Journal of Military Ethics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/smil20

Saying 'No!' to Lethal Autonomous Targeting

Noel Sharkey [a]
[a] Department of Computer Science, University of Sheffield, UK
Version of record first published: 16 Dec 2010.

Journal of Military Ethics,
Vol. 9, No. 4, 369–383, 2010

Saying 'No!' to Lethal Autonomous Targeting

NOEL SHARKEY
Department of Computer Science, University of Sheffield, UK

ABSTRACT *Plans to automate killing by using robots armed with lethal weapons have been a prominent feature of most US military forces' roadmaps since 2004. The idea is to have a staged move from 'man-in-the-loop' to 'man-on-the-loop' to full autonomy. While this may result in considerable military advantages, the policy raises ethical concerns with regard to potential breaches of International Humanitarian Law, including the Principle of Distinction and the Principle of Proportionality. Current applications of remote piloted robot planes or drones offer lessons about how automated weapons platforms could be misused by extending the range of legally questionable, targeted killings by security and intelligence forces. Moreover, the alleged moral disengagement by remote pilots will only be exacerbated by the use of autonomous robots. Leaders in the international community need to address the difficult legal and moral issues now, before the current mass proliferation of development reaches fruition.*

KEY WORDS: robotics, military robotics, drones, ethics, distinction, discrimination, proportionality

Introduction

Robots in warfare are an inevitable stage in the evolution of weapons development designed to separate fighters from their foes. Weapons technology has evolved to enable killing from ever increasing distances throughout the history of warfare. From stones to pole weapons to bows and arrows, from cannon to aerial bombing to cruise missiles, killing has become ever easier. Battlefield robots may not change the character of war but they will change the way that wars are fought.

In 2010, the ultimate distance weapons in the sights of military organisations throughout the world are armed robots. To be clear, the robots discussed here are mobile platforms that can carry multiple weapons systems, bombs, cameras, sensors or other payloads. They can be recalled, refuelled and re-weaponised. Operationally, they can be sent into the theatre on a fishing expedition without a preset goal. It is this type of flexibility that separates them from other robots such as cruise missiles and torpedoes.

Correspondence Address: Noel Sharkey, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello, Sheffield S1 4DP, UK. E-mail: noel@dcs.shef.ac.uk

1502-7570 Print/1502-7589 Online/10/040369–15
http://www.informaworld.com
DOI: 10.1080/15027570.2010.537903

Of course there are some exceptional systems such as loitering munitions that fall between the cracks. For example, the Israeli Harpy and Harop made by Israel Aircraft Industries are a cross between a missile and an Unmanned Aerial Vehicle (UAV). They are launched like a missile but hover over an area, select their own targets (hostile radar signatures) as they arise and then impact them like a missile. Thus they have some mission flexibility but not reusability.

Many thousands of robots have been used during both the Iraq and Afghanistan conflict (up from 150 in 2004) for tasks such as bomb disposal and cave clearance on the ground and for surveillance from the air. The undisputed success of UAVs (now more commonly known as Remote Piloted Aircraft, RPA, or Drones) for gathering intelligence in the Iraq/Afghanistan conflicts has made them a showcase for military powers everywhere. There is an almost insatiable military demand for UAVs. Troops do not like to move without them. This same demand is reflected in more than 40 other countries that are either developing or purchasing military robots. These include Russia, China, Israel, Pakistan, Korea, India and Iran.

The focus of this paper is on ethical concerns about the application of armed robots in areas with mixed combatant and civilian populations. It is unclear whether International Humanitarian Law (the Laws of War) will eventually need to be amended to accommodate emerging technologies. But for now we need to scrutinise the mapping between the applications of the new technologies and the current laws to ensure that they are preserved and followed.

The main ethical concerns expressed here are about the planned use of the armed autonomous 'decision making' robots that have appeared in all of the US military's roadmaps since 2004. However, before examining the issues surrounding autonomous lethal force in detail it will be useful to look at some of the ethical problems that have arisen for man-in-the-loop or Remote Piloted Drones. There are lessons to be learned that will highlight some of the dangers of fully autonomous armed robots.

Armed 'Man-in-the-Loop' Systems

The United States leads the field with its armed drones. The Predator MQ-1 was first armed by the Central Intelligence Agency (CIA) in 2001 with its now standard two Hellfire missiles. Nowadays 'pilots' of the 432nd Air Expeditionary Wing fly official military missions over Iraq and Afghanistan from Creech Air Force Base in the Nevada desert, thousands of miles away from the action. The Royal Air Force 39th Squadron also participates at Creech. Each plane has a two-person crew: a remote pilot and a sensor operator who sit at what looks like games consoles making decisions about the application of lethal force. The planes are flown around the clock in shifts and it is easy for pilots to take a break from 'battle' at any time or even go home for dinner with their families.

