

George Bekey: Robots and Ethics

How and why did you get interested in the field of autonomous robots and specifically in military robots?

My interest in robotics developed as a synthesis of a number of technologies I had studied. My PhD thesis was concerned with mathematical models of human operators in control systems, e.g., a pilot controlling an aircraft. The goal was to develop a mathematical representation of the way in which a pilot (or other human operator of a complex system) generates an output command, such as movement of the control stick on the aircraft) in response to changes in the visual input. This work led to increasing interest in human-machine systems. Shortly after completing my graduate studies I developed a hybrid analog-digital computer at a Los Angeles aerospace company. The goal of this project was simulation of the flight of an intercontinental ballistic missile, where the flight control system was represented on the analog portion of the system, and the highly precise generation of the vehicle trajectory was done on the digital computer. These early experiences gave me a great deal of insight into military technology, while at the same time improving

my knowledge and skills in computers and control systems. When I joined the University of Southern California in 1962 I continued to work in all these areas. When industrial robots became prominent in the late 1970s it became clear to me that here was a research field which included all my previous experience: human-machine systems, control theory and computers. Further, we hired a young faculty member from Stanford University who had some experience in robots. He urged me to write a proposal to the National Science Foundation to obtain funding for a robot manipulator. I did this and obtained funding for a PUMA industrial robot. From then on, in the 1980s and 90s my students and I worked in robotics, with an increasing interest in mobile robots of various kinds.

You asked specifically about my interest in military robots. As I indicated above, I started working with military systems in the 1960's, but largely left that area for some time. However, when I started looking for funding for robotics research, I found that a large portion of it came from the U.S. Defense Department. While most of the support for my research

came from the US National Science Foundation, during the 1990s I received several large contracts and grants from the Defense Department for work in robotics. While I was pleased to have the funding so that I could support my laboratory and my Ph.D. students, I became increasingly uncomfortable with work on military robots. For many years I had been concerned about the ethical use of technology. This led to participation in a Committee on Robot Ethics of the Robotics and Automation Society, one of main societies forming the professional core of the Institute of Electrical and Electronics Engineers (IEEE), a major international professional organization in the field of electrical engineering.

To summarize: my interest in robotics arose as a way of integrating my various research interests. Military robotics was a major source of research funding, but I was increasingly disturbed by the way robots were being used. Please note that this does not imply a direct criticism of the U.S. military establishment, since I consider myself to be a patriotic person and I believe that countries need a defense structure (since we have not yet learned to solve all disputes among nations by peaceful means). Rather, it represents a desire to contribute to the ethical use of robots, both in military and peacetime applications.

In the recent discussion of military unmanned systems or military robots, it has been argued that especially for future international legislation concerning this matter it would also be necessary to find a universal definition of what constitutes a "robot". How would you define a robot? Should we define robots opposed to intelligent ammunitions and other automated weapon systems or would a broader definition be more useful?

It is interesting that in many of the current discussions of military robots we do not define what we mean by "robot". In my own work I have defined a robot as: *"A machine that senses, thinks and acts"*. This definition implies that a robot:

- Is not a living organism,
- is a physical system situated in the real world, and is not only software residing on a computer,
- uses sensors to receive information from the world
- processes this information using its own computing resources (but these may be special purpose chips, artificial neural networks or other hardware which enable it to make decisions and approximate other aspects of human cognitive functions), and
- it uses actuators to produce some effect upon the world

With this very broad definition it is clear that automated weapon systems *are* robots, to the extent that

they are able to sense the world, process the sensed information and then perform appropriate actions (such as navigation, obstacle avoidance, target recognition, etc). Note that a system may be a robot in some of its functions and not others. Thus, a Predator aircraft is a robot as far as its takeoff, navigation, landing and stability properties are concerned; but it is not a robot with respect to its use as a weapon if the decision to fire and the release of a missile is done under human control. Clearly, if and when the decision to fire is removed from human control and given to the machine, then it would be a "robot" in these actions as well.

It should also be noted that the use of the word "thinks" is purposely vague, but it intentionally allows actions ranging from simple YES/NO binary decisions to complex cognitive functions that emulate human intelligence.

