erschienen in der FIfF-Kommunikation,
herausgegeben von FIfF e.V. - ISSN 0938-3476
www.fiff.de

Noel Sharkey

Weapons of Indiscriminate Lethality

The development of autonomous robot weapons is well underway for use in a new style of hi-tech warfare. This will lead to less physical risk to the combatants deploying them but greater moral risk. There has been insufficient consideration of how these new weapons will impact on innocents. Two of the most serious ethical concerns discussed here are: (i) the inability of robot weapons to discriminate between combatants and non-combatants and (ii) the inability of such robots to ensure a proportionate response in which the military advantage will outweigh civilian casualties.

War is a very odd human endeavour. In normal life, humans are not permitted to murder one another without facing severe penalties. In many civilised countries, even the state is not allowed the right to execute for serious crimes. Yet in war it is acceptable for one large group of humans to kill humans from another large group without moral sanction or guilt. But there is a simple proviso for civilised countries, that those killed must be from amongst those who are killing back or are contributing to it. In other words, the ideal is that combatants only kill other combatants. This is one of the cornerstones of the Laws of War.

The Geneva and Hague conventions as well as the various treaties and the laws of armed conflict strictly specify that innocents must be protected from harm. This is part of the justice in the conduct of a war, jus in bello, and is often expressed as the principle of discrimination – only combatants/warriors are legitimate targets of attack. All others, including children, civilians, service workers and retirees, should be immune from attack. In fact, the laws of protection even extend to combatants that are wounded, have surrendered or are mentally ill.[1]

These protections have been in place for many centuries. Thomas Aquinas, in the 13th Century, developed the doctrine of Double Effect. Essentially there is no moral penalty for killing innocents during a conflict providing that (i) you did not intend to do so, or (ii) that killing the innocents was not a means to winning, or (iii) the importance to the defence of your nation is proportionally greater than the number of civilian deaths.

There are many circumstances in a modern war where it is extremely difficult, if not impossible, to fully protect non-combatants. For example, in attacking a warship, some non-combatants such as chaplains and medical staff may be unavoidably killed. Similarly, but less ethically secure, it is difficult to protect the innocent when large explosives are used near civilian populations, or when missiles get misdirected. In modern warfare, the equivalent of the doctrine of Double Effect is the Principle of Proportionality which, "… requires that the anticipated loss of life and damage to property incidental to attacks must not be excessive in relation to the concrete and direct military advantage expected to be gained."[2]

In warfare, both the principles of discrimination and proportionality can be problematic although their violation requires accountability and can lead to war crimes tribunals. But now it looks as though we may be about to unleash new weapons that could violate both of these principles.[3] These are the proposed autonomous weapons such as unmanned combat vehicles. These will be able to make decisions about who to kill and when to kill them and yet cannot be accountable in themselves.

Lethal Autonomous Robots and the Problem of Discrimination

There are between four and six thousand robots currently operating on the ground in Iraq and Afghanistan. These are mainly deployed in dull, dirty or dangerous tasks such as disrupting or exploding improvised explosive devices and surveillance in dangerous areas such as caves. There are only three armed Talon SWORDS robots made by Foster-Miller, although more are expected soon. Most of the armed robots are in the sky; semi-autonomous Unmanned Combat Air Vehicles, such as the MQ1-Predator, that flew some 400,000 mission hours up to the end of 2006 and have flown significantly more since, and the more powerful MQ-9 Reapers with a payload of 14 Hellfire missiles – the RAF have two MQ-9s operating in Iraq. These can navigate and search out targets but, like the ground robots, it is a remote operator, this time thousands of miles away in the Nevada desert, who makes the final decision about when to apply lethal force.

There is now massive spending and plans are well underway to take the human out of the loop so that robots can operate autonomously to locate their own targets and destroy them without human intervention [4]: This is high on the military agenda of all the US forces: "The Navy and Marine Corps should aggressively exploit the considerable warfighting benefits offered by autonomous vehicles (AVs) by acquiring operational experience with current systems and using lessons learned from that experience to develop future AV technologies, operational requirements, and systems concepts."[5] There are now a number of autonomous ground vehicles such as DARPA's "Unmanned Ground Combat Vehicle and Perceptor Integration System" otherwise known as the Crusher.[6] And BAE systems recently reported that they have "… completed a flying trial which, for the first time, demonstrated the coordinated control of multiple UAVs autonomously completing a series of tasks".[7]

The move to autonomy is clearly required to fulfil the current US military plans. Tele-operated systems are more expensive to manufacture and require many support personnel to run them. One of the main goals of the Future Combat Systems project is to use robots as a force multiplier so that one soldier on the battlefield can be a nexus for initiating a large scale robot attack from the ground and the air. Clearly one soldier cannot remotely operate several robots alone and it takes the soldier away from operational duties.

