

The social impact of intelligent artefacts

Richard S. Rosenberg

Received: 5 March 2007 / Accepted: 18 June 2007 / Published online: 19 July 2007
© Springer-Verlag London Limited 2007

Abstract The simplistic assumption that replacing humans by intelligent artifacts or introducing such artifacts, or robots, into all aspects of human society will necessarily benefit society at large must be continually re-evaluated. Clearly, contributing factors will involve concerns of efficiency, the role of work as a component in human self-worth, the distribution of wealth generated by advanced technologies, the potential for growing divisions in society resulting from gross inequities in income and from the loss of work as a central fact of life, and of course, the unpredictability of regular and widespread interactions between humans and artificially intelligent programs and devices. The focus of this paper is an exploration of the future relationship between people and their robots and the anticipated role of ethics.

Introduction

South Korea's obsession with technology has led it to consider what may be the first government-backed ethical code for robots.

...

In a move sure to delight science fiction fans, the *Robot Ethics Charter* is likely to be modeled on the instructions devised by the American writer Isaac Asimov in his series, *I, Robot*. Spencer (2007, A8).

R. S. Rosenberg (✉)
Department of Computer Science, University of British Columbia,
201-2366 Main Mall, Vancouver, BC, Canada V6T 1Z4
e-mail: rosen@cs.ubc.ca

Although the current focus of much research in artificial intelligence (AI) is typically on robots, intelligent behaviour can of course be manifested in a variety of activities. Another area of interest is the use of intelligent programs in data mining, the analysis of vast amounts of information to determine useful patterns of commercial or other activities. Such patterns are of considerable value to direct marketers and government agencies and represent a serious threat to individual privacy. Along this direction as well, there is interest in natural language understanding programs that could be used by filtering and blocking software on the Internet to restrict access, in an intelligent manner, to “undesirable” Web sites. Smart vision programs could be used to block access to images that may be too sexually explicit for some tastes. The current interest in controlling and limiting access, as a means to make the Internet a safe place for business and society at large will be aided by the diffusion of intelligent programs that replace inefficient humans in these difficult and time-consuming tasks.

What follows is by necessity speculative. Its purpose is to raise questions and to argue that a serious and ongoing analysis of the social impact of intelligent artifacts is not only necessary but long overdue. There is no shortage of discussions and analyses of the societal impact of medical and biological advances, ranging from cloning to genetic engineering and to a host of reproductive technologies. It seems obvious that society must understand the implications of these technologies because of their intimate relation to the very basis of our existence. But except for a flourishing science fiction literature, there is very little in the way of a public discourse on the societal implications of intelligent machines, although this situation may be changing.

Perhaps because of the gradual increase in the non-standard applications of computers, AI has appeared to be somewhat esoteric and not of immediate concern except perhaps for security, military and financial applications. However, increasingly sophisticated robots with planning and vision abilities are making their presence felt but with little public impact, so far. The fact that there is so little public debate, if any, about such innovations lends weight to the critical position that AI is a prime example of the technological imperative. That is, innovations diffuse because of an implicit and perhaps internal logical motivation, not because an open and democratic process has determined that they are beneficial for society as a whole, even though there may be temporary dislocations and even harm to some segments. Perhaps the prime, and some would say the most extreme, advocate of this position is Ellul (1967).

Given the existing importance of robotics in production and the current attempts to exploit limited but extensive niches in the home, it is not surprising that considerable effort is being devoted to the development of intelligent robots. If the economic motivations are obvious then so is the question about the social role of robots and their impact on work and workers. It is claimed that robots will relieve people of boring, repetitious, and even dangerous work, thereby freeing them to realize their potential in more meaningful ways. But of course the motivation to replace people with machines has rarely been based on humanitarian concerns. If we look back at the beginnings of the Industrial Revolution, more specifically, the first decade of the nineteenth century, the activities of some of the workers have received

considerable attention. These workers, called Luddites, who in so-called blind opposition to automation in the weaving industry attempted to destroy the new machinery, have been portrayed as irrational and their name has come to stand for opposition to progress. However, new historical evaluations have cast their motivations and activities in a new light. As the historian Noble (1995) notes,

According to these revisionist interpretations, the Luddites who resisted the introduction of new technologies were not against technology per se but rather against the social changes that the new technology reflected and reinforced. Thus, the workers of Nottingham, Yorkshire and Lancashire were not opposed to hosiery and lace frames, the gig mill and shearing frames, larger spinning jennies, or even power looms. Rather, in a postwar period of economic crisis, depression, and unemployment much like our own, they were struggling against the efforts of capital, using technology as a vehicle, to restructure social relations and the patterns of production at their expense.

How will political structures respond to such dramatic changes in the workplace and elsewhere? Are such changes desirable and how could this question be explored and ultimately answered? Note that it is often the case when utopias are described, albeit in science fiction, that very little space is devoted to the process of moving from here to there. We are usually presented with the description of some world resembling ours but with a number of unexpected features some desirable, some abhorrent. Thus, making machines smart, obviously an extremely difficult task may not be that much more difficult than determining how they should be used effectively and humanely. The use of the passive construction in the previous sentence conceals the crucial question, namely, who takes the responsibility for maintaining a livable society? What are the appropriate ethical principles?