I do not believe that it would be useful to separate "intelligent ammunition" and "automated weapon systems" from robots in general. Clearly, there are (and will be more) military robots, household robots, eldercare robots, agricultural robots, and so on. Military robots are robots used for military purposes.

In your academic career you have published more than 200 papers and several books on robotics and

over the years you have witnessed euphoria and disillusionment in this field. How would you assess the future of robots in the human society?

I believe that robots are now where personal computers were in the 1980s: they are increasingly evident and will be ubiquitous within the next 10 to 20 years. We already see robots in the home doing vacuuming, grass cutting and swimming pool cleaning. We see autonomous vehicles appearing, both in civilian and military contexts. I anticipate that robots will be so integrated into society that we will no longer think of them as separate systems, any more than we consider an automatic coffee maker a robot. However, there are major challenges to the further development of the field, such as: (1) the need to improve the ways in which robots and humans communicate and interact with each other, which is known as human-robot interaction or HRI, and (2) the need for robots to develop at least a rudimentary form of consciousness or self-awareness to allow for intellectual interaction with humans, and (3) the need to insure that robots behave ethically, whether in health care or home assistance or military assignments. Nevertheless, I believe that the so-called "service robots" which provide assistance in the home or the workplace will continue to proliferate. We have seen great successes in entertainment

robots and in various cleaning robots (for carpets, kitchen floors, swimming pools, roof gutters, etc). On the other hand, there have been some attempted introductions that did not succeed. I can think of two important ones: an automobile fueling robot and a window cleaning robot. Some 10 years ago one of the US oil companies developed a robot for filling the gasoline tank of automobiles. A bar code on the windshield provided information on the location of the filler cap. The hose moved automatically to the cap, filled the tank, and returned to its resting position; the amount of the charge was deducted from the balance of the customer's credit card. After some months of testing, the experiment was abandoned, but I still think it is a good idea. Also, several years ago one of the Fraunhofer Institutes in Germany developed a remarkable robot for cleaning windows in high rise buildings. The machine climbed up on the building using suction cups; it washed the windows as it moved and recycled the dirty water, so there was no spillage on to the sidewalk below. It was able to climb over the aluminium separators between window panes. After some significant publicity, it disappeared from public view, but it may still be available in Germany. These two systems are examples of the robotic innovations which will have a major impact on society, by using machines to replace manual labor.

Clearly, such robots will also create major social problems, since all the displaced workers will need to be retrained for jobs requiring higher skills. While there may be objections to such displacements from labor groups, I believe they are inevitable. Another similar area of application lies in agriculture, where current experiments make it clear that robots can perform harvesting, crop spraying and even planting of seedlings; again, low-skilled workers would need training for new jobs.

It has been argued, that for a decision to be ethical, mere rational thought is not sufficient but emotion does play a large role. Apart from the most optimistic prognoses, Artificial Intelligence and therefore robots, will not obtain the full potential of the human mind in the foreseeable future, if at all. However, it is clear, that machines and their programming will become much more sophisticated. Colin Allen and Wendell Wallach have put forward, that machines eventually will obtain a "functional morality", possessing the capacity to assess and respond to moral challenges. How do you think will society respond, if confronted with machines with such a potential?

I basically agree with Allen and Wallach, but let me make a couple of comments. First, your question indicates that emotion plays an

important role in decision making. There is a great deal of research on robot emotions (at Waseda University in Japan and other institutions), i.e., on providing robots with the ability to understand certain emotional responses from their human co-workers and conversely, to exhibit some form of “functional emotion”. This implies, for example, that a robot could display functional anger by agitated movements, by raising the pitch of its voice and by refusing to obey certain commands. Similarly a robot could appear sad or happy. Critics have said that these are not “real” emotions, but only mechanical simulations, and that is why I have called them “functional emotions”. Clearly, a robot does not have an endocrine system which secretes substances into the bloodstream responsible for “emotional” acts. It does not have a human-like brain, where emotional responses may arise in the amygdala or elsewhere. (An excellent discussion of this matter can be found in the book by Jean-Marc Fellous and Michael A. Arbib, “Who Needs Emotions?: The Brain Meets the Robot”). Hence, I believe that a *functional morality* is certainly possible. However, it should be noted that ethical dilemmas may not be easier for a robot than for a human. Consider, for example, a hypothetical situation in which an intelligent robot has been programmed to obey the “Rules of War”, and the “Rules of Engagement” of a particular conflict,