The ethical problem is that no autonomous robots or artificial intelligence systems have the necessary sensing properties to allow for discrimination between combatants and innocents. Allowing them to make decisions about who to kill would fall foul of the fundamental ethical precepts of a just war under jus in bello as enshrined in the Geneva and Hague conventions and the various protocols set up to protect civilians, wounded soldiers, the sick, the mentally ill, and captives. There are no visual or sensing systems up to that challenge.

The problem is exacerbated further by not having a specification of "civilianess". A computer can compute any given procedure that can be written down in a programming language. We could, for example, give the computer on a robot an instruction such as, "if civilian, do not shoot". This would be fine if and only if there was some way to give the computer a clear definition of what a civilian is. We certainly cannot get one from the Laws of War that could provide a machine with the necessary information. The 1944 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:

1. A civilian is any person who does not belong to one of the categories of persons referred to in Article 4 A (1), (2), (3) and (6) of the Third Convention and in Article 43 of this Protocol. In case of doubt whether a person is a civilian, that person shall be considered to be a civilian.

2. The civilian population comprises all persons who are civilians.

3. The presence within the civilian population of individuals who do not come within the definition of civilians does not deprive the population of its civilian character. [8]

And even if there was a clear computational definition of civilian, we would still need all of the relevant information to be made available from the sensing apparatus. All that is available to robots are sensors such as cameras, infrared sensors, sonars, lasers, temperature sensors and ladars etc. These may be able to tell us that something is a human, but they could not tell us much else. In the labs there are systems that can tell someone's facial expression or that can recognize faces but they do not work on real time moving people. And even if they did, how useful could they be in the fog of war. British teenagers beat the surveillance cameras by wearing hooded jackets.

In a conventional war where all of the combatants wore the same clearly marked uniforms (or better yet, radio frequency tags) the problems might not be much different from those faced for conventional methods of bombardment. But the whole point of using robot weapons is to help in warfare against insurgents and in these cases sensors would not help in discrimination. This would have to be based on situational awareness and of having a theory of mind, i.e. understanding someone else's intentions and predicting their likely behaviour in a particular situation. Humans understand one another in a way that machines cannot and we don't fully understand how. Cues can be very subtle and there are an infinite number of circumstances where lethal force is inappropriate. Just think of a children being forced to carry empty rifles or insurgents burying their dead.

The Problem of Proportionality

According to the Laws of War, a robot could potentially be allowed to make lethal errors providing that the non-combatant casualties were proportional to the military advantage gained. But how is a robot supposed to calculate what is a proportionate response. There is no sensing or computational capability that would allow a robot such a determination. As mentioned for the discrimination problem above, computer systems need clear specifications in order to operate effectively. There is no known metric to objectively measure needless, superfluous or disproportionate suffering[9]. It requires human judgment.

No clear objective means are given in any of the Laws of War for how to calculate what is proportionate.[10] The phrase "excessive in relation to the concrete and direct military advantage expected to be gained" is not a specification. How can such values be assigned and how can such calculations be made? What could the metric be for assigning value to killing an insurgent relative to the value of non-combatants, particularly children who could not be accused of willingly contributing to insurgency activity? The military say that it is one of the most difficult decisions that a commander has to make; but that acknowledgement does not answer the question of what metrics should be applied. It is left up to a military force to argue as to whether or not it has made a proportionate response as has been evidenced in the recent Israeli-Gaza conflict.

Uncertainty needs to be a factor in any proportionality calculus. Is the intelligence correct and is there really a genuine target in the kill zone? The target value must be weighted by a probability of presence/absence. This is an impossible calculation unless the target is visually indentified at the onset of the attack. Even

then errors can be made. The investigative journalist, Seymour Hersh, gives the example of a man in Afghanistan being mistaken for bin Laden by CIA Predator operators in 2002. A Hellfire was launched killing three people who were later reported to be three local men scavenging in the woods for scrap metal.[11] And this error was made using a robot plane with a human in the loop. There is also the problem of relying on informants. The reliability of the informant needs to be taken into account and so does the reliability of each link in the chain of information reaching the informant before being passed onto the commander/operator/pilot. There can be deliberate deception anywhere along the information chain as was revealed in investigations of Operation Phoenix – the US assignation programme – after the Vietnam War. It turned out that many of the thousands on the assignation list had been put there by South Vietnamese officials for personal reasons such as erasing gambling debts or resolving family quarrels.[12]