While biologists must explain, why research on cloning, for example, will have long-term benefits, such concerns rarely arise in AI. It seems to be taken for granted that benefits to society of intelligent artifacts are so obvious that critical review is unnecessary. This viewpoint must be challenged and this paper is meant to serve that purpose. Furthermore, traditional ethics, defining appropriate human behaviour in a variety of situations must be extended to human interaction with advanced technologies on the horizon, a difficult and daunting task...

Some current robot projects

In the early days of AI before robots were seen as a challenge to human autonomy, predictions of the inevitable dominance of computers were frequently offered for public consumption. For example, consider the following, expressing Arthur Samuel's view of the economic possibilities arising from his successful checker-playing program: Samuel (1959, p. 223).

As a result of these experiments one can say with some certainty that it is now possible to devise learning schemes which will greatly outperform an average

person and such learning schemes may eventually be economically feasible to apply to real-life problems Samuel (1959, p. 223).

It is not uncommon for AI researchers to predict an extraordinary future based on minimally suggestive research results. Therefore, it is not surprising that predictions made with respect to rapidly improving robots are extravagant as well. Witness the views of the *eminence grise* of AI, Minsky (1970), namely,

In from three to eight years we will have a machine with the general intelligence of an average human being. I mean a machine that will be able to read Shakespeare, grease a car, play office politics, tell a joke, have a fight. At that point the machine will begin to educate itself with fantastic speed. In a few months it will be at genius level and a few months after that its powers will be incalculable Minsky (1970).

The media have regularly offered stories on the steady progress made by robot designers in extending the range of applications. The robot vehicle competition sponsored by the US Department of Defence Advanced Project Agency has received extensive coverage as a possible future application of robotics, especially in a battlefield domain Orenstein (2005). Recently, another emerging robot activity was disclosed, namely, “Some robots are destined to rove the surface of Mars. Others, like Hyperactive Bob, will work in fast-food restaurants” Kanellos (2007). Some of the abilities incorporated in Hyperactive Bob are as follows: Kanellos (2007).

Pittsburgh’s Hyperactive Technologies has come up with a system, based on the computer vision and artificial intelligence systems employed by robots, to manage the kitchens at so-called quick-service restaurants.

The vision system in Hyperactive Bob essentially scans the parking lot for incoming cars. It then cross-references traffic patterns against data about the restaurant—the bell curve of orders, the time of day, cooking times, the current amount of food in the restaurant’s warming bins—and issues cooking orders to the employees manning the grill or the deep fat fryer. There isn’t a mechanical humanoid assembling chicken sandwiches behind the counter. Instead, Hyperactive Bob combines machine intelligence with human activity.

The range of abilities required in this domain is surprisingly great and the fact that a product is now available for commercial use is a measure of the progress achieved in transferring research results to the marketplace.

The chief technology reporter for the New York Times wrote a column in mid-2006, in which he described the extension of robot applications “Into Daily Life” Markoff (2006). Although there is considerable doubt about extravagant claims made by AI practitioners—a condition with a long history in AI—the final quotation seems to capture the current sentiment, “It’s time to build an A.I. robot”, said Andrew Ng, a Stanford computer scientist and a leader of the project, called Stanford AI Robot, or Stair. “The dream is to put a robot in every home” Markoff (2006).

This theme of robots in the home is particularly present in South Korea, which is often described as the most wired country on earth. Here are some recent predictions

for the diffusion of robots into many aspects of South Korea's everyday life Onishi (2006).

By 2007, networked robots that, say, relay messages to parents, teach children English and sing and dance for them when they are bored, are scheduled to enter mass production. Outside the home, they are expected to guide customers at post offices or patrol public areas, searching for intruders and transmitting images to monitoring centers.

If all goes according to plan, robots will be in every South Korean household between 2015 and 2020. That is the prediction, at least, of the Ministry of Information and Communication, which has grouped more than 30 companies, as well as 1,000 scientists from universities and research institutes, under its wing. Some want to move even faster.

...

Late last year, three types of robots were distributed to 64 randomly selected households, as well as two post offices, with mixed results, Mr. Oh said. In October, a second phase in the testing will put robots in 650 households and 20 public places.

By 2007, the networked robots are expected to be on the market. Yujin Robot started developing prototypes in 2004 and has sold 100, mostly to universities and research institutes, said Shin Kyung Chul, the company's president. It is the leader in making small, \$500 robots that move around the house using sensors, vacuuming or sweeping. They have become popular gifts for newlyweds.

There are two further indications of how accepting we seem to have become of the idea of robots playing a meaningful role in society, and particularly in our homes and on the battlefield. In an overview of military applications in the magazine *Popular Science*, the following selections are of some concern: Lerner (2005).

But will they ever actually kill on their own? Right now, the party line is that there will always be a human in the loop before lethal action is taken—that a robot will never decide on its own to fire a gun or cannon, or light a missile.

