Countering the Frankenstein Complex

Lee McCauley
Department of Computer Science
University of Memphis
mccauley@memphis.edu

Introduction

At 50 years old, the fields of artificial intelligence and robotics capture the imagination of the general public while, at the same time, engendering a great deal of fear and skepticism. Hollywood and the media exacerbate the problem while some well known authors and scientists lend credence to it. This fear is much older than the relatively modern concepts of AI and robotics. Isaac Asimov recognized this deep-seated misconception of technology and created the Three Laws of Robotics intended to demonstrate how these complex machines could be made safe. For robots to be accepted in everyday society, we must examine the underlying fear of intelligent robots and reevaluate Asimov's response. Not because he understood how to build robots, but to discover if there are lessons to be learned from his insights into society's response to intelligent technology. Asimov knew that fear would be the greatest barrier to success and, consequently, implanted all of his fictional robots with the Three Laws of Robotics. Above all, these laws served to protect humans from almost any perceivable danger. Asimov believed that humans would put safeguards into any potentially dangerous tool and saw robots as just advanced tools.

The "Frankenstein Complex"

In 1920 a Czech author by the name of Karel Capek wrote the widely popular play R.U.R. which stands for Rossum's Universal Robots. The word "robot" comes from the Czech word "robota" meaning 'drudgery' or 'servitude' (Jerz, 2002). As typifies much of science fiction since that time, the story is about artificially created workers that ultimately rise up to overthrow their human creators. Even before Capek's use of the term 'robot', however, the notion that science could produce something that it could not control had been explored most acutely by Mary Shelly under the guise of Frankenstein's monster (Shelley, 1818). The full title of Shelley's novel is "Frankenstein, or The Modern Prometheus." In Greek mythology Prometheus brought fire (technology) to humanity and, consequently, was soundly punished by Zeus. In medieval times, the story of Rabbi Judah Loew told of how he created a man from the clay (in Hebrew, a 'golem') of the Vltava river in Prague and brought it to life by putting a shem (a tablet with a Hebrew inscription) in its mouth. The golem eventually went awry, and Rabbi Loew had to destroy it by removing the shem. This fear of man broaching, through technology, into "God's realm" and being unable to control his own creations is referred to as the "Frankenstein Complex" by Isaac Asimov in a number of his essays (most notably (Asimov, 1978)).

The "Frankenstein Complex" is alive and well. Hollywood seems to have rekindled the love/hate relationship with robots through a long string of productions that have, well, gotten old. To make the point, here is a partial list: Terminator (all three); I, Robot; A.I.: Artificial Intelligence; 2010: a Space Odyssey; Cherry 2000; D.A.R.Y.L; Blade Runner; Short Circuit; Electric Dreams; the Battlestar Galactica series; Robocop; Metropolis; Runaway; Screamers; The Stepford Wives; and Westworld. Even though several of these come from Sci-Fi stories, the fact remains that the predominant theme chosen when robots are on the big or small screen involves their attempt to harm people or even all of humanity. This is not intended as a critique of Hollywood, to the contrary. Where robots are concerned, the images that people can most readily identify with, those that capture their imaginations and tap into their deepest fears, involve the supplanting of humanity by its metallic offspring.

Even well respected individuals in both academia and industry have expressed their belief that humans will engineer a new species of intelligent machines that will replace us. Ray Kurzweil (1999; , 2005), Kevin Warwick (2002), and Hans Moravec (1998) have all weighed in on this side. Bill Joy, co-founder of Sun Microsystems, expressed in a 2000 Wired Magazine article (Joy, 2000) his fear that artificial intelligence would soon overtake humanity and would, inevitably, take control of the planet for one purpose or another. The strongest point in their arguments hinges on the assumption that the machines will become too complicated for humans to build using standard means and will, therefore, relinquish the design and manufacture of future robots to intelligent machines themselves. Joy argues that robotics, genetic engineering, and nanotechnology pose a unique kind of threat that the world has never before faced,

"robots, engineered organisms, and nanobots share a dangerous amplifying factor: They can self-replicate. A bomb is blown up only once – but one bot can become many, and quickly get out of control." Clearly, Joy is expressing the underpinnings of why the public at large continues to be gripped by the Frankenstein Complex.

