The Mind and the Machine. On the Conceptual and Moral Implications of Brain-Machine Interaction

Maartje Schermer

Abstract Brain-machine interfaces are a growing field of research and application. The increasing possibilities to connect the human brain to electronic devices and computer software can be put to use in medicine, the military, and entertainment. Concrete technologies include cochlear implants, Deep Brain Stimulation, neurofeedback and neuroprosthesis. The expectations for the near and further future are high, though it is difficult to separate hope from hype. The focus in this paper is on the effects that these new technologies may have on our 'symbolic order'—on the ways in which popular categories and concepts may change or be reinterpreted. First, the blurring distinction between man and machine and the idea of the cyborg are discussed. It is argued that the morally relevant difference is that between persons and non-persons, which does not necessarily coincide with the distinction between man and machine. The concept of the person remains useful. It may, however, become more difficult to assess the limits of the human body. Next, the distinction between body and mind is discussed. The mind is increasingly seen as a function of the brain, and thus understood in bodily and mechanical terms. This raises questions concerning concepts of free will and moral responsibility that may have far reaching consequences in the field of law, where some have argued for a revision of our criminal justice system, from retributivist to consequentialist. Even without such a (unlikely and unwarranted) revision occurring, brain-machine interactions raise many interesting questions regarding distribution and attribution of responsibility.

M. Schermer (✉)
Medical Ethics and Philosophy of Medicine, ErasmusMC,
Room AE 340, PO Box 2040, 3000 Rotterdam,
The Netherlands
e-mail: m.schermer@erasmusmc.nl

Introduction

Within two or three decades our brains will have been entirely unravelled and made technically accessible: nanobots will be able to immerse us totally in virtual reality and connect our brains directly to the Internet. Soon after that we will expand our intellect in a spectacular manner by melting our biological brains with non-biological intelligence. At least that is the prophecy of Ray Kurzweil, futurist, transhumanist and successful inventor of, amongst other things, the

electronic keyboard and the voice-recognition system.[1] He is not the only one who foresees great possibilities and, what's more, has the borders between biological and non-biological, real and virtual, and human and machine, disappear with the greatest of ease. Some of these possibilities are actually already here. On 22 June 2004, a bundle of minuscule electrodes was implanted into the brain of the 25-year-old Matthew Nagel (who was completely paralysed due to a high spinal cord lesion) to enable him to operate a computer by means of his thoughts. This successful experiment seems to be an important step on the way to the blending of brains and computers or humans and machines, that Kurzweil and others foresee. With regard to the actual developments in neuroscience and the convergence of neurotechnology with information, communication- and nanotechnology in particular it is still unclear how realistic the promises are. The same applies to the moral and social implications of these developments. This article offers a preliminary exploration of this area. The hypothesis is that scientific and technological developments in neuroscience and brain-machine interfacing challenge—and may contribute to shifts in—some of the culturally determined categories and classification schemes (our '*symbolic order*'), such as body, mind, human, machine, free will and responsibility (see the Introduction to this issue: Converging Technologies, Shifting Boundaries)

Firstly I will examine the expectations regarding the development of brain-machine interfaces and the forms of brain-machine interaction that already actually exist. Subsequently, I will briefly point out the moral issues raised by these new technologies, and argue the debate on these issues will be influenced by the shifts that may take place in our symbolic order—that is, the popular categories that we use in our everyday dealings to make sense of our world—as a result of these developments. It is important to consider the consequences these technologies might have for our understanding of central organizing categories, moral concepts and important values. Section four then focuses on the categories of human and machine: are we all going to become cyborgs? Will the distinction between human and machine blur if more and more artificial components are built into the body and brain? I will argue that the answer depends partly on the context

in which this question is asked, and that the concept of the person may be more suitable here than that of the human. Section five is about the distinction between body and mind. I argue that as a result of our growing neuroscientific knowledge and the mounting possibilities for technological manipulation, the mind is increasingly seen as a component of the body, and therefore also more and more in mechanical terms. This put the concept of moral responsibility under pressure. I will illustrate the consequences of these shifts in concepts and in category-boundaries with some examples of the moral questions confronting us already.

Developments in Brain-Machine Interaction

Various publications and reports on converging technologies and brain-machine interaction speculate heatedly on future possibilities for the direct linkage of the human brain with machines, that is: some form of computer or ICT technology or other. If the neuro-sciences provide further insight into the precise working of the brain, ICT technology becomes increasingly powerful, the electronics become more refined and the possibilities for uniting silicones with cells more advanced, then great things must lie ahead of us—or so it seems. The popular media, but also serious governmental reports and even scientific literature, present scenarios that are suspiciously like science fiction as realistic prospects: the expansion of memory or intelligence by means of an implanted chip; the direct uploading of encyclopaedias, databases or dictionaries into the brain; a wireless connection from the brain to the Internet; thought reading or lie detection via the analysis of brain activity; direct brain-to-brain communication. A fine example comes from the report on converging technologies issued by the American National Science Foundation:

> 'Fast, broadband interfaces directly between the human brain and machines will transform work in factories, control automobiles, ensure military superiority, and enable new sports, art forms and modes of interaction between people. [...] New communication paradigms (brain-to-brain, brain-machine-brain, group) could be realized in 10–20 years.' [39]

It is not easy to tell which prospects are realistic, which to a certain extent plausible and which are total

---

[1] See his website www.kurzweilAI.net for these and other future forecasts

nonsense. Some scientists make incredible claims whilst others contradict them again. These claims often have utopian characteristics and seem to go beyond the border between science and science fiction. Incidentally, they are frequently presented in such a way as to create goodwill and attract financial resources. After all, impressive and perhaps, from the scientific point of view, exaggerated future scenarios have a political and ideological function too—they help to secure research funds[2] and to create a certain image of these developments, either utopian or dystopian, thus steering public opinion.

Uncertainty about the facts—which expectations are realistic, which exaggerated and which altogether impossible—is great, even amongst serious scientists [12]. Whereas experts in cyberkinetic neurotechnology in the reputable medical journal, The Lancet, are seriously of the opinion that almost naturally-functioning, brain-driven prostheses will be possible, the editorial department of the Dutch doctors' journal, Medisch Contact, wonders sceptically how many light-years away they are [1, 23]. It is precisely the convergence of knowledge and technology from very different scientific areas that makes predictions so difficult. Although claims regarding future developments sometimes seem incredible, actual functioning forms of brain-machine interaction do in fact exist, and various applications are at an advanced stage of development. Next, I will look at what is currently already possible, or what is actually being researched and developed.

