

7: THE FUTURE OF ROBOTS

Donald A. Norman

Science fiction can be a useful source of ideas and information, for it is, in essence, detailed scenario development. Writers who have used robots in their stories have had to imagine in considerable detail just how they would function within everyday work and activities. Isaac Asimov was one of the earliest thinkers to explore the implications of robots as autonomous, intelligent creatures, equal (or superior) in intelligence and abilities to their human masters. Asimov wrote a sequence of novels analyzing the difficulties that would arise if autonomous robots populated the earth. He realized that a robot might inadvertently harm itself or others, both through its actions or, at times, through its lack of action. He therefore developed a set of postulates that might prevent these problems; but, as he did so, he also realized that they were often in conflict with one another. Some conflicts were simple: given a choice between preventing harm to itself or to a human, the robot should protect the human. But other conflicts were much more subtle, much more difficult. Eventually, he postulated three laws of robotics (laws one, two and three) and wrote a sequence of stories to illustrate the dilemmas that robots would find themselves in, and how the three laws would allow them to handle these situations. These three laws dealt with the interaction of robots and people, but as his story line progressed into more complex situations, Asimov felt compelled to add an even more fundamental law dealing with the robots relationship to humanity itself. This one was so fundamental that it had to come first; but, because he already had a law labeled One, this fourth law had to be labeled Zero.

Asimov's vision of people and of the workings of industry was strangely crude. It was only his robots that behaved well. When I reread his books in preparation for this chapter, I was surprised at the discrepancy between my fond memories of the stories and my response to them now. His people are rude, sexist, and naïve. They seem unable to converse unless they are insulting each other, fighting, or jeering. The U.S. Robots and Mechanical Men Corporation doesn't fare well either. It is secretive, manipulative, and allows no tolerance for error: make one mistake and the company would fire you. Asimov spent his entire life in a university: maybe that is why he had such a weird view of the real world.

Nonetheless, his analysis of the reaction of society to robots – and of robots to humans -- was interesting. He thought society would turn against robots; and, indeed, he wrote that “most of the world governments banned robot use on earth for any purpose other than scientific research between 2003 and 2007.”¹ (Robots, however, were allowed for space exploration and mining; and in Asimov's stories, these activities are widely deployed in the early 2000s, which allow the robot industry to survive and grow.) The Laws of Robotics are intended to reassure humanity that robots will not be a threat and will, moreover, always be subservient to humans.

Today, even our most powerful and functional robots are far from the stage of Asimov's. They do not operate for long periods without human control and assistance. Even so, the laws are an excellent tool for examining just how robots and humans should interact.

Asimov's Four Laws of Robots²

Zeroth Law: A robot may not injure humanity, or, through inaction, allow humanity to come to harm.

First law: A robot may not injure a human being, or, through inaction, allow a human being to come to harm, unless this would violate the Zeroth Law of Robotics.

Second Law: A robot must obey orders given it by human beings, except where such orders would conflict with the Zeroth or First Law.

Third Law: A robot must protect its own existence as long as such protection does not conflict with the Zeroth, First, or Second Law.

Many machines already have key aspects of the laws hard-wired into them. Let's examine how these laws are implemented.

The Zeroth Law -- that "a robot may not injure humanity, or, through inaction, allow humanity to come to harm," is beyond current capability, for much the same reasons that Asimov did not need this law in his early stories: to determine just when an action -- or lack of action -- will harm all humanity is truly sophisticated, probably beyond the abilities of most people.

The first law -- that "a robot may not injure a human being, or, through inaction, allow a human being to come to harm, unless this would violate the Zeroth Law of Robotics," could be labeled "safety." It isn't legal, let alone proper, to produce things that can hurt people. As a result, all machines today are designed with multiple safeguards to minimize the likelihood that they can harm by their action. Liability laws guarantee that robots -- and machines in general -- are outfitted with numerous safeguards to prevent their actions from harming people. Industrial and home robots have proximity and collision sensors. Even simple machines such as elevators and garage doors have sensors that stop them from closing on people. Today's robots try to avoid bumping into people or objects. Lawn mower and vacuum cleaner robots have sensing mechanisms that cause them to stop or back away whenever they bump into anything or come too close to an edge, such as a stairway. Industrial robots are often fenced off, so that people can't get near them when they are operating. Some have people detectors, so they stop when they detect someone nearby. Home robots have many mechanisms to minimize the chance of damage; but at the moment, most of them are so underpowered that they couldn't hurt even if they tried to. Moreover, the lawyers are very careful to guard against potential damage. One company sells a home robot that can be used to teach children by reading books to them, and that can also serve as a home sentinel, wandering about the house, taking photographs of unexpected encounters and notifying its owners, by email if necessary (through its wireless internet connection, attaching the photographs along with the message, of course). Despite these intended applications, the robot comes with stern instructions that it is not to be used near children, nor is it to be left unattended in the house.