...

But some observers sense that robots' lack of emotions will eventually be taken advantage of "Part of the process of creating soldiers is disinhibiting people from killing", says John Pike, director of GlobalSecurity.org, a military-policy think tank. "Robots have no such inhibition. They will kill without pity".

...

A fully-autonomous weaponized unmanned ground vehicle—one that can distinguish enemy combatants or vehicles on its own and attack them without direct human command—remains decades away, but analysts are convinced that it will arrive. Such a vehicle will need to be able to differentiate among friendly forces, enemy combatants, and civilians with unquestionable reliability. It will also require the ability to act from mission objectives, combat tactics, and all military protocols and rules of engagement.

As a foreshadowing of ethical issues, which will certainly arise, one obvious question is will the availability of intelligent battlefield robots increase the likelihood of military engagements because the risks to human life and limb will be reduced? And of course, for the purposes of this paper, the ethical issues associated with the military deployment of robots, assume a critical position.

Finally, in response to a question about further opportunities for home robots, Colin Angle, the chief executive officer of iRobot Corporation, the maker of the home robot vacuum cleaner, Roomba, replied as follows: Jones (2006).

The number of people over 65 in our country is growing rapidly. What our aging population typically wants to do is live independently in their current situation for longer periods of time. Central to making that possible in the future is going to be practical and affordable robots that can do basic tasks for them.

This motivation also drives the development of domestic robots in such countries as South Korea and Japan, as well. Thus we are witnessing the accelerating development of robots as they evolve from their origins in the factory to the relatively unconstrained environment of the home and the marketplace. The aging of the population in industrial and post-industrial countries places increasingly great demands on the development and diffusion of robots to serve the elderly, the bed-ridden, the immobile and many others. These needs are real and increasing and a ready market awaits. However, the real fear exists that unless this next generation of robots is equipped with effective safety controls governed, dare I say, by comprehensive ethical principles, many unfortunate accidents and other unpleasant events will occur. Are such ethical principles available?

Exploration of possibly relevant ethical principles

In this section, it will be necessary to review, however briefly, Isaac Asimov's three laws of robotics and their legacy. Suggested modifications and additions will also be explored in the context of accepted ethical principles for human behaviour. The famous three laws, motivated, by Asimov's concerns with the fictional interaction of humans and robots, are given as follows: Asimov (1968).

The 1940 Laws:

- First Law: A robot may not injure a human being, or, through inaction, allow a human being to come to harm.
- Second Law: A robot must obey orders given it by human beings, except where such orders would conflict with the First Law.
- Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Asimov detected, as early as 1950, a need to extend the first law, which protected individual humans, so that it would protect humanity as a whole. As such, the Zeroth Law appeared in 1985, namely, "A robot may *not injure humanity*, or, through inaction, allow humanity to come to harm". This law places the survival of

humanity as a whole, or even a greater number of humans, above the life of an individual. It seems clear that robots will face moral challenges not unfamiliar to humans.

Many science fiction writers have used these “laws” to explore a variety of ethical issues to the point that it seems hardly possible to propose alternative approaches. Nevertheless, the Australian computer scientist, Roger Clarke, has suggested a number of modifications meant to deal with certain gaps and limitations in the original formulation. In Clarke’s proposal, a Meta-Law is added to ensure that the actions of all robots are subject to the Laws. This Meta Law is “A robot may not act unless its actions are subject to the Laws of Robotics” and is added to the Zeroth Law and Clarke’s modifications of the three original laws. In Clarke’s reformulation, the Extended Set of Laws appears as follows: Clarke (1994).

- The Meta-Law

A robot may not act unless its actions are subject to the Laws of Robotics.

- Law Zero

A robot may *not injure humanity*, or, through inaction, allow humanity to come to harm.

- Law One

A robot may *not injure a human being*, or, through inaction, allow a human being to come to harm, unless this would violate a higher-order Law.

- Law Two

A robot must *obey orders given it by human beings*, except where such orders would conflict with a higher-order Law.

A robot must *obey orders given it by superordinate robots*, except where such orders would conflict with a higher-order Law.

- Law Three

A robot must *protect the existence of a superordinate robot* as long as such protection does not conflict with a higher-order Law.

A robot must *protect its own existence* as long as such protection does not conflict with a higher-order Law.

- Law Four

A robot must *perform the duties for which it has been programmed*, except where that would conflict with a higher-order law.

- The Procreation Law

A robot may *not take any part in the design or manufacture of a robot* unless the new robot’s actions are subject to the Laws of Robotic.

One question immediately arises. Does this set do the job? Unfortunately, this question is ill-formed and history shows that constructing an adequate set of

principles to govern human behaviour is probably an impossible task. If reproducing these “laws” or other formulations can limit the damage potential of future robots, it is a task worth doing but the complexity of robots to come may pose an insurmountable challenge.