Existing Brain-Machine Interactions

The first category of existing brain-machine interaction is formed by the sensory prostheses. The earliest form of brain-machine interaction is the cochlear implant, also known as the bionic ear, which has been around for about 30 years. This technology enables deaf people to hear again, by converting sound into electrical impulses that are transmitted to an electrode implanted in the inner ear, which stimulates the auditory nerve directly. While there have been fierce discussions about the desirability of

the cochlear implant, nowadays they are largely accepted and are included in the normal arsenal of medical technology (e.g. [5]). In this same category, various research lines are currently ongoing to develop an artificial retina or 'bionic eye' to enable blind people to see again.

A second form of brain-machine interaction is Deep Brain Stimulation (DBS). With this technique small electrodes are surgically inserted directly into the brain. These are connected to a subcutaneously implanted neurostimulator, which sends out tiny electrical pulses to stimulate a specific brain area. This technology is used for treatment of neurological diseases such as Parkinson's disease and Gilles de la Tourette's syndrome. Many new indications are being studied experimentally, ranging from severe obsessive-compulsive disorders, addictions, and obesity to Alzheimer's disease and depression. The use of this technique raises a number of ethical issues, like informed consent from vulnerable research subjects, the risks and side-effects, including effects on the patient's mood and behaviour [38].

More spectacular, and at an even earlier stage of development, is the third form of brain-machine interaction in which the brain controls a computer directly. This technology, called neuroprosthetics, enables people to use thought to control objects in the outside world such as the cursor of a computer or a robotic arm. It is being developed so that people with a high spinal cord lesion, like Matt Nagel mentioned in the introduction, can act and communicate again. An electrode in the brain receives the electrical impulses that the neurons in the motor cerebral cortex give off when the patient wants to make a specific movement. It then sends these impulses to a computer where they are translated into instructions to move the cursor or a robot that is connected to the computer. This technology offers the prospect that paraplegics or patients with locked in syndrome could move their own wheelchair with the aid of thought-control, communicate with others through written text or voice synthesis, pick up things with the aid of artificial limbs et cetera.

In future, the direct cortical control described above could also be used in the further development of artificial limbs (robotic arms or legs) for people who have had limbs amputated. It is already possible to receive the signals from other muscles and control a robotic arm with them (a myoelectrical prosthesis);
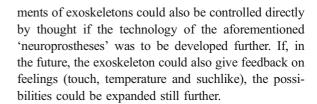
---

[2] A lot of research in the field of brain-machine interaction and other converging technologies is carried out by DARPA, the American Ministry of Defence research institute. In 2003, for example, DARPA subsidized research into brain-machine interfaces to the tune of 24 million dollars [9, 35].

whether the patient's own remaining nerves can be connected directly to a prosthesis to enable it to move as though it is the patient's own arm is now being examined. Wireless control by the cortex would be a great step in prosthetics, further enabling patient rehabilitation. Next to motor control of the prosthesis, tactile sensors are being developed and placed in artificial hands to pass on the feeling to the patient's own remaining nerves, thus creating a sense of touch. It is claimed that this meeting of the (micro)technological and (neuro)biological sciences will in the future lead to a significant reduction in invalidity due to amputation or even its total elimination [23, 24].

In the fourth form of brain-machine interaction, use is made of neurofeedback. By detecting brain activity with the aid of electroencephalography (EEG) equipment, it can be made visible to the person involved. This principle is used, for instance, in a new method for preventing epileptic attacks with the aid of Vagal Nerve Stimulation (VNS). Changes in brainwaves can be detected and used to predict an oncoming epileptic attack. This 'warning system' can then generate an automatic reaction from the VNS system which stimulates the vagal nerve to prevent the attack. In time, the detection electrodes could be implanted under the skull, and perhaps the direct electrical stimulation of the cerebral cortex could be used instead of the vagal nerve [17]. Another type of feedback system is being developed by the American army and concerns a helmet with binoculars that can draw a soldier's attention to a danger that his brain has subconsciously detected enabling him to react faster and more adequately. The idea is that EEG can spot 'neural signatures' for target detection before the conscious mind becomes aware of a potential threat or target [50].

Finally, yet another technology that is currently making rapid advances is the so-called exoskeleton. Although this is not a form of brain-machine interaction in itself, it is a technology that will perhaps be eminently suitable for combination with said interaction in the future. An exoskeleton is an external structure that is worn around the body, or parts of it, to provide strength and power that the body does not intrinsically possess. It is chiefly being developed for applications in the army and in the health care sector.[3] Theoretically, the move-

ments of exoskeletons could also be controlled directly by thought if the technology of the aforementioned 'neuroprostheses' was to be developed further. If, in the future, the exoskeleton could also give feedback on feelings (touch, temperature and suchlike), the possibilities could be expanded still further.

Ethical Issues and Shifts in Our Symbolic Order

The developments described above raise various ethical questions, for instance about the safety, possible risks and side effects of new technologies. There are also speculations as to the moral problems or dangers that may arise in connection with further advances in this type of technologies. The European Group on Ethics in Science and New Technologies (EGE), an influential European advisory body, warns for the risk that ICT implants will be used to control and locate people, or that they could provide third parties with access to information about the body and mind of those involved, without their permission EGE [9]. There are also concerns about about the position of vulnerable research subjects, patient selection and informed consent, the effects on personal identity, resource allocation and about the use of such technologies for human enhancement Over the past few years, the neuroethical discussion on such topics has been booming (e.g. [2, 7, 10, 12, 14, 46, 20]).

It has been argued that while these ethical issues are real, they do not present anything really new [8]. The point of departure of this article, however, is that it is not so easy to deal adequately with the moral questions raised by these new technologies because they also challenge some of the central concepts and categories that we use in understanding and answering moral questions. Hansson, for example, states that brain implants may be "reason to reconsider our criteria for personal identity and personality changes" [20; p. 523]. Moreover, these new technologies may also change some elements of our common morality itself, just like the birth control pill once helped to change sexual morality [28]. In brief: new technologies not only influence the ways we can act, but also the symbolic order: our organizing categories and the associated views on norms and values.

