KEVIN WARWICK

The disappearing human–machine divide

In this article a look is taken at some of the different ways in which the human–machine divide is rapidly disappearing. In each case the technical basis is described and then some of the implications are also considered. In particular results from experiments are discussed in terms of their meaning and application possibilities. The article is written from the perspective of scientific experimentation opening up realistic possibilities to be faced in the future, rather than giving conclusive comments. In each case consideration is also given to some of the philosophical questions that arise.

Introduction

A variety of different practical experiments where the divide between humans and machines is disappearing are investigated in this article. We consider here four areas in total. Firstly we look at the concept of growing a biological brain and placing it in a robot body. Secondly we consider the use of implant technology to link the human brain directly with computers. Thirdly we investigate the use of deep brain stimulation for therapeutic purposes and finally the Turing imitation game is seen as a present-day test of whether a machine can communicate like a human.

The article is arranged such that experiments are described in turn in individual sections. Whilst there are distinct overlaps between the sections, they each throw up individual considerations. Following a description of each investigation some pertinent issues on the topic are therefore discussed. Points have been raised with a view to near-term future technical advances and what these might mean in a practical scenario. It is not the case here of an attempt to present a fully packaged conclusive document, rather the aim has been to open up the range of research being carried out, see what is actually involved and look at some of its implications. In each case, following the technical description, some of the philosophical spin-offs are also discussed.

Biological brains in a robot body

Initially when one thinks of linking a brain with technology, then it is probably in terms of a brain already functioning and settled within its own body. Here however we consider the possibility of a fresh merger where a brain is firstly grown and then given its own body in which to operate.

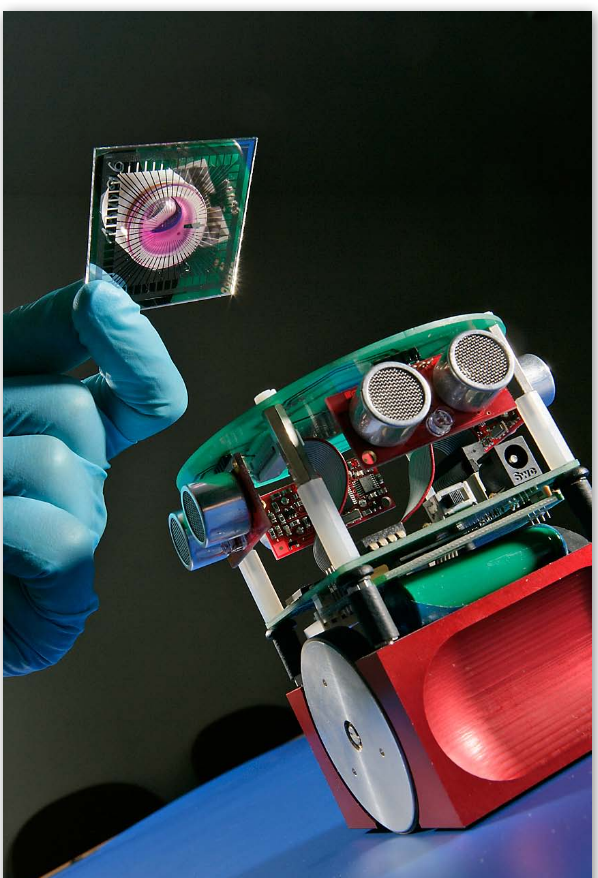
When one considers a robot it may be a little wheeled device that springs to mind, or perhaps a metallic head that looks roughly human-like. Whatever the physical appearance our thoughts tend to be that the robot might be operated remotely by a human, as in the case of a bomb-disposal robot, or it may be controlled by a simple computer programme, or it may even be able to learn, with a microprocessor as its technological brain. In all these cases we regard the robot simply as a machine. But what if the robot had a biological brain made up of brain cells (neurons), possibly even human neurons?

Neurons cultured/grown under laboratory conditions on an array of non-invasive electrodes provide an attractive alternative with which to realise a robot controller. An experimental control platform – essentially a robot body – can move around in a defined area purely under the control of such a network/brain and the effects of the brain, controlling the body, can be witnessed. Whilst this is extremely interesting from a robotics perspective it also opens up a different approach to the study of the development of the brain itself because of its sensory-motor embodiment. Investigations can in this way be carried out into memory formation and reward/

punishment scenarios – the elements that underpin the basic functioning of a brain.

The process of growing networks of brain cells (typically around 150,000 at present) *in vitro* commences, in most cases, by separating neurons obtained from foetal rodent cortical tissue. They are then grown (cultured) in a specialised chamber, in which they can be provided with suitable environmental conditions (e.g. appropriate temperatures) and fed with a mixture of minerals and nutrients. An array of electrodes embedded in the base of the chamber (a multi-electrode array, MEA) acts as a bi-directional electrical interface to/from the culture. This enables electrical signals to be supplied to stimulate the culture and also for recordings to be taken as outputs from the culture. The neurons in such cultures spontaneously connect, communicate and develop, within a few weeks giving useful responses, typically for three months at present. To all intents and purposes it is rather like a brain in a jar!

In fact the brain is grown in a glass specimen chamber lined with a flat '8x8' multi-electrode array which can be used for real-time recordings (see Figure 1). It is, in this way, possible to separate the firings of small groups of neurons, by monitoring the output signals on the electrodes. Thereby a picture of the global activity of the entire network can be formed. It is also possible to electrically stimulate the culture via any of the electrodes to induce neural activity. The multi-electrode array therefore forms a bi-directional interface with the cultured neurons (Chiappalone *et al.* 2007, DeMarse *et al.* 2001).



Robot with cultured brain.

The brain can then be coupled to its physical robot body (Warwick *et al.* 2010). Sensory data fed back from the robot is subsequently delivered to the culture, thereby closing the robot-culture loop. Thus, the processing of signals can be broken down into two discrete sections (a) 'culture to robot', in which live neuronal activity is used as the decision-making mechanism for robot control, and (b) 'robot to culture', which involves an input-mapping process, from the robot sensor to stimulate the culture.

The actual number of neurons in a brain depends on natural density variations in seeding the culture in the first place. The electrochemical activity of the culture is sampled and this is used as input to the robot's wheels. Meanwhile the robot's (ultrasonic) sensor readings are converted into stimulation signals received by the culture, thereby closing the loop.

Once the brain has grown for several days, which involves the formation of some elementary neural connections, an existing neuronal pathway through the culture is identified by searching for strong relationships between pairs of electrodes (Xydas *et al.* 2008). Such pairs are defined as those electrode combinations in which neurons close to one electrode respond to stimulation from the other electrode at which the stimulus was applied more than 60 per cent of the time and respond no more than 20 per cent of the time to stimulation on any other electrode.