In October 2007, the Predator fleet was joined by the larger and more powerful MQ-9 Reaper with its larger payload capacity of 1,700kg – up to

14 Hellfire missiles or a mixture of missiles and bombs. The number of Reapers flying in operations doubled after their first year and General Atomics have reported difficulties in keeping up with the demand. By March 2009, the number of armed flying drones in the field was reported to be 195 Predators and 28 Reapers. The number has grown dramatically since.

Anticipated increases can be seen in the training figures for Predator crews. In 2005, 32-person crews consisting of a pilot and a sensor operator were being trained per year, whereas in 2008, 160 crews were trained and in late 2008, an additional $412m was added to the budget for training more non-pilot operators. In 2009, the number of remote pilot operators trained outnumbered the number of conventional pilots. Figures for the size of the current fleet of Predators and Reapers were not publicly available at the time of writing but in May 2009 the Obama administration budgeted $5.5 billion to add to the fleet in 2010. The military demand for the armed UAVs is highlighted by the more than a million mission hours flown by the MQ-1 by 2010 (up from 250,000 in June 2007).

Moral Disengagement

Flying a UAV from thousands of miles away from the battlefield clearly alleviates one of the two fundamental obstacles that war-fighters must face – fear of being killed. As Daddis (2004) pointed out, fear is one of the greatest obstacles to a soldier's effectiveness in battle. It is obvious that the greater the distance from the enemy, the less fear will play in the action. Many battles throughout history have been lost by men bolting in panic as fear swept through the ranks (Holmes 2003).

Certainly, the remote pilots of attack planes like the Reaper and the Predator have no need to worry about their personal safety. Sitting in cubicles thousands of miles away from the action they can give a mission their full attention without worrying about being shot at. They are in highly secure home ground where no pilot has ever been safer.

The second fundamental obstacle is *resistance to killing*. Grossman (1995) argues that 'not firing' is not cowardice but really a compulsion of ordinary men not to kill. In his book *Acts of War*, Holmes (2003) argues that the *hit rates* in a number of historical battles show that many soldiers were not prepared to fire directly at the enemy. This is also borne out by the analyses and findings of Marshall ([1947]2000]) from his interviews of soldiers after World War II.

Similarly, Royakkers and van Est (2010) discuss the emotional and moral disengagement of UAV pilots, or 'cubicle warriors' as they calls them, from the consequences of their actions in the field. The crux is that the type of control system used encourages a 'Playstation' mentality that makes an operator more careless about decisions to kill. Royakkers and van Est suggest that new recruits may have been playing videogames for many years and may not see a huge contrast with that of being a cubicle warrior. They provide examples from Peter Singer's book *Wired for War* in which a young

pilot is reported to have said, 'It's like a video game. It can get a little bloodthirsty. But it's fucking cool' (Singer 2009b: 308–309). And in a further example, a young pilot says, 'The truth is, it wasn't all I thought it was cracked up to be. I mean, I thought killing somebody would be this life-changing experience. And then I did it, and I was like "All right, whatever."... Killing people is like squashing an ant. I mean, you kill somebody and it's like "All right, let's go get some pizza."' (ibid.: 391–392).

In contrast to this view, there are reports of remote pilots suffering from different types of stress than onboard pilots. After an attack the operator gets to see the aftermath, using high-resolution cameras. This is more than a conventional fighter pilot ever sees and, at least on the surface, it would seem that remote operators would be less morally disengaged than the crew of a high altitude bomber.

However, for an article appearing in the *Air Force Times* (Lindlaw 2008), five remote pilots interviewed said that they had had not been particularly troubled by their missions although they can sometimes make for a strange existence. Colonel Chris Chambliss, a commander of the 432nd Wing at Creech, said that on only four or five occasions had sensor operators gone to see the Chaplain or their supervisors and that this was a very small proportion of the total number of remote operators.

All of this is anecdotal evidence and proper scientific research needs to be conducted to find out what kind of stresses, if any, the operators are facing. It is well known that making good decisions is more difficult under stressful conditions. Longitudinal studies would be required as the operators may start out being stressed by graphic scenes but then become inured with sufficient exposure. This is a bit like the process trainee surgeons and nurses go through. They are known often to faint when they start out but quickly get habituated to horrific medical events.

Killing distance does not have to be physical distance. It can be psychological as well. Grossman (1985) cites Clausewitz and the French Colonel Ardant du Picq (1821–70) for expounding at length that the vast majority of deaths in battle occurred when the victors chased the losing side in retreat. Du Picq suggests that Alexander the Great lost fewer than 700 men over all his battles because his army was never pursued by a victorious enemy. Although UAV operators have high-resolution images on large screens, the cameras are looking down from above and it can be very hard to make out the faces of their victims. Is this so different from attacking retreating troops?

