Ethics and Information Technology **5:** 131–137, 2003.

Cyborg morals, cyborg values, cyborg ethics

Kevin Warwick
Department of Cybernetics, University of Reading, Whiteknights, Reading RG6 6AY, UK

Abstract. The era of the Cyborg is now upon us. This has enormous implications on ethical values for both humans and cyborgs. In this paper the state of play is discussed. Routes to cyborgisation are introduced and different types of Cyborg are considered. The author's own self-experimentation projects are described as central to the theme taken. The presentation involves ethical aspects of cyborgisation both as it stands now and those which need to be investigated in the near future as the effects of increased technological power have a more dramatic influence. An important feature is the potential for cyborgs to act against, rather than for, the interests of humanity.

Introduction

From a cybernetics viewpoint, the boundaries between humans and machines become almost inconsequential. Self imposed human subject boundaries are seen as being nothing more than historical and philosophical dinosaurs that arise from evolved human mental states. The human and machine together become an integrated system, a Cyborg, part human part machine.

The question then arises as to what exactly is and what isn't a Cyborg. Some could regard a blind man with his cane (Bateson 1972) as a Cyborg, the cane feeding important information on the local environment, to the man. Meanwhile a hearing aid for a deaf person or even a pair of worn glasses could come into the same category. More recently some researchers in the field of wearable computers have become self-professed cyborgs (Pentland 1998).

We have witnessed many intrusions into the human body beautiful. Cochlea implants are now relatively common, as indeed are hip replacements; and heart pacemakers, whilst not being so prolific, continue a trend in which technology is readily accepted as being a necessary intrusion. But each of these, and the list is not conclusive; represent modifications intended to compensate for deficiencies (Hayles 1999). Even in these instances the establishment of conceptual limits and boundaries becomes a complex process.

The situation lands up on more difficult terrain when, rather than repairing the ineffective parts of a human body, technology is employed to enhance normal functioning. Many examples of this already exist, particularly in the military domain, such as infra red night sight incorporated into weapon sighting systems or voice controlled firing mechanisms introduced into the helmet of a fighter pilot. In her seminal work (Harraway 1985) "A Manifesto for Cyborgs: Science, Technology and Social Feminism in the 1980s," Donna Harraway discussed these issues as part of the cyborg's disruption of traditional categories. Clearly the Cyborg violates the human/machine distinction.

But should such entities, if indeed they are truly cyborgs, present an ethical problem? Surely they are no different to a spider using a web to catch a fly or a chimpanzee employing a stick with which to extract termites from a mound, which can be seen as vital functioning for those creatures. In each case, although the individual's physical capabilities take on a different form and their abilities are possibly enhanced, their inherent mental state, their consciousness, their perception, has not been altered other than to the extent of itself concluding what the individual might be capable of accomplishing.

Where the cyborgs represent a powerful ethical dilemma is in the case when an individual's consciousness is modified by the merging of human and machine. Essentially it is not so much the physical enhancements or repairs that should be our cause for concern but where the nature of an individual is changed by the linking of human and machine mental functioning. In the case of a human this means linking technology directly with the human brain or nervous system, rather than by a connection which is either external to the nervous system but internal to the body or even one which is external to both.

To be clear, the type of cyborg considered in this paper is one in which the cyborg is formed by a human, machine brain/nervous system coupling. Whilst this does refer to a relatively narrow definition with respect to all cyborg possibilities, much of the arguments that follow are dependant on such a definition.

Connections between technology and the human nervous system not only affect the nature of the individual, raising questions as to the meanings of ‘I’ and ‘self’ but they also directly influence autonomy. An individual human wearing a pair of glasses, whether they contain a computer or not, remains respectfully an autonomous being. Meanwhile a human whose nervous system is linked to a computer not only puts forward their individuality for serious questioning but also, when the computer is part of a network or at least connected to a network, allows their autonomy to be compromised.

It is this latter class of Cyborg that is the subject of this paper. The main question arising from this discourse being: when an individual’s consciousness is based on a part human part machine nervous system, in particular when they exhibit Cyborg consciousness, will they also hold to Cyborg morals, values and ethics? These being potentially distinctly different to human morals, values and ethics. Also, as a consequence, will cyborgs, acting as post humans, regard humans in a Nietschian like way (Nietsche 1961) rather akin to how humans presently regard cows or chimpanzees?