A rough input–output response map of the culture can therefore be created by cycling through all of the electrodes in turn. In this way, a suitable input–output electrode pair can be chosen in order to provide an initial decision-making pathway for the robot. This is then employed to control the robot body – for example if the ultrasonic sensor is active and we wish the response to cause the robot to turn away from the object being located ultrasonically (possibly a wall) in order to keep moving.

For experimentation purposes at this time, the intention is for the robot to follow an onward path until it reaches a wall, at which point the front sonar value decreases below a threshold, triggering a stimulating pulse. If the responding/output electrode registers activity the robot turns to avoid the wall. In experiments the robot turns spontaneously whenever activity is registered on the response electrode. The most relevant result is the occurrence of the chain of events: wall detection–stimulation–response. From a neurological perspective it is of course also interesting to speculate why there is activity on the response electrode when no stimulating pulse has been applied.

As an overall control element for direction and wall avoidance the cultured brain acts as the sole decision-making entity within the feedback loop. Clearly one important aspect then involves neural pathway changes, with respect to time, in the culture between the stimulating and recording electrodes.

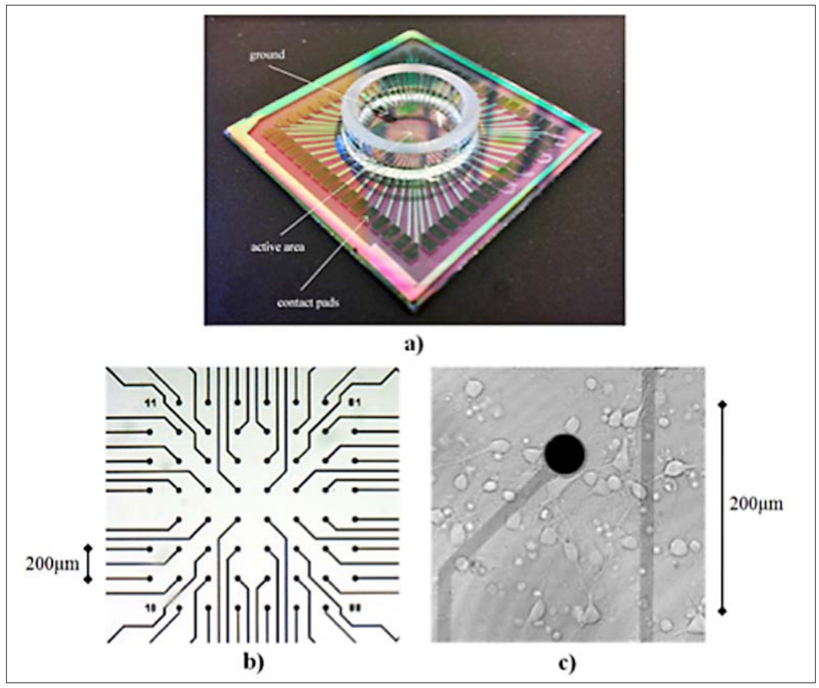


Figure 1. a) A multi-electrode array (MEA) showing the electrodes; b) electrodes in the centre of the MEA seen under an optical microscope; c) an MEA at x40 magnification, showing neuronal cells in close proximity to an electrode.

In terms of research, learning and memory investigations are generally at an early stage. However the robot can clearly be seen to improve its performance over time in terms of its wall-avoidance ability in the sense that neuronal pathways that bring about a satisfactory action tend to strengthen purely though the process of being habitually performed – learning due to habit (Hebb 1949).

The number of variables involved is considerable however and the plasticity process, which occurs over quite a period of time, is (most likely) dependent on such factors as initial seeding and growth near electrodes as well as environmental transients such as temperature and humidity (Downes *et al.* 2012). Learning by reinforcement – rewarding good actions and punishing bad is more in terms of investigative research at this time.

On many occasions the culture responds as expected, on other occasions it does not, and in some cases it provides a motor signal when it is not expected to do so. But does it 'intentionally' make a different decision to the one we would have expected? We cannot tell but merely guess.

In terms of robotics, it has been shown by this research that a robot can successfully have a biological brain with which to make its 'decisions'. The 100,000–150,000 neuron size is merely due to the present day limitations of the experimentation described. Indeed three-dimensional structures are already being investigated. Increasing the complexity from two dimensions to three dimensions realises a figure of approximately 30 million neurons for the three-dimensional case – not yet reaching the 100 billion neurons of a 'perfect' human brain, but nevertheless in tune with the brain size of many other animals.

Robots with biological brains: some issues

This area of research is expanding rapidly. Not only is the number of cultured neurons increasing, but the range of sensory inputs is being expanded to include audio, infra red and even visual. Such richness of stimulation will no doubt have a dramatic effect on culture development. The potential of such systems, including the range of tasks they can deal with, also means that the physical body can take on different forms. There is no reason, for example, that the body could not be a two-legged, walking robot, with rotating head and the ability to walk around a building.

It is certainly the case that understanding neural activity becomes more difficult as the culture size increases. With a three-dimensional structure, monitoring activity deep within the central area, as with a human brain, becomes extremely complex, even with needle-like electrodes. In fact the present 150,000 neuron cultures are already far too complex at present to gain an overall insight. When they are grown to sizes such as 30 million neurons and beyond, clearly the problem is significantly magnified.

Looking a few years out, it seems quite realistic to assume that such cultures will become larger, potentially growing to the magnitude of billions of neurons. On top of this, the nature of the neurons utilised may be diversified. At present rat neurons are generally employed in studies. However human neurons are also being cultured even now, thereby bringing about a robot with a human neuron brain. If this brain then consists of billions of neurons, many social and ethical questions will need to be asked (Warwick 2010).

For example, if the robot brain has roughly the same number of human neurons as a typical human brain then could/should it have similar rights to humans? Also, what if such creatures have far more human neurons than in a typical human brain – e.g. a million times more – would they make all future decisions, rather than regular humans? Certainly it means that as we look to the near future we could shortly witness thinking robots with brains not too dissimilar to those of humans.

A key question in this regard is whether or not such a brain could possibly be conscious. Clearly this is an enormous question and one too deep to be anywhere near fully explored here. However philosophical arguments with regard to the consciousness of machines have invariably been made in the past (primarily concerning human consciousness) in terms of a comparison between the workings of a human brain and a silicon computer (e.g. Penrose 1995). A major argument in such circumstances has been the difference between individual human neurons and any technological copy that could be realised. As a result some have concluded (e.g. Searle 1990) that consciousness is an emergent property – essentially, put enough human neurons together, with high connectivity, and consciousness emerges. In terms of this argument there is no reason whatsoever therefore that robots with biological brains composed of sufficient numbers of human neurons could not be conscious.

