Noel Sharkey:
Moral and Legal Aspects of Military Robots

How and why did you get interested in the field of robots, especially in military robots and their ethical challenges?

I have been working and conducting research and moving around the fields of Psychology, Cognitive Science, Artificial Intelligence, Engineering, Philosophy, Computer Science and Robotics for about 30 years. I am probably best known in the academic world for my work on neural network learning. A big motivation for me has been questions about the nature of mind that started when I was a teenager – I still haven't found the answers. Robotics became a favourite because it is so rich in challenges in a great variety of areas from sensing and control to construction and everything in between.

My background is not in ethics. I have had a private interest in ethical issues such as the treatment of animals, torture and mistreatment of humans, human rights, social justice and equality, and universal rights for children as long as I can remember and always like to dabble in philosophy but not professionally. I have no pretentions to being a moral philosopher and don't even have a coherent moral theory (yet). So it has all been a very sharp learning curve.

Most of my research now gets classed as applied ethics and I would describe myself as an ethical mongrel – a dash of virtue ethics with a bit of duty ethics, a drop of the deontological, and a healthy helping of consequentialism. I have a sense of what I think is fair and just and loot and plunder from the great ethical thinkers of the past. I am not ashamed to admit that I still have an incredible amount to learn.

I came into the area of robot ethics and the ethics of emerging technologies through the backdoor. I gained a high public profile in the UK through involvement in popular BBC TV programmes about robots and also from some major museum robotics projects – doing science in the public eye. This gave me great access to the public and led to a passion for public engagement and to a Research Council fellowship (EPSRC) with a remit to both encourage more young people into science and engineering and to engage with the public about issues of concern within my expertise.

Engagement means not just talking *at* the public but taking their point of view seriously and feeding it back to the appropriate bodies and policy makers and using the media to affect change on their behalf. I am very committed to the idea that senior academics[1] have a responsibility to the public. What I have found so attractive about public dialogue is that I most often learn more from them than they do from me.

My discussions with the public from around the world began to become more about the ethical issues in the application of technology and journalists were beginning to ask me about military robots. (What may seem surprising to some people is that personal conversations with journalists can provide a lot of solid information.) So I began to read all of the US military plans for robots and unmanned systems that I could get hold of.

From my knowledge of the limitations of AI and robotics, this set extreme alarm bells ringing. I was quite shocked by what I read – particularly the push toward autonomous systems applying lethal force. I felt a very strong urge to give priority to reading and writing and letting the public know about the dangers of this area. So I immersed myself in a study of military issues, the laws of war, Just War theory and the Geneva conventions and the various protocols as well as the legal aspects.

This opened the debate considerably for me and led to more focussed discussions and talks with a great number of people including the military themselves. I have been researching and writing both newspaper and journal articles about the issues ever since (as well as about a number of other ethical issues).

Although an increasingly number of people is beginning to express doubts, you are one of the people in the field, who for quite some time have been openly critical about the use of autonomous military systems with lethal potential. How do you see your role in the discussion of unmanned military systems?

I like to see myself as an unelected representative speaking on behalf of the public, to express their concerns and to inform them and policy makers about the issues involved.

I take opportunities to highlight the problems as they arise. Thinking about it now, I guess that there have been five major components to my role:

(i) providing a sanity check on the limitations of what the technology can do and is unlikely to be able to do soon;
(ii) keeping the issues in the public eye through the media and keeping a dialogue flowing;

(iii) discussing the ethical issues with the military;
(iv) bringing the issues to the attention of policy makers and trying to get international discussion going;
(v) keeping abreast of new developments both in the military and in other areas that might be useful to the military; keeping up to date with military plans, calls for proposals and new deployment and updating the public.

A lot of my time is taken up with these activities.

Unmanned military systems, though yet not fully autonomous, are a reality on the battlefields of today. What are the main ethical and legal challenges concerning the present use these of military systems?