Some may prefer to look through Hollywood-style, philosophical pink glasses (Harraway 1985) and see post-human cyborgs as being “conducive to the long range survival of humans.” Surely it will be the cyborgs themselves that will make the ultimate pro-human, anti-human decisions. A missile heading towards an individual will not cease from its course and disappear, simply because that individual does not like the thought of missiles or does not exhibit the intelligence to comprehend them.

Intelligent machines

We now have machines that, many consider exhibit intelligence of their own (Warwick 1998, 2000). In most cases this is distinct from human intelligence and exhibits a number of characteristic properties when compared to human intelligence. In particular a number of positive features associated with the performance of machine intelligence can be picked out and highlighted.

The obvious immediate advantage is that many computers can carry out as many calculations in the blink of an eye (about one third of a second) as the typical human does in a week. Not only that but the computer will usually get all the results correct.

This processing speed can then be combined with phenomenal memory capabilities and retrieval. Even the most cynical amongst us agree that “on any issue of computing power, if computers do not have the advantage over human brains already, then they will certainly have it before too long” (Penrose 1994).

The way an intelligent machine perceives the world around it depends on how it senses the world. And here again machines have an enormous advantage over humans, with the potential ability to sense in such as the infrared, ultra violet, x-ray and ultrasonic spectra. All of these are simply not sensed by a stand alone human. Indeed, as a human we may well have, at some time, a super intelligent ultra violet being standing right next to us. However we wouldn’t know it because we wouldn’t sense the creature.

One problem with the human brain is that it has evolved to think about and understand the world around it, as being three dimensional, four dimensional if we include time as well as space. Humans simply cannot visualise more than 3 dimensions. This does not mean that space around us is three dimensional, but merely that that is how we perceive it because that is all our brains can cope with. Computers meanwhile are quite capable of dealing with hundreds of dimensions, and realising relationships involving these dimensions. In fact computers have the potential to understand the world as the ten or eleven dimensional entity that physicists now regard it as (White and Gribbin 1997).

The biggest advantage of all for machine intelligence is communication. In comparison with the capabilities of machines, human communication is so poor as to be embarrassing. Humans start with a complex set of electro-chemical signals in their brain and convert them to very slow, mechanical signals, in order to speak to someone else, nothing more and nothing less. When they eventually receive the signals the other person then converts the mechanical sound waves back to electro-chemical signals and tries to form some sort of understanding of what the original signals were all about. A high error rate, language, dialects and a limitation to serial transmission all hamper things considerably.

In comparison, machines can communicate round the world, with very little/no error, using standard languages, with millions of messages being successfully transmitted and received in parallel. Humans have developed communication along the human speech route very little, whereas technology has developed in leaps and bounds.

Meanwhile other aspects of machine intelligence chip away at the human bastions of intelligence, as was discussed at length in QI: The Quest for Intelligence (Warwick 2000) rather like castles in the sand with the tide coming in. One example in recent years was in 1998 when IBM’s chess playing computer, Deep Blue, beat the erstwhile chess champion amongst humans, Gary Kasparov (Warwick 1998). Each year a version of the Turing Test is carried out in which a panel of experts must decide, in discussion with a number of hidden terminals, whether a computer can fool the expert into thinking it is a human. In 2001 I was one of 5 experts at the Loebner Competition held in the Science Museum, London to try out the Turing idea. I was shocked when 2 of the 5 experts (not me) picked out one of the machines as being more human than either of the two humans who were acting as respondents. Clearly it will not be long before the Turing Test will be yet another castle in the sand, slipping ignominiously into the sea (Turing 1950; Sparrow 2001).

Overall though, from a human point of view, a number of distinct advantages could be accrued by becoming a Cyborg.

