from making us stupid, these technologies are the only things that will keep us smart.

2 3

Subject

Boundaries Between Humans and Machines

Humanoid Robotics: Ethical Considerations

David Bruemmer

purpose: informative

May 30, 2006

The Idaho National Laboratory

http://www.inl.gov/adaptiverobotics/humanoidrobotics/ethicalconsiderations.shtml

Intro

The world's population of real humans continues to steadily grow. One might ask why we would want to make a machine that looks, thinks and emotes like a human when we have plenty of humans already, many of whom do not have jobs or good places to live. It is important to re-emphasize that humanoids cannot and will not ever replace humans. Computers and humans are good at fundamentally different things. Calculators did not replace mathematicians. They did change drastically the way mathematics was taught. For example, the ability to mentally multiply large numbers, although impressive, is no longer a highly valued human capability. Calculators have not stolen from us part of what it means to be human, but rather, free our minds for more worthy efforts. As humanoids change the contours of our workforce, economy

Thesis

Thesis { and society, they will not encroach on our sovereignty, but rather enable us to explore and further realize the very aspects of our nature we hold most dear.

supporting { So why should we have intelligent, emotion exhibiting humanoids? Emotion is often considered a debilitating, irrational characteristic. Why not keep humanoids, like calculators, merely as useful gadgetry? If we do want humanoids to be truly reliable and useful, they must be able to adapt and develop. Since it is impossible to hard-code high-utility, general-purpose behavior, humanoids must play some role as arbiters of their own development. One of the most profound questions for the future of Humanoid Robotics is, "How we can motivate such development?" Speaking in purely utilitarian terms, emotion is the implementation of a motivational system that propels us to work, improve, reproduce and survive. In reality, many of our human "weaknesses" actually serve powerful biological purposes. Thus, if we want useful, human-like robots, we will have to give them some motivational system. We may choose to call this system "emotion" or we may reserve that term for ourselves and assert that humanoids are merely simulating emotion using algorithms whose output controls facial degrees of freedom, tone of voice, body posture, and other physical manifestations of emotion.

example { Most likely, two distinct species of humanoids will arise: those that respond to and illicit our emotions and those we wish simply to do work, day in and day out, without stirring our feelings. Some ethicists believe this may be a difficult distinction to maintain. On the other hand, many consider ethical concerns regarding robot emotion or intelligence to be moot. According to this line of reasoning, no robot really feels or knows anything that we have not (albeit indirectly) told them to feel or know. From this perspective, it seems unnecessary to give a second thought to our treatment of

Most likely, two distinct species of humanoids will arise: those that respond to and illicit our emotions and those we wish simply to do work, day in and day out, without stirring our feelings. Some ethicists believe this may be a difficult distinction to maintain. On the other hand, many consider ethical concerns regarding robot emotion or intelligence to be moot. According to this line of reasoning, no robot really feels or knows anything that we have not (albeit indirectly) told them to feel or know. From this perspective, it seems unnecessary to give a second thought to our treatment of humanoids. They are not 'real.' They are merely machines.

At their onset, all technologies seem artificial, and upset the perceived natural way of things. With the rise of the Internet, we coined the notion of a "virtual world" as a means to distinguish a new, unfamiliar arena from our usual daily life. This once clean distinction between a "real world" and "virtual world" already seems ephemeral. To someone who spends 10 hours a day logged into Internet chat rooms, the so-called "virtual world" is as real to them as anything else in their lives. Likewise, the interactions humans have with humanoids will be real because we make them so. Many years from now, our children will be puzzled by the question, "Does the robot have 'real' intelligence?" Intelligence is as intelligence does. As we hone them, enable them to self-develop, integrate them into our lives and become comfortable with them, humanoids will seem (and be) less and less contrived. Ultimately, the most relevant issue is not whether a robot's emotion or intelligence can be considered 'real,' but rather the fact that, real or not, it will have an effect on us.

The real danger is not that humanoids will make us mad with power, or that humanoids will themselves become super intelligent and take over the world. The consequences of their introduction will be subtler. Inexorably, we will interact more with machines and less with each other. Already, the average American worker spends astonishingly large percentages of his/her life interfacing with machines. Many return home only to log in anew. Human relationships are a lot of trouble, forged from dirty diapers, lost tempers and late nights. Machines, on the other hand, can be turned on and off. Already, many of us prefer to forge and maintain relationships via e-mail, chat rooms and instant messenger rather than in person. Despite promises that the Internet will take us anywhere, we find ourselves—hour after hour—glued to our

chairs. We are supposedly living in a world with no borders. Yet, at the very time we should be coming closer together, it seems we are growing further apart. Humanoids may accelerate this trend.

