing research and innovation systems for IP-intensive neuroscience-specific applications (EGSG 2007).

Missed opportunities

A particular economy may fail to gain desired leadership in a successfully anticipated developmental direction, with benefits accruing instead to competitors making different choices at earlier stages. There are many historic examples of this, including what are conceded on all sides to have been misconceived British choices in the 1960s of nuclear reactor design. Again, this is as true of different applications within the field of neuroscience as of possible non-neuroscience alternatives that a general prioritising of science-based approaches might obscure. For instance, there are queries whether the driving objectives of UK neuroscience research appropriately

balance preventive public health or individualised private health? (see also Section 3.2) Does a 'preventive' focus rest on practices that address entire populations (implying collective public provision) or proprietary products targeting specific sub-populations (implying more readily privately-appropriable revenues)? Either way, structural shifts in global markets may significantly affect the prospects for differing patterns of neuroscience research (see also Section 3.1). These may substantively affect competitiveness in presently undetermined ways (Williams and Edge 2006).

Forced tramlines

This is like 'foregone futures', but can play out globally or locally. In contrast to reduced competitiveness from historical or politically innocent decision-making, this syndrome is exacerbated by the deliberate exercise of power within research and innovation systems. Either way, economies may be forced by narrow incumbent interests incorrectly to anticipate and commit to a particular innovation trajectory, that turns out in actuality to be inferior to contrasting alternatives. An example is the early neglect by the centralised UK power generating system of large national renewable energy capabilities and resources. In similar ways, there are possibilities that current driving motivations behind neuroscience research (such as pursuing 'rents' on IP or enhanced military applications), may 'lock in' to certain research and innovation pathways. The greater the diffidence displayed in research governance or regulatory appraisal, to explicitly examine power relations and vested interests *within* research and innovation systems, the greater the exposure to this syndrome (SPRU 2009).

Convenient blinkers

Choices among possible technological pathways may be further impeded or misguided by the contrived invisibility of unforeseen, or unintended, side effects. Conventional risk assessment focuses on a typically rather limited set of outcomes. These are circumscribed by the remits of regulatory institutions (eg on health rather than equity) and by what can be documented and quantified in currently-accredited evidence. These institutionalised blinkers may create further vulnerability to strategic manipulation of the kind occurring in 'forced tramlines'. In this way, society may be presented with an unbalanced picture of the most attractive alternatives. By the time gaps or bias becomes evident, it may be too late easily to change course.

Examples of this syndrome may be seen retrospectively in persistent global dependencies on technologies like fossil fuels and carcinogenic and stratospheric ozone-depleting halogenated hydrocarbons. These short-sighted attitudes towards what are now acknowledged to be prohibitive health or environmental effects have thus contributed to systematic neglect of superior substitute technologies. In this way, there are dangers that neuroscience research in areas like incapacitating chemical weapons (see Sections 2.2 and

3.2) and other forms of manipulation or interference with brain function may yield side effects not only in direct use, but also in less visible indirect ways. Crucial here is that the mere existence (even positing) of some of these technologies may—even without tangible physical effects—exert certain kinds of indirect social impact. Falling outside the conventional scope of regulatory attention, there is currently no effective provision for the taking of responsibility for these kinds of wider implications (Wynne 1992).

Shared delusion

A particular technological pathway, despite being widely adopted, may still fail to achieve the expected (or promised) aims on any reasonable timescale. This may be entirely independent of countervailing pressures or competing innovations. It may be seen, for instance, in experience hitherto with nuclear fusion power. Here, the expected commercialisation horizon has remained at a steady three decades or so for the past fifty years. Yet commitments are sustained on a roughly constant scale irrespective of this, over periods that see entire cycles in the rise and fall of other technologies. Where a significant proportion of current investment in neuroscience rests on some of the more colourful 'magic bullet' claims made on behalf of applications in fields like cognitive enhancement (see Section 2.2 and Box 2), national security or criminal justice (see Sections 3.2, 3.3 and Box 1), then it would be prudent to keep a wary eye open for prospects of this kind of syndrome (Leach *et al.* 2010).