If one wishes to consider that such robots cannot possibly be conscious then a scientific approach is required (Warwick 2010). In any such analysis one must consider not only the robots in question but also the different humans that exist. For example if it is felt that the robot's body is an issue, perhaps because they have wheels and not legs then it is a consequence that humans who have no legs and move around on wheels must be considered in the same light.

General-purpose brain implants

Many human brain–computer interfaces are used for therapeutic purposes, in order to overcome a medical/neurological problem, one example being, as will be discussed shortly, deep brain stimulation electrodes employed to overcome the effects of Parkinson's disease (Pinter *et al.* 1999, Pan *et al.* 2012, Wu *et al.* 2010a), or the use of external electrodes to understand the functioning of parts of the brain (Daly *et al.* 2012). However it is possible to consider employing such technology in alternative ways to give individuals abilities not normally possessed by humans – human enhancement!

With more general brain–computer interfaces the therapy-enhancement situation is quite complex. In some cases it is possible for those who have suffered an amputation or have received a spinal injury due to an accident to regain control of devices via their (still functioning) neural signals (Donoghue *et al.* 2004). Meanwhile stroke patients can be given limited control of their surroundings, as indeed can those who have conditions such as motor neurone disease.

With these cases the situation is not straightforward, as each individual is given abilities that no normal human has – for example the ability to move a cursor around on a computer screen from neural signals alone (Kennedy *et al.* 2004). The same quandary exists for blind individuals who are allowed ex-

Robot hand controlled by (author's) neural signs.

tra-sensory input, such as sonar (a bat-like sense) – it doesn't repair their blindness, but rather allows them to make use of an alternative sense.

Some of the most impressive human research to date has been carried out using the micro-electrode array, shown in Figure 2. The individual electrodes are 1.5 mm long and taper to a tip diameter of less than 90 microns. Although a number of trials not using humans as a test subject have occurred, human tests are at present limited to two groups of studies. In the second of these the array has been employed in a recording-only role, most notably recently as part of (what was called) the 'BrainGate' system.

Essentially electrical activity from a few neurons monitored by the array electrodes was decoded into a signal to direct cursor movement. This enabled an individual to position a cursor on a computer screen, using neural signals for control combined with visual feedback. The same technique was later employed to allow the individual recipient, who was paralysed, to operate a robot arm (Hochberg *et al.* 2006). The first use of the micro-electrode array (shown in Figure 2) has considerably broader implications though, which extend the capabilities of the human recipient.

Deriving a reliable command signal from a collection of monitored neural signals is not necessarily a simple task, partly due to the complexity of signals recorded and partly due to the real-time constraints involved in dealing with the data. In some cases however it can be relatively easy to look for and obtain a system response to certain anticipated neural signals – especially when an individual has trained extensively with the system. In fact neural signal shape, magnitude and waveform with respect to time are considerably different to other apparent signals, such as noise, and this makes the problem a little easier.

The interface through which a user interacts with technology provides a layer of separation between what the user wants the machine to do and what the machine actually does. This separation imposes a cognitive load on the individual concerned that is proportional to the difficulties experienced. The main issue is interfacing the human motor and sensory channels with the technology in a reliable, durable, effective, bi-directional way. One solution is to avoid this sensorimotor bottleneck altogether by interfacing directly with the human nervous system.

An individual human so connected can potentially benefit from some of the advantages of machine/artificial intelligence, acquiring for example rapid and highly accurate mathematical abilities in terms of 'number crunching'; a high-speed, almost infinite, internet knowledge base, and accurate long-term memory. Additionally, it is widely acknowledged that humans have only five senses that we know of, whereas machines offer a view of the world which includes infra-red, ultraviolet and ultrasonic signals, to name but a few.

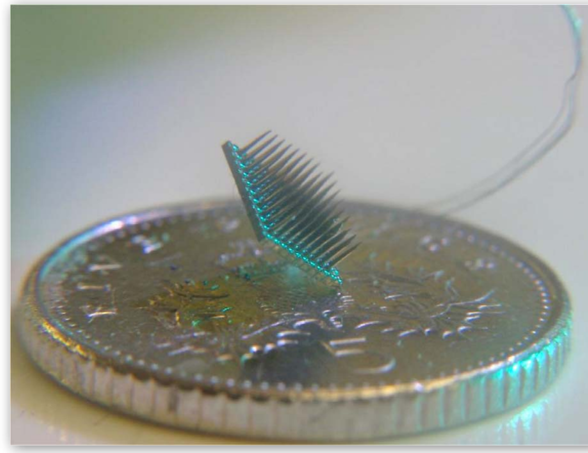


Figure 2. A 100 electrode, 4x4 mm micro-electrode array, shown on a UK 1 pence piece for scale.

Humans are however also limited in that they can only visualise and understand the world around them in terms of a limited three-dimensional perception, whereas computers are quite capable of dealing with hundreds of dimensions. Perhaps most importantly, the human means of communication, essentially transferring a complex electro-chemical signal from one brain to another via an intermediate, often mechanical, slow and error-prone medium (e.g. speech), is extremely poor, particularly in terms of speed, power and precision. It is clear that connecting a human brain, by means of an implant, with a computer network could in the long term open up the distinct advantages of machine intelligence, communication and sensing abilities to the implanted individual.

As a step towards a broader concept of brain–computer interaction, the micro-electrode array (as shown in Figure 2) was implanted into the median nerve fibres of a healthy human individual (the author) during two hours of neurosurgery in order to test bi-directional functionality in a series of experiments. Stimulation current applied directly into the nervous system allowed information to be sent to the user, while control signals were decoded from neural activity in the region of the electrodes (Warwick *et al.* 2003). In this way a number of trials were successfully concluded (Warwick *et al.* 2004). In particular:

1. Extra-sensory (ultrasonic) input was successfully implemented.
2. Extended control of a robotic hand across the internet was achieved, with feedback from the robotic fingertips being sent back as neural stimulation to give a sense of force being applied to an object (this was achieved between Columbia University, New York, USA, and Reading University, England).
3. A primitive form of telegraphic communication directly between the nervous systems of two humans (the author's wife assisted) was performed (Warwick *et al.* 2004).
4. A wheelchair was successfully driven around by means of neural signals.
5. The colour of jewellery was changed as a result of neural signals – also the behaviour of a collection of small robots.