A lot of effort has gone into implementation of the safety provision of the first law. Most of this work can be thought of as applying to the visceral level, where fairly simple mechanisms are used to shut down the system if safety regulations are violated.

The second part of the law -- do not allow harm through inaction -- is quite difficult to implement. If determining how a machine's actions might affect people is difficult, trying to determine how the lack of an action might have an impact is

even more so. This would be a reflective level implementation, for the robot would have to do considerable analysis and planning to determine when lack of action would lead to harm. This is beyond most capabilities today.

Despite the difficulties, some simple solutions to the problem do exist. Many computers are plugged into “non-interruptible power supplies” to avoid loss of data in cases of power failure. If the power failed and no action was taken, harm would occur, but in these cases, when the power fails, the power supply springs into action, switching to batteries, converting the battery voltage to the form the computer requires. It can also be set to notify people and to turn off the computer gracefully. Other safety systems are designed to act when normal processes have failed. Some automobiles have internal sensors that watch over the path of the car, adjusting engine power and braking to ensure that the auto keeps going as intended. Automatic speed control mechanisms attempt to keep a safe distance from the car in front, and lane-changing detectors are under investigation. All of these devices safeguard car and passengers when inaction would lead to accident.

Today, these devices are simple and the mechanisms built in. Still, one can see the beginnings of solutions to the inaction clause of the first law, even in these simple devices.

The second law — that “a robot must obey orders given it by human beings, except where such orders would conflict with the Zeroth or First Law,” is about obeying people, in contrast to the first, which is about protecting them. In many ways, this law is trivial to implement, but for elementary reasons. Machines today do not have an independent mind, so they must obey orders: they have no choice but to follow the commands given them. Machines do what they are commanded; and if they fail, they face the ultimate punishment: they are shut off and sent to the repair shop.

Can a machine disobey the second law in order to protect the first law? Yes, but not with much subtlety. Command an elevator to take you to your floor, and it will refuse if it senses that a person or object is blocking the door. This, however, is the most trivial of ways to implement the law, and it fails when the situation has any sophistication. Actually, in cases where safety systems prevent a machine from following orders, usually a person can override the safety system to permit the operation to take place anyway. This has been the cause of many an accident in trains, airplanes, and factories. Maybe Asimov was correct: we should leave some decisions up to the machines.

Some automatically deployed safety systems are an example of the “through inaction” clause of the law. Thus, if the driver of an automobile steps on the brakes rapidly, but only depresses the brake pedal halfway, most autos would only slow halfway. The Mercedes Benz, however, considers this “harm through inaction,” so when it detects a rapid brake deployment, it puts the brakes on full, assuming that the owner really wants to stop as soon as possible. This is a combination of the first and second laws: the first law, because it prevents harm to the driver; and the second law because it is violating the “instructions” to apply the brakes at half strength. Of course, this may not really be a violation of the instructions: the robot assumes that full power was intended, even if not commanded. Perhaps the robot is invoking a new rule: “Do what I mean, not what I say,” an old concept from some early artificial intelligence computer systems³.

Although the automatic application of brakes in an automobile is a partial implementation of the second law, the correct implementation would have the auto examine the roadway ahead and decide for itself just how much speed, braking, or steering ought to be applied. Once that happens, we will indeed have

a full first and second law implementation. Once again, this is starting to happen. Some cars automatically slow up if too close to the car in front, even if the driver has not acted to slow the vehicle.

We don't yet have the case of conflicting orders, but soon we will have interacting robots, where the requests of one robot might conflict with the requests of the human supervisors. Then, determining precedence and priority will become important.