Society Can Breath Easy

The likelihood of a robotic uprising that destroys or subjugates humanity is quite low. The primary argument that robots will take over the world is that they will eventually be able to design and manufacture themselves in large numbers thereby activating the inevitability of evolution. Once evolution starts to run its course humanity is out of the loop and will eventually be rendered superfluous. On the surface this seems like a perfectly logical argument that strikes right at the heart of the Frankenstein Complex. However, there are several key assumptions that must hold in order for this scenario to unfold as stated.

First, there is the underlying assumption that large numbers of highly intelligent robots will be desired by humans. At first, this might seem reasonable. Why wouldn't we want lots of robots to do all the housework, dangerous jobs, or any menial labor? If those jobs require higher-order intelligence to accomplish, then we already have a general purpose machine that is cheaply produced and in abundance – humans. If they do not require higher-order intelligence, then a machine with *some* intelligence can be built to handle that specific job more economically than a highly intelligent general robot. In other words, we may have smarter devices that take over some jobs and make others easier for humans, but those devices will not require enough intelligence to even evoke a serious discussion of their sentience. Therefore, we will see the mass production of *smart enough* devices, but not general purpose robots or artificial intelligences. This is not to say that we won't create in some lab a human-level artificial intelligence. We will do it because we can. These will be expensive research oddities that will get a lot of attention and raise all of the hard philosophical questions, but their numbers will be low and they will be closely watched because of their uniqueness.

Second, the assumption is made that evolution will occur on an incredibly fast time scale. There are a couple of ways that this might come about. One argument goes that since these machines will be produced at such a high rate of speed, evolution will happen at a predacious rate and that humans will not be able to control. How fast would intelligent robots evolve? Let's compare with human evolution. In 2006, there will be more than 130 million human births on the planet. In 1998 just fewer than 16.5 million personal computers were manufactured in the U.S. While computer components are built all around the world, the vast majority are assembled in the U.S. For argument's sake, let's say that the world's production of computers is twice that, some 33 million. Let's also assume that that number has quadrupled since 1998 to 132 million computers manufactured worldwide in one year, about equal to our number of humans produced. Computers are moderately complex machines created at a rate at least as fast as our future intelligent robots might be. Evolution works, outside of sexual reproduction, by making mistakes during the copying of one individual – a mutation. If we allow that our manufacturing processes will make mistakes on par with biological processes, then the evolution of our reproducing machines would be roughly equal to that of human evolution – if one discounts the huge effect of genetic crossover via sexual reproduction. Even if we allowed for the jumpstarting of the process by already having a fairly intelligent robot running the manufacturing show, this would be comparable to starting with an Australopithecus and waiting to come up with a Homo sapiens sapiens. Furthermore, each robot produced must have all of the knowledge, capability, resources, time and motivation to build more robots, otherwise the mutations don't propagate and evolution goes nowhere. Why would we give our house cleaning robot the ability to reproduce on its own? To sum up, if we start with a fairly intelligent seed robot that can reproduce, and it builds copies of itself, and each one of the copies builds copies of themselves on and on to create large numbers of reproducing robots, then it will take thousands of years for the process to create any meaningful changes whatsoever, much less a dominant super species. There are no likely circumstances under which this sort of behavior would go on unchecked by humans.