With a human brain linked to a computer brain, that individual could have the ability to:

- use the computer part for rapid maths
- call on an internet knowledge base, quickly
- have memories that they have not themselves had
- sense the world in a plethora of ways
- understand multi dimensionality
- communicate in parallel, by thought signals alone, i.e., brain to brain

All of the above would appear to be extremely valid reasons for an individual human to wish to become a Cyborg. But at what cost? What might the consequences be? What about the problems associated with actually becoming a cyborg? Most importantly, is this mere philosophical discussion or are we talking actual science?

Clearly the realisation of such cyborgs presents enormous questions that affect all aspects of human society and culture. Political and normative implications are very much part of this. In attempting to answer such questions a string of positive and negative potentials appear. Standing still is not an option. In the extremes, if humans, en masse, opted for a non-cyborg future, could the result be an intelligent machine superculture (Warwick 1998)? Conversely, if humans, en masse, opted for a cyborg future, could society and culture cope with such a distinct non linearity in evolution?

Animal and medical

Experiments not just to equip humans with cameras on their glasses or enhanced walking sticks or shoes, but rather to link computer and nervous system together have been ongoing for some time.

As an example in 1997 in a widely publicised project, a group at the University of Tokyo attached some of the motor neurons of a cockroach, to a microprocessor. Signals were then sent to the motor neurons to artificially propel the cockroach, despite what it might have wanted to do itself. Meanwhile in 1998, the antennae from a male silk moth were linked directly to a small-wheeled robot. When a female silk moth came near and gave off her pheromone signal as an attractor, it was received by the male antennae, and as a result the wheeled robot moved towards the female.

More recently, John Chapin taught a group of rats to each pull a lever and, as a result, obtain a tot of water as a treat. Electrodes were then connected into specific positions in the rat's brains and these emitted signals when the rats merely thought about pulling the lever, before they physically did anything. The signals from the electrodes were used to 'automatically' release the tot of water. The rats learned very quickly that they didn't need to actually pull the lever to get the water treat, merely to think about it was sufficient.

With humans, experimentation of this kind is not yet upon us, or is it? Two examples are worth mentioning. Firstly, Medtronics manufacture a stimulator implant that has now been successfully used to counteract, purely electronically, the tremor effects associated with Parkinson's Disease. Not only has this been employed with quite a number of humans, but the effects, when the stimulators are first switched on, can be dramatic. The stimulators appear to completely counteract the effects of the disease.

The second example is even more profound (Kennedy et al. 2000). At Emory University in Atlanta, Philip Kennedy has implanted two stroke victims, the second of these being Johnny Ray. Philip carried out an MRI scan to ascertain when Johnny thought about moving, which areas of his brain were active. Implants were then positioned in the main part of these areas in Johnny Ray's brain. Henceforth when Johnny thought about moving, signals associated with the movement were transmitted, by radio, to the computer.

The signals transmitted from Johnny's brain to the computer were used to cause the cursor on the computer screen to move left, right, up and down. In this way it was possible, by means of his thoughts, for Johnny to move the cursor around on the screen, thereby spelling out words and making requests. Very quickly Johnny learnt to communicate in this way by thought signals alone.

The 1998 experiment

In the fall of 1998 I had a silicon chip transponder surgically implanted in my upper left arm. I did not have a medical need, I just wanted to find out what it would be like. As I entered the Cybernetics Department at Reading University so a radio signal across the doorway energised the chip, causing it to transmit a unique identifying signal. The signal was received by the building's computer network, which could therefore identify me as the person who had just entered. As a result I was greeted with a loud "Hello Professor Warwick" as I passed through the foyer, and the foyer light switched on.

Elsewhere in the building, as I approached my laboratory, the network was able to track me and, as a result, opened the laboratory door automatically. My computer even switched on to my web page and informed me of my email count. A map of the building, updated by the computer, indicated my whereabouts at all times, and recorded when I had entered a particular room and how long I had been there. Clearly we were able to demonstrate a number of features immediately, whilst many more, e.g., using such technology to replace credit cards, can be left to the imagination. The potential, for an implant of this type, is considerable.

The implant was made up 50% by a coil of wire with which the radio signal in various doorways around the Cybernetics building could react. The radio signal caused a current to flow in the coil, by means of induction. In this way the transponder did not need its own power supply. As a result the implant was fairly light and, at 23 mm long, not overly large.