See no evil

A particular technology may realise its initial promise, but this very feasibility may itself create opportunities for deliberate or inadvertent misuse. This raises the very simple and familiar challenge of contending (sometimes perverse) human intentions. Although readily foreseeable in the same terms as benign uses, malign applications are typically understated in regulatory assessment, sometimes for legal reasons. Yet easily anticipated effects may be of a magnitude that seriously jeopardises overall benefits. This is exemplified by the paradox that military aims are at the same time so prominent and so under-scrutinised in global research. Roughly one third of worldwide human effort in research and innovation is devoted—directly or indirectly—to refining ways to perpetrate premeditated organised violence. Yet the prospect of violent intentions is often regarded as an irrelevant factor in appraisal of supposedly 'neutral' technologies. With an uncertain proportion of world neuroscience research directed towards military ends, the picture here is particularly obscure (see Section 3.2). However, this 'see no evil' syndrome has long been acute with 'dual use' technologies in the nuclear, biological, and chemical sector and is now frequently raised of transgenics and synthetic biology. Misuse even of ostensibly non-military applications of neuroscience may hold profound security implications, for instance in prospects of 'remote brain imaging', 'brain fingerprinting' and 'transcranial magnetic stimulation' (Voß *et al.* 2006) (see Section 3.2 and Box 4) Some of these issues will be addressed as part of

Module 3 of the *Brain Waves* project on neurosience, conflict and security.

Unwise aspirations

This syndrome raises a familiar and innocent human foible: wishful thinking. A particular technology may deliver on its professed aims, but these may themselves turn out to be questionable. For instance, it is precisely the success of visions of affordable autonomous individual mobility that have led to domination of the urban automobile and associated pollution and gridlock. Despite undoubted benefits of increased material affluence, similar concerns are voiced about intensification of 'the consumer society'. Such issues may also arise in neuroscience. For instance, a companion section in this volume (see Section 3.2) asks whether enhanced memory faculties in old age are actually preferable to 'benign neglect'? Private provision for expensive forms of cognitive enhancement may also seriously aggravate social inequalities (see also Sections 3.2 and 3.3). Such innovation pathways may be judged problematic not because of 'side effects', but in terms of the intended aims themselves. Concerns compound even if—especially if—these innovations turn out to be 'successful'. In neuroscience and technology, as elsewhere, 'we should be careful what we wish for' (EGSG 2007).

Three principles

Taken together, these challenges for the governance of neuroscience and neurotechnology may seem dauntingly complex and intractable. Yet there does exist a well developed body of understanding and practice suggesting a series of practical policy responses. Although there are no panaceas, these may offer important ways systematically to maximise potential benefits and minimise adverse impacts. This is despite the fact that what constitutes a 'benefit' or 'impact' may be uncertain or contested (ESTO 1999; EEA 2001; EGSG 2007).

Perhaps the best way to express these responses is through three principles for 'social appraisal' of neuroscience and neurotechnology: *responsibility*, *precaution* and *engagement*. These move beyond narrow, technical, compartmentalised notions of 'risk analysis', 'regulatory assessment' or 'ethical deliberation'. They allow 'social appraisal' to develop in ways that transcend entrenched institutional and cultural divides. In particular, they equally avoid the perils of cornucopian notions of 'foresight' and apocalyptic ideas of 'precaution'. They help navigate a path between technocratic forms of 'expert analysis' and the more romantic aspirations to 'public participation'. Going beyond organisational remits and formal procedures, these principles are shared responsibilities of a range of actors and practices—not just 'regulators' or 'decision makers'. Rather than sidelining critical politics, they actively promote more mature political debate. Taken together, they offer more rigorous and democratically accountable paths for the pursuit of neuroscience and neurotechnology (Voß *et al.* 2006).

Responsibility

Governance of neuroscience and neurotechnology should be demonstrably independent from vested institutional, economic and political interests. These include academic disciplines, networks and institutes. Rather than treating appraisal as the exclusive preserve of specialist expertise (including ethics and social science), this requires open, proactive, accountable and transparently political oversight of innovation systems and regulatory structures. It should be recognised in various ways that developers, regulators, and scientists themselves are responsible and accountable for the indirect and 'unanticipated' impacts of neuro-innovations as well as their directly foreseeable effects. For instance, this may raise issues around the additive or synergistic effects of neuroscience applications in combination with other technologies, as well as possibilities for inadvertent or malign misuse.

Deliberation over such issues should take place at the earliest 'upstream' stages in innovation and policy processes, thus helping foster more benign pathways before 'lock-in' occurs. It should include account of industrial trends and cumulative behavior (as often missed in case-by-case regulation of individual applications). It should scrutinise the reversibility, 'agility' and 'resilience' of associated institutional or industrial commitments in the event of negative experience. The more reversible the commitments, the more relaxed might be the attitude to approval. Above all, such deliberation should explicitly reflect on the implications of the exercise of power within research and innovation systems (Voß *et al.* 2006).