Once again, these are easy cases. Asimov had in mind situations where a car would refuse to drive: "I'm sorry, but the road conditions are too dangerous tonight." We haven't yet reached that point – but we will. Asimov's second law will be useful.

Least important of all the laws, so Asimov thought, was self-preservation – "a robot must protect its own existence as long as such protection does not conflict with the Zeroth, First, or Second Law" -- so it is numbered three, last in the series. Of course, given the limited capability of today's machines, where laws one and two seldom apply, this law is of most importance today, for we would be most annoyed if our expensive robot damaged or destroyed itself. As a result, this law is easy to find in action within many existing machines. Thus, sensors are built in to robot vacuum cleaners to prevent them from falling down stairs. Similarly, they – and robot lawn mowers – have bump and obstacle detectors to avoid damage by bumping into things. In addition, many robots monitor their energy state and either go into "sleep" mode or return to a charging station when their energy level drops. Resolution of conflicts with the other laws is not well handled, except by the presence of human operators who are able to override safety parameters when circumstances warrant.

Asimov's Laws cannot be fully implemented until machines have a powerful and effective capability for reflection, including meta-knowledge (knowledge of its own knowledge) and self-awareness of its state, activities, and intentions. These raise deep issues of philosophy and science as well as complex implementation problems for engineers and programmers. Progress in this area is happening, but slowly.

Even with today's rather primitive devices, having some of the capabilities would be useful. Thus, in cases of conflict, there could be sensible overriding of the commands. Automatic controls in airplanes could look ahead to determine the implications of the path they are following so that they would continue on it unless a mountain was in the way. Planes have flown into mountains while on automatic control, so the capability would have saved lives. In actuality, many automated systems already are beginning to do this kind of checking.

Even today's toy pet robots have some self-awareness. Consider a robot whose operation is controlled both by its "desire" to play with its human owner, but also to make sure that it doesn't exhaust its battery's power. When low on energy, it will therefore return to its charging station, even if the human wishes to continue playing with it.

The greatest hurdles to our ability to implement something akin to Asimov's Laws are his underlying assumptions of autonomous operation and central control mechanisms which may not apply in today's systems.

Asimov's robots worked as individuals. Give a robot a task to do, and off it would go. In the few cases where he had robots work as a group, one robot was always in charge. Moreover, he never had people and robots working together as a team. We are more likely to want cooperative robots, systems in which people and robots or teams of robots work together, much as a group of human workers can

work together at a task. Cooperative behavior requires a different set of assumptions than Asimov had. Thus, cooperative robots need rules that provide for full communication of intentions, current state, and progress.

Asimov's main failure, however, was his assumption that someone had to be in control. When he wrote his novels, it was common to assume that intelligence required a centralized coordinating and control mechanism with a hierarchical organization structure beneath it. This is how armies have been organized for thousands of years: armies, governments, corporations, and other organizations. It was natural to assume that the same principle applied to brains as well. But this is not the way of nature. Many natural systems, from the actions of ants and bees, to the flocking of birds, and even the growth of cities and the structure of the stock market, occur as a natural result of the interaction of multiple bodies, not through some central, coordinated control structure. Modern control theory has moved away from this assumption of a central command post. Distributed control is the hallmark of today's systems. As a result, robots are likely to work together in groups, with cooperative, interacting control structures. Asimov assumed a central decision structure for each robot that decided how to act, guided by his laws. In fact, that is probably not how it will work: the Laws will be part of the robot's architecture, distributed throughout the many modules of its mechanisms: lawful behavior will emerge from the interactions of the multiple modules. This is a modern concept, not understood while Asimov was writing, so it is no wonder he missed this development in our understanding of complex systems.⁴

Still, Asimov was ahead of his time. Asimov was indeed thinking far ahead to the future: his stories were written in the 1940s and 50s, but in his novel *I, robot*, he quotes the three laws of robotics from the 2058 edition of the *Handbook of Robotics*, thus, he looked ahead more than 100 years. By 2058, we may indeed need his laws. Moreover, as the analyses indicate, the laws are indeed relevant, and many systems today follow them, even if inadvertently. The difficult aspects have to do with damage due to lack of action, as well as with properly assessing the relative importance of following orders versus damage or harm to oneself, others, or humanity.