The concepts and categories we, as ordinary people, use to classify our world to make it manageable and comprehensible are subject to change. These categories

---

[3] In the former case, these applications might enable soldiers to carry heavy rucksacks more easily, in the latter they could, for example, help a nurse to lift a heavy patient on his or her own.

also play an important part in moral judgement, since they often have a normative next to a descriptive dimension. Categories such as human and machine, body and mind, sick and healthy, nature and culture, real and unreal are difficult to define precisely and the boundaries of such notions are always vague and movable. Time and again it takes 'symbolic labour' to reinterpret these categories and to re-conceptualise them and make them manageable in new situations. In part, this symbolic labour is being done by philosophers who explicitly work with concepts and definitions, refining and adjusting them; in part it is also a diffuse socio-cultural process of adaptation and emerging changes in symbolic order.[4] Boundaries are repeatedly negotiated or won, and new concepts arise where old ones no longer fit the bill An example is the new concept of 'brain dead' which arose a few decades ago as a consequence of the concurrent developments in elec-troencephalography, artificial respiration and organ transplantation. Here the complex interplay of tech-nology, popular categories of life and death, and scientific and philosophical understandings of these concepts is clearly demonstrated [27; p. 16].

Morality, defined as our more or less shared system of norms, values and moral judgements, is also subject to change It is not a static 'tool' that we can apply to all kinds of new technologically induced moral problems. Technological and social developments influence and change our morality, although this does not apply equally to all its elements. Important values such as justice, well-being or personal autonomy are reasonably stable, but they are also such abstract notions that they are open to various and changing interpretations. The norms we observe in order to protect and promote our values depend on these interpretations and may require adjustment under new circumstances. Some norms are relatively fixed, others more contingent and changing [6]. The detailed and concrete moral rules of conduct derived from the general norms are the most contin-gent and changeable. The introduction of the notion brain death, for example, led to adaptations in ethical norms and regulations. Likewise, the new develop-ments in genomics research are now challenging

and changing existing rules of informed consent as well as notions of privacy and rules for privacy protection [33].

In the field of brain-machine interaction we can therefore also expect that certain fixed categories that we classify our world with and that structure our thinking, will evolve alongside the new technologies. This will have consequences for the ethical questions these technologies raise *and* for the way in which we handle both new and familiar moral issues. A first shift that can be expected concerns the distinction between human and machine. This distinction might fade as more parts of the body can be replaced with mechanical or artificial parts that become more and more 'real'. Secondly, we might expect a blurring of boundaries between our familiar concepts of body and mind when neuroscience and neurotechnologies increasingly present the brain as an ordinary part of our body and the mind as simply one of its 'products'. The following sections analyse these possible shifts in the symbolic order and the associated moral questions in more detail.

Symbolic Order in Motion: The Human Machine

The blurring of the boundary between human and machine brought about by brain-machine interaction forms the first challenge to the familiar categories with which we think. The more artificial parts are inserted in and added to the human body, the more uncertainty there is about where the human stops and the machine begins. Instead of human or machine, we increasingly seem to be looking at cyborgs: human *and* machine in a single being.

For a long time it was easy to distinguish between people and the tools, machines or devices that they used. Gradually, however, our lives have become increasingly entangled with machines—or, in the broader sense, with technology—and we have become dependent on them for practically every facet of our daily lives. Increasingly, parts of the human body itself are replaced or supplemented by technology.[5] Of

---

[4] How these processes interact with one another and how socio-cultural changes influence philosophical thinking and vice versa is an interesting and complicated question, that I cannot start to answer here.

[5] Not that this is an entirely new phenomenon, seeing that all sorts of bodily prostheses have existed for centuries; the first artificial leg dates back to 300 before Christ. Other prostheses that we more or less attach to our bodies are, for instance, spectacles, hairpieces and hearing aids. But the insertion of external parts *into* the human body is more recent.

course, the notion that the human body works as a machine has been a *leitmotiv* in western culture since Descartes; this vision has enabled modern medicine while the successes achieved substantiate the underlying beliefs about the body at the same time. The emergence of transplantation medicine was a clear step in the development of popular views on the body as a machine. Since the first kidney transplantation in 1954 and the first heart transplantation in 1967, lungs, liver, pancreas and even hands and faces have become transplantable, thus enforcing the image of the human body as a collection of replaceable parts. Some have criticised transplantation medicine because of the ensuing mechanization and commodification of the human body.

Besides organs, more and more artefacts are now being implanted in the human body: artificial heart valves, pacemakers, knees, arterial stents and subcutaneous medicine pumps. Prostheses that are attached to the body, such as artificial limbs, are becoming increasingly advanced, and are no longer easy to detach—unlike the old-fashioned wooden leg. Experiences of patients who wear prostheses seem to indicate that people rapidly adapt to using them and fuse with them to the extent that they perceive them as natural parts of themselves. Artificial parts are rapidly included in the body scheme and come to be felt as 'ones own'.[6]

In a certain sense, then, we are familiar with the perception of the body as a sort of machine, and with the fact that fusing the human body with artificial parts is possible. Do technologies like neuroprostheses, artificial limbs and exoskeletons break through the boundary between human and machine in a fundamentally new, different fashion? Should the conceptual distinction between human and machine perhaps be revised? Many publications, both popular and more academic, suggest that the answer has to be yes. A

---

[6] More attention has recently been paid to the importance of the body for the development and working of our consciousness. In the Embodied Mind model, body and mind are seen far more as interwoven than in the past (e.g. [15]). This can have even further implications for brain-machine interaction; if for example, neuroprostheses change our physical, bodily dealings with the world, this may also have consequences for the development of the brain and for our consciousness. Neuroscientists have even claimed that: 'It may sound like science fiction but if human brain regions involved in bodily self-consciousness were to be monitored and manipulated online via a machine, then not only will the boundary between user and robot become unclear, but human identity may change, as such bodily signals are crucial for the self and the 'I' of conscious experience' [4]

notion that is often used in this connection is that of the cyborg: the human machine.