In the opening page of Chap. 15, Ethics and Professionalism, of the book, *The Social Impact of Computers*, Rosenberg (2004, p. 657) the following appears:

- Why be moral? Socrates.
- Do the right thing! Spike Lee.
- You know what’s right! My mother.
- “Ethics has to do with what my feelings tell me is right or wrong”.
- “Ethics has to do with my religious beliefs”.
- “Being ethical is doing what the law requires”.
- “Ethics consists of the standards of behaviour our society accepts”.
- “I don’t know what the word means”.

No deep thoughts are being expressed; the purpose of this quotation is just to remind us that there are no generally accepted ethical principles and what is accepted does not constitute an algorithm for governing behaviour. As such the task of constructing effective limitations on the behaviour of robots, which will move and perform their tasks among us, is likely impossible and at best can only be approached incrementally and with modest expectations of success. However, it is both necessary and worthwhile to explore some of the relevant issues.

The issue at hand is to review, however briefly, possible approaches to embedding within robots a subsystem to prevent or at least limit possible harm caused by misbehaviour or by behaviour resulting from unanticipated situations. On the one hand, this is an engineering problem, i.e. dealing with minor or major failures in performance but for robot behaviour the stakes are substantially higher. There is a positive correlation between an increase in the complexity of behaviour and the resources needed to limit negative outcomes. Clarke’s reformulation of Asimov’s laws is a start but more needs to be done.

Biology and ethical behaviour

The biological roots of ethical behaviour have been receiving considerable attention recently. For example, consider the following news item: Brown (2006, p. A11).

Temporarily disrupting the decision-making portion of the human brain can induce people to behave selfishly, so that self-interest trumps fairness in dealings with others, according to a recent study in the journal *Science*.

...

When scientists used electrical impulses to disrupt the region called the right prefrontal cortex, selfish urges won out over the sense of fairness in a money-sharing exercise known as the Ultimatum Game.

“Civil courage and fair behaviour often stand in diametric opposition to one’s own economic advantage and thus require control and the suppression of egoistical impulses”, Knoch [Daria, the study’s lead author] said.

That there are regions of the brain involved in morality and ethical decision-making may not be surprising but how these have evolved to operate in the real world is of considerable interest. Another example of the apparent relationship between morality and brain structures is described as follows: Carey (2007).

Damage to an area of the brain behind the forehead, inches behind the eyes, transforms the way people make moral judgments in life-or-death situations, scientists are reporting today. In a new study, people with this rare injury expressed increased willingness to kill or harm another person if doing so would save others' lives.

The findings are the most direct evidence to date that humans' native revulsion for hurting others relies on a part of neural anatomy, one that likely evolved before the brain regions responsible for analysis and planning.

...

"I think it's very convincing now that there are at least two systems working when we make moral judgments", said Joshua Greene, a psychologist at Harvard who was not involved in the study. "There's an emotional system that depends on this specific part of the brain, and another system that performs more utilitarian cost-benefit analyses which in these people is clearly intact".

The possibility of ethical decision-making by robots has received considerable attention recently. Witness the appearance last year of a special issue of the journal *Intelligent Systems*, published by the Institute of Electrical and Electronics Engineers (IEEE), on *Machine Ethics*. The editors note that, "We believe that the time has come for adding an ethical dimension to at least some machines" (Anderson and Anderson 2006, p. 10). They define machine ethics as, "concerned with how machines behave towards human users and other machines [and a] goal of machine ethics is to create a machine that's guided by an acceptable ethical principle or set of principles in the decisions it makes about possible courses of actions it could take". Such an investigation could have the important by-product of improving current studies into ethics in general. Thus, several papers in this collection make a variety of contributions to the particular study of machine ethics in particular and human ethics in general. A concluding remark notes that, "Ethics experts continue to make progress towards consensus concerning the right way to behave in ethical dilemmas. The task for those working in machine ethics is to codify these insights, perhaps even before the ethics experts do so themselves".

That the brain is involved in decision-making, moral or otherwise is obvious. That certain parts of the brain seem to be involved in specific tasks that require more formal reasoning while others have a stronger emotional component is more interesting. Clearly evolution has played a central role but for the present purposes, it makes sense to assume that it will be possible to incorporate into the reasoning component of robots, ethical decision-making that will limit much of the possible harm that may result from actions of robots among us.

A newborn child raised in any culture will, without any difficulty, or apparent conscious effort acquire the ability to speak and understand the associated language. One explanation posed by the linguist Noam Chomsky is the existence

of a language organ, referred to as a universal grammar. Consider the following brief introduction to Chomsky's conception of a universal grammar: Chomsky (1977, p. 2).

The class of possible human languages is, I assume, specified by a genetically determined property, apparently species-specific in important respects. Any proposed linguistic theory - in particular, EST [Chomsky's own Extended Standard Theory] - may be regarded as an attempt to capture this property, at least in part. Thus a linguistic theory may be understood as a theory of the biological endowment that underlies the acquisition and use of language; in other terms as a theory of universal grammar (UG), where we take the goal of UG to be the expression of those properties of human language that are biologically necessary.