As machines become more capable, as they take over more and more human activities, working autonomously, without direct supervision, they will get entangled in the legal system, which will try to determine fault when accidents arise. Before this happens, it would be useful to have some sort of ethical procedure in place. There already are some safety regulations that apply to robots, but they are very primitive⁵. We will need more.

It is not too early to think about the future difficulties that intelligent and emotional machines may give rise to. There are numerous practical, moral, legal, and ethical issues to think about. Most are still far in the future, but that is a good reason to start now – so that when problems arrive, we will be ready.

THE FUTURE OF EMOTIONAL MACHINES AND ROBOTS: IMPLICATIONS AND ETHICAL ISSUES

The development of smart machines that will take over some tasks now done by people has important ethical and moral implications. This point becomes especially critical when we talk about robots, machines that interact with us in an anthropomorphic manner, machines that might have emotions, and, as I discuss

in the epilogue, machines to which people might bond or, at the least, form strong emotional attachments.

What is the role of emotional robots? How will they interact with us? Do we really want machines that are autonomous, self-directed, with a wide range of behavior, a powerful intelligence, and affect and emotion? I think we do, for they can provide many benefits. Obviously, as with all technologies, there are dangers as well. We need to ensure that people always maintain oversight and control, that they serve human needs appropriately.

Will robot teachers replace human teachers? No, but they can complement them. Moreover, they could be sufficient in situations where there is no alternative – to enable learning while traveling, or while in remote locations, or when one wishes to study a topic for which there is not easy access to teachers. Robot teachers will help make lifelong learning a practicality. They can make it possible to learn no matter where one is in the world, no matter the time of day. Learning should take place when it is needed, when the learner is interested, not according to some arbitrary, fixed school schedule and, even then, only for the first fraction of a person's life. Robot instructors can help.

Will robots help out in the house? Yes. Will there be robot pets? Yes, but only for those who want them, only for those with a true need that cannot be satisfied in any other way.

Many are bothered by these possibilities, so much so that they reject them out of hand as unethical, immoral. Although I do not do so, I do sympathize with their concerns. However, I see the development of intelligent machines as both inevitable and beneficial. Where will there be benefits? In such tasks as doing dangerous tasks, driving automobiles, piloting commercial vessels, in education, medicine, and in taking over routine work. Where might there be moral and ethical concerns? Pretty much the same list of activities. Let me review the items on this list in more detail.

Consider some of the benefits. Robots could be – and to some extent already are – used in dangerous tasks, where people's lives are at risk. This includes such things as search and rescue operations, exploration, and mining. What are the problems? The major ones come from use of machines to enhance illegal or unethical activities: robbery, murder, and terrorism.

Will robot cars replace the need for human drivers? I hope so. Every year, tens of thousands of people are killed, hundreds of thousands seriously injured, through motor vehicle accidents. Wouldn't it be nice if automobiles were as safe as commercial aviation? Here is where automated vehicles can be a wonderful saving. Moreover, automated vehicles could drive more closely to one another, helping to dissolve traffic congestion, and they could drive more economically, helping to solve some of the energy issues associated with driving

Driving an automobile is deceptively simple: most of the time it takes little skill. As a result, many are lulled into a false sense of security and self-confidence. But when danger arises, it does so rapidly, and then the distracted, the semiskilled, the untrained, and those temporarily impaired by drugs, alcohol, illness, fatigue, and sleep deprivation are often incapable of reacting properly in time. Even commercial, well-trained drivers have accidents: automated vehicles will not reduce all accidents and injuries, but they stand a good chance of dramatically reducing the present toll. Yes, some people truly enjoy the sport of driving, but these could be accommodated on special roads, recreational areas, and race tracks. Automation of everyday driving would lead to loss of jobs for drivers of commercial vehicles, but with a saving of life, overall.

I foresee robots as tutors. Many people argue against this, stating that only human beings would have the care and attention required to teach children. But this criticism of machine teaching is an emotional argument, with little evidence to back it up (a strange criticism for me to make, given this book, I must admit). For all of history, humankind have relied upon human teachers, but what is the result? Yes, many of us are well educated, but a much larger number are not. The fact that almost every nation in the world believes its children to be ill-educated, unprepared for the modern world does not lead me to have much faith in the power of human teachers, even though I, myself, am a teacher. It isn't that people are bad teachers: it is that the practical reality of trying to find sufficient highly qualified people to act as teachers for the entire world, to pay enough money to train and support them, seems beyond the capability of the nations of the world.