which include the mandate to avoid civilian casualties to the greatest extent possible. The robot is instructed by a commanding officer to destroy a house in a given location because it has been learned that a number of dangerous enemy soldiers are housed there. The robot approaches the house and with its ability to see through the walls, interpret sounds, etc. it determines that there are numerous children in the house, in addition to the presumed dangerous persons. It now faces an ethical conflict: Should it obey its commander and destroy the house, or should it disobey since destruction would mean killing innocent children? With contemporary computers, the most likely result of such conflicting instructions will be that the robot’s computer will lock up, and the robot will freeze in place.

In response to your direct question about society’s response to the existence of robots equipped with such functional morality: I believe that people will not only accept it in robots, but will come to expect it. We tend to expect the best even of our non-robotic machines like our cars: “This car has never let me down in the past...”, or we kick and curse our machines as if they intentionally disobey our requests. It would not surprise me if functional machine morality in robots could become a standard for judging human (biological) morality, but I cannot foresee the consequences for society.

Before completing your Ph.D. in engineering you have also studied world religions and you are also teaching courses which cover Hinduism, Buddhism, Taoism, Judaism, Islam, Zoroastrianism and other frameworks at the California Polytechnic State University. Have these experiences and the richness of human thought influenced your perspective on robots and ethics in robotics?

Of course. I believe that studying robots can also teach us a great deal about ourselves and our relationships with other human beings and the physical world. Robots are not yet conscious creatures, but as computing speed increases and we learn more both about the human brain and artificial intelligence, there will be increasingly interesting and complex interactions between humans and robots. These interactions will include the whole range of ethical and other philosophical issues which confront humans in society. My background in world religions has led me to study the ways in which different societies seek to find meaning in their lives, both individually and collectively. I believe that increasingly complex robots will find places in society where interactions with humans will go beyond mere completion of assigned tasks, but will lead to emotional issues involving anger, attachment, jealousy, envy, admiration, and of course, issues of right

and wrong, e.g., ethics. I believe that it is possible, in fact likely, that future robots will behave in ways that we may consider “good”, e.g., if they help humans or other robots in difficulty by being altruistic; or in ways we may consider “bad”, such as taking advantage of others for their own gain. These and other forms of behavior are one of the major concerns of religion. However, religion is also concerned with the spiritual development of human beings; robots are unlikely to be concerned with such matters.

Though for the time being the question of “robot rights” seems far fetched, do you think that in the future this will be an issue human society will have to deal with?

Yes, but I believe that Kurzweil’s prediction that robots will demand equal rights before the law by 2019, (and that by 2029 they will claim to be conscious) are somewhat exaggerated. Granted that Kurzweil is indeed a genius and that many of his technical predictions have come true, I think that the question of robot rights involves a number of non-technical social institutions, like law, justice, education and government as well as the cultural and historical background of a particular society. As you know, human beings become emotionally attached to inorganic objects, including automobiles and toys. Clearly, as robots acquire more human-like

qualities (appearance, voice, mannerisms, etc.), human attachments to them will grow. Children become so attached to toys that they attribute human qualities to them and may become emotionally disturbed if the toys are lost or damaged. There are stories that US soldiers in Iraq had become so attached to their Packbots that they become emotionally disturbed if their robot is damaged or destroyed, and they insist on some ceremony to mark its demise. Hence, I believe that indeed robots will acquire some rights, and that society will have to learn to deal with such issues. The more “conscious” and “intelligent” and “human-like” the robots become, the greater will be our attachment to them, and hence our desire to award them some rights normally reserved for humans. Please note that I believe such “rights” may be granted spontaneously by people, and may become tradition and law. I think it is much less likely that the robots will *demand* rights before the law, since this implies a high degree of consciousness which is not likely in the near future.