Not surprisingly, the details of this theory have been much debated but the necessity of an underlying language facility, genetically based must be accepted. It is therefore not surprising that researchers committed to a genetic basis for ethical behaviour would be influenced by the ongoing research in linguistics. Given the complexity and power of grammars for natural language, based in part on their recursive structure, it is not unreasonable to attempt to model the process of acquiring a system of ethical rules by exposure to examples as well as non-examples of ethical behaviour.

One interesting version of an approach to the acquisition ethical behaviour that is based on a built-in "universal moral grammar" is that proposed by Marc Hauser, a professor in the departments of psychology, organismic and evolutionary biology, and biological anthropology at Harvard University, where he is also co-director of the Mind, Brain and Behaviour Program and director of the Cognitive Evolution Laboratory. Professor Hauser has proposed the idea that in parallel with Chomsky's version of a universal grammar, referred to above, to account for the basic human facility in natural language, there is a universal moral grammar that is the foundation of human moral behaviour. Marc Hauser is one of the more prominent researchers in this area. A genetic basis to ethical behaviour, structured as a grammar, would of course provide a possible way to explore the acquisition of such behaviour by suitably equipped robots immersed in a given culture. Hauser's own words are given as follows, in a brief summary of his theory: Ross (2006).

In brief, I argue that we are endowed with a moral faculty that delivers judgments of right and wrong based on unconsciously operative and inaccessible principles of action. The theory posits a universal moral grammar, built into the brains of all humans. The grammar is a set of principles that operate on the basis of the causes and consequences of action. Thus, in the same way that we are endowed with a language faculty that consists of a universal toolkit for building possible languages, we are also endowed with a moral faculty that consists of a universal toolkit for building possible moral systems.

A few more details of this theory follow: Ross (2006).

By grammar I simply mean a set of principles or computations for generating judgments of right and wrong. These principles are unconscious and inaccessible. What I mean by unconscious is different from the Freudian unconscious. It is not only that we make moral judgments intuitively, and without consciously reflecting upon the principles, but that even if we tried to uncover those principles we wouldn't be able to, as they are tucked away in the mind's library of knowledge. Access comes from deep, scholarly investigation.

...

And in the same way that the unconscious but operative principles of language do not dictate the specific content of what we say, if we say anything, the unconscious but operative principles of morality do not dictate the specific content of our moral judgments, nor whether we in fact choose to help or harm others in any given situation.

Thus, under this theory, it may be possible to uncover a core moral grammar that could serve to make the behaviour of intelligent robots consistent with the environment in which they are used. Such a research program will be long and difficult and in the interim, robots will increasingly be diffused into populations that may be harmed by unpredictable behaviours.

A critic of Hauser's views is the philosopher, Richard Rorty, who wrote a review, in 2006, of Hauser's book, *Moral Minds*. Some of Rorty's criticisms are given as follows: Rorty (2006).

He [Marc Hauser] holds that "we are born with abstract rules or principles, with nurture entering the picture to set the parameters and guide us toward the acquisition of particular moral systems". Empirical research will enable us to distinguish the principles from the parameters and thus to discover "what limitations exist on the range of possible or impossible moral systems.

...

Biologists, he thinks, are in a position to amplify this voice. For they have discovered evidence of the existence of what Hauser sometimes calls "a moral organ" and sometimes "a moral faculty". This area of the brain is "a circuit, specialized for recognizing certain problems as morally relevant.

...

Knowing more details about how the diodes in your computer are laid out may, in some cases, help you decide what software to buy. But now imagine that we are debating the merits of a proposed change in what we tell our kids about right and wrong. The neurobiologists intervene, explaining that the novel moral code will not compute. We have, they tell us, run up against hard-wired limits: our neural layout permits us to formulate and commend the proposed change, but makes it impossible for us to adopt it. Surely our reaction to such an intervention would be, "You might be right, but let's try adopting it and see what happens; maybe our brains are a bit more flexible

than you think”. It is hard to imagine our taking the biologists’ word as final on such matters, for that would amount to giving them a veto over utopian moral initiatives.

The final sentence seems to sum up Rorty’s concerns with respect to the dangers inherent in giving biologists “a veto over utopian moral initiatives”. In a letter to the *New York Times Book Review*, Hauser, not surprisingly, objects strenuously to Rorty’s criticisms. Among his objections are the following remarks: Hauser (2006).

Contrary to Rorty’s portrayal, the idea that we are born with a universal moral grammar, like the idea that we are born with a universal grammar for language, doesn’t deny the role of culture, nor does it make certain outcomes inevitable. Rather, it lays out a framework for exploring which aspects of our biology enable us to acquire a particular moral system, and which aspects of our moral principles are universal and which vary across cultures. In the same way that something about the biology of humans enables us, uniquely, to acquire certain types of languages, something about our biology must enable us, but not other animals, to acquire certain types of moral systems.

...

Contrary to Rorty, this doesn’t mean that scientists should override parents, judges or teachers. But we all can benefit from a better understanding of the way that biological processes guide our unconscious, moral judgments and thereby exert powerful influences over how we perceive the world. By recognizing their influence we can better design our legal and educational systems.