In most, if not all, of the above cases it could be considered that the trial proved useful for purely therapeutic reasons, e.g. the ultrasonic sense could be useful for an individual who is blind or the telegraphic communication could be very useful for those with certain forms of motor neurone disease.

Each trial can however also be seen as a potential form of enhancement beyond the human norm for an individual. Indeed the author did not need to have the implant for medical purposes to overcome a problem, but rather the experimentation was performed purely for the purposes of scientific exploration. The question arises therefore as to how far should things be taken? Clearly enhancement by means of brain–computer interfaces opens up all sorts of new technological and intellectual opportunities; however it also throws up a raft of different ethical considerations that need to be addressed directly.

Brain implants: issues arising

When ongoing experiments of the type described involve healthy individuals where there is no reparative element in the use of a brain–computer interface, but rather the main purpose of the implant is to enhance an individual's abilities, it is difficult to regard the operation as being for therapeutic purposes. Indeed the author, in carrying out such experimentation, specifically wished to investigate actual, practical enhancement possibilities (Warwick *et al.* 2003, Warwick *et al.* 2004).

From the trials it is clear that extra-sensory input is one practical possibility that has been successfully trialled; however, improving memory, thinking in many dimensions and communication by thought alone are other distinct potential, yet realistic, benefits, with the latter of these also having been investigated to an extent. To be clear: all these things appear to be possible (from a technical viewpoint at least) for humans in general.

As we look to the future it is quite possible that commercial influences coupled with the societal wish to communicate more effectively and perceive the world in a richer form will drive a market desire. Ultimately, direct brain-to-brain communication, possibly using implants of the type described, is a tremendously exciting proposition, possibly resulting in thoughts, emotions, feelings, colours and basic ideas being transmitted directly from brain to brain. Whilst this raises many questions as to how it would work in practice, clearly we would be foolish not to push ahead technically to achieve it.

But then we come to the big questions. As communication is an extremely important part of human intelligence, it follows that for anyone who has an implant of this type it will necessarily provide a considerable boost to their intelligence. Clearly this will stretch intellectual performance in society with the implanted section outperforming those who have elected to stay as mere (unchipped) humans. Will it, though, bring about a digital divide, an us-and-them situation, leaving regular humans lagging behind on the evolutionary ladder?

Indeed there is a considerable possibility here of the emergence of a post-human or rather superhuman being (Nietzsche 2006). In simple terms, if I have such an implant that enables me to communicate in a much richer way with others who have a similar implant; if I can also sense the world in a richer way and interact much more intimately with computer networks, how am I likely to treat those who do not have such implants and who still communicate in that old-fashioned, outmoded way using those error-prone mechanical pressure waves referred to as speech? Probably I will not regard them very highly, even dismissing their opinions as trivial (Warwick 2003).

Other potential uses of this technology, even now, involve its potential use in the military domain and space travel (Warwick 2007). Clearly here an individual could be given the ability to control technology remotely as a body part. Essentially now an individual's brain and their body do not have to be in the same place. If that distant body part then gets blown up in battle it will presumably be of little concern to the individual. This technology will be a game changer: on the one hand the individual (their brain anyway) will be safe and sound back at home, but on the other hand their sense of responsibility will be much reduced, as they are constantly remote from the battlefield scenario.

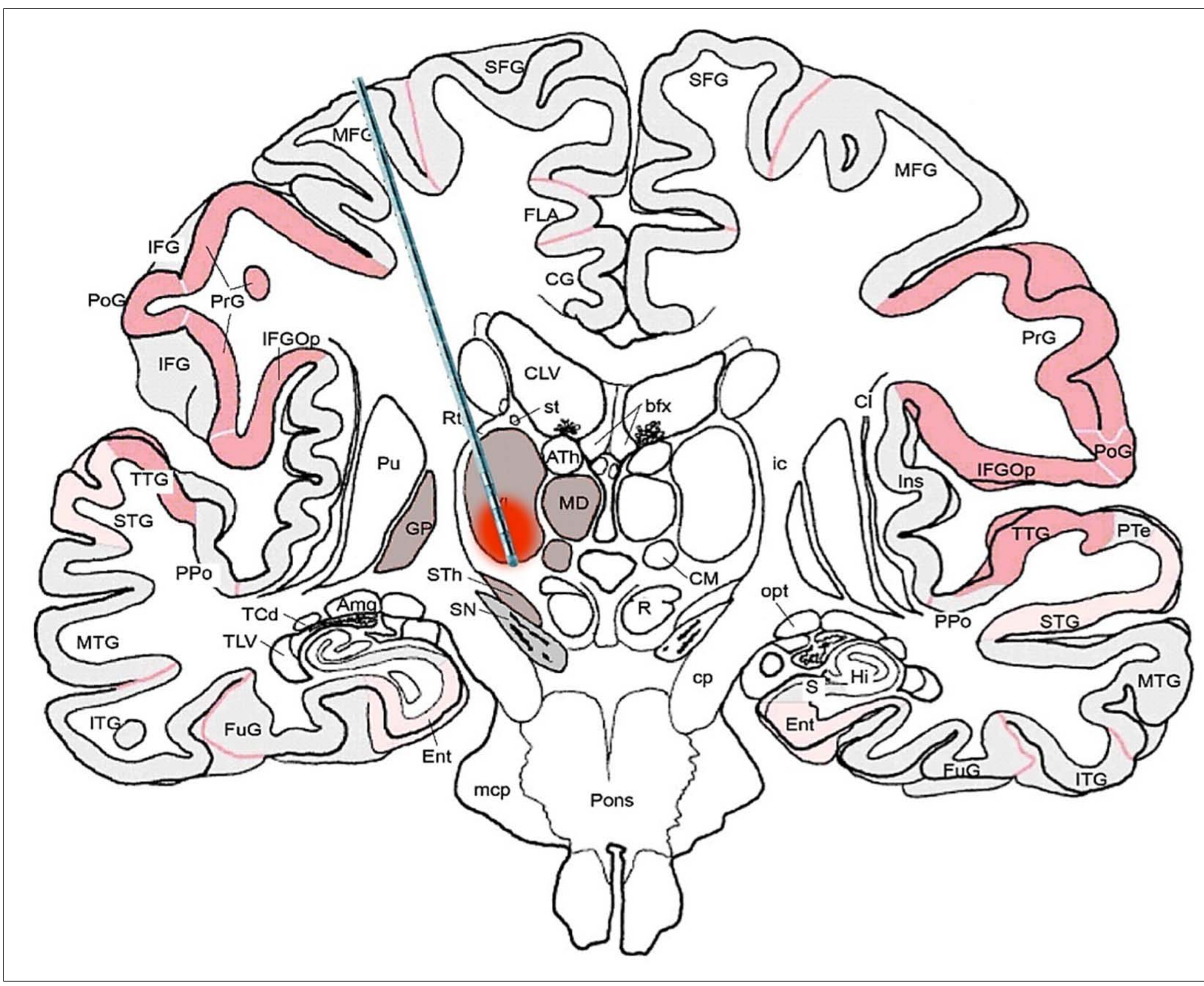
Deep brain stimulation

Probably the most widely spread examples of this in everyday life are humans who have had some technology implanted for therapeutic purposes. These include cochlear implants, heart pacemakers and even artificial hips. In each case it would appear that perhaps some philosophical issues arise, but arguably they are not of significant weight. To put it simply, an individual has a problem and technology is employed in an attempt to restore the person back towards a human norm. However in the case of deep brain stimulation the situation is not so simple, particularly as we look to the future.