The part that most excites me, however, is the potential for changing the way we teach. Today's model is far too often that of a pedant lecturing at the front of the classroom, forcing students to listen to material they have no interest in, that appears irrelevant to their daily lives. Lectures – and textbooks – are the easiest way to teach from the point of view of the teacher, but the least effective for the learner. The most powerful learning takes place when well-motivated students get excited by a topic and then struggle with the concepts, learning how to apply them to issues they care about. Yes, struggle: learning is an active, dynamic process, and struggle is a part of it. But when students care about something the struggle is enjoyable. This is how great teaching has always taken place – not through lecturing, but through apprenticeship, coaching, and mentoring. This is how athletes learn. This is the essence of the attraction of video games, except that in games, what students learn is of little practical value. These methods are well known in the learning sciences, where they are called problem-based, inquiry-learning, or constructivist.

Here is where emotion plays its part. Students learn best when motivated, when they care. They need to be emotionally involved, to be drawn to the excitement of the topic. This is why examples, diagrams and illustrations, videos and animated illustrations are so powerful. Learning need not be a dull and dreary exercise, not even learning of what are normally considered dull and dreary topics: every topic can be made exciting, every topic excites the emotions of someone – why not excite everyone? It is time for lessons to become alive, for history to be seen as a human struggle, for students to understand and appreciate the structure of art, music, science, and mathematics. How can these topics be made exciting? By making them relevant to the lives of each individual student. This is often most effective by having students put their skills to immediate application. Developing exciting, emotionally engaging, and intellectually effective learning experiences is truly a design challenge worthy of the best talent in the world.

Robots, machines, or computers can be of great assistance in instruction by providing the framework for motivated, problem-based learning. Computer learning systems can provide simulated worlds in which students can explore problems in science, literature, history, or the arts. Robot teachers can make it easy to search the world's libraries and knowledge bases. Human teachers will no longer have to lecture, but instead can spend their time as coaches and mentoring, helping teach not only the topic, but also how best to learn, so that the students will maintain their curiosity thorough life, as well as the ability to teach themselves when necessary. Human teachers are still essential, but they can play a different, much more supportive and constructive role than they do today.

Moreover, although I believe strongly that we could develop efficient robot tutors, perhaps as effective as Stephenson's *Young Lady's Illustrated Primer* (see page 000), we would not have to abandon human teachers: automated tutors -- whether books, machines, or robots -- should act as supplements to human

instruction. Even Stephenson writes in his novel that his star pupil knew nothing of the real world and of real people because she had spent far too much time locked up in the fantasy world of the book.

Robots in medicine? Yes, they could be used in all its aspects. In medicine, however, as in many other activities, I foresee this as a partnership, where well-trained human medical personnel work with specialized robotic assistants to increase the quality and reliability of care.

Laser surgery on eyes is now close to complete machine control, and any activity where great precision is required is a candidate for machine operation. Machine diagnosis is more tricky, and I suspect that skilled physicians will always be involved, but that they would be aided by dynamic, intelligent machines that can assess a large database of prior cases, medical records, medical facts, and pharmaceutical information. This assistance is already required, as the amount of information – and the rapid addition of new information – becomes overwhelming to practicing physicians. Moreover, as we get better diagnostic tools – more efficient analyses of body fluids and physiological records, DNA analyses, and various body scans -- where some of the information will be routinely collected and sent from a patient's home or even place of work to the medical office, only a machine could keep up with the information. People are excellent at synthesis, at dynamic, creative decisions, at seeing the whole, global picture, whereas machines are superb at rapid search through large numbers of cases and information files, without being subject to the biases that accompany human memory. The team of trained physician and robotic assistant would be far superior to either working alone.

One common fear, of course, is that robots will take over many routine jobs from people, therefore leading to great unemployment and turmoil. Yes, more and more machines and robots will take over jobs, not only of lower-skilled workers, but increasingly of much routine work of all kinds, including some management. Throughout history, each new wave of technology has displaced workers, but the total result has been increased life span and quality of living for everyone, including, in the end, increased jobs – although of a different nature than before. In transitional periods, however, people are displaced and unemployed, for the new jobs that result often require skills very distant from those of the people who have been displaced. This is a major social problem that must be addressed.