There is another scenario that would increase the rate of evolution. Humans could build a robot or AI with the sole task of designing and building a new AI that is better at designing and building AI which builds another AI, etc., etc. This directed sort of evolution is likely to be much quicker and is also likely to be something that an AI researcher might try. This would, ultimately, also be a very expensive endeavor. Either the AI is custom built in hardware with each successive version or it is created in a virtual manner and run within some larger system. This system would likely need to be quite large if the AI is intended to

be truly intelligent. As the versions become more and more adept and complex, the system that houses the AI would need to become increasingly complex and ultimately a proprietary machine would need to be created whose purpose would be to run the AI. We are then back to the custom hardware versions and progression from there. Another problem with this notion is that the very first AI to begin the process, since we are not using evolutionary processes, will need a great deal of knowledge regarding the nature of intelligence in order to effectively guide the development. Solving the problem of creating a truly intelligent machine, therefore, is almost a "catch 22;" we would have to already know how to create an intelligent machine before we could create one. One might still argue that this could be implemented using some form of learning or genetic algorithm based on some general intelligence measure. An evolutionary version of this possibility would need to simulate whole populations of human-level intelligent entities and, therefore, be even more massive. Even if this is implemented at some point in the future, it is not something that will be accessible by your everyday hacker due to the cost and will, therefore, be relegated to a small number of academic institutions or corporations. It is, therefore, the responsibility of those few researchers to consider the ramifications of what they create.

The third assumption underlying our doomsday of reproducing robots is that humans would never actually check to see if the robots produced deviated from the desired output. Especially if they are being mass produced, this seems quite out of the question. Approximately 280 cars were sacrificed to crash tests alone in 2006 just by the Insurance Institute for Highway Safety and National Highway Traffic Safety Administration. Every model sold in the United States undergoes a huge battery of tests before it is allowed on the streets. Why would robots be any less regulated? This further reduces the chances of evolutionary style mutations. Of course there will still be defects that crop up for a given robot that did not show up in the tests just as with automobiles. Also, just as with automobiles, these defects will be dealt with and not passed on to future generations of robots.

The Future

Many well known people have told us that the human race is doomed to be supplanted by our own robotic creations. Hollywood and the media sensationalize and fuel our fears because it makes for an exciting story. However, when one analyzes the series of improbable events that must occur for this to play out, it becomes obvious that we are quite safe. Is there still a possibility that we will be overrun by our metallic progeny? There is still the possibility of technology misuse and irresponsibility on the part of robotics and AI researchers that, while not resulting in the obliteration of humanity, could be disastrous for the people directly involved. For this reason, Bill Joy's call for scientists and engineers to have a Hippocratic Oath (Joy, 2000) has been taken up for roboticists and researchers of artificial intelligence (McCauley, 2006). The Roboticist's Oath calls for personal responsibility on the part of researchers and to instill in their creations the spirit of three rules stemming from Isaac Asimov's original Three Laws of Robotics.

The future will be filled with smart machines. These will get smarter but not sentient, not alive. A small number of robots in labs may achieve human-level or better intelligence, but these will be closely studied oddities. Can the human race still destroy itself? Sure, but not through artificial intelligence. Humanity must always be wary of its power and capability for destruction. It must also not fear the future with or without intelligent robots.

References

Asimov, I. (1978). The Machine and the Robot. In P. S. Warrick, M. H. Greenberg & J. D. Olander (Eds.), *Science Fiction: Contemporary Mythology*: Harper and Row.

Jerz, D. G. (2002). R.U.R. (Rossum's Universal Robots) [Electronic Version]. Retrieved June 7, 2006 from http://jerz.setonhill.edu/resources/RUR/.

Joy, B. (2000). Why the future doesn't need us. *Wired*.

Kurzweil, R. (1999). *The Age of Spiritual Machines*: Viking Adult.

Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*: Viking Books.

McCauley, L. (2006). AI Armageddon and the Three Laws of Robotics (pp. 23): University of Memphis.

Moravec, H. P. (1998). *Robot: Mere Machine to Transcendent Mind*. Oxford: Oxford University Press.

Shelley, M. (1818). *Frankenstein, or The Modern Prometheus*. London, UK: Lackington, Hughes, Harding, Mavor & Jones.

Warwick, K. (2002). *I, Cyborg*: Century.