It does appear to be worthwhile to explore the possible biological substrate for moral reasoning in anticipation of the diffusion of intelligent robots into our lives. Whether around the corner, so to speak, or in the near future, it will be necessary for us to deal with robots, which appear to be intelligent but perhaps with serious, unpredictable limitations.

Chips implanted in humans

With the emphasis on external, intelligent, artificial agents, i.e. robots, one should not overlook the potentially serious issues raised by the implantation of chips into humans. The rapidly growing use of Radio Frequency Identification chips (RFID) is on the verge of initiating considerable concern about the possible loss of personal autonomy. For example consider the following story: Libbenga (2006).

A Cincinnati video surveillance company, CityWatcher.com, now requires employees to use VeriChip human implantable microchips to enter a secure data centre. Until now, the employees entered the data centre with a VeriChip housed in a heart-shaped plastic casing that hangs from their keychain”.

This chip is an RFID tag encapsulated in a glass case, approximately the size of a grain of rice. The tag is injected into the triceps area of the arm and can be read by radio eavesdropping from distances of a few centimeters to several meters. It may

be helpful to say more about this technology because of its potential impact on a variety of aspects of everyday life. Consider the following description: (Lockton and Rosenberg 2005, pp. 221–222).

In its simplest form, it consists of a tag (microchip) and a reader. The tag is composed of an electronic circuit which stores data and an antenna which broadcasts this data by radio wave in response to a query signal from a nearby reader. The reader also contains an antenna which receives the radio signal, and also has a demodulator which transforms the analog radio data into digital data suitable for any computer processing which will then be done. “Active” RFID tags include a battery, allowing them to constantly transmit the data stored on the circuit, whereas

It is useful to present some of the current applications of this relatively simple technology in order to appreciate what the future may hold, especially with respect to the implantation of the next generation of far more complex devices that will need to be governed by some form (ethical) of limitations as they interact with their human hosts (Lockton and Rosenberg 2005, p. 222).

Tens of millions of pets worldwide have been “chipped” in order to facilitate their identification at animal shelters. It is estimated that in the United States alone, 6000 animals were reunited with their owners due to their RFID tags every month in 2003 [Hines, 2004]. Animal tagging is also being done to attempt to prevent the spread of disease; at least 20 million livestock have been tagged in order to track outbreaks of Mad Cow and other diseases, and Portuguese legislators have ordered that the nearly two million dogs in that country be chipped and registered in a national database in an effort to control the spread of rabies. It is certainly not just animals which are being tagged, however. According to Texas Instruments vice-president David Slinger, the revolution in RFID usage began in 1993 with an effort to deter auto theft. A chip was added to the ignition key of vehicles, and a transponder was incorporated into the steering column; if the wrong key (or no key) was used to start the car, it would be immobilized. 7 of 10 cars now have this feature, and Ford is reporting that theft rates on their oft-targeted Mustang line are down 75%.

Nearly 40 million Americans already carry RFID tags [Garfinkel, 2004], whether they are embedded in car keys, building access devices, or speed payment fobs. However, the RFID industry did not reach its current level of prominence until an order was given by the largest, most powerful retailer in the world, WalMart.

Implanting chips into humans in order to provide possibly tamper-proof identification is a serious step on the road to a massive violation of privacy rights. Fortunately, privacy advocates have targeted this technology and provincial and federal privacy commissioners in Canada and elsewhere have issued guidelines, voluntary ones of course. For example, the Office of the Information and Privacy Commissioner of Ontario has issued guidelines, based on the following principles: Information and Privacy Commissioner/Ontario (2006).

- (1) Focus on RFID Information Systems, not Technologies.
- (2) Privacy and Security Must be Built in from the Outset—at the Design Stage.
- (3) Maximal Individual Participation and Consent.

These will probably be insufficient in the face of a growing technological imperative. Although the most immediate concerns are with privacy issues and this is as it should be, the implantation of increasingly complex chips into humans raises some of the same issues as the increasing diffusion of robots. That these implanted devices can communicate with other such implants as well with robots raises the stakes for achieving an effective set of incorporated ethical principles.

Before concluding this section, we should at least mention the dedicated efforts of Professor Kevin Warwick, of the University of Reading, to implant chips into his body in order to conduct experiments on himself. In 2002, “a one hundred electrode array was surgically implanted into the median nerve fibres of the left arm of Professor Kevin Warwick” Warwick (2005). Some of the experiments carried out using this much more complex chip are described as follows: Warwick (2005).

A number of experiments have been carried out using the signals detected by the array, most notably Professor Warwick was able to control an electric wheelchair and an intelligent artificial hand, developed by Dr Peter Kyberd, using this neural interface. In addition to being able to measure the nerve signals transmitted down Professor Warwick’s left arm, the implant was also able to create artificial sensation by stimulating individual electrodes within the array. This was demonstrated with the aid of Kevin’s wife Irena and a second, less complex implant connecting to her nervous system.