One type of stress that keeps getting mentioned in the media as if to counter the game mentality notion is that of remote pilots not getting a chance to wind down together and bond in their mutual operations. They return home to their families after their shifts have finished and even go to parents evening at school. The problem is that they have no debriefing about their day on the battlefield and cannot talk to their families about it. This is a very different issue from whether or not they are morally engaged or disengaged from the killer strikes.

Targeted Killings

Up until now the discussion has centered on conventional forces using armed drones to support troops on the ground in the war zones of Afghanistan and Iraq. But there is a second armed UAV 'Air Force' controlled by the CIA out of Langley in Virginia. The first ever US kill with a drone was in Yemen on 3 November 2002 when a CIA-operated MQ-1 killed a known Al-Qaeda leader and five other men travelling in the same vehicle.[1] This was controversial at the time but was considered, by Department of Defense lawyers, to be a legitimate defensive pre-emptive strike against Al-Qaeda. Since then, the use of drones for targeted killings or 'decapitation strikes' in states that are not at war with the US has been rising at an accelerating pace. It is what the *Asia Times* has called 'the most public "secret" war of modern times' (Turse 2010).

According to estimates published on the New America Foundation website,[2] the US drone attacks in Pakistan have risen from nine between 2004 and 2007 up to 85 by mid-October in 2010. The estimates are shown in Table 1. The number of civilian deaths has been very difficult to estimate and have ranged from as few as 20 to more than a thousand. There are many different counts and no objective way to decide between them. The New America Foundation gives a figure of 1 in 3 civilian to insurgent ratio while the Brookings Institute study in 2009 indicated that only 1 in 10 of those killed was an insurgent.

The legality of targeted killing is at best questionable. The word decapitation is often used to mean cutting off the leaders of an organisation or nation fighting a war from the body of their war-fighters. The aim of the current spate of aerial decapitation strikes was initially to target Al-Qaeda and Taliban leaders without risk to US military personnel. The idea is to eventually leave only replacements from the shallowest end of the talent pool that will be ineffective and easy to defeat. With less risk, the targeted leaders do not have to be at the highest level. This explains the increasing number of death-delivering attacks in Pakistan using Predators and Reapers.

These individually targeted killings come despite the banning in the US of all politically motivated killing of individuals since the famous Church commission report on the CIA political assassinations in 1975. In 1976, President Gerald Ford issued a presidential executive order that 'no person employed by or acting on behalf of the United States Government shall

Table 1. US drone attacks in Pakistan 2004–2010

		Numbers killed		
Year	estimated strikes	high estimate	low estimate	leaders killed
2004–2007	9	109	86	3
2008	34	296	263	11
2009	53	709	413	10
2010*	86	719	426	10

*Until 10 October 2010
Source: See endnote 2.

engage in, or conspire to engage in, assassination'. This is not codified in the US law but is an executive order that the president can change at will and without public notice. President Ronald Reagan issued Executive Order (EO) 12333, which expanded the ban to include persons acting on *behalf* of the US government and all subsequent presidents have supported it. However, EO 12333 does not limit lawful self-defence options against legitimate threats to the national security of US citizens. (Parks 1989)

During wartime, a combatant is considered to be a legitimate target at all times. If a selected individual is sought out and killed it is not termed an assassination. According to a Memorandum on EO 12333, which is said to be consistent with United Nations (UN) Charter 51, 'a decision by the President to employ clandestine, low-visibility, or overt military force would not constitute assassination if US military forces were employed against the combatant forces of another nation, a guerrilla force, or a terrorist or other organization whose actions pose a threat to the security of the United States'.

But the real legal question is, does the CIA have a right to assassinate *alleged* insurgent combatants without due process. Seymour Hersh (2002), whose writings were one of the main motivations for the Church commission, complained that, 'the targeting and killing of individual al-Qaeda members without juridical process has come to be seen within the Bush Administration as justifiable military action in a new kind of war, involving international terrorist organizations and unstable states'. The insurgents have been redefined as combatants, but without receiving the rights of prisoners of war and without being given the chance to surrender or to face trial. This move, in combination with Article 51, provides legal cover for the right to assassinate insurgent combatants.

The legality of the targeted killings has also been challenged at the UN General Assembly meeting in October 2009 by Philip Alston, UN special rapporteur on Extrajudicial Killings. He made a request for US legal justification about how the CIA is accountable for the targets that they are killing. The US turned down the request by refusing to comment on what it said were covert operations and a matter of national security.