There are many ethical issues and challenges facing us with the use of unmanned systems (autonomous or even man in the loop) that I have written about that are too lengthy to repeat here. The most pressing concern is the protection of the lives of innocents regardless of nationality, religious affiliation or ideology. Allowing robots to make decisions about who to kill would fall foul of ethical precepts of a Just War under jus in bello.

In particular armed autonomous robots are against the spirit of the law set down in the Geneva convention under the Principle of Distinction and the Principle of Proportionality. These are two of the cornerstone of Just War Theory.

The principle of distinction is there to protect civilians, wounded soldiers, the sick, the mentally ill, and captives. The law, simply put, is that we must discriminate between combatants and non-combatants and do everything in our power to protect the latter. In a nutshell the ethical problem is that no autonomous robots or artificial intelligence systems have the necessary sensing and reasoning capabilities to discriminate between combatants and innocents. We do not even have a clear definition anywhere in the laws of war as to what a civilian is. The 1949 Geneva Convention requires the use of common sense while the 1977 Protocol 1 essentially defines a civilian in the negative sense as someone who is not a combatant.

There is also the Principle of Proportionality which holds that civilian casualties are often unavoidable in warfare and that the number of civilian deaths should be proportional to the military advantage gained. But there is no objective measure available for a computational system to calculate such proportionality. It is down to a commander's militarily informed opinions and experience. I have written about the big problems of proportionality calculations for humans never mind machines.

Yes, humans do make errors and can behave unethically, but they can be held accountable. Who is to be held responsible for the lethal mishaps of a robot? Certainly not the machine itself. There is a long causal chain associated with robots: the manufacturer, the programmer, the designer, the department of defence, the generals or admirals in charge of the operation, the operators, and so on.

There are a number of ill specified dimensions in the Laws of War about the protection of innocents that are muddied incredibly by insurgent warfare. In history, state actors have often behaved reciprocally – you bomb our civilians and we will bomb yours. This is morally reprehensible but gets worse when we consider non-state actors. Who are their civilians? It is like asking who are the civilians of any arbitrary group such the railway workers or the bakers.

I have recently been thinking through the idea of a proportionality calculation based on a variant of the philosopher John Rawls' "original position", which was for a thought experiment about the principles of justice in a free and fair society. Rawls' notion is that representatives of citizens are placed behind a "veil of ignorance", that deprives them of information about the individuating characteristics of the citizens they represent. This lack of information forces them to be objective about the fairness of the social contract they are attempting to agree upon. Crudely put, it is a little like you cutting a pie knowing that I will have first choice of portion.

My "veil of ignorance for proportionality judgments" would similarly deprive the decision maker of information of the nationality, religion and ideology of the innocents that are likely to be killed. To take an extreme example, a baby in my country has as much right to protection as a baby in a country where insurgents are fighting. Through the veil of ignorance, the baby would expect a better chance of survival.

Ok, so the baby example is a bit emotive, but there is another example I can use taken from a drone strike of a village last year that I have written about elsewhere. The aim of the strike was to kill an al-Qaeda leader; a number of children were among the dead. In a newspaper article, senior US military were reported to say that they knew there was a high risk of killing the children, but the leader was such a "high value" target, it was worthwhile. (Subsequent DNA analysis of the corpses showed that the target had not been present in the village.)

To turn this into a concrete 'veil of ignorance' example, imagine that the commander in charge of the strike had just been informed that a

party of US school children may be visiting the village. Would the calculation of military advantage change?

This is a preview of an idea that I am working on for a paper and it needs more thought and discussion.

It seems unlikely to win the "hearts and minds" of people with military robots. Do you think that – for certain roles – unmanned military systems do have an eligibility in armed conflicts?

First, I think that you are absolutely right about the hearts and minds issue. I have been heartened recently by reports from the new head of the armed forces in Afghanistan, Lt General Stanley McCrystal. He seems to have really grasped the idea that killing civilians means creating many more insurgents and is actually fulfilling their goals for them. He sees that for every innocent killed, a number of their family members will take up arms. I won't take up time with his position here, but it is well worth checking out. It is a pragmatic rather than an ethical approach, but it highly correlates with the ethical and may have more impact.