Warwick’s motivation is complex. He believes we are limited because our biological sensors can barely sample our rich environment. So implants could give us direct access to a number of hitherto invisible signals. For Warwick’s purposes, the key question is “Will we evolve into a cyborg community? Linking people via chip implants to superintelligent machines seems a natural progression—creating, in effect, superhumans” Warwick (2000). And for our present purposes, what will be the source of ethical behaviour for this new species of cyborgs?

Conclusions and final comments

Given our inability to construct ethical codes that are universally applicable and adhered to, why should we believe that we (and our intelligent artifacts) can be successful in devising foolproof codes that prevent these artifacts from turning on us? Consider the following concerns:

1. A realistic evaluation must be attempted of current and near-future prospects for AI applications at home, in the workplace, and in the government.
2. A similar evaluation is necessary of the impact of computer-related technologies in the workplace, balancing benefits against perceived problems, including deskilling, monitoring, job loss, restricted promotion paths, break-

- down of traditional social organizations in the office, limited entry level opportunities and health-related concerns.
3. The implications of partially realized intelligent systems in terms of the requirements placed on humans to accommodate to their, the systems', inadequacies must be considered. In the haste to introduce AI into the workplace, pressures may be placed on people to work with systems, which, while advertised as intelligent, are seriously deficient in many areas.
 4. Of particular interest is the role of AI, embedded in intelligent artifacts, such as robots or in seemingly ordinary computer systems, in decision-making, whether in financial institutions, in the executive suite, or in diverse military situations such as autonomous land vehicles, pilots' aid, aircraft carrier battle management (all of which are components of post-9/11 military initiatives) or in surprisingly complex home environments...
 5. Intelligent systems may find ready application in intelligence activities such as automatic interpretation of tape recordings and the cross-correlation of electronic files, namely data mining. Added to current threats to privacy, the availability of such powerful mechanisms could increase real and anticipated assaults on individual privacy.

Traditional ethics, defining appropriate human behaviour in a variety of situations must be extended to human interaction with advanced technologies on the horizon, a difficult and daunting task. Ethical behaviour by humans must regularly be reinforced, given differences in ethnicity and religion, as well changing values in society. Adding intelligent artifacts to the mix may present insurmountable difficulties in the short run, at least. But the burden remains on us, researchers and implementers, to carefully craft systems which do not violate accepted ethical principles, probably including the hoary three laws of robotics formulated by Asimov (1968), the science fiction visionary.

The various formulations of Asimov's Laws are promising but the limitations inherent in attempting to formalize ethical behaviour must be recognized. What then are the practical options? As with all machines that we use, basic safety provisions are built in and warnings are provided in order to limit the possibility of accidents resulting from misuse. The stakes are considerably raised as the range and quality of behaviour of these machines increases. As the complexity of tasks that robots, for this is what these machines really are, increases our concern must also increase. Given critical tasks such as to help the infirm carry out their normal activities and to make sure that children are safe, among others, robots will soon be welcomed into our homes. Will we be ready?

One final vision expressed by Professor Warwick may be instructive or somewhat scary Warwick (2000).

Anything a computer link can help operate or interface with could be controllable via implants: airplanes, locomotives, tractors, machinery, cash registers, bank accounts, spreadsheets, word processing, and intelligent homes. In each case, merely by moving a finger, one could cause such systems to operate.

Linking up in this way could allow for computer intelligence to be hooked more directly into the brain, allowing humans immediate access to the Internet, enabling phenomenal math capabilities and computer memory. Will you need to learn any math if you can call up a computer merely by your thoughts? Must you remember anything at all when you can access a world Internet memory bank?

...

Yet once a human brain is connected as a node to a machine - a networked brain with other human brains similarly connected - what will it mean to be human? Will we evolve into a new cyborg community? I believe humans will become cyborgs and no longer be stand-alone entities.

So if robots are us and in the future we will interact with them in a “natural” way, the deep issues of robot ethics will be finessed. Whether biological or technological, sentient beings will belong to the same genus. Of course this vision may not appeal to many and if that is the case, we, people, must initiate efforts to understand our uniqueness and to ensure that technology remains a tool not a partner. As such, ethics for robots will be redefined as safety regulations, however complex they may be.

Epilogue

More than 40 years ago, the British novelist and playwright, Frayn (1965) wrote a witty and prescient novel, *The Tin Men*. In it he describes attempts by Professor Macintosh, the head of the ethics department at the William Morris Institute of Automation, to build an ethical robot. The basic experiment is to place the test robot, Samaritan, on a raft with some other object, in test tank of water. The problem for the robot is to determine whether or not to sacrifice itself by jumping overboard to save the test object, because otherwise, the raft will sink and both will be lost. Thus, the experimental method is described as follows: Frayn (1965, p. 23).

His first attempt, Samaritan I, had pushed itself overboard with great alacrity, but it had gone overboard to save anything which happened to be next to it on the raft, from seven stone of lima beans to twelve stone of wet seaweed. After many weeks of stubborn argument, Macintosh had conceded that the lack of discrimination in the response was unsatisfactory and he had abandoned Samaritan I and developed Samaritan II, which would sacrifice itself only for an organism at least as complicated as itself.