A rebuttal by Harold Koh (2010), Legal Adviser, US Department of State, stated, 'US targeting practices including lethal operations conducted by UAVs comply with all applicable law including the laws of war.' However, there are no independent means of determining how the targeting decisions are being made. It remains unclear as to what type and level of evidence is being used to sentence non-state actors to death by Hellfire missile without right to appeal or right to surrender. It is also unclear what other methods were exhausted or attempted to bring the targets to justice. The whole process sits behind an unjustified cloak of national secrecy.

The CIA use of drones was also questioned at a Senate Hearing in a prepared statement by US law Professor Kenneth Anderson (2010). Anderson said that Koh 'nowhere mentions the CIA by name in his defense of drone operations. It is, of course, what is plainly intended when speaking of self-defense separate from armed conflict. One understands the hesitation of senior lawyers to name the CIA's use of drones as lawful when the official

position of the US government, despite everything, is still not to confirm or deny the CIA's operations.'

However, the Director of the CIA, Leon Panetta (2009), has been very open about the CIA use of drones. He told the Pacific Council on International Policy in 2009 that, 'it's the only game in town in terms of confronting and trying to disrupt the al-Qaeda leadership.' Revealing the CIA's intentions on the expansion of targeted drone kills, Panetta went on to say of Al-Qaeda that, 'If they're going to go to Somalia, if they're going to go to Yemen, if they're going to go to other countries in the Middle East, we've got to be there and be ready to confront them there as well. We can't let them escape. We can't let them find hiding places.'

This expansion of targeted killing is just what was of concern to the UN special rapporteur on Extrajudicial Killings. A subsequent report by Alston (2010) to the UN General Assembly discusses drone strikes as violating international and human rights law because both require transparency about the procedures and safeguards in place to ensure that killings are lawful and justified: 'a lack of disclosure gives States a virtual and impermissible license to kill'. Some of Alston's arguments revolve around the notion of 'the right to self-defence' and whether the drone strikes are legal under Article 51. He also examines some of the complex legal issues surrounding proportionality the different conditions that pertain when the military action is called 'self-defence'.

Panetta and many others have made the case that armed UAVs are more accurate and will kill less civilians than a B-52 bomber in attacking the tribal regions in Pakistan. But as a former CIA operative told me, there is no way that Pakistan or other state actors not at war with the US could 'turn a blind eye' to the bomber strikes as they do now for drones. It can be argued that it is their perceived precision and accuracy that allows them to penetrate areas and kill people in ways that would not previously have been available without major political and legal obstacles.

Attacking with a remote piloted vehicle is no different under the Laws of War than attacking with a manned helicopter gunship or even with a rifle. The worry is that the nature of an unmanned vehicle with no risk to military personnel, an ability to hover over an area for very many hours, and its perceived accuracy may lead to considerable expansion of potential targets.

I qualified 'accuracy' with the word 'perceived' here both because the attacks are often on buildings where the targets cannot be verified and, if the Brookings or New America Foundation figures discussed above are anything to go by, the accuracy is not as high as some might imagine. One of the oft-cited targeting methods of the CIA is to determine the locations of people through their cell phones; switch on your phone and you are dead. But a recent copyright lawsuit between two companies sheds doubt on the accuracy of this targeting method (Stein 2010). A small company called Intelligent Integration Systems alleges that one of its client companies, Netezza, reverse-engineered their software, Geospatial, on a tight deadline for the CIA. The court heard that the illegal version of the software could produce locations

that were out by as much as 40 feet and that the CIA had knowingly accepted the software.

But even if targeting was 100 per cent accurate, how can we be sure that alleged insurgents are 'guilty as charged'. Information about target identity and their role and position is heavily dependent on the reliability of the intelligence on which it is based. There are lessons that should have been learned from the Vietnam War investigations of Operation 'Phoenix' in which thousands were assassinated. It turned out that many of those on the assassination list had been put there by South Vietnamese officials for personal reasons such as erasing gambling debts or resolving family quarrels

Philip Alston (2010) reports that during a mission to Afghanistan he found out how hard it was for forces on the ground to obtain accurate information. 'Testimony from witnesses and victims' family members showed that international forces were often too uninformed of local practices, or too credulous in interpreting information, to be able to arrive at a reliable understanding of a situation.' He suggests that, 'States must, therefore, ensure that they have in place the procedural safeguards necessary to ensure that intelligence on which targeting decisions are made is accurate and verifiable.'

From *in*-the-Loop to *on*-the-Loop to *out-of*-the-Loop

Since 2004 all of the roadmaps and plans of the US forces has made clear the desire and intention to develop and use autonomous battlefield robots. Fulfilment of these plans to take the human out of the loop is well underway. The end goal is that robots will operate autonomously to locate their own targets and destroy them without human intervention (Sharkey 2008a). The Committee on Autonomous Vehicles in Support of Naval Operations (2005) wrote: 'The Navy and Marine Corps should aggressively exploit the considerable war-fighting 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.' We can only hope that the 'lessons learned' will include ethical and moral lessons. The ethical problems will only be amplified and added to by the use of autonomous robots.