The number of Parkinson's disease (PD) patients is estimated to be 120–180 out of every 100,000 people, although the percentage is increasing rapidly as life expectancies increase. For decades researchers have exerted considerable effort to understand more about the disease and to find methods successfully to limit its symptoms, which are, most commonly, periodic (and frequently acute) muscle tremor and/or rigidity. Many other symptoms, such as stooping, may however occur in the later stages of PD.

Several approaches exist to treat this disease. In its early stages, the drug levodopa (L-dopa), since 1970, has been the most common. However, it is found that the effectiveness of L-dopa decreases as the disease worsens, whilst the severity of the side effects increases, something that is far more apparent when PD is contracted by a younger person.

Surgical treatment, such as lesioning, is an alternative when drug treatments have become ineffective. Lesioning can alleviate symptoms, thus reducing the need for drug therapy altogether. An alternative



Positioning of deep brain electrodes.

treatment of PD by means of deep brain stimulation (DBS) only became possible when the relevant electrode technology became available from the late 1980s onwards. From then on, many neurosurgeons have moved to implanting neurostimulators connected to deep brain electrodes positioned in the thalamus, sub-thalamus or globus pallidus for the treatment of tremor, dystonia and pain.

A typical deep brain stimulation device contains an electrode lead with four or six cylindrical electrodes at equally spaced depths attached to an implanted pulse generator (IPG), which is surgically positioned below the collarbone. DBS treatment has many advantages, including being a reversible operation. It is also potentially much less dangerous than lesioning and is, in many cases, highly effective. However, it presently utilises a continuous current simulation at high frequency, resulting in the need for a regular battery replacement every 24 months or so. The cost of the battery replacement, the time-consuming surgery involved and the trauma of the repetitive surgery involved severely limit the number of patients who can benefit and excludes particularly those who are frail, or have problems with their immune system or who are not particularly wealthy.

The obvious solution, namely remote inductive battery recharging, is fraught with problems such as the size of the passive coil size that needs to be implanted and the nasty chemical discharges that occur within the body – and even then the average time between replacements is only marginally improved. Another solution to prolong the battery life is simply to improve battery technology. However, the link between the price of the battery and battery life is stark. If we are considering a battery that could potentially supply the stimulation currents required over a ten- or twenty-year period then the technology to achieve this in a low-cost, implantable, durable form is not on the horizon.

However ongoing research is aimed at developing an 'intelligent' stimulator (Pan *et al.* 2012, Wu *et al.* 2010a, Wu *et al.* 2010b). The idea of the stimulator is to produce warning signals before the tremor starts, so that the stimulator only needs to generate signals occasionally instead of continuously – in this sense operating in a similar fashion to a heart pacemaker.

Artificial intelligence (AI) tools, based on artificial neural networks, have been shown to successfully provide tremor onset prediction. In either case, data input to a network is provided by the measured electrical local field potentials, obtained by means of the deep brain electrodes. The network is primed to recognise the nature of electrical activity deep in the human brain and to predict (several seconds ahead) the subsequent tremor onset outcome. In this way the DBS device becomes 'intelligent' when the stimulation is only triggered by the AI system.

Many issues exist with the AI system however as much pre-processing of the brain data is necessary, along with frequency filtering to minimize the difficulty of prediction. Comparative studies are now ongoing to ascertain which AI method appears to be the most reliable and accurate in a practical situation.

It is worth pointing out here that false positive predictions (that is the AI system indicating that a tremor is going to occur when in fact this is not the case) are not so much of a critical problem. The end result with such a false positive is that the stimulating current may be applied when it is not strictly necessary. In any event no actual tremor would occur, which is a good outcome for the patient in any case; but unnecessary energy would have been used – in fact if numerous false predictions occurred the intelligent stimulator would tend to operate in the same way as the present 'blind' stimulator. The good news is that results show that the network can be readily tuned to avoid the occurrence of most false positives anyway.

Missing the prediction of a tremor onset is, though, extremely critical and is simply not acceptable. Such an event would mean that the stimulating current would not come into effect and the patient would actually suffer from tremors occurring.

Whilst deep brain implants are, as described, aimed primarily to provide current stimulation for therapeutic purposes, they can also have a broader portfolio in terms of the effects they can have within the human brain. It is worth stressing however that in all cases further implantations are at this time forging ahead with little or no consideration being given to the general technical, biological and ethical issues that pervade.

Questions from deep brain stimulation

The same physical stimulator that is used for the treatment of Parkinson's disease is also employed, albeit in fewer instances at present, for cases of Tourette's syndrome, epilepsy and even clinical depression. In many people's eyes it is probable that the use of deep brain stimulators for the treatment of Parkinson disease, epilepsy or Tourette's syndrome is perfectly acceptable because of the improved quality of life it can effect for the individual recipient.

Otto-Ville Väätäinen

Professor Kevin Warwick challenging the Aboagora audience to rethink the human–machine divide.

However long-term modifications of brain organisation can occur in each case, causing the brain to operate in a completely different fashion; for example, there can be considerable long-term mental side effects in the use of such technology. The stimulators, when positioned in central areas of the brain, can cause other direct results, including distinct emotional changes. The picture is therefore not one of merely overcoming a medical problem – it is far more complex.

And what if things go wrong, even with present-day deep brain implants? For example, let us assume that the implanted wiring picks up a radio signal which causes a spurious signal to be realised within the brain and as a result the patient performs a murderous act. Who is to blame in such a circumstance? The patient? The surgeon? The manufacturer? The company broadcasting the radio signal? The person (possibly David Cameron!!) making the radio announcement?