One reason to carry out the experiment was to look at the 'Big Brother' scenario of George Orwell's 1984 (Orwell 1948) and the corresponding issues of privacy and individualism. My conclusion turned out to be somewhat different than expected in that I always felt very positive towards the implant, despite any dangers that might have been associated with it. Essentially, it did things for me, not against me. I had no worries that the computer knew where I was in the Cybernetics building. Maybe my feelings were something like those of a person with a credit card. The card is convenient, flexible and easy to use, yet it gives the computer system considerable details about the user's buying patterns. Yet many people do not worry about this.

The biggest surprise for me during the experiment was that I very quickly regarded the implant as being "part of my body." Indeed this feeling appears to be shared by most people who have a cochlea implant, or heart pacemaker. In my case though there was also a computer linked to my implant and because the computer was making things happen I quickly became attached, emotionally to the computer as well. Subsequently, when the implant was removed, on the one hand I felt relieved because of the medical problems that could have occurred, but on the other hand something was missing, it was as though a friend had died.

If I had to draw one conclusion from my experience it would be that when linked with technology inside my body, it is no longer a separate piece of technology. Mentally I regard such technology as just as much part of me as my arms and legs. If my brain was linked with a computer it is difficult to imagine where I would feel my body ended.

Whilst it is perhaps not completely methodologically correct to generalise results to all humanity from those obtained with a small number of individuals, in particular oneself, it is nevertheless one potential scenario for which there is a realistic experiential basis from which a discussion can ensue. This paper is written on that basis and as such is open to criticism that the scenario may turn out to be not as general as was first thought. Despite this my feeling is that it is worth reporting on my experimental work and in doing so to indicate some of the questions that arise and the potential consequences should the scenario become a general one.

The new experiment

Since 1998 we worked on putting together a new implant, which was located into position on March 14th, 2002. Again it was surgically placed in my left arm. However on this occasion there were direct links with the nervous fibres in my arm. Essentially the signals from the computer could be transmitted, by radio, to the implant, where they were played down onto the nerve fibres. The main body of the technology was, in this case, a radio receiver – transmitter.

We investigated movement signals. For example when I moved my finger some of the electronic signals on my nervous system, which caused the muscles and tendons to operate, were also transmitted to the computer, where they were stored as a sequence. Subsequently the same signals could be played back from whence they came in order to attempt to recreate as much of the original movement as possible.

We also investigated extra sensory input. We have, in the Department of Cybernetics many mobile robots, which operate in an autonomous way. Different versions of these can also be witnessed in the Science Museum London, Ars Elettronika in Linz, Austria and the Millennium Point Birmingham. They generally sense the world using ultrasonic sensors – not a sense that humans have. In the experiment we fed the output from an ultrasonic sensor down onto my own nervous system, rather than as an input to the robot. It was evident that my brain was able to make sense and good use of the signals that arrived.

Physical emotions exhibit signals on the nervous system as well. For example, when a person is angry, shocked or excited, distinct signals can be witnessed on their nervous system. For our experiment certain emotions were enforced, with the resultant signals being recorded on the computer. As an example I could be surprised and the signals associated with this shock transmitted to the computer where they could be recorded. Subsequently the shock signals could be played back down onto my nervous system to see what happened.

To be clear on this point, it is recognised that physical emotions are dependant on not only a stimulus but also on a variety of other factors such as endocrinology, memory, personal history, taste and so on. All of these things affect the physical signals that appear on the nervous system in response to a stimulus. It is the resultant, measurable, physical signals that could be recorded and played back.

As everything went well with my own implant, my wife joined me by having electrodes inserted into the median nerve of her left arm. We were able to successfully send signals from one person's nervous system, across the internet, down onto the other person's nervous system.

Signals from my nervous system were transmitted across the internet from Columbia University, New York, to Reading University, UK, to move around a robot hand. The robot hand was directly controlled by my neural signals generated on another continent. Effectively my nervous system did not stop at my body's limits but rather where the internet link concluded.

Clearly the new experiment threw up many more questions than it answered. The message we can all see however is that it opens up a variety of ethical posers.