Before delving into these problems it is important to clarify what is meant by 'robot autonomy' here. This is often confused with science fiction notions of robots with minds of their own with the potential to turn on humanity. The reality is very different. The autonomous robots being discussed for military applications are closer in operation to your washing machine than to a science fiction *Terminator*. The way the term 'autonomy' is used in robotics should not be confused with how the term is used in philosophy, politics, individual freedom or in common parlance. It is more related to the term automatic. An automatic robot carries out a pre-programmed sequence of operations or moves in a structured environment. A good example is a robot arm painting a car.

An autonomous robot is similar except that it operates in open or unstructured environments. To do this, the robot is controlled by a

programme takes inputs from its sensors and adjusts the speed and direction of its motors and actuators as specified. If the goal of a robot is to avoid objects and the sensors detect an object, the programme would adjust the motors so that the robot moves to avoid it. For example if the left-hand sensors detect the object, the robot would move right, and if the right-hand sensors detect the object, the robot would move left.

This leads us to another often-misunderstood topic, the robot decision process. This should not be confused with human decision making except by analogy. A computer decision process can be as simple as, IF object on left, THEN turn right OR IF object on right, THEN turn left, ELSE continue. Alternatively, the activity on a sensor may activate a different sub-programme to help with the decision. For example, to get smoother passage through a field laden with objects, a sub-programme could be called in to calculate if a turn to the left would result in having to negotiate more obstacles that a turn to the right.

Programmes can become very complex through the management of several sub-programmes and decisions about which sub-programme should be initiated in particular circumstances. But the bottom line for decision making by a machine, whether is it using mathematical decision spaces or artificial intelligence reasoning programmes, is the humble IF/THEN statement.

The other important thing to say about autonomy is that a system does not have to be exclusively autonomous or exclusively remote operated. The US Army, Navy and Air Force have all at some point discussed the classification of robot control on a continuum from totally human operated to fully autonomous. Each has separate development programmers and each has its own operational definitions of the different levels of robot autonomy. The Army has ten levels while the Air Force has four. The Navy goes for a characterisation in terms of mission complexity but points to three different classes of autonomous robot vehicle: (i) *scripted* in which the robot carries out a pre-planned script of the 'point, fire and forget' variety; (ii) *supervised* in which some or all of the functions of planning, sensing, monitoring and networking are carried out with the help of human operators to make decisions; (iii) *intelligent* in which attributes of human intelligence are used in software to make decisions, perceive and interpret the meaning of sensed information, diagnose problems, and collaborate with other systems.

The first of the Navy classifications, *scripted,* is the closest to 'automatic' and includes cruise missiles, torpedoes and automated weapons such as the Phalanx Close In Weapon System. The *supervised* category refers to the both the current 'man-in-the-loop' systems such as the Predator and Reaper control and the newly proposed 'man-on-the-loop' systems. And the *intelligent* category refers to fully autonomous operation although frankly I am not sure exactly what they mean by 'attributes of human intelligence in software'. Perhaps there is a little slippage into wishful thinking here.

There are at least four reasons that make an autonomous robot militarily desirable including: (i) remote operated systems are more expensive to manufacture and require many support personnel to run them; (ii) it is possible to jam either the satellite or radio link or take control of the system.

(iii) one of the military goals is to use robots as force multipliers so that one human can be a nexus for initiating a large-scale robot attack from the ground and the air; (iv) the delay time in remote piloting a craft via satellite (approximately 1.5 seconds) means that it could not be used for interactive combat with another aircraft. At a press briefing in December 2007, Dyke Weatherington, deputy director of the US Department of Defense's Unmanned Aerial Systems Task Force, said, 'Certainly the roadmap projects an increasing level of autonomy . . . to fulfill many of the most stressing requirements. Let me just pick one for example. Air-to-air combat – there's really no way that a system that's remotely controlled can effectively operate in an offensive or defensive air combat environment. That has to be-the requirement of that is a fully autonomous system.'[3]

The most recent United States Air Force Unmanned Aircraft Systems Flight Plan 2009–2047[4] opens the strategy for a staged move from current remote piloted systems to fully autonomous systems. It represents a push to shrink the role of the man in the loop. To begin with, autonomous operation will be mainly for tasks such as take-off, landing and re-fuelling. As unmanned drones react in micro- or nanoseconds the 'humans will no longer be "in the loop" but rather "on the loop" – monitoring the execution of certain decisions. Simultaneously, advances in AI [artificial intelligence] will enable systems to make combat decisions and act within legal and policy constraints without necessarily requiring human input' (ibid.: 41).