Cyborgs

The term 'cyborg'—derived from cybernetic organism— was coined in 1960 by Manfred Clynes and Nathan Kline, American researchers who wrote about the ways in which the vulnerable human body could be technologically modified to meet the requirements of space travel and exploration. The figure of the cyborg appealed to the imagination and was introduced into popular culture by science fiction writers, filmmakers, cartoonists and game designers; famous cyborgs include The Six Million Dollar Man, Darth Vader and RoboCop. In the popular image, the cyborg thus stands for the merging of the human and the machine.

In recent literature, both popular and scientific, the cyborg has come to stand for all sorts of man-machine combinations and all manner of technological and biotechnological enhancements or modifications of the human body. With the publication of books like *I, Cyborg* or *Cyborg Citizen*, the concept now covers a whole area of biopolitical questions. Everything that is controversial around biotechnological interventions, that raises moral questions and controversy, that evokes simultaneous horror and admiration, is now clustered under the designation 'cyborg' [18, 25, 48].

The concept of the cyborg indicates that something is the matter, that boundaries are transgressed, familiar categorizations challenged, creating unease and uncertainty. For Donna Haraway, well-known for her Cyborg Manifesto [21], the concept of the cyborg stands for all the breaches of boundaries and disruptions of order, not merely for the specific breaking through of the distinction between human and machine which concerns me here. The term cyborg can thus be used to describe our inability to categorize some new forms of human life or human bodies. The use of the term compels us to delay categorization—in familiar terms of human or machine—at least for the moment and so creates a space for further exploration.

Monsters

Following Mary Douglas, Martijntje Smits has called these kind of entities that defy categorization and challenge the familiar symbolic order *monsters* [44]. Smits discusses four strategies for treating these

monsters, four ways of cultural adaptation to these new entities and the disruption they bring about.

The first strategy, *embracing the monster*, is clearly reflected in the pronouncements of adherents of the transhumanist movement. They welcome all manner of biotechnological enhancements of humans, believe in the exponential development of the possibilities to this end and place the cyborg, almost literally, on a pedestal. The second strategy is the opposite of the first and entails *exorcizing the monster*. Neo-Luddites or bioconservatives see biotechnology in general and the biotechnological enhancement of people in particular, as a threat to the existing natural order. They frequently refer to human nature, traditional categories and values and norms when attacking and trying to curb the new possibilities. To them, the cyborg is a real monster that has to be stopped and exorcized.

The third strategy is that of *adaptation of the monster*. Endeavours are made to classify the new phenomenon in terms of existing categories after all. Adaptation seems to be what is happening with regard to existing brain-machine interaction. The conceptual framework here is largely formed by the familiar medical context of prostheses and aids. The designation of the electrodes and chips implanted in the brain as neuroprostheses, places them in the ethical area of therapy, medical treatment, the healing of the sick and support of the handicapped. As long as something that was naturally present but is now lost due to sickness or an accident is being replaced, brain-machine interaction can be understood as therapy and therefore accepted within the ethical limits normally assigned to medical treatments. However, for non-medical applications the problem of classifications remains. Prostheses to replace functions that have been lost may be accepted relatively easily, but how are we going to regard enhancements or qualitative changes in functions such as the addition of infrared vision to the human visual faculty? Are we only going to allow the creation of cyborgs for medical purposes, or also for military goals, or for relaxation and entertainment?

Finally, the fourth strategy is *assimilation of the monster*, whereby existing categories and concepts are adjusted or new ones introduced.[7] In the following I

---

[7] The distinction between adaptation and assimilation is not very clear—it depends on what one would wish to call an 'adaptation' or 'adjustment' of a concept.

will suggest that the concept of the person—in the sense in which it is used in ethics, rather than in common language—may be useful for this purpose.

Morality of Persons

In the empirical sense, cyborgs, or blends of human bodies with mechanical parts, are gradually becoming less exceptional. It therefore seems exaggerated to view people with prostheses or implants as something very exceptional or to designate them as a separate class. And this raises the question of why we should really worry about the blurring of the distinction between human and machine? This is not merely because the mixing of the flesh with steel or silicone intuitively bothers us, or because the confusion about categories scares us. More fundamentally, I believe this is because the distinction between the human and the machine also points to a significant *moral* distinction. The difference between the two concepts is important because it indicates a moral dividing line between two different normative categories. For most of our practices and everyday dealings the normative distinction between human and machine matters. You just treat people differently to machines—with more respect and care—and you expect something else from people than you expect from machines— responsibility and understanding, for example. Human beings deserve praise and blame for their actions while machines cannot. The important question is therefore whether brain-machine interfaces will somehow affect the moral status of the people using them [13]. Do we still regard a paralysed patient with a brain implant and an exoskeleton as a human being, or do we see him as a machine? Will we consider someone with two bionic legs to be a human being or a machine?

I belief that in part this also depends on the context and the reasons for wanting to make the distinction. In the context of athletic competition, the bionic runner may be disqualified because of his supra-human capacities. In this context, he is 'too much of a machine' to grant fair competition with fully biological human beings. However, in the context of everyday interaction with others, a person with bionic legs is just as morally responsible for his actions as any other person. In this sense he clearly is human and not a machine. This is because, with regard to moral status, the human being as an acting, responsible moral agent is identified more with the mind than with the body. The mind is what

matters in the moral sense. Whether this mind controls a wheelchair with the aid of hands, or electrical brain-generated pulses, is irrelevant to the question of who controls the wheelchair: the answer in both cases is the person concerned. Whether someone is paralysed or not does not alter the question of whether he or she is a person or not; it will of course affect the kind of person he or she is but whether he or she is a person depends on his or her mental capacities. Ethical theories consider the possession of some minimal set of cognitive, conative and affective capacities as a condition for personhood. This means that, ethically speaking, under certain conditions, intelligent primates or Martians could be considered persons while human babies or extremely demented old people would not. Whatever the exact criteria one applies, there is no reason to doubt the fact that someone who is paralysed, someone who controls a robot by remote or someone who has a DBS electrode is a person. Certain moral entitlements, obligations and responsibilities are connected to that state of 'being a person'. This notion therefore helps to resolve the confusion surrounding the cyborg. Rather than classifying him as either man or machine, we should be looking at personhood. Personhood is what really matters morally and this is not necessarily affected by brain-machine interfaces. As long as they do not affect personhood, brain-machine interfaces are no more special than other types of prosthesis, implants or instrumental aids that we have already grown used to.