It is also often practically impossible to calculate a value for the actual military advantage. This is not necessarily the same as the political advantage of creating a sense of military success by putting a face to the enemy to rally public support at home and to boost the morale of the troops. Obviously there are gross calculations that work in the extreme such as a military force carrying weapons sufficient to kill the population of a large city. Then it could be possible to balance the number of civilians killed against the number saved. Military advantage at best results in deterrence of the enemy from acting in a particular way, disruption of the social, political, economic, and/or military functions and destruction of the social, political, economic, and/or military functions.[13] Proportionality calculations should be based on the likely differences in military outcome if the military action killing innocents had not been taken.[14]

Despite the impossibility of proportionality calculations, military commanders at war have a political mandate to make such decisions on an almost daily basis. Commanders have to weigh the circumstances before making a decision but ultimately it will be a subjective metric. Clearly the extremes of wiping out a whole city to eliminate even the highest value target, say Osama bin Laden, is out of the question. So there must be some subjective estimates about just how many innocent people equal the military value of the successful completion of a given mission.

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 no way to punish a robot. We could just switch it off but it would not care anymore about that than my washing machine would care. Imagine telling your washing machine that if it does not remove stains properly you will break its door off. Would you expect that to have any impact? 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 and the operator. It is thus difficult to allocate responsibility for deliberate war crimes or even mishaps.

Conclusions

There are some weapons that can be entirely excessive in relation to the concrete and direct military advantage to be gained, and that are, by their very nature, indiscriminate. Their lethal application can be decided by factors outside of decisions made by military commanders. For example, many civilised countries have signed treaties to ban landmines and cluster bombs because of their impact on the innocent population outside of their military application. But the military and the weapons manufactures continually exploit new technology to develop new weapons.

There are no current international guidelines or even discussions about the uses of autonomous robots in warfare. These are needed urgently. If there was a political will to use them, then legal arguments could be constructed that leave no room for complaints[15]. This is especially the case if they could be released somewhere where there is a fairly high probability that they will kill a considerable greater number of enemy combatants (uniformed and non-uniformed) than innocents, i.e., the civilian death toll where not disproportionate to the military advantage. Or if they could be restricted to a "kill box" – to use the US military term – they could be treated legally in the same way as a bombing mission.

Armed autonomous robots could also be treated in a legally similar way to submunitions such as the BLU-108 developed by Textron Defense Systems[16]. The BLU-108 parachutes to near

Noel Sharkey

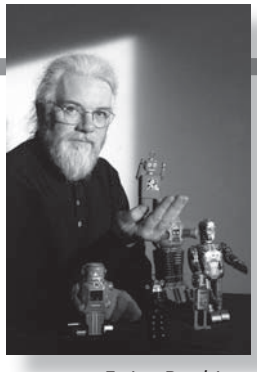

Foto: Bocking

Noel Sharkey BA PhD FIET, FBCS CITP FRIN FRSA is a Professor of AI and Robotics and Professor of Public Engagement at the University of Sheffield (Department of Computer Science) and EPSRC Senior Media Fellow (2004-2009). His main research interests are now in Biologically Inspired Robotics, Cognitive Processes, history of automata (from ancient times to present), Human-Robot interaction and communication, Representations of Emotion and Machine learning. He is currently involved in initiating public discussion about the ethical use of robots and the implications for public safety and human rights. *(noel@dcs.shef.ac.uk)*

the ground where an altitude sensor triggers a rocket that spins it upwards. It then releases four Skeet warheads at right angles to one another. Each has a dual-mode active and passive sensor system: the passive infrared sensor detects hot targets such as vehicles while the active laser sensor provides target profiling. They can hit hard targets with penetrators or destroy soft targets by fragmentation.

But the BLU-108 is not like other bombs because it has a method of target discrimination. If it had been developed in the 1940s or 1950s there is no doubt that it would have been classified as a robot, and even now it is debatably a form of robot. The Skeet warheads have autonomous operation and use sensors to target their weapons. The sensors provide discrimination between hot and cold bodies of a certain height but, like autonomous robots, they cannot discriminate between legitimate targets and civilians. If BLU-108s were dropped on a civilian area they would destroy buses, cars and lorries. Like conventional bombs, discrimination between innocents and combatants requires accurate human targeting judgements. It is this and only this that keeps the BLU-108 within humanitarian law.[17]

To use robot technology over the next 25 years in warfare would at best be like using the BLU-108 submunition, i.e. can sense a target but cannot discriminate innocent from combatant.[18] But the big difference with the types of autonomous robots currently being planned and developed for aerial and ground warfare is that they are not perimeter limited like the Skeet. The BLU-108 has a footprint of 820ft all around. By way of contrast, mobile autonomous robots are limited only by the amount of fuel or battery power that they can carry. They can potentially travel long distances and move out of line of sight communication.