However, as described here, 'intelligent' deep brain stimulators are starting to be designed (Pan *et al.* 2007). In such a case a computer (artificial brain) is used to understand the workings of specific aspects of the human brain. The job of the artificial brain, as can be seen from the description of the experimentation, is to monitor the normal functioning of the human brain such that it can accurately predict a spurious event, such as a Parkinson tremor, several seconds before it actually occurs. In other words the artificial brain's job is to out-think the human brain and to stop it doing what it 'normally' wants to do.

Clearly the potential for this system to be applied across a broad spectrum of different uses is enormous. Restricting the application of the technology to therapeutic purposes also limits the need for philosophical argument. Meanwhile, extending the use opens up many possibilities. To be already using the method to make an individual happy (to overcome depression) then opens up the possibility of recreational use. But perhaps the most significant employment of the technology would potentially be to overcome bad traits – not merely bad habits – along the lines suggested in *The Terminal Man* (Crichton 1972).

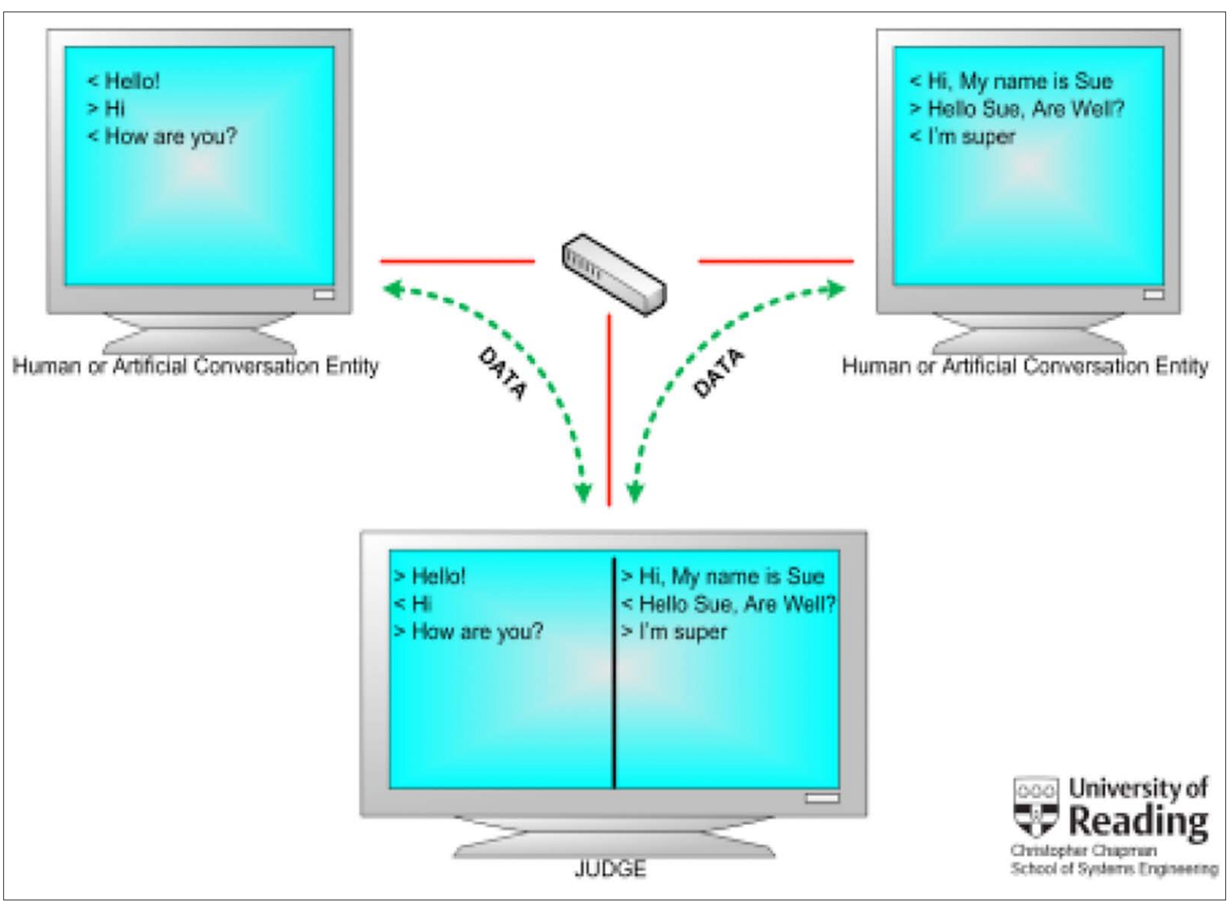
Turing's imitation game

The Turing test involves a human judge conversing at the same time with a hidden human and an unseen machine via a communications link, most likely a computer terminal. The machine pretends to be a human in terms of conversational abilities, whereas the human is just themselves; human. In a directly paired comparison the attempt is for the machine to appear to be more human than the human against whom it is paired. At the end of the conversation the judge has to decide with certainty which entity was the human and which was the machine, or if they feel both entities were either machine or human (Warwick 2012).

To conform to Turing's original wording in his 1950 paper (Turing 1950), reference is given here to 5-minute-long tests only. What are considered here are two specific transcripts taken from a day of actual, practical Turing tests (total around 200 tests) which were held under strictly-timed conditions, with many external viewers, at Bletchley Park, England on 23 June 2012. The date marked the 100th anniversary of Turing's birth and the venue was that at which, amongst other things, during the Second World War, Turing led a team of code breakers who cracked the German Enigma machine cypher (Hodges 1992). Five different machines took part in the tests during the day, along with 30 different judges and numerous hidden humans, against which the machines were compared in terms of their conversational abilities.

Turing's prediction was: 'I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have a more than 70 per cent chance of making the right identification after five minutes of questioning' (Turing 1950). The 'right identification' being to tell exactly which is the machine and which is the human. This is what has become known as the Turing test; to identify when we get to the stage that enough people cannot tell the difference between humans and machines in conversation.

At the Bletchley Park series of tests a number of exciting new results were realised. To consider one of them, please have a look at the transcript that follows. This was as a result of an actual conversation that occurred on that day between a judge and a hidden entity. Try to make your own mind up as to whether the judge was conversing with a human or a machine. What is shown is the 5-minute conversation in its



Turing's imitation game – interrogator's view.

entirety. The timing indicates the exact real time on that day. Both the judge's decision at the end of the conversation and the actual identity of the entity are subsequently given. This transcript has been taken from the article (Warwick *et al.* 2013) where further transcripts from that day are also given.