New Views on Physical Integrity?

Nevertheless, brain-machine interfaces may in some cases cause new moral issues. A concrete example that can illustrate how shifting catagories can affect concepts and ethics is that of physical integrity. How should this important ethical and legal principle be interpreted when applied to cyborgs? The principle itself is not under discussion. We want to continue to guard and protect physical integrity. The question is, however, how to define the concept 'body' now that biological human bodies are becoming increasingly fused with technology and where to draw the line between those plastic, metal or silicone parts that *do* belong to that body and the parts that do not.

In the spring of 2007 the Dutch media paid attention to an asylum seeker who had lost an arm as a result of torture in his native country and had received a new myoelectrical prosthesis in the Netherlands. He just got used to the arm and was trained in using it naturally when it became apparent that there were problems with the insurance and he would have to return the prosthesis. Evidently, according to the regulations a prosthesis does not belong to the body of the person in question and it does not enjoy the protection of physical integrity. However, the loss of an arm causes a great deal of damage to the person, whether the arm is natural or a well-functioning prosthesis. If prostheses become more intimately connected to and integrated with the body (also through tactile sensors) such that they become incorporated in the body scheme and are deemed a natural part of the body by the person concerned, it seems there must come a point at which such a prosthesis should be seen as *belonging* to the (body of the) person concerned from the moral and legal point of view. It has even been questioned whether the interception of signals that are transmitted by a wireless link from the brain to a computer or artificial limb, should perhaps also fall under the protection of physical integrity [29]

Symbolic Order in Motion: Body-Mind

In the previous section I assumed the distinction between body and mind to be clear-cut. The common view is that the mind controls the body (whether this body is natural or artificial) and that the mind is the seat of our personhood, and of consciousness, freedom and responsibility. In this section I examine how this view might change under the influence of new brain-machine interactions and neuroscientific developments in general and what implications this may have for ethics. I will concentrate on DBS, since this brain-machine technique has at present the clearest impact on human mind and behaviour.[8] Of course, however, our categories and common views will not change because of one single new technique—rather, it is the whole constellation of neuroscientific research and (emerging) applications that may change the ways in which we understand our minds and important related concepts.

---

[8] By contrast, in the case of the neuroprosthesis discussed in the previous section, it is mainly the mind that influences the body, through the interface.

The Mind as Machine

Neuroprostheses and other brain-machine interactions call into question the demarcation between body and mind, at least in the popular perception. Technologies such as neuroprosthetics and DBS make very clear the fact that physical intervention in the brain has a direct effect on the mind of the person in question. By switching the DBS electrode on or off, the behaviour, feelings and thoughts of the patient can be changed instantly. Thoughts of a paralysed person can be translated directly into electrical pulses and physical processes. As a result of neuroscience and its applications the human mind comes to be seen more and more as a collection of neurones, a system of synapses, neurotransmitters and electrical conductors. A very complex system perhaps, but a physical system nonetheless, that can be connected directly to other systems.

For some, this causes moral concern, since it may lead us to see ourselves merely in mechanical terms:

'The obvious temptation will be to see advances in neuroelectronics as final evidence that man is just a complex machine after all, that the brain is just a computer, that our thoughts and identity are just software. But in reality, our new powers should lead us to a different conclusion: even though we can make the brain compatible with machines to serve specific functions, the thinking being is a being of very different sorts.' [26; p. 40–41]

I believe this change in our popular view of the mind that Keiper fears is actually already taking place. Neuroscientific knowledge and understanding penetrate increasingly into our everyday lives, and it is becoming more normal to understand our behaviour and ourselves in neurobiological terms. This shift is for example noticeable in the rise of biological psychiatry. Many psychiatric syndromes that were still understood in psychoanalytical or psychodynamic terms until well into the second half of the twentieth century, are now deemed biological brain diseases. The shift is also noticeable in the discussion on the biological determinants of criminal behaviour (and opportunities to change such behaviour by intervening in the brain) or in the increased attention for the biological and evolutionary roots of morality. Also in popular magazines and books, our behaviour and ourselves are increasingly presented as the direct result of our brains' anatomy and physiology.

Scientific and technological developments have contributed to this shift. The development of EEG in the first half of the last century revealed the electrical activity of the brain for the first time, thus creating the vision of the brain as the wiring of the mind. The development of psychiatric drugs in the second half of the last century also helped naturalize our vision of the mind, picturing the brain as a neurochemical 'soup', a collection of synapses, neurotransmitters and receptors [22]. More recently the PET scan and the fMRI have made it possible to look, as it were, inside the active brain. The fact that fMRI produces such wonderful pictures of brains 'in action' contributes to our mechanical view of the relation between brain and behaviour. Certain areas of the brain light up if we make plans, others if an emotional memory is evoked; damage in one area explains why the psychopath has no empathy, a lesion in another correlates with poor impulse control or hot-headedness. While neurophilosophers have warned against the oversimplified idea that images are just like photographs that show us directly how the brain works, these beautiful, colourful images appeal to scientists and laymen alike [41].

According to Nikolas Rose, we have come to understand ourselves increasingly in terms of a biomedical body, and our personalities and behaviour increasingly in terms of the brain. He says that a new way of thinking has taken shape: 'In this way of thinking, all explanations of mental pathology must 'pass through' the brain and its neurochemistry—neurones, synapses, membranes, receptors, ion channels, neurotransmitters, enzymes, etc.' [40; p. 57]

We are experiencing what he calls a 'neurochemical reshaping of personhood' [40; p. 59]. Likewise, Mooij has argued that the naturalistic determinism of the neurosciences is also catching on in philosophy and has now spread broadly in the current culture 'that is to a large extent steeped in this biological thinking, in which brain and person more or less correspond' [34; p.77].

The mind is being seen more and more as a physical, bodily object (the 'the mind = the brain' idea), and given that the human body, as described above, has long been understood in mechanical terms, the equal status of the mind and brain means that the mind can also be understood in mechanical terms. As the basic distinction between mind and machine seems to drop away the distinction between human and machine once more raises its head, but now on

a more fundamental and extremely relevant level, morally speaking. If in fact our mind, the seat of our humanity, is also a machine, how should we understand personhood in the morally relevant sense? How can we hold on to notions such as free will and moral responsibility?