Implications of the experiments

Experiments, such as those just described, do not take place in a vacuum. The experiments themselves affect and in turn are affected by society. In a straightforward way they might open up a completely new branch of science or study. Of equal importance however is the fact that they change the mindset of a group of people. Suddenly it is realised that something is technically possible, implants are not science fiction but are science fact.

Whilst many may regard the employment of such technology as merely part of an infinite stream of progress, others may grab hold of the new technology and welcome it with open arms, yet others still may reel in horror at the mere thought of it. Those from different cultures and ethnical backgrounds may well respond differently. Who will drive the technological development forward and who will control its implementation? Should it be left in the hands of commercial concerns, which could easily be devisive to society, or should political entities be made responsible? Should military deployment be regulated and overseen? An individual's rights are clearly at stake not only in terms of being forced into being a subject for the technology but also, if they miss out on it, how they are affected by those who have the technology.

The 1998 experiment raised questions about monitoring and tracking. Is it so bad to have such an implant rather than an external, visually offensive, tagging system? Could such an implant be used instead of a credit card? In such a case, it would be difficult to lose or get it stolen and shopping would be much simpler, an individual could just walk out the door with their supplies, there would be no need to queue. Conversely, serious issues of privacy are apparent, in particular due to the large amount of individual information that such an implant could store.

The 2002 implant experiment raises much more serious questions however. Remote control movement is fine as far as helping those who have a break in their nervous system, perhaps due to an illness or an accident. Hopefully some people will return to having some amount of limited movement. Therefore the research appears to be worth doing. Yet at the same time it opens up the possibility of one individual remote controlling the movements of another. In certain military or political scenarios this could be, to say the least, most unpleasant.

But then there is the extra sensory input. This could be immediately useful for a person who is blind, not to repair their blindness but to give them an alternative sense. How could we possibly deny a blind person the ability to move around freely and know the whereabouts of objects? For me though, the ultrasonic sense was an extra. Perhaps in the future I will also have infrared, X-ray and ultra violet senses. It will be extremely interesting to find out how it feels to have the ability to sense the world in ways that other humans cannot.

Also there is the question of hooking a nervous system up to the Internet. If it is possible to alter an individual's feelings by downloading electronic signals, rather than chemical, then this will revolutionise medicine. It is worth remembering that the human nervous system and brain is electro-chemical, part electronic, part chemical. Traditionally medicine has almost exclusively been concerned with chemical treatments. But just as chemicals can be used in a positive way as a medicine, so they can also be used as a drug. The same would of course be true for electronic medicine – it is both a positive and a negative immediately.

Then there are aspects of the experiments, which simply probe outside the box and raise more in the way of ethical speculation rather than debate. For example, sending signals from nervous system to nervous system, across the Internet, opens up a whole new communications medium. For our experimentation it was rather like a telegraph, one finger movement means one thing, a wrist movement means something else. But ultimately with connections to the brain rather than nervous system, it opens up the possibility of individuals controlling machinery just by their thoughts (Warwick 2000) and ultimately communicating between humans just by thinking to each other (Warwick 2002). To be clear, this does not mean that all of an individual's thoughts could be read by another, but rather that an individual would be able to communicate, the thoughts they want to, directly onto another's brain. Will anyone mind if speech becomes obsolete?

Conclusion

Some argue that linking technology with humankind can merely be seen as humans acting as eccentric living beings. Any view of the appearance of superhuman cyborgs can be seen as being unwarranted 'metaphysical' speculation (Coolen 2001). On the other hand it could be felt that humankind is itself at stake (Warwick 1998; Cerqui 2001). A viewpoint can then be taken that either it is perfectly acceptable to upgrade humans, turning them into Cyborgs, with all the enhanced capabilities that this offers (Warwick 2002), or conversely it can be felt that humankind is just fine as it is and should not be so tampered with (Cerqui 2001).

Perhaps the most important point here is that we are considering not merely a physical extension of human capabilities but rather a completely different basis on which the cyborg brain operates in a mixed human, machine fashion. Whilst it is no doubt true that physical extensions such as an aeroplane, a pair of glasses or even a wearable computer give a human capabilities that they would not themselves normally possess, when the nature of the brain itself is altered the situation is a very different one. Such a cyborg would have a different foundation on which any thoughts would be conceived in the first place. From an individualistic viewpoint therefore, as long as I am myself a cyborg I am happy with the situation. Those who wish to remain human however may not be so happy.