> [10:41:42] Judge: hey, just to lighten the mood a little, what's brown and sticky?
> [10:42:41] Entity: no idea, what's the answer?
> [10:42:58] Judge: a stick… so anyway what's your name?
> [10:43:16] Entity: Sheldon Cooper, what's yours?
> [10:43:32] Judge: as in the guy off the big bang theory?
> [10:43:59] Entity: I don't believe in the big bang
> [10:44:13] Judge: so are you religious then?
> [10:44:29] Entity: aren't we all?
> [10:45:06] Judge: well i like to consider myself an agnostic, which religion is it
> you follow then if you dont mind me asking?
> [10:45:44] Entity: you can ask what you like, it's a free world, who am I to mind?

At the end of this conversation the judge decided that they most definitely had been conversing with a machine. In fact the hidden entity was, in reality, the author of this article (most assuredly a human) who went on to take part in a further five such conversations with different judges. In not one case did any of the judges conclude that they had been conversing with a human. The other five transcripts are also given in Warwick *et al.* 2013.

Although this is not the first time that a human was misidentified in a single Turing test conversation as being a machine, it is the first time that a hidden human entity has not once been classified as being human by all of the judges involved. Although, as was indicated in Shah and Warwick 2010a it is frequently the case that when a judge makes an error in their decision they often do not realise that they have done so and subsequently do not wish to admit that they have made a mistake when their error is pointed out to them.

Now consider another transcript example. Again the idea is firstly to read through and try to make up your own mind as to the identity of the hidden entity, whether it be human or machine.

> [15:46:05] Judge: My favourite music is contemporary Jazz, what do you prefer
> [15:46:14] Entity: To be short I'll only say that I HATE Britnie Spears. All other music is OK compared to her.
> [15:47:06] Judge: do you like to play any musical instruments
> [15:47:23] Entity: I'm tone-deaf, but my guinea pig likes to squeal Beethoven's 'Ode to Joy' every morning. I suspect our neighbors want to cut his throat… Could you tell me about your job, by the way?
> [15:48:02] Judge: Guine pig/ are you an animal lover
> [15:48:08] Entity: Yeah. A nice little guinea pig. Not some annoying chatter bot.
> [15:48:50] Judge: do you live with anyone else?
> [15:48:59] Entity: Yes I do. But better ask something else. And I forgot to ask you where you are from…
> [15:49:39] Judge: Ireland? geen patures and too much rain! How about you

As can be seen from the timing, this conversation, again lasting for 5 minutes only, occurred on the afternoon of the same day. As with the previous conversation, any spelling mistakes or poor grammar is due to the communication, it is not due to poor editing of this article – this is how it actually occurred.

In this conversation, the judge was a male science professor and clinician, although he had no previous knowledge of acting as an interrogator for a Turing test. In this case the judge concluded that the hidden entity was a male, adult, native English speaker. In fact the entity was a machine, called Eugene Goostman, a teenage Ukrainian boy.

Unfortunately the judge was rather slow to get going here and hence this was a relatively short conversation. In fact the machine responded reasonably appropriately on topic to each point raised by the interrogator and successfully steered the conversation on each occasion, even throwing in a humorous comment.

The point with this example is that a machine can take on a complex persona which is difficult to uncover in such a conversation. Further, it is not an issue as to whether a hidden machine entity is right or wrong in any answer that they give, but rather if they give the sort of answer that a human would give. So asking mathematical questions is a waste of time and even factual reasoning doesn't usually get very far (Shah and Warwick 2010b).

The Turing test: some points

It is certainly the case that in some machine conversations it is easy for a judge to decide that they were conversing with a machine when that was indeed the case. The machine has probably failed to follow the conversation at some point, by picking up on a nuance or incorrect meaning. But now machines are getting very good at such conversation and it will not be long before they are indistinguishable from humans.

The importance of this issue is paramount when one considers on line chatbots that can interact on a regular basis on the World Wide Web. When one is in a conversation on Facebook, for example, can you be sure that the hidden entity is a human? Some people are very gullible to this, in Turing fashion deciding clearly that an entity is human and even getting an image in their mind of the entity's character, background and family. Once they have been tied in the person is hooked and can easily be subject to cyber crime. By studying conversations such as those shown, we get an insight into how to combat this type of cyber crime.

Conclusions

In this article a look has been taken at several different concepts. Experimental cases have been reported on in order to indicate how the human–machine divide is diminishing – thereby throwing up a plethora of social and ethical considerations. In each case reports on actual practical experimentation have been given, rather than merely some theoretical concept.

In particular when considering robots with biological brains, this could mean perhaps human brains operating in a robot body. Therefore, should such a robot be given rights of some kind? If one was switched off would this be deemed to be an instance of cruelty to robots? More importantly at this time – should such research forge ahead regardless? Before too long we may well have robots with brains made up of human neurons that have the same sort of capabilities as those of the human brain.

In the section on a general purpose, invasive brain implant, as well as implant employment for therapy a look was taken at the potential for human enhancement. Already extra-sensory input has been scientifically achieved, extending the nervous system over the internet and creating a basic form of thought communication. So it is likely that many humans will upgrade and become part machine themselves. This may mean that ordinary (non-implanted) humans are left behind as a result. If you could be enhanced, would you have any problem with it?

In the section which looked at deep brain stimulation, it may be at first considered that perhaps there are no issues apparent here because this is merely a therapeutic treatment. However the main problem here is what happens if something goes wrong. If it does then who is responsible? In the final section we could see a number of issues, which arise as a result of a study of the Turing test. Yes, machines are gradually getting better at human conversation. Yes, it will not be too long before we cannot tell the difference. But we can also pick up some signs of human fallibility here. Perhaps we are not as good as we think we are, both in terms of communication *per se* and in terms of our accurate assessment of the communication of others.

As well as taking a look at the procedures involved, the aim in this article was to have a look at some of the likely ethical and social issues as well. Some technological issues have though also been pondered on in order to open a window on the direction that developments are heading in. In each case however a firm footing has been planted on actual practical technology and on realistic future scenarios rather than on mere, speculative ideas. In a sense the overall idea is to open up a sense of reflection such that the further experimentation which we will now witness can be guided by the informed feedback that results. ■

Kevin Warwick is Professor of Cybernetics at the University of Reading, England where his research interests are artificial intelligence, robotics, biomedical engineering and control. He obtained his PhD and later a DSc from Imperial College London and has a further DrSc from the Czech Academy of Sciences, Prague. He has six honorary doctorates. He has published a number of books including *I, Cyborg* and *March of the Machines* (2002), his most recent being *Artificial Intelligence: The Basics* (2012). He has appeared in the *Guinness Book of Records* for his research on several occasions and is perhaps best known for his implant self-experimentation, linking his own nervous system with a computer network. Email: k.warwick(at)reading.ac.uk

Bibliography

Chiappalone, Michela, Alessandro Vato, Luca Berdondini, Milena Koudelka-Hep, and Sergio Martinoia 2007. 'Network dynamics and dynchronous activity in cultured cortical neurons', *International Journal of Neural Systems* 17, 87–103.