Neuroscientific Revisionism

A recent notion amongst many neuroscientists and some neurophilosophers is that our experience of having a self, a free will or agency, is based on a misconception. The self as a regulating, controlling authority does not exist, but is only an illusion produced by the brain.[9] From this notion it seems to follow that there is no such thing as free will and that there can therefore be no real moral responsibility. Within philosophy revisionists, who claim that our retributive intuitions and practices are unwarranted under determinism, claim that this view obliges us to revise our responsibility-attributing practices, including our legal system. Revisionism implies the need to replace some of our ordinary concepts with new ones. It has, for example, been suggested to substitute blame with 'dispraise' [43] or to eliminate concepts connected to desert like blame, moral praise, guilt and remorse, altogether [47]. On a revisionist account, praise, blame and punishment are just devices that modify conduct, and that can be more or less effective, but not more or less deserved.

Greene and Cohen assume that because of the visible advances in the neurosciences—and I take brain-machine interfaces to be part of those—the naturalistic deterministic view on human behaviour will by degrees be accepted by more and more people, and revisionism will catch on [19]. To their way of thinking, our moral intuitions and our folk psychology will slowly adapt to the overwhelming evidence the neurosciences present us with. The technologies enabled on the basis of neuroscientific understanding, such as DBS, neurofeedback, psychiatric drugs, and perhaps also intelligent systems or intelligent robots, can contribute to this. Little by little we will hold people less responsible and liable for their actions, according to Greene and Cohen, but will see them increasingly as determined beings who can be regulated, more or less effectively, by sanctions or rewards. They allege that questions concerning free will and responsibility will lose their power in an age in which the mechanistic nature of the human decision process will be totally understood. This will also have consequences for the legal system. 'The law will continue to punish misdeeds, as it must for practical reasons, but the idea of distinguishing the truly, deeply guilty from those who are merely victims of neuronal circumstances will, we submit, seem pointless.' ([19; p. 1781])

Greene and Cohen, like other revisionists, advocate a shift in the nature of our criminal justice system, from a retributive to a consequentialistic system. This means a shift from a system based on liability and retribution to one based on effects and effectiveness of punishment. A consequentialistic system of this kind is, in their opinion, in keeping with the true scientific vision of hard determinism and the non-existence of free will. Greene and Cohen recognize that many people will intuitively continue to think in terms of free will and responsibility. What is more, they think that this intuitive reflex has arisen through evolution and is deeply rooted in our brains. We can hardly help thinking in these sorts of terms, despite the fact that we know better, scientifically speaking. Nonetheless, Greene and Cohen insist that we should base important, complex matters such as the criminal justice system[10] on the scientific truth about ourselves and not allow ourselves to be controlled by persistent, but incorrect, intuitions.

---

[9] 'Obviously we have thoughts. Ad nauseam, one might say. What is deceptive, is the idea that these thoughts control our behaviour. In my opinion, that idea is no more than a side effect of our social behaviour. […] The idea that we control our deeds with our thoughts, that is an illusion', says cognitive neuroscientist Victor Lamme, echoing his colleague Wegner. [30; p. 22, 49].

[10] Likewise, moral views on responsibility may change. An instrumental, neo-behaviouristic vision on morality and the moral practice of holding one another responsible might arise. Holding one another responsible may still prove a very effective way of regulating behaviour, even if it is not based on the actual existence of responsibility and free will. As long people change their behaviour under the influence of moral praise and blame, there is no reason to throw the concept of responsibility overboard. From this point of view, there would be no relevant difference anymore between a human being and any other system that would be sensitive to praise and blame, such as an intelligent robot or computer system. If such a system would be sensitive to moral judgements and respond to them with the desired behaviour on this view they would qualify as much as moral actors as human beings would.

Moral Responsibility Reconsidered

A whole body of literature has accumulated refuting this thesis and arguing that new neuroscientific evidence need not influence our moral and legal notions of responsibility (e.g. [37]). This literature reflects the dominant position in the determinism debate nowadays, that of compatibilism. According to compatibilism determinism is reconcilable with the existence of a free will, and with responsibility. As long as we can act on the basis of reasons and as long as we are not coerced, we are sufficiently free to carry responsibility and the naturalistic neuroscientific explanatory model of behaviour is therefore not necessarily a threat to our free will and responsibility, according to the compatibilist. The question is, however, whether the compatibilist's philosophical argumentation also convinces the average layman or neuroscientist, certainly in the light of new experimental findings and technical possibilities. How popular views on this topic will develop remains to be seen.

At the moment even adherents of the revisionist view seem convinced that we will never be able to stop thinking, or even think less, in terms of intentions, reasons, free will and responsibility. It seems almost inconceivable *not* to hold one another responsible for deeds and behaviour [45].

Nevertheless, neuroscientific research does challenge our view of ourselves as rational, autonomous and moral beings [31]. Research shows us, for example, that many if not most of our actions automatic, unconsciously initiated and only some of our actions are deliberate and consciously based on reasons. Our rationality, moreover, is limited by various biases, like confirmation bias, hyperbolic discounting, false memories et cetera. New findings in neuroscience, such as the fact that immaturity of the frontal lobes impedes the capacities for reasoning, decision making and impulse control in adolescents, or that exercise of self-constraint eventually leads to exhaustion of the capacity for self-control (ego-depletion), do necessitate us to re-think the ways in which or the degrees to which we are actually morally responsible is specific situations and circumstances (see for example the series of articles on addiction and responsibility in AJOB Neuroscience 2007).

A more naturalized view of the human mind could thus still have important consequences, even if we do not jettison the notion of moral responsibility altogether. More grounds for 'absence of criminal responsibility' could, for example, be acknowledged in criminal law, whereby new technologies could play a role. Functional brain scans might provide more clarity on the degree to which an individual has control over his or her own behaviour.

Prosthetic Responsibility?