In a following experiment, Samaritan is placed on a suspended raft that already has Macintosh’s principal assistant in place. The raft is dropped onto the water and slowly begins to sink. The robot quickly measures the size of the assistant’s head, computes for a moment and then rolls off the raft. It must be quickly retrieved because in order to carry out the experiment, Samaritan must not know in advance that it will be saved.

I am not suggesting that the research program for exploring robot ethics follow this experimental model but...

References

- Anderson M, Anderson SL (2006) Machine ethics. *IEEE Intell Syst* 24(4):10–11
- Asimov I (1968) Runaround. In *I Robot*, Granada, London, pp 33–51. The laws were originally published in 1942
- Brown E (2006) Electrical impluses can key brain to be selfish. *The Vancouver sun*. Available at <http://mitpress.mit.edu/e-books/Hal/chap2/two2.html>
- Carey B (2007) Study finds brain injury changes moral judgment. *The New York Times*. Available at <http://www.nytimes.com/2007/03/21/health/21cnd-brain.html>. Cited 6 Oct 2006
- Chomsky N (1977) *Essays on form and interpretation*. North-Holland, New York
- Clarke R (1994) Asimov's laws of robotics. Available at <http://www.anu.edu.au/people/Roger.Clarke/SOS/Asimov.html>
- Ellul J (1967) *The technological society*. Vintage Books, New York
- Frayn M (1965) *The tin men*. Ace Books, New York
- Garfinkel S (2004) RFID rights. *Technology Review*. Available at http://www.technologyreview.com/articles/04/11/wo_garfinkell10304.asp
- Hauser M (2006) Moral minds. Letter to the *New York Times* Book Review. Available at <http://www.nytimes.com/2006/09/10/books/Letters2.html>
- Hines M (2004) Portuguese pooches to get radio-tagged, *CNet*. Available at http://news.com.com/Portuguese+pooches+to+get+radio+tagged/2100-7343_3-5281608.html
- Information and Privacy Commissioner/Ontario (2006) Privacy guidelines for RFID information systems. Available at <http://www.ipc.on.ca/docs/rfidguidelines.pdf>
- Jones J (2006) Making robots for the home or battlefield. *The New York Times*. Available at <http://www.nytimes.com/2006/08/12/technology/12interview.html>
- Kanellos M (2007) For fast food help, call in the robots, *Cnet NEWS.COM*. Available at http://news.com.com/For+fast-food+help%2C+call+in+the+robots/2100-11394_3-6170097.html
- Lerner P (2005) Robots go to war. *Popular Science*. Available at <http://www.popsoci.com/popsoci/printerfriendly/technology/generaltechnology/f59ca60b2e948010vgnvcm100004eebcddrcrd.html>
- Libbenga J (2006) Video surveillance outfit chips workers. *The Register*. Available at http://www.theregister.co.uk/2006/02/10/employees_chipped/
- Lockton V, Rosenberg RS (2005) RFID: the next serious threat to privacy. In: *Proceedings of the 6th international, conference on computer ethics: philosophical enquiry. Ethics of New Information Technology*, University of Twente, Enschede, The Netherlands, 17–19 July, pp 41–54
- Minsky M (1970) Interview, *Life Magazine*. Available at <http://mitpress.mit.edu/e-books/Hal/chap2/two2.html>. Cited 6 Oct 2006
- Markoff J (2006) Brainy robots start stepping into daily life. *The New York Times*. Available at <http://www.nytimes.com/2006/07/18/technology/18brain.html>
- Noble DF (1995) *Progress without people. New Technology, Unemployment, and the Message of Resistance, Between the Lines*, Toronto [note that the quoted material was first published in 1983]
- Onishi N (2006) In a wired South Korea, robots will feel right at home. *The New York Times*. Available at <http://www.nytimes.com/2006/04/02/world/asia/02robot.htm>
- Orenstein D (2005) Stanford team's win in robot car race nets \$2 million prize. *Stanford Report*. Available at <http://news-service.stanford.edu/news/2005/october12/stanleyfinish-100905.html>
- Rorty R (2006) Born to be good. *The New York Times* Book Review. Available at <http://www.nytimes.com/2006/08/27/books/review/Rorty.t.html>
- Ross G (2006) American scientist author interview: the bookshelf talks with mark hauser. Available at <http://www.americanscientist.org/template/InterviewTypeDetail/assetid/52880.htm>
- Rosenberg RS (2004) *The social impact of computers*. Elsevier Academic, San Diego
- Samuel AL (1959) Some studies in machine learning using the game of checkers. *IBM J* 210–229. Available at <http://www.research.ibm.com/journal/rd/033/ibmrd0303B.pdf>
- Spencer R (2007) South Korea considers ethics code for robots. *The Vancouver Sun* March 08, A8
- Warwick K (2000) *Cyborg 1.0*, *Wired Magazine*, Issue 8.02. Available at http://www.wired.com/wired/archive/8.02/warwick_pr.html
- Warwick K (2005) The next step towards true cyborgs? Available at <http://www.kevinwarwick.com/Cyborg2.htm>