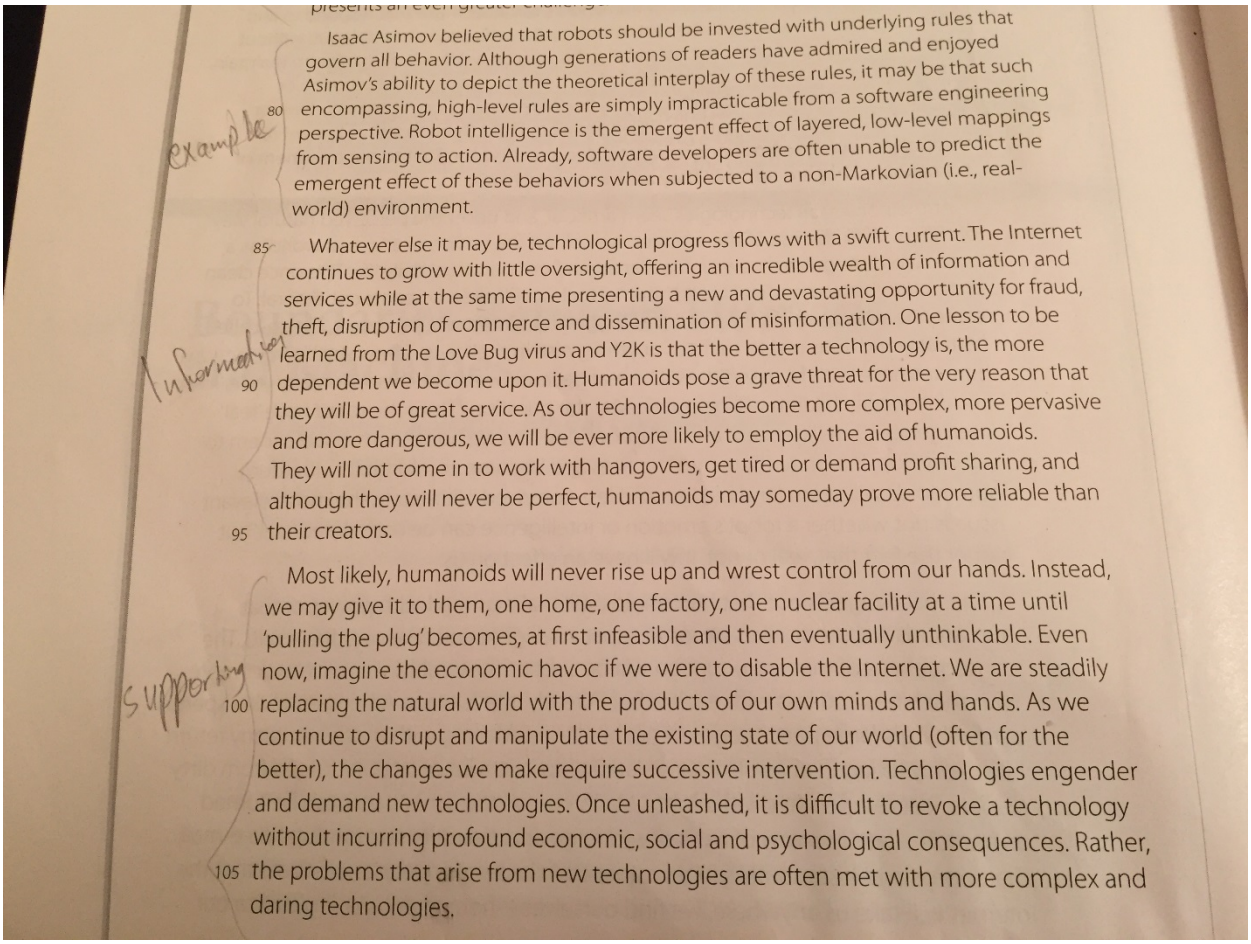
If it is hard to imagine how humans could develop an emotional connection to a robot, consider what the effects would be of systematically imparting knowledge, personality and intentions to a robot over a sustained period of time. It may well be that much of the software for intelligent humanoid robot control is developed under an Open Source paradigm, which means that thousands or even millions of developers will be able to modify the software of their own or other people's robots. Source code aside, humanoids will be given the ability to develop and learn in response to the input they receive. Could a cruel master make a cruel humanoid? Will people begin to see their robots as a reflection of themselves? As works of art? As valuable tools? As children? If humanoids learn "bad behavior," whom should we hold responsible? The manufacturer? The owner? The robot? Or the surrounding environment as a collective whole? The ethical question of nature vs. nurture is relevant for humanoids as well as humans. It will be hard enough to monitor the software and mechanical 'nature' of humanoids (i.e., the state in which humanoids emerge from the factory crate). 'Nurture' presents an even greater challenge.