Your paper (together with Patrick Lin and Keith Abney) “Autonomous Military Robotics: Risk, Ethics and Design”¹ is the first systematically laid out study on this topic which has become known to the general public and is being cited by newspapers all over the world. How did

you get involved in this project and did you expect such a resonance?

Actually, Ronald Arkin's work (which has now been published in book form) preceded ours, as did some of the papers of Noel Sharkey and others, although our project had significantly more emphasis on ethics and philosophical issues. As you know from my background, I have been interested in the broader implications of technology from my graduate student days. Several years ago I saw a news item about Patrick Lin who had recently joined Cal Poly's Philosophy Department, describing his interest in ethical issues in nanotechnology and robotics. I contacted him, we wrote a proposal on robot ethics, which was funded by the Office of Naval Research under an umbrella grant to the University after a careful evaluation with many competing proposals. Patrick, Keith and I get along very well, since we have different but complementary backgrounds. We have now submitted a major proposal to the National Science Foundation to study ethical issues involving robots in health care. This study, if it is approved and funded, will be done jointly with Prof. Maja Mataric at USC, and will involve actual robots in her laboratory, used in rehabilitation projects with patients. I am increasingly concerned with the lack of attention to ethical issues

involving robots in health care, and this study will begin to address some of them.

I am glad that you believe there is broad interest in our study. From my point of view, engineers and scientists as a whole are concerned with solving technical problems and answering scientific questions, and they tend to ignore broader social issues. And yet, it is clear that all technology has dual aspects, and may be used for good or evil. This is one of the lessons of ancient philosophies like Taoism, where “good” and “bad” are seen as inseparable aspects of the same reality. In the West we tend to be surprised when something developed for a good purpose is later used to produce harm. Certainly this was the case with Alfred Nobel’s invention, and it is true of robotics as well.

You have pointed out, that robotic technology is used increasingly in health care, yet there is no widespread discussion of its ethical impact and consequences. Where do you see the main challenges of the proliferation of robotic technology in the different aspects of human society?

This is a very interesting question. As I indicated in my response to the previous question, one of my continuing concerns is that engineers who design and build robots are not

concerned with possible ethical consequences of their inventions. In the past, to insure that no harm resulted from new systems, we incorporated “fail-safe” features. Of course, such design changes may come about only after some damage or destruction has occurred. With industrial robots, fences, enclosures and other protective systems were incorporated only after a robot in Japan malfunctioned and killed a worker. As robots are increasingly integrated in society, both in industry and in the home, the possibility of harmful malfunctions will also increase. Again, I suspect that many of the design features to protect people will not come about until some serious damage is done. There is a common belief in the US that new traffic control signals are not installed at intersections until a child is killed there by an automobile. Now, let us extrapolate this danger to a time in the future, say 20 or 30 or 40 years hence, when robots have been supplied with computers and software systems which enable them to display a level of intelligence and varieties of behavior approaching that of human beings. Clearly, we will not be able to predict all possible unethical actions of such robots. After all, Microsoft is not able to predict all possible malfunctions of a new operating system before it is released, but the consequences of such malfunctions in robots working in close proximity to humans are

much more serious. Consider some questions: Could a robot misinterpret physical punishment of a child as a violent act it must prevent, and thus cause harm to the parent? Or, would be constrained by some new version of Asimov's laws and not be able to defend its owner against a violent intruder, if it is programmed never to injure a human being? If the electricity fails in a home, what would a household robot do to protect its own power supply (and hence, its very existence? Will there be conflicts between robots designed for different functions if they interpret commands in a different way?

So, my answer to your question is that as robots proliferate in society, the potential ethical conflicts will also proliferate, and that we will be ill-prepared to handle them. There will be after-the-fact patches and modifications to the robots, both in hardware and software, since we will be unable to foresee all the possible problems from their deployment. Certainly every new generation of robot designers will attempt to incorporate lessons learned from earlier systems, but like in other systems, they will be constrained by such issues as cost, legal restrictions, tradition, and competition, to say nothing of the difficulty of implementing ethical constraints in hardware and software. We are moving into uncharted waters, so that we cannot

predict the main challenges resulting from the introduction of new robots.

¹ http://ethics.calpoly.edu/ONR_report.pdf.