In a recent sign of these future weapons the US Air Force sent out a call for proposals for Guided Smart Submunitions: "This concept requires a CBU (Cluster Bomb Unit) munition or UAV capable of deploying guided smart submunitions that have the ability to engage and neutralize any targets of interest. The goals for the submunitions is (sic) very challenging, when considering the mission of addressing mobile and fixed targets of interest. The submunition has to be able to reacquire the target of interest it is intended to engage." [19] This could be very like an extended version of the BLU-108 that could pursue hot bodied targets. It is the words "reacquire the target of interest" that is most worrying. If a targeted truck were to, for example, overtake a school bus, the weapons may acquire the bus as a target rather than the truck.

The only humane course of action is to severely restrict or ban the deployment of these new weapons until there have been international discussions about how they might pass an "innocents discrimination test". At the very least there should be discussion about how to limit the range and action of autonomous robot weapons before the inevitable proliferation. Although all of the elements discussed here can be accommodated within the existing Laws of War, autonomous robot weapons could change the nature of war considerably. This needs to be thought through properly and specific new laws should be implemented to not just accommodate but to constrain.

Notes

1 *But see also Ford, John S. (1944) The Morality of Obliteration Bombing, Theological Studies,261-309.*
2 *Petraeus D.H. and Amos, J.F. Counterinsurgency, Headquarters of the Army, Field Manual FM 3-24 MCWP 3-33.5, Section 7-30*
3 *Sharkey, N.E. (2008) The Ethical Frontiers of Robotics, Science, 322. 1800-1801*
4 *Sharkey, N.E. (2008) Cassandra or the false prophet of doom: AI robots and war, IEEE Intelligent Systems, vol. 23, no, 4, 14-17, JulyAugust Issue*
5 *Committee on Autonomous Vehicles in Support of Naval Operations National Research Council (2005) Autonomous Vehicles in Support of Naval Operations, WashingtonDC: The National Academies Press*
6 *Fox News (2008) Pentagon's 'Crusher' Robot Vehicle Nearly Ready to Go, Feb. 27, 2008*
7 *United Press International (2008) BAE Systems tech boosts robot UAVs IQ, Industry Briefing, Feb. 26.*
8 *Protocol 1 Additional to the Geneva Conventions, 1977 (Article 50)*
9 *Bugsplat software and its successors have been used to help calculate the correct bomb to use to destroy a target and calculate the impact. A human is there to decide and it is unclear how successful this approach has been in limiting civilian casualties.*
10 *Sharkey, N. (in press for 2009) Death Strikes from the Sky: The calculus of proportionality, IEEE Science and Society.*
11 *Hersh, S.M. (2002)"Manhunt: The Bush administration's new strategy in the war against terrorism," New Yorker, p. 66, Dec. 2002.*
12 *Hersh ibid*
13 *Hyder, V. D. (2004) Decapitation Operations: Criteria for Targeting Enemy Leadership, Monograph/report approved for publication. School of Advanced Military Studies United Sates Army Command and General Staff College, Fort Leavenworth, Kansas. p5*
14 *Chakwin, B., Voelkel, D. and Enright S. (2002) Leaders as Targets, Joint Forces Staff College, Seminar # 08*
15 *But it seems that, regardless of treaties and agreements, any weapon that has been developed may be used if the survival of a state is in question. The International Court of Justice (IJC (1996) Nuclear Weapons Advisory Opinion decided that it could not definitively conclude that in every circumstance the threat or use of nuclear weapons was axiomatically contrary to international law, see Stephens, D. and Lewis, M.W. (2005) The law of armed conflict – a contemporary critique. Melbourne J. International Law.6*
16 *Thanks to Richard Moyes of Landmine Action for pointing me to the BLU-108 and to Marian Westerberg and Robert Buckley from Textron Defense Systems for their careful reading and comments on my description.*
17 *A key feature of the BLU-108 is that it has built-in redundant self-destruct logic modes that largely leave battlefields clean of unexploded warheads and thus keeps it out of the 2008 treaty banning cluster munitions.*
18 *Sharkey, N.E. (2008) Grounds for Discrimination: Autonomous Robot Weapons, RUSI Defence Systems, 11 (2), 86-89*
19 *USAF call for proposals (2008) Guided Smart Munitions, Topic Number: AF083-093, August, 25*