I have no ethical issues against the use of unmanned systems for protecting soldiers in their normal functioning. Improvised explosive devices on roadsides kill very many soldiers and even the relatively crude robots deployed for disrupting these are of great benefit. I would much prefer to see some of the large budgets that are going into armed predators and reapers being used to develop better explosives detection – detection of explosive at a distance.

There are precedents for weapon systems, which have been banned from the battlefields, either because they lack the ability to discriminate or they cause unnecessary suffering. Could these international treaties act as guidance for how to cope with the questions surrounding the use of unmanned military systems?

Yes, these treaties are useful in setting out guidance. They are not binding and countries can give notice to no longer be signatories. Also not everyone signs up to them. For example China, Russia and the US were not among the 150 countries banning cluster munitions. However, although the US also did not sign up for the landmine treaty, they behave as if they did. These treaties are useful in setting out moral standards.

A similar treaty for unmanned systems or even armed unmanned systems would be much more difficult – at the very least there would be definitional problems. For example is a cruise missile an unmanned system? There are often academic debates about what is considered to

be a robot. I have my own ideas of course, but we will need a consensus. I would very much like to see, at the very least, some serious international debate about the possibility of setting up an unmanned systems arms control treaty. Proliferation now seems inevitable given the military advantages that have recently been showcased.

You could argue that unmanned systems are already covered under the Geneva Convention and the various treaties etc., and this is true in a general sense. But I think that very serious consideration needs to be given specifically to the detailed implications of these new weapons and how they will impact on civilians as they are developed further.

Some argue that robot weapons are just the same as other distance weapons and are just a later stage in the evolution started by the slingshot. I think that robots could be a new species of weapon. As they develop further they could become stand-ins for soldiers or pilots at ever greater distances. Unlike missiles or other projectiles, robots can carry multi-weapon systems into the theatre of operations and act flexibly once in place.

I am currently working on the idea of setting up a Committee for Robot Arms Control and would welcome any supporters of robot arms control reading this to get in touch.

Do you think that concepts to integrate ethical decision making capacities in automated systems, like for example Ronald C. Arkin's "Ethical Governor", will in the end result in systems that can be used in compliance with the laws of armed conflict and/or ethical considerations?

Ron's intentions are good and he has very important things to say. His motivation is based on his concerns about the ethical behaviour of soldiers in battle. He was shocked, like many of us, by the US Surgeon General's report of a survey of US troops in Iraq. He also, like me believes that autonomous armed robots seem to be inevitable. However, I have serious misgivings about his enterprise.

He says that robots don't get angry and will not seek revenge. I agree, but they will also not feel sympathy, empathy, compassion, remorse or guilt. I believe that these are needed for the kinds of moral judgements required in fighting a just war – particularly urban insurgent warfare.

One of the main problems that I see for the ethical governor is the discrimination problem. There is absolutely no point, apart from research purposes, in having a system of rules about ethical behaviour if the input does not tell them the right information to operate with.

We have a side bet running about the timescale for solving the discrimination problem. Ron believes we will have the discrimination technology in operation within the next twenty-five years and I think that he is being overly optimistic. Whichever of us is wrong will buy the other a pint of beer.

Another problem that I have with systems like this in general, is that they are too black and white – too absolute about the rules. The ethical governor is a deontological system. In war we often need consequentialist ethics (with clear moral underpinnings) – there are very many circumstances in war where behaving on the basis of the consequences of an action is more important than blind rule following. The principle of proportionality is intrinsically a consequential problem for a start.

In relation to this last point is that the Geneva Convention and all its associated bit are not written with computer programming in mind. To turn it into "if then rules", will require considerable interpretation.

Soldiers need to do a lot of reasoning about moral appropriateness (even if they are absolutist, they need reasoning to plug into their moral judgements). There are heart-warming reports of troops in the current Middle East conflict responding appropriately in a variety of situations such as letting insurgents pass with a coffin and taking off their helmets as a mark of respect.