The idea of a human on-the-loop means that the human will be in executive control overall to call in or call off the robots rather than being in control of each individually. In other words, the robots will be essentially autonomous: 'SWARM technology will allow multiple MQ-Mb aircraft to cooperatively operate in a variety of lethal and non-lethal missions at the command of a single pilot' (ibid.: 39). Such a move will require decisions being made by the swarm, human decision making will be too slow and not able to react to the control of several aircraft at once. With the increasing pace of the action and with the potential of several aircraft to choose targets at the same time, it will not be possible to have the human make all of the decisions to kill.

The main ethical concern is that allowing robots to make decisions about the use of lethal force could breach both the Principle of Distinction and the Principle of Proportionality as specified by International Humanitarian Law (Sharkey 2008b). Currently and for the foreseeable future no autonomous robots or artificial intelligence systems have the necessary properties to enable discrimination between combatants and civilians or to make proportionality decisions.

Under the Principle of Distinction only combatants/warriors are legitimate targets of attack. All others, including children, civilians, service workers and retirees, should be immune from attack. The same immunity covers combatants that are wounded, have surrendered or are mentally ill.[5] The Principle of Proportionality applies in circumstances where it is not fully

possible to protect non-combatants in an action. It requires that the loss of life and damage to property incidental to attacks must not be excessive in relation to the concrete and direct military advantage.

The discrimination between civilians and combatants is problematic for any robot or computer system. First, there is the problem is the specification of 'civilianness.' A computer can compute any given procedure that can be written as a programme. 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 precise specification of what a civilian is. The Laws of War do not help. The 1949 Geneva Convention requires the use of common sense to determine the difference between a civilian and combatant 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.
   Protocol 1 Additional to the Geneva Conventions, 1977 (Article 50)

Even if a clear definition of civilian did exist, it would have to be couched in a form that enabled the relevant information to be extracted from the sensing apparatus. All that is available to robots are sensors such as cameras, infrareds, sonar, lasers, temperature sensors and ladars etc. While these may be able to tell us that something is a human or at least animal, they could not tell us much about combat status. There are systems that can identify a face or a facial expression but they do not work well on real time moving people. And even if they did, how useful could they be in the fog of war or from the air? British teenagers beat surveillance cameras just by wearing hooded jackets.

In a war with non-uniformed combatants, knowing who to kill would have to be based on situational awareness and on having human understanding of other people's intentions and their likely behaviour. In other words, human inference is required. Humans understand one another in a way that machines cannot. Cues can be very subtle and there are an infinite number of circumstances where lethal force is inappropriate. Just think of children being forced to carry empty rifles or of insurgents burying their dead.

The problem has been put clearly by Major Daniel Davis (2008), a combat veteran of Iraq 1991 and Afghanistan 2005, who writes, 'Suggesting that within the next 12-plus years technology could exist that would permit life-and-death decisions to be made by algorithms is delusional. A machine

cannot sense something is wrong and take action when no orders have been given. It doesn't have intuition. It cannot operate within the commander's intent and use initiative outside its programing. It doesn't have compassion and cannot extend mercy.' Davis also quotes Colonel Lee Fetterman who has a high regard for remote controlled robots:

> The function that robots cannot perform for us — that is, the function we should not allow them to perform for us – is the decide function. Men should decide to kill other men, not machines. This is a moral imperative that we ignore at great peril to our humanity. We would be morally bereft if we abrogate our responsibility to make the life-and-death decisions required on a battlefield as leaders and soldiers with human compassion and understanding. This is not something we would do. It is not in concert with the American spirit. (ibid.)

Turning to the Principle of Proportionality, there is no way for a robot to perform the human subjective balancing act required to make proportionality decisions. No clear objective methods are provided for calculating what is proportionate in the Laws of War (Sharkey 2009). The phrase 'excessive in relation to the concrete and direct military advantage expected to be gained' is not a specification. It is also practically impossible to calculate a value for 'the actual military advantage'. 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? Proportionality calculations should be based on the likely differences in military outcome if the military action killing innocents had not been taken (Chakwin et al. 2002).

Commanders have to weigh the circumstances before making a decision but ultimately this is a subjective matter. 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. There has to be circumstantial and subjective estimates about just how many innocent people are equal to the military value of the successful completion of a given mission. A computer system could be used to assist commanders to determine which weapons would cause the minimum of collateral damage. However, after the calculation is completed, a decision must still be made about whether the minimal number of civilian casualties and damage to civilian property is proportional to the military advantage gained.