Example

Isaac Asimov believed that robots should be invested with underlying rules that govern all behavior. Although generations of readers have admired and enjoyed Asimov's ability to depict the theoretical interplay of these rules, it may be that such encompassing, high-level rules are simply impracticable from a software engineering perspective. Robot intelligence is the emergent effect of layered, low-level mappings from sensing to action. Already, software developers are often unable to predict the emergent effect of these behaviors when subjected to a non-Markovian (i.e., real-world) environment.

Example

Whatever else it may be, technological progress flows with a gentle current. The robot

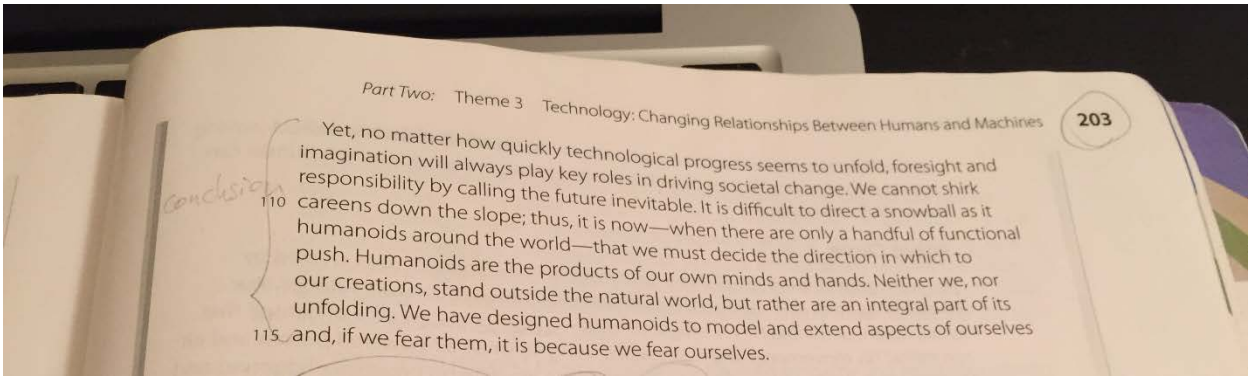


presents an even greater challenge.

Isaac Asimov believed that robots should be invested with underlying rules that govern all behavior. Although generations of readers have admired and enjoyed Asimov's ability to depict the theoretical interplay of these rules, it may be that such encompassing, high-level rules are simply impracticable from a software engineering perspective. Robot intelligence is the emergent effect of layered, low-level mappings from sensing to action. Already, software developers are often unable to predict the emergent effect of these behaviors when subjected to a non-Markovian (i.e., real-world) environment.

Whatever else it may be, technological progress flows with a swift current. The Internet continues to grow with little oversight, offering an incredible wealth of information and services while at the same time presenting a new and devastating opportunity for fraud, theft, disruption of commerce and dissemination of misinformation. One lesson to be learned from the Love Bug virus and Y2K is that the better a technology is, the more dependent we become upon it. Humanoids pose a grave threat for the very reason that they will be of great service. As our technologies become more complex, more pervasive and more dangerous, we will be ever more likely to employ the aid of humanoids. They will not come in to work with hangovers, get tired or demand profit sharing, and although they will never be perfect, humanoids may someday prove more reliable than their creators.

Most likely, humanoids will never rise up and wrest control from our hands. Instead, we may give it to them, one home, one factory, one nuclear facility at a time until 'pulling the plug' becomes, at first infeasible and then eventually unthinkable. Even now, imagine the economic havoc if we were to disable the Internet. We are steadily replacing the natural world with the products of our own minds and hands. As we continue to disrupt and manipulate the existing state of our world (often for the better), the changes we make require successive intervention. Technologies engender and demand new technologies. Once unleashed, it is difficult to revoke a technology without incurring profound economic, social and psychological consequences. Rather, the problems that arise from new technologies are often met with more complex and daring technologies.



Yet, no matter how quickly technological progress seems to unfold, foresight and imagination will always play key roles in driving societal change. We cannot shirk responsibility by calling the future inevitable. It is difficult to direct a snowball as it careens down the slope; thus, it is now—when there are only a handful of functional humanoids around the world—that we must decide the direction in which to push. Humanoids are the products of our own minds and hands. Neither we, nor our creations, stand outside the natural world, but rather are an integral part of its unfolding. We have designed humanoids to model and extend aspects of ourselves and, if we fear them, it is because we fear ourselves.