Due to their ability to directly influence complex human behaviour by intervening in the brain, brain-machine interfaces may raise interesting issues of responsibility, even when we reject revisionism, as can be illustrated by the following case of a 62 year old Parkinson patient treated with DBS.[11]

After implantation of the electrodes, this patient became euphoric and demonstrated unrestrained behaviour: he bought several houses that he could not really pay for; he bought various cars and got involved in traffic accidents; he started a relationship with a married woman and showed unacceptable and deviant sexual behaviour towards nurses; he suffered from megalomania and, furthermore, did not understand his illness at all. He was totally unaware of any problem. Attempts to improve his condition by changing the settings of the DBS failed as the manic characteristics disappeared but the patient's severe Parkinson's symptoms reappeared. The patient was either in a reasonable motor state but in a manic condition lacking any self reflection and understanding of his illness, or bedridden in a non-deviant mental state. The mania could not be treated by medication [32].

Who was responsible for the uninhibited behaviour of the patient in this case? Was that still the patient himself, was it the stimulator or the neurosurgeon who implanted and adjusted the device? In a sense, the patient was 'not himself' during the stimulation; he behaved in a way that he never would have done without the stimulator.[12] That behaviour was neither

---

[11] This case also been discussed by [3, 16, 42].

[12] Of course, this problem is not exclusive for DBS; some medications can have similar effects. However, with DBS the changes are more rapid and more specific and can be controlled literally by a remote control (theoretically, the patients behaviour can thus be influenced without the patient's approval once the electrode is in his brain). These characteristics do make DBS different from more traditional means of behaviour influencing, though I agree with an anonymous reviewer that this is more a matter of degree than an absolute qualitative difference.

the intended nor the predicted result of the stimulation and it therefore looks as though no one can be held morally responsible for it. However, in his non-manic state when, according to his doctors, he was competent to express his will and had a good grasp of the situation, the patient chose to have the stimulator switched on again. After lengthy deliberations the doctors complied with his wishes. To what extent were his doctors also responsible for his manic behaviour? After all, they knew the consequences of switching on the stimulator again. To what extent was the patient himself subsequently to blame for getting into debt and bothering nurses?
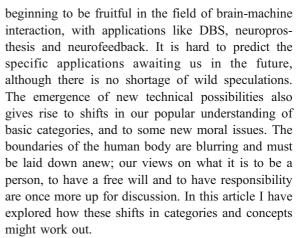
For such decisions, the notion of 'diachronic responsibility' [36] can be of use, indicating that a person can take responsibility for his future behaviour by taking certain actions. Suppose, for example, that DBS would prove an effective treatment for addiction, helping people to stay off drugs, alcohol or gambling, could it then rightly be considered a 'prosthesis for willpower' [11], or even a prosthesis for responsibility? I believe that technologies that enable us to control our own behaviour better—as DBS might do in the case of addiction, or in the treatment of Obsessive Compulsive Disorder—can be understood in terms of diachronic responsibility and self-control, and thus enhance autonomy and responsibility [42].

Future applications of brain-machine interaction may raise further questions: suppose a doctor would adjust the settings of DBS without consent from the patient and cause behaviour the patient claimed not to identify with—who would then be responsible? As Clausen has pointed out, neuroprostheses may challenge our traditional concept of responsibility when imperfections in the system lead to involuntary actions of the patient [7]. Likewise, if the wireless signals of a neuroprosthesis were incidentally or deliberately disrupted, it would be questionable who would be responsible for the ensuing 'actions' of the patient.

Clearly, even without major shifts in our views on free will and responsibility, brain-machine interfaces will require us to consider questions of responsibility.

Conclusion

The convergence of neuroscientific knowledge with bio-, nano-, and information technology is already beginning to be fruitful in the field of brain-machine interaction, with applications like DBS, neuroprosthesis and neurofeedback. It is hard to predict the specific applications awaiting us in the future, although there is no shortage of wild speculations. The emergence of new technical possibilities also gives rise to shifts in our popular understanding of basic categories, and to some new moral issues. The boundaries of the human body are blurring and must be laid down anew; our views on what it is to be a person, to have a free will and to have responsibility are once more up for discussion. In this article I have explored how these shifts in categories and concepts might work out.

I have argued that the distinction between human and machine, insofar as it concerns a morally relevant distinction, does not have to be given up immediately because increasingly far-reaching physical combinations are now being made between human and mechanical parts. Depending on the context and the reasons we have for wanting to make a distinction, we will draw the line between human and machine differently. In the context of sports, a bionic limb may disqualify its user for being too much of a 'machine' while in another context such a limb may be qualified as an integral part of a human being and be protected under the right to physical integrity. Important general moral questions that lie behind the confusion about categories of human and machine concern moral responsibility and moral status. The concept of a person, as used in ethical theory to designate moral actors, is more precise and more useful in this context than the general category of the 'human' or the poly-interpretable notion of the 'cyborg'.

In the most radical scenario of shifts in our symbolic order, the concept of 'person' may also come under pressure. As I have shown based on Greene and Cohen's vision, the person as a being with a free will and moral responsibility, and as a moral actor, should, according to some, disappear from the stage altogether. Implementing such a neuroreductionistic vision on the mind and free will would have clear consequences for criminal law: it would have to be revised to a consequentialist, neo-behaviouristic system. People would then barely be considered to be morally responsible beings but be seen as systems that respond to praise and blame in a mechanical fashion. I believe it is unlikely that such a shift in our popular views will come about, because the intuitive appeal of

the notion of responsibility, and because there are many good arguments to resist this shift. Even if we do not jettison responsibility altogether, however, brain-machine interactions raise many interesting questions regarding distribution and attribution of responsibility.

A general lesson for ethics of emerging technologies is that such technologies necessitate renewed consideration and reinterpretation of important organizing concepts and distinctions that are crucial to moral judgement. The symbolic labour required to answer such conceptual and normative questions is at least as important for the development of converging technologies as the technical-scientific labour involved.