In the heat of battle, both the principles of discrimination and proportionality can be problematic. Humans do make errors and can behave unethically but they can be held accountable. Armed autonomous robots could violate both of these principles but could not be held accountable (Sharkey 2008c). There is no way to punish a robot. We could threaten to switch it off but that would be like telling your washing machine that if it does not remove stains properly you will break its door off. This leaves the question about who is responsible along the long causal chain that includes: the manufacturer for the description they gave, the programmer, the designer, the Department of Defense, the generals or admirals in charge of the operation and the operator. Can we hold someone

accountable? If not, then, legally speaking, these robots cannot be used (Sparrow 2007).

Some Lessons to Learn

It is quite likely that autonomous robots will come into operation in a piecemeal fashion. Research and development is well underway and the fielding of autonomous robot systems may not be far off. However, to begin with they are likely to have assistive autonomy on board such as flying or driving a robot to a target destination and perhaps even selecting targets and notifying a human (the Israeli autonomous ground robot, Guardium, already has this functionality). This will breed public trust and confidence in the technology – an essential requirement for progression to autonomy (Dahm 2010). The big worry is that allowing such autonomy will be a further slide down a slippery slope to give machines the power to make decisions about whom to kill.

Apart from the legal and ethical problems of discrimination and proportionality for fully autonomous armed robots, there are lessons to be learned from the current application of remote piloted vehicles. We discussed some of the controversy about the issue of moral disengagement of the operators and the Playstation mentality (Royakkers & van Est 2010). While this is an interesting and plausible narrative, it is not possible to be conclusive because the evidence, outside of the anecdotal, is just not available. It must be noted, nonetheless, that the power of graphic detail to desensitise should not be underestimated. The narrative about moral disengagement is even more compelling when it comes to autonomous robot warriors. As Cumming (2006) puts it, 'developing technologies of this sort also have the potential to provide for the creation of moral buffers that allow humans to act without adequately considering the consequences'.

Another lesson can be learned from the use of the remote piloted armed aircraft by the CIA for targeted killings. At best, such targeted killings are legally questionable in terms of the Laws of War as discussed above. Apart from problems with verifying intelligence about targeting and the lack of a transparent judicial process, there is concern that the use of drones has greatly expanded the range of targets that can be picked off by the CIA. Autonomous aircraft that can hover on their own for extended periods and select their own targets could greatly exacerbate the problems. The Defense Advanced Research Projects Agency Vulture programme has just awarded a contract to Boeing (teamed with QinetiQ) to develop a heavier-than-air platform capable of keeping 1,000 lb of payload with 5 kilowatts of power aloft for five years. These developments need to be watched very carefully.

It is difficult to design and develop these new technologies but they are not so difficult to copy. We have already seen the rapid proliferation of drones in response to their perceived military success in the Iraq/Afghanistan conflicts. If armed autonomous robots are fielded in the same way with or without a human in the weapons loop, it will not be long before there is similar proliferation. If it turns out to be militarily advantageous to keep a human

out of the loop – for example in circumstances when it is not possible to maintain a radio or satellite link – then they will be used by all.

Notes

[1] Israel may have been using armed drones for longer but denied it for several years despite eyewitness testimonies. It cannot be verified here.

[2] The Year of the Drones, New America Foundation website: http://counterterrorism.newamerica.net/drones, accessed 15 October 2010; Internet.

[3] DoD Press Briefing with Mr Weatherington from the Pentagon Briefing Room, US Dept. of Defense, 18 December 2007, www.defenselink.mil/transcriptstranscript.aspx?transcriptid=4108; Internet.

[4] United States Air Force Unmanned Aircraft Systems Flight Plan 2009–2047, Headquarters of the United States Air Force, Washington, DC, 18 May 2009.

[5] But see also Ford 1944.

References

Alston, P. (2010) Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Philip Alston to the UN Human Rights Council, 14th session, A/HRC/14/24/Add.6, 28 May 2010.

Anderson, K. (2010) Written Testimony to the US House of Representatives Committee on Oversight and Government Reform, Subcommittee on National Security and Foreign Affairs, Subcommittee Hearing: 'Drones II', 28 April 2010, accessed 3 November 2010, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1619819; Internet.

Chakwin, B., Voelkel, D. & Enright S. (2002) Leaders as Targets, Joint Forces Staff College, Seminar # 08, accessed 3 November 2010, available at http://www.stormingmedia.us/13/1361/A136124.html; Internet.

Committee on Autonomous Vehicles in Support of Naval Operations National Research Council (2005) *Autonomous Vehicles in Support of Naval Operations* (Washington DC: The National Academies Press).

Cumming, M. L. (2006) Automation and Accountability in Decision Support System Interface Design, *Journal of Technology Studies*, 32(1), pp. 23–31.

Daddis, A. (2004) Understanding Fear's Effect on Unit Effectiveness, *Military Review*, July–August, pp. 22–27.