In research prioritization and funding, as well as innovation, development and regulation, measures should be enacted to avoid the undue privileging of neuroscience over other socially feasible and beneficial applications. For instance, appraisal should explicitly and symmetrically compare an array of diverse alternative technology and policy options and potential substitutes for each neuroscience application. This should begin with prioritised societal needs, rather than favoured 'solutions' in search of applications. It should also afford comparable attention to organizational and behavioural innovations, as well as IP-intensive scientific or technology-based options. Greater value might be placed on diversity in research and development portfolios, such as to provide a variety of complementary options for addressing each focal problem. This involves a premium on increased international collaboration and co-ordination in neuroscience research (EEA 2001; SPRU 2009).

Precaution

Rather than simply assuming claimed benefits, regulation of neuroscience and neurotechnology should require justification of the social aims of neuroscience applications and their alternatives. This helps address otherwise opposing concerns, in that developers are concerned that regulation tends systematically to neglect the importance of benefits, while critics are concerned at

credulous acceptance that supplier advocacy equates with legitimate societal benefit. In examining pros and cons, appraisal should go beyond probabilistic methods, to scrutinise a wider range of uncertainties, sensitivities and scenarios than is normal in risk assessment or cost-benefit analysis. It should deliberately search for 'blind spots', gaps in knowledge and divergent scientific views of a kind not conventionally publicised by expert advisory committees. In specialist deliberations as in public communications, it should be clearly emphasised that 'absence of evidence' is not 'evidence of absence' of harm (ESTO 1999).

Where evidence is lacking, appraisal should attend to proxies of possible harm as well as manifest harm itself. For instance, rapid dissemination or eventual ubiquity of a neuroscience innovation are not in themselves hazardous, yet remain relevant to the scale of possible unforeseen harm. Rather than presuming 'proof' to be singular or self-evident, appraisal should deliberate over levels of proof, who should bear the burden of persuasion and who should resource the gathering of evidence and conduct of analysis. It should openly acknowledge that such matters are intrinsically value-based and therefore political in nature. Treating them more explicitly and distinctly, offers to improve both democratic and professional accountability.

Likewise, appraisal should avoid over-reliance on glamorous theoretical modeling at the expense of scientifically less-rewarding (yet crucial) monitoring and surveillance. This also helps defuse the otherwise polarised 'all-or-nothing' political dynamics between unconditional approval of 'safe' innovations or irrevocable 'banning' of questionable products. Where residual uncertainties remain over otherwise favourable innovations, then these may be addressed by deliberately-designed provision for liability, responsibility or product stewardship (ESTO 1999; EEA 2001).

Engagement

Principles of responsibility and precaution should be implemented in research and innovation governance in a fashion that is subject to open, accessible and democratically-accountable public engagement. There exists a wide variety of methods, procedures, and institutions for facilitating this. Different approaches will be appropriate in contrasting contexts. But certain basic features span the differences. For instance, design and implementation of governance procedures should draw on relevant non-expert knowledge and experience as well as a diversity of specialist expertise. Depending on context, this requires regulation to go beyond the most immediately-implicated neuroscience disciplines to include patients, families, teachers, healthcare staff, consumers and local communities. Although specialists are often also all these things, their engagement specifically with neuroscience and technology issues is conditioned by their roles as specialists—involving interests and evaluative perspectives as well as technical knowledge. This is as true of social scientific and ethical expertise, as it is of the natural science and engineering specialisms (EGSG 2007; Stirling 2008).

Particular attention should be paid to the interests and perspectives of 'stakeholder groups' who hold themselves to be possibly affected, whether or not these groups are formally organised or recognised. The point here is to engage with a diversity of relevant values and interests—including those of academic institutions, research networks and specialist disciplines. As has been demonstrated in regulatory experience like that with bovine spongiform encephalopathy (BSE), even apparently uninformed and spurious sensibilities like the 'yuk factor' can sometimes contain important intuitions. If considered in a measured fashion rather than rejected as irrational, such considerations can be indirectly informative, or serve to prompt valuable attention to otherwise neglected issues. The point here is not simply to foster 'trust', 'credibility or 'acceptance', but substantively to influence the nature and direction of the scientific and technological pathways that are pursued.