In the past, most of the jobs replaced by automation have been low-level jobs, jobs that did not require much skill or education to perform. In the future, however, robots are apt to replace some highly skilled jobs. Will film actors be replaced by computer generated characters that sound and act just as realistic, but are much more under the control of the director? Will robot athletes compete, if not with humans, then perhaps in their own leagues -- but thereby leading to the demise of human leagues? Such a situation might very well happen with chess tournaments and leagues, now that computer chess players can beat even the best human players. What about jobs such as accounting, bookkeeping, drafting, stock keeping, or even simple management jobs: will these be replaced? Yes, all this is possible; some of it has already started. Robot musicians? The list of potential activities is large, along with the dangers of social upheaval.

When robots are used for activities such as space exploration, dangerous coal mining, or search and rescue missions to help people in dangerous situations, or even when they do simple things around the house, such as lawn mowing, vacuum cleaning, and chores, there is not apt to be much resistance. But when they start taking over some of the jobs of high-paid professionals, or displacing large amounts of people from routine activities, then this does become a legitimate concern, one for which I do not have a satisfactory answer.

I believe that we should welcome machines that eliminate the dreary tedium of many jobs – the dull shuffling of paperwork probably being even more demeaning than many of the low-paid, routine service jobs. This welcome, of course, assumes that machines will free people to engage in more creative activities, where they can apply their abilities both more pleasurably and, effectively.

I have visited many parts of the world where poverty, continual hunger and starvation, and high death rates have made me doubt the benefits of today's systems. I have seen silk factories in India where young girls are locked into buildings, forced to weave from early morning till evening, locked in so that they cannot leave – or even escape the building if there is fire – without someone from the outside unlocking the doors. My study of history has taught me that this inequity, brutality, and callous treatment of so many is not unusual, and long predates the development of modern technology.

Yes, I see the downside of the deployment of intelligent machines and robots, but I also see the downside of no deployment. Call me an optimist, if you wish, but I believe that in the end, the human ingenuity that we show in creating these powerful devices will also serve us in ways to create more enriching, more enlightened activities for all of us. Optimism does not blind me to the inequities and problems of today's life: optimism reflects my belief that we can overcome them in the future. Yes, we still have poverty, starvation, political inequity, and wars, but these result more from the evils of people than from our technologies. I do not see why the introduction of smart, emotional robots and machines will change this situation, either for the worse or for the better. To change evil, we must confront it directly. It is a social, political, and human problem: not a technological one. This, of course, does not minimize this problem nor does it absolve us from working toward a solution. But the solution must be social and political, not technological.

The story becomes even more complex if I expand the view beyond the short-term horizon. At some point, robots and other machines are apt to become truly autonomous. This is a long time away, perhaps centuries, but it will happen. Then, there will indeed be major disruptions of life when much or all human work can be done by robots: farming, mining, manufacturing, distribution, and sales. Education and medicine. Even many aspects of art, music, literature, and entertainment. Robots may manufacture themselves. At that point, the relationship between natural animals and robots becomes exceedingly complex. The complexity will be amplified because many humans will actually be cyborgs – part human, part machine. Artificial implants already exist, mostly as medical prostheses; but some people are talking about having them implanted artificially, the better to enhance natural capabilities. Strength, athletic ability, sensory capability, memory and decision making could all be aided by implanted, electronic, chemical, mechanical, biological, or nanotechnology devices. Steroids are used by athletes to enhance their existing strength, and laser treatment of the cornea has been done by some athletes and pilots to enhance normal acuity. The artificial lenses in my eyes – necessary to remove cataracts – have provided me with far better vision than I have ever experienced before, with the sole problem being that my eyes can only focus at a distance. But someday, artificial lenses will be able to focus, probably even better than natural ones, perhaps providing telescopic in addition to normal vision. When this happens, even people who do not have cataracts might wish to have their normal lenses replaced by these more effective ones. Even more striking artificial enhancement is possible. All of these possibilities raise complex ethical issues, but these truly go beyond the boundaries of this book.