Dahm, W. J. A. (2010) 'Technology Horizons': A Vision for Air Force Science & Technology, Volume 1, AF/ST-TR-10-01, accessed 14 October 2010, available at http://www.airforce-magazine.com/SiteCollectionDocuments/TheDocumentFile/Strategy and Concepts/TechnologyHorizonsVol1_2010.pdf; Internet.

Davis, D. (2007) Who Decides: Man or Machine? Armed Forces Journal, 11, accessed 3 November 2010, available at http://www.armedforcesjournal.com/2007/11/3036753; Internet.

Ford, J. S. (1944) The Morality of Obliteration Bombing, *Theological Studies*, 5, pp. 261–309.

Grossman, D. (1995) *On Killing: The Psychological Cost of Learning to Kill in War and Society* (Boston: Little, Brown).

Hersh, S. M., (2002) Manhunt: The Bush Administration's New Strategy in the War against Terrorism, *New Yorker*, 66 (December), accessed 3 November 2010, available at http://www.newyorker.com/archive/2002/12/23/021223fa_fact; Internet.

Holmes, R. (2003) *Acts of War: The Behaviour of Men in Battle* (London: Cassell).

Koh, H. (2010) Speech to The American Society of International Law, Washington, DC. March 25, 2010, accessed 3 November 2010, available at http://www.state.gov/s/l/releases/remarks/139119.htm; Internet.

Lindlaw, S. (2008) UAV operators suffer war stress, *Air Force Times*, 8 August 2008, accessed 10 October 2010, available at http://www.airforcetimes.com/news/2008/08/ap_remote_stress_080708/; Internet.

Marshall, S.L. A. ([1947] 2000) *Men against Fire: the Problem of Battle Command* (Oklahoma City: Oklahoma University Press).

Panetta, L. E. (2009) Director's Remarks at the Pacific Council on International Policy, 18 May 2009, accessed 15 October 2010, available at https://www.cia.gov/news-information/speeches-testimony/directors-remarks-at-pacific-council.html; Internet.

Parks, H. W. (1989) Memorandum on Executive order 12333, Reproduction Department of the Army, office of the Judge Advocate General of the Army, Washington DC 20310–220.

Royakkers, L. & van Est, R. (2010) The Cubicle Warrior: the Marionette of Digitalized Warfare, *Ethics and Information Technology*, 12, pp. 289–296.

Sharkey, N. (2008a) Cassandra or the False Prophet of Doom: AI Robots and War, *IEEE Intelligent Systems*, 23(4), pp. 14–17.

Sharkey, N. (2008b) Grounds for Discrimination: Autonomous Robot Weapons, *RUSI Defence Systems*, 11(2), October, pp. 86–89.

Sharkey, N. (2008c) The Ethical Frontiers of Robotics, *Science*, 322, pp. 1800–1801.

Sharkey, N. (2009) Death Strikes from the Sky: The Calculus of Proportionality, *IEEE Science and Society*, Spring issue, pp. 16–19.

Singer, P. W. (2009) *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (New York: The Penguin Press).

Sparrow, R. (2007) Killer Robots, *Journal of Applied Philosophy*, 24, pp. 62–77.

Stein, J. (2010) CIA Drones Could be Grounded by Software Suit, *Washington Post*, SpyTalk, 11 October 11, 2010, accessed 14 October 2010, available at http://blog.washingtonpost.com/spy-talk/2010/10/cia_drones_could_be_grounded_b.html; Internet.

Turse, N. (2010) Drone Surge: Today, Tomorrow and 2047, Asia Times, 26 January 2010, accessed 15 October 2010, available at http://www.atimes.com/atimes/South_Asia/LA26Df01.html; Internet.

Biography

Noel Sharkey is a Professor of Artificial Intelligence (AI) and Robotics and a Professor of Public Engagement at the University of Sheffield (Department of Computer Science), UK. His research and teaching span academic disciplines, including engineering, philosophy, psychology, cognitive science, linguistics, artificial intelligence, computer science and robotics. Previously he worked at Yale in AI, Stanford in Psychology, Essex in Linguistics and Cognitive Science and Exeter in Computer Science. He is on the editorial board of several journals and is Editor-in-Chief of the *Journal of Connection Science*. His work has ranged from computational modelling of psychological processes, to neural computing and machine learning, to monitoring safety critical systems, to robotics. He appears regularly on television, radio and in the press.

As well as more than 150 academic articles and books, Professor Sharkey writes for national newspapers and magazines and has created thrilling robotics museum exhibitions and mechanical art installations. His core research is now on ethical applications of robotics in domains such as the military, child care, elder care, policing, medicine and crime. He is cofounder of the International Committee for Robot Arms Control and has been awarded a two-year Leverhulme Research Fellowship (September 2010) to conduct an ethical and technical appraisal of Robots on the Battlefield.