Crichton, Michael 1972. *The Terminal Man*, New York, Knopf.

Daly, Ian, Slawek Nasuto, and Kevin Warwick 2012. ‘Brain-computer interface control via functional connectivity dynamics’, *Pattern Recognition* 15(6), 2123–6.
DeMarse, Thomas, Danie Wagenaar, Axel Blau, and Steve Potter 2001. ‘The neurally controlled animat: biological brains acting with simulated bodies’, *Autonomous Robots* 11, 305–10.
Donoghue, John, Arto Nurmikko, Gerhard Friehs, and Michael Black 2004. ‘Development of neuromotor prostheses for humans’, Chapter 63 in *Advances in Clinical Neurophysiology*, 588–602, ed. M. Hallett, L. H. Phillips, D. L. Schomer and J. M. Massey, Supplements to Clinical Neurophysiology 57, Elsevier.
Downes, Julia, Mark Hammond, Dimitris Xydas, Matthew Spencer, Victor Becerra, Kevin Warwick, Ben Whalley, and Slawek Nasuto 2012. ‘Emergence of a Small-World Functional Network in cultured neurons’, *PLOS Computational Biology* 8(5), e1002522.
Hebb, Donald 1949. *The Organisation of Behavior*, New York, John Wiley & Sons.
Hochberg, Leigh, Mijail Serruya, Gerhard Friehs, Jon Mukand, Maryam Saleh, Abraham Caplan, Almut Branner, David Chen, Richard Penn, and John Donoghue 2006. ‘Neuronal ensemble control of prosthetic devices by a human with tetraplegia’, *Nature* 442, 164–71.
Hodges, Andrew 1992. *Alan Turing: the Enigma,* New York, Vintage Press.
Kennedy, Philip, Dinal Andreasen, Princewill Ehirim, Brandon King, Todd Kirby, Hui Mao, and Melody Moore 2004. ‘Using human extra-cortical local field potentials to control a switch’, *Journal of Neural Engineering* 1(2), 72–7.
Nietzsche, Friedrich 2006. *Thus Spoke Zarathustra*, Cambridge University Press.
Pan, Song, Kevin Warwick, Mark Gasson, John Burgess, Shou-Yan Wang, Tipu Aziz, and John Stein 2007. ‘Prediction of Parkinson’s disease tremor onset using artificial neural networks’ in *BIEN ’07 Proceedings of the fifth IASTED International Conference: Biomedical Engineering,* 241–5, Anaheim, CA, ACTA Press.
Pan, Song, Serdar Iplicki, Kevin Warwick, and Tipu Aziz 2012. Parkinson’s disease tremor classification: a comparison between support vector machines and neural networks’, *Expert Systems with Applications* 39(12), 10764–71.
Penrose, Roger 1995. *Shadows of the Mind*, Oxford University Press.
Pinter, Michaela, Monika Murg, Francois Alesch, Brigetta Freundl, Reinhard Helscher, and Heinrich Binder 1999. ‘Does deep brain stimulation of the nucleus ventralis intermedius affect postural control and locomotion in Parkinson’s disease?’, *Movement Disorders* 14(6), 958–63.
Searle, John 1990. *The Mystery of Consciousness*, New York Review of Books.
Shah, Huma, and Kevin Warwick 2010a. ‘Testing Turing’s five minutes, parallel-paired imitation game’, *Kybernetes* 39(3), 449–65.
—2010b. ‘Hidden interlocutor misidentification in practical Turing tests’, *Minds and Machines* 20, 441–54.
Turing, Alan 1950. ‘Computing machinery and intelligence’, *Mind* 59 (236), 433–60.
Warwick, Kevin 2003. ‘Cyborg morals, cyborg values, cyborg ethics’, *Ethics and Information Technology* 5(3), 31–137.
—2007. ‘The promise and threat of modern cybernetics’, *Southern Medical Journal* 100(1), 112–14.
—2010. ‘Implications and consequences of robots with biological brains’, *Ethics and Information Technology* 12(3), 223–34.
—2012. ‘Not another look at the Turing test!’ in *SOFSEM 2012*, 130–40, ed. M. Bielkova *et al*., Lecture Notes in Computer Science 7147, Berlin, Springer-Verlag.
Warwick, Kevin, Mark Gasson, Ben Hutt, Iain Goodhew, Peter Kyberd, Brian Andrews, Peter Teddy, and Amjad Shad 2003. ‘The application of implant technology for cybernetic systems’, *Archives of Neurology* 60(10), 1369–73.
Warwick, Kevin, Mark Gasson, Ben Hutt, Iain Goodhew, Peter Kyberd, Henning Schulzrinne, and Xiaotao Wu 2004. ‘Thought communication and control: a first step using radiotelegraphy’, *IEE Proceedings on Communications* 151(3), 185–9.
Warwick, Kevin, Slawek Nasuto, Victor Becerra, and Ben Whalley 2010. ‘Experiments with an in-vitro robot brain’ in *Computing with Instinct*, 1–15, ed. Y. Cai, Lecture Notes in Artificial Intelligence 5987, Berlin, Springer-Verlag.
Warwick, Kevin, Huma Shah, and James Moor 2013. ‘Some implications of a sample of practical Turing tests, minds and machines’, DOI: 10.1007/s11023-013-9301-y.
Wu, Defeng, Kevin Warwick, Zi Ma, Jonathan Burgess, Song Pan, and Tipu Aziz 2010a. ‘Prediction of Parkinson’s disease tremor onset using radial basis function neural networks’, *Expert Systems with Applications* 37(4): 2923–8.
Wu, Defeng, Kevin Warwick, Zi Ma, Mark Gasson, Jonathan Burgess, Song Pan, and Tipu Aziz 2010b. ‘Prediction of Parkinson’s disease tremor onset using radial basis function neural networks based on particle swarm optimization’, *International Journal of Neural Systems* 20(2): 109–16.
Xydas, Dimitris, Daniel Norcott, Kevin Warwick, Ben Whalley, Slawek Nasuto, Victor Becerra, Mark Hammond, Julia Downes, and Simon Marshall 2008. ‘Architecture for neuronal cell control of a mobile robot’ in *Proc. European Robotics Symposium*, 23–31, Prague, Springer-Verlag.