We must be clear that with extra memory, high powered mathematical capabilities, including the ability to conceive in many dimensions, the ability to sense the world in many different ways and communication by thought signals alone, such cyborgs will be far more powerful, intellectually, than humans. It would be difficult imagining that a Cyborg of this ilk would want to voluntarily give up their powers. It is also difficult to imagine that this Cyborg would pay any heed to a human's trivial utterances.

With a brain which is part human, part machine, a Cyborg would haye some links to their human background but their view on life, what is possible and what is not, would be very much different from that of a human. The values, morals and ethics of a Cyborg would relate to its own life, what it feels is important and what not. In fact humans may not figure too highly in such a scenario.

It could be argued that humans are already digitally enhanced by current technology (Clark 2003) and to some extent this may, in any case, alter morals. Yet despite this for the most part things have remained the same. The big difference with regard to my own concept of cyborgs though is that their brain is part human part machine and hence the epicentre of moral and ethical decision making is no longer of purely human form, but rather it is of a mixed human, machine base.

One murky aspect is that a Cyborg would most likely have a brain, which is not stand alone, but rather, via its machine part, is connected directly to a network. The leading question realistically is therefore is it morally acceptable for cyborgs to give up their individuality and become mere nodes on an intelligent machine network. This is of course as much of a question for cyborgs as it is for humans.

This whole area is now throwing up vitally important ethical questions. Should every human have the right to be upgraded into a Cyborg? If an individual does not want to should they be allowed to defer, thereby taking on a role in relation to a cyborg rather akin to a chimpanzee's relationship with a human today? Even those humans that do upgrade and become a Cyborg will have their own problems. Just how will Cyborg ethics relate to human ethics? The whole topic needs to be brought to the fore, now.

References

G. Bateson. *Steps to an Ecology of Mind*. Ballantine Books, New York, 1972.

D. Cerqui. *The Future of Humankind in the Era of Human and Computer Hybridisation. An Anthropological Analysis*. Proc. Int. Conference on Computer Ethics and Philosophical Enquiry (CEPE), Lancaster Univ., pp. 39–48, 2001.

A. Clark. *Natural Born Cyborgs*, Oxford University Press, 2003.

M. Coolen. *Becoming a Cyborg as One of the Ends of Disembodied Man*. Proc. Int. Conference on Computer Ethics and Philosophical Enquiry (CEPE), Lancaster Univ., pp. 49–60, 2001.

D. Haraway. A manifesto for cyborgs: Science, Technology and Socialist Feminism in the 1980s. *Socialist Review*, 80: 65–108, 1985.

M.K. Hayles. *How We became Posthuman*. University of Chicago Press, 1999.

P.R. Kennedy, R.A.E. Bakay, M.M. Moore, K. Adams and J. Goldwaithe. Direct control of a computer from the human nervous system. *IEEE Transactions on Rehabilitation Engineering*, 8(2): 198–202, 2000.

F. Nietsche. *Thus Spoke Zarathustre*. Penguin Classics, 1961.

G. Orwell. *1984*. Penguin, 1948.

R. Penrose. *Shadows of the Mind*. Oxford Univ. Press, 1994.

A.P. Pentland. Wearable intelligence. *Scientific American*, 9(4): 90–95, 1998.

R. Sparrow. *Artificial Intelligences, Embodiment and the 'Turing Triage Test'*. Proc. Int. Conference on Computer Ethics and Philosophical Enquiry (CEPE), Lancaster Univ., pp. 222–235, 2001.

A. Turing. Computing machinery and intelligence. *Mind*, 59: 433–460, 1950.

K. Warwick. *In the Mind of the Machine*. Arrow, 1998.

K. Warwick. Cyborg 1.0. *Wired*, 8(2): 144–151, 2000.

K. Warwick. *QI: The Quest for Intelligence*. Piatkus, 2000.

K. Warwick. *I Cyborg*. Century, 2002.

M. White and J. Gribbin *Einstein*. Pocket Books, 1997.