It is not just a case of a conditional rule like "if combatant then fire". My worry is that there are a very large, possibly infinite set of exceptions that we could not predict in advance to programme into a computer. I know that current AI systems do not have the required reasoning abilities and I am not sure when or if they will ever have them.

The final problem that I have with such systems (and have to stop myself rambling on forever) is they may be used to push the development of armed autonomous systems with the promise of "don't worry, everything will be OK soon". With a potential (or apparent) technological solution in sight, it may allay political opposition to deployment of autonomous killers.

If armed autonomous robot systems are inevitable, work like this will be needed. In my view the ethical governor will raise more problems than it will solve and that is the only way to make progress. However, a preferable choice for me would be to have the money spent on better ethical training of the troops and more effective monitoring strategies as well as greater accountability.

Concerning not only military applications but all aspects of human

society, from care for the elderly to entertainment, where do you see robots and artificial intelligence in the foreseeable future?

I have written quite a lot about the areas of military, policing, eldercare and companionship, robot nannies and medical robotics, but there are many more ethical issues in other areas of robotics and emerging technologies – these are just the areas that I have thought about in detail.

There is a lot of cash to be made in robotics and they are becoming cheaper to manufacture all the time, and so we could see them entering human society in fairly large numbers soon. It is hard to predict their scope and tasks as there are many creative entrepreneurs and developers.

I am not expecting any great leaps in AI or the ability of robots to think for themselves but I am expecting a lot of very clever applications. Many of these will be welcome and perhaps take away the drudgery of some of our duller work although I don't think they will cause unemployment any more than the computer did.

All trends suggest to me that robots could enter our lives in many ways that we cannot predict – much like the internet and the web did. I do think, though, that there will be many ethical tradeoffs to deal with over the benefits and disadvantages of using robots and I suspect that there will be human rights issues as well. With large commercial interests in play, I guess that the biggest worry is that we will be confronted by very many novel applications before we have had time to think them through.

The keyword "human enhancement". Which kind of new challenges are you expecting in this field?

Wow! This is a very big question in the disguise of a small one. This is not my true area of expertise but I have looked into some of the issues for the UK think tank 2020HealthOrg. Our report will be released as a green paper in the House of Commons this year. At present it is difficult to sort the facts from the hopes, the fantasy and there are large commercial interests at stake.

I am going to be short and general about this one.

One person's enhancement can be another person's alleviation of a serious disability. For that reason, if for none other, there is great hope for brain and nervous system implants, new drugs and drug delivery implants. There is some great work, for example, in overcoming tremors in Parkinson's disease.

These applications have their own ethical issues. One specific to the UK, for example, is about whether the tax payer should foot the bill. The application for illness is not really classed as enhancement, but it is not always easy to draw the line. For example, what if we can enhance the ability to learn, and we use it to bring people with learning difficulties towards the norm. That in itself will, by definition, change the norm and so more people will need to be enhanced and so on.

One of the important ethical concerns is about deceit – the secret use of enhancement to gain advantage (think Olympic doping). For example a device (or drug) may be used to temporarily enhance intelligence to do better on examination, entrance tests or to deceive a potential employer. Let us be clear that legislation is unlikely to stop this practice any more than it stops the use of illegal drugs at present.

Another issue that concerns me is the inequity that could be created between the wealthy and the poor. The wealthy have big enough advantages as it is with the education system. At least now, those with strong analytical skills from a poor background can still work their way into top jobs. Expensive enhancement could see an end to that.

I will finish with one bit of speculation. A big issue that people of the future might have to face is "disenhancement". If we have the technology to enhance people cognitively and physically, we could turn it around to do the opposite. We have all heard of psychiatric drugs being used on political dissidents in the former Soviet Union. Political landscapes can change rapidly as well as treatment of criminals and what constitutes a crime (political or otherwise). We could end up with some very powerful tools to constrain people's thoughts and actions.

I have no doubt that we will be hearing a lot more about the ethical issues associated with implants over the coming years.

[1] I say "senior academics" because it is not a well rewarded career move although that is beginning to change in the UK.