References

1. Anonymus (2007) Kunstarm met gevoel. Med Contact 62:246
2. Bell EW, Mathieu G, Racine E (2009) Preparing the ethical future of deep brain stimulation. Surg Neurol. doi:10.1016/j.surneu.2009.03.029
3. Berghmans R, de Wert G (2004) Wilsbekwaamheid in de context van elektrostimulatie van de hersenen. Ned Tijdschr Geneeskd 148:1373–75 [Competence in the context of DBS]
4. Blanke O, Aspell JE (2009) Brain technologies raise unprecedented ethical challenges. Nature, l458, 703
5. Blume S (1999) Histories of cochlear implantation. Soc Sci Med 49:1257–68
6. van den Burg W (2003) Dynamic Ethics. J Value Inq 37:13–34
7. Clausen J (2008) Moving minds: ehical aspects of neural motor prostheses. Biotechnol J 3:1493–1501
8. Clausen J (2009) Man, machine and in between. Nature 457:1080–1081
9. EGE—European Group on Ethics in Science and New Technologies (2005) Ethical aspects of ICT implants in the human body. opinion no. 20. Retrieved from http://ec.europa.eu/european_group_ethics/docs/avis20_en.pdf
10. Ford PJ (2007) Neurosurgical implants: clinical protocol considerations. Camb Q Healthc Ethics 16:308–311
11. Ford P, Kubu C (2007) Ameliorating or exacerbating: surgical 'prosthesis' in addiction. Am J Bioeth 7:29–32
12. Foster KR (2006) Engineering the brain. In: Illes J (ed) Neuroethics. Defining issues in theory, practice and policy. Oxford University, Oxford, pp 185–200
13. Gillet G (2006) Cyborgs and moral identity. J Med Ethics 32:79–83
14. Glannon W (2007) Bioethics and the brain. Oxford University, Oxford
15. Glannon W (2009) Our brains are not us. Bioethics 23:321–329
16. Glannon W (2009) Stimulating brains, altering minds. J Med Ethics 35:289–292
17. Graham-Rowe D (2006) Catching seizures before they occur. Retrieved September 15, 2009, from http://www.technologyreview.com/biotech/17124/
18. Gray CH (2001) Cyborg ctizen: politics in the posthuman age. Routledge, New York
19. Greene J, Cohen J (2004) For the law, neuroscience changes nothing and everything. Phil Trans Roy Soc Lond B 359:1775–1785
20. Hansson SO (2005) Implant ethics. J Med Ethics 31:519–525
21. Haraway D (1991) A cyborg manifesto: science, technology, and socialist-feminism in the late twentieth century. In: Haraway D (ed) Simians, cyborgs and women: the reinvention of nature. Routledge, New York, pp 149–181
22. Healy D (2000) The creation of psychopharmacology. Harvard University, Cambridge
23. Hochberg L, Taylor D (2007) Intuitive prosthetic limb control. Lancet 369:345–346
24. Hochberg L et al (2006) Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature 442:164–171
25. Hughes J (2004) Citizen cyborg: why democratic societies must respond to the redesigned human of the future. Westview, Cambridge
26. Keiper A (2006) The age of neuroelectronics. The New Atlantis Winter 2006:4–41
27. Keulartz JM, Korthals MS, Swierstra T (eds) (2002) Pragmatist ethics for a technological culture. Kluwer Academic, Dordrecht
28. Keulartz J, Schermer M, Korthals M, Swierstra T (2004) Ethics in technological culture: a programmatic proposal for a pragmatist approach. Sci Technol human values 29(1):3–29
29. Koops B, van Schooten H, Prinsen B (2004) Recht naar binnen kijken: een toekomstverkenning van huisrecht, lichamelijke integriteit en nieuwe opsporingstechnieken, eJure, ITER series 70. Retrieved from http://www.ejure.nl/mode=display/downloads/dossier_id=296/id=301/Deel_70_Koops.pdf
30. Lamme V (2006) De geest uit de fles. Dies rede. University of Amsterdam, Amsterdam
31. Levy N (2007) Neuroethics. Cambridge University, Cambridge
32. Leentjes AFG et al (2004) Manipuleerbare wilsbekwaamheid: een ethisch probleem bij elektrostimulatie van de nucleus subthalamicus voor een ernstige ziekte van Parkinson. Ned Tijdschr Geneeskd 148:1394–1397 [Manipulable competence]
33. Lunshof JE, Chadwick R, Vorhaus DB, Church GB (2008) From genetic privacy to open consent. Nat Rev Genet 9:406–411
34. Mooij A (2004) Toerekeningsvatbaarheid. Over handelingsvrijheid. Boom, Amsterdam [Accountability. On the freedom to act]
35. Moreno JD (2004) DARPA on your mind. Cerebrum 6:91–99

36. Morse S (2007) Voluntary control of behavior and responsibility. Am J Bioeth 7:12–13
37. Morse SJ (2008) Determinism and the death of folk psychology: two challenges to responsibility from neuroscience. Minnesota Journal of Law, Science and Technology 9:1–35
38. Rabins P et al (2009) Scientific and ethical issues related to deep brain stimulation for disorders of mood, behavior and thought. Arch Gen Psychiatry 66:931–37
39. Roco MC, Bainbridge WS (2002) Converging technologies for improving human performance. National Science Foundation, Arlington
40. Rose N (2003) Neurochemical selves. Society 41:46–59
41. Roskies A (2008) Neuroimaging and inferential distance. Neuroethics 1:19–30
42. Schermer M (2007) Gedraag Je! Ethische aspecten van gedragsbeïnvloeding door nieuwe technologie in de gezondheidszorg. NVBE, Rotterdam [Behave yourself! Ethical aspects of behaviour-influencing new technologies]
43. Smart JJC (2003) Free will, praise and blame. In: Watson G (ed) Free will, 2nd edn. Oxford University, New York, pp 58–71
44. Smits M (2006) Taming monsters: the cultural domestication of new technology. Tech Soc 28:489–504
45. Strawson P (2003) Freedom and resentment. In: Watson G (ed) Free will, 2nd edn. Oxford University, New York, pp 72–93
46. Synofzik M, Schlaepfer TE (2008) Stimulating personality: ethical criteria for deep brain stimulation in psychiatric patients for enhancement purposes. Biotechnol J 3:1511–1520
47. Vargas M (2005) The revisionist's guide to responsibility. Philos Stud 125:399–429
48. Warwick K (2004) I, cyborg. Century, London
49. Wegner DM (2002) The illusion of conscious will. MIT, Cambridge
50. Weinberger S (2007) Pentagon to merge the next-gen binoculars with soldiers' brain. Retrieved September 15 2009 from http://www.wired.com/gadgets/miscellaneous/news/2007/05/binoculars