But this book does focus upon emotions and their role in the development of artificial devices and the way that human beings emotionally attach themselves to their belongings, their pets, and to one other. Robots might act like all of these.

At first, robots will be belongings, but ones with clear personal attachment, for if a robot is with you for a large part of your life, able to interact, to remind you of your experiences, to give advice, or even just comic relief, there will be strong emotional attachments. Even today's robot pets, crude though they may be, have already invoked strong emotions among their owners. In the decades to come, robot pets may take on all the attributes of real pets and, in the minds of many people, be superior. Today people abuse and abandon their pets. Many communities have bands of stray abandoned cats or dogs scavenging. Might the same happen with robotic pets? Who is legally responsible for their care and maintenance? What if a robot pet injures someone? Who is legally responsible? The robot? The owners? The designer or manufacturer? With real pets, the owner is responsible.

And finally, what happens when robots act as independent, sentient beings, with their own hopes, dreams, and aspirations? Will something akin to Asimov's Laws of Robotics be necessary? Will they be sufficient? If robot pets can cause damage, what might an autonomous robot do; and if a robot causes damage, injury, or death, who is to blame, and what is the recourse? Asimov concluded in his novel *I, Robot*, that robots will indeed take over, that mankind will lose its own say in its future. Science fiction? Yes, But all future possibilities are fiction before they are fact.

We are in a new era. Machines are already smart, and they are getting smarter. They are developing motor skills. And soon they will have affect and emotion. The positive impact will be enormous. The negative consequences are also real. This is how it is with all technology: it is a two-edged sword always combining potential benefits with potential deficits. But there is no reason why we cannot control our technology. The future, after all, is up to us.

REFERENCES

Asimov, I. (1950). *I, Robot*. London: D. Dobson. (Reprinted numerous times, and including: Asimov, I. (1983). *The Foundation trilogy: Foundation, Foundation and Empire, Second Foundation; The stars, like dust; The naked sun; I, robot*. New York N Y: Octopus/Heinemann.)

Asimov, I. (1983). *The Foundation trilogy: Foundation, Foundation and Empire, Second Foundation; The stars, like dust; The naked sun; I, robot*. New York N Y: Octopus/Heinemann.

Asimov, I. (1985). *Robots and empire* (1st ed.). Garden City N Y: Doubleday.

Clarke, R. (1993). Asimov's Laws of Robotics: Implications for Information Technology, Part 1. *IEEE Computer*, 26 (12), 53-61.
<http://www.anu.edu.au/people/Roger.Clarke/SOS/Asimov.html>

Clarke, R. (1994). Asimov's Laws of Robotics: Implications for Information Technology, Part 2. *IEEE Computer*, 27 (1), 57-66.
<http://www.anu.edu.au/people/Roger.Clarke/SOS/Asimov.html>

Industrial Robots and Robot System Safety. Occupational Safety and Health Administration, US Department of Labor, OSHA Technical Manual (TED 1-0.15A). (1999). <http://www.osha.gov/SLTC/machineguarding/publications.html> Section V entitled "Control and Safeguarding Personnel" outlines specific means for safeguarding robot systems.

Johnson, S. (2001). *Emergence: the connected lives of ants, brains, cities, and software*. New York: Scribner.

END NOTES

¹ “most of the world governments banned robot use.” (Asimov, 1950)

² “Asimov’s Four Laws of Robots.” Roger Clarke, in his writings and on his authoritative website (Clarke, 1993, 1994), dates the origins of laws one, two, and three from Clarke’s discussion with science fiction author and editor John Campbell in 1940 (Asimov, 1950, 1983). The zeroth law was added 45 years later, in 1985 (Asimov, 1985).

³ “Do what I mean, not what I say,.” Note that DWIM (Do What I Mean) is a very old concept: Warren Teitelman introduced it into the command interpretation system of the LISP computer programming system in 1972. When it works, it is very, very nice.

⁴ Central control. A good review of the work on emergent systems, that is, against central control, is in Johnson’s book “Emergence.” (Johnson, 2001).

⁵ “There already are some safety regulations that apply to robots.” (*Industrial Robots and Robot System Safety. Occupational Safety and Health Administration, US Department of Labor, OSHA Technical Manual (TED 1-0.15A)*, 1999)