To this end, public engagement means being as rigorous in asking questions in social appraisal of neuroscience, as in deriving the answers. It is only in this way that there can be confidence over the institutional frameworks and procedures through which appraisal and governance are undertaken. Here, the point is not that public engagement delivers apparently simple prescriptive findings to policy making—with the task of decision makers then reduced merely to accepting or rejecting. Depending on the context, the role of public engagement may lie more in 'opening up' a diversity of possible choices, or in conveying to political decision makers a plurality of equally-reasonable possible policies—each with their associated conditions for adoption. It is in this way that public engagement of all kinds can avoid itself becoming a 'technical fix'—and inform and enrich (rather than subvert) democratic politics. Only through 'plural and conditional' signals to policy, may public engagement simultaneously support both scientific rigour and democratic accountability in the governance of neuroscience (Stirling 2008; Leach *et al.* 2010).

References and further reading

BIS (2009) Stilgoe J (ed), *The Road Ahead: Public Dialogue on Science and Technology*, Department for Business, Innovation and Skills. London: HMSO.

EEA (2001) Gee D, Harremoes P, Keys J, MacGarvin M, Stirling A, Vaz S, and Wynne B *Late Lesson from Early Warnings: the precautionary principle 1898–2000*. Copenhagen: European Environment Agency.

EGSG (2007). Expert Group on Science and Governance (Felt U, Wynne B, Callon M, Goncalves M, Jasanoff S, Jepsen M, Joly P-B, Konopasek Z, May S, Neubauer C, Rip A, Siune K, Stirling A, and Tallachini M (2007) *Science and Governance: taking European Knowledge Society Seriously*, Report to the European Commission DG Research.

ESTO (1999) European Science and Technology Observatory, *On Science and Precaution in the Management of Technological Risk: Volume I—a synthesis report of case studies*. Seville: European

Commission Institute for Prospective Technological Studies.

Leach M, Scoones I and Stirling A (2010) Dynamic Sustainabilities: Technology, Environment, Social Justice. London: Earthscan.

SPRU (2009) Scrase I, Stirling A, Geels F, Smith A and Van Zwanenberg P, *Transformative Innovation: A report to the Department for Environment, Food and Rural Affair.* Brighton: University of Sussex, Science and Technology Policy Research.

Stirling A (2008) Opening Up and Closing Down: power, participation and pluralism in the social appraisal of technology, *Science Technology and Human Values*, Vol 33, No 2, pp 262–294.

Stirling A (2009) Science, Precaution and the Politics of Technological Risk: converging implications in evolutionary and social scientific perspectives. *Annals of the New York Academy of Sciences*, Vol 1128, pp 95–110.

Voß J, Bauknecht D and Kemp R (eds) (2006) *Reflexive Governance for Sustainable Development*. Cheltenham: Edward Elgar.

Williams R and Edge D (1996) The Social Shaping of Technology. *Research Policy*, Vol 25, pp 865–899.

Wynne B (1992) Uncertainty and Environmental Learning: reconceiving science and policy in the preventive paradigm. *Global Environmental Change*, pp 111–127.

The Royal Society

The Royal Society is a Fellowship of more than 1400 outstanding individuals from all areas of science, mathematics, engineering and medicine, who form a global scientific network of the highest calibre. The Fellowship is supported by over 140 permanent staff with responsibility for the day-to-day management of the Society and its activities. The Society encourages public debate on key issues involving science, engineering and medicine, and the use of high quality scientific advice in policy-making. We are committed to delivering the best independent expert advice, drawing upon the experience of the Society's Fellows and Foreign Members, the wider scientific community and relevant stakeholders.

We are working to achieve five strategic priorities:

- **Invest** in future scientific leaders and in innovation
- **Influence** policymaking with the best scientific advice
- **Invigorate** science and mathematics education
- **Increase** access to the best science internationally
- **Inspire** an interest in the joy, wonder and excitement of scientific discovery

For further information

The Royal Society
Science Policy Centre
6–9 Carlton House Terrace
London SW1Y 5AG

T +44 (0)20 7451 2500
F +44 (0)20 7451 2692
E science.policy@royalsociety.org
W royalsociety.org

ISBN: 978-0-85403-879-4
Issued: January 2011 Report 01/11 DES2015

Founded in 1660, the Royal Society is the independent scientific academy of the UK, dedicated to promoting excellence in science

Registered Charity No 207043

ISBN 978-0-85403-879-4

Price £20