CANADIAN FORCES COLLEGE / COLLÈGE DES FORCES CANADIENNES
JSCP 34 / CCEM 34

Master of Defence Studies Research Project

FUTURE WARFARE OR FUTURE FOLLY?

AUTONOMOUS WEAPON SYSTEMS ON THE FUTURE BATTLEFIELD: AN ASSESSMENT OF ETHICAL AND LEGAL IMPLICATIONS IN THEIR POTENTIAL USE

By Lieutenant Colonel NS Makin Royal Artillery

25 April 2008

Abstract

This paper considers the ethical and legal issues surrounding the deployment of autonomous weapon systems on the future battlefield. System definitions, potential military advantages in the use of autonomous battlefield systems and employment opportunities on the battlefield are discussed along with the rationale behind autonomous system proliferation. How autonomous systems could be incorporated within future doctrine is considered. Ethical issues are assessed within the principles of Just War theory and 'jus in bello', emphasizing discrimination and responsibility. The legal environment within which autonomous weapon systems could operate is described and, in particular, the significance of creating collateral damage is discussed. Lessons are drawn from the Ottawa Treaty banning the use of anti-personnel mines, an example of the consequences of ethical and legal failure. Although considerable technical challenges exist in the development of fully autonomous weapon systems that can be deployed onto the battlefield of the future, it is considered more likely than not that these challenges will ultimately be overcome. However, if these systems fail to perform adequately in either the ethical or legal domains discussed within this paper then the wisdom of their future use is questioned.

Contents

Introduction

Throughout history, the implementation of military technology has given rise to a range of enhanced capabilities such as the introduction of gunpowder, the longbow, machine guns and armoured vehicles. Such technology has enabled weapon systems to be developed capable of influencing activities across a wider battlespace with increased lethality and precision. Technology continues to enhance military capabilities to find, fix and strike further. Moreover, the deployment of autonomous battlefield systems offers the potential to augment military capabilities and conduct activities which soldiers cannot perform. Automation which involves equipment instead of manual labour already exists on the battlefield of today. It ranges from relatively simple mechanical automatic

weapons through to complex digitised robot... m6(kensitynsugh to aer wleh tigtisgltowerdaos wereovey hsccis)potenhantie

A key issue with the implementation of autonomous systems is the level of autonomy at which these systems are capable of operating and how that may change the human/technology/machine relationship. In particular, who takes or is given responsibility for decision making and subsequent actions. Consequently the employment of autonomous weapon systems brings with it a number of ethical and legal issues. This paper reviews these issues. It defines autonomous weapon systems, describes the key system components and assesses their advantages and potential utility. Ethical issues are reviewed through an analysis of key principles associated with 'jus in bellum' or Justice in War: notably discrimination and responsibility. This paper considers that responsibility (for the consequences of the actions of autonomous battlefield weapon systems) and discrimination (as to what constitutes a legitimate target) are two essential areas that need addressing in order for such systems to gain military and political (as well as public) support. Unintended consequences, such as autonomous battlefield systems resulting in a lowering of a nation's threshold to conduct warfare, are also discussed. An assessment of legal issues follows which includes a comparison with the Ottawa Treaty banning the use of anti-personnel mines. A failure to consider ethical and legal issues when designing autonomous weapons should result in a similar ban being created.

Autonomous weapon systems offer a range of benefits and capabilities which could significantly enhance military operations on the future battlefield. However, such systems must be designed, developed and deployed in accordance with ethical and legal principles. Failure to ensure this could jeopardize meeting accepted legal and ethical norms for the conduct of battle and nullify the benefits of developing such systems in the first place.

Chapter 1 – Setting the Scene

This chapter sets the conditions for the rest of the paper to assess the ethical and legal issues surrounding autonomous battlefield systems. It provides definitions, describes the essential components of autonomous systems, the advantages and potential utility of such systems and assesses how their potential widespread implementation on the battlefield of the future is reflected in concepts that will eventually become doctrine. In considering the potential utility of such systems it shows why autonomous battlefield weapons systems are the focus of considerable military Research and Development (R & D) funding and effort.

1.1 Technology and history

Technology has had a significant influence over the conduct of warfare and there are numerous historical examples of how the implementation of military technology has given rise to enhanced capabilities (such as the introduction of gunpowder and the longbow). Military technology has also affected the status of the soldier on the battlefield. In particular the introduction of the machine gun can be considered to have made 'soldiers secondary to machines, robbing them of primacy on the battlefield'[1]. Furthermore, the combined use of tanks and infantry towards the end of the First World War challenged attrition trench warfare of the preceding decades. Radar, computers, global positioning systems, satellites, lasers and many other technologies have and continue to change the conduct of modern warfare. Steven Gaffney remarks that 'victory

[1] John Ellis, *The Social History of the Machine Gun* (London: Pimlico, 1976) 50.

has often been determined by the quality of weapons and tactics before the battle had even begun, regardless of the quality of the individual warriors'[2]. Such weapons allow targets to be identified at greater range, with improving definition and clarity, throughout the spectrum of environmental conditions and to be engaged by more accurate weapons systems which offer greater reduction in collateral damage and fratricide. Consequently military technology will continue to enhance the ability of the military to find, fix and strike from greater standoff distances. In future wars the quality of the combatant's automated systems developed in times of peace may be the deciding factor in battle[3]. Furthermore, it is likely that 'investments in science and technology … will continue to be the essential underpinning for maintaining superior Land Force warfighting capabilities'[4]. The military historian Max Boot, however, adds caution. He notes that technology alone has not, historically, created an overriding advantage. To be fully effective, it needs to be synchronised with tactics, organisations, training and leadership.[5]

---

[2] Steven Gaffney, *Battlefield Automation: A Luddite's View* (US Marine Corps, 1991) 2,

www.globalsecurity.org/military/library/report/1991/gsj.html. Internet accessed 12 December 2007.

[3] *Ibid*, …, 2.

[4] Regan Reshke, *Brave New Conflicts: Emerging Global Technologies and Trends* (Canadian Forces Department of National Defence Research Paper, 2007) 1,

http://lfdts.army.mil.ca/DLCD-DCSFT/pubs/jadex/paper1/JADEX-Paper-1-en.pdf Internet accessed 9 January 2008.

[5] Max Boot, *War Made New: Technology, Warfare, and the Course of history 1500 to Today* (New York: Gotham Books, 2006) 10.

1.2 Definitions and system components

This section considers a number of definitions associated with autonomous weapon systems. First is military technology which can be defined as 'the application of practical or mechanical sciences for military purposes'[6]. Within this broad definition sits automation, a key system characteristic, which is common on the battlefield of today. Automation can be defined as the 'use of automatic equipment instead of manual labour[7]'. This general definition is further refined by Barbara Adams and Lora Bruyn who, in their literature review of automated systems, offer the following:

> 'A device or system that accomplishes (partially or fully) a function that was previously carried out (partially or fully) by a human operator and
>
> Automation is any sensing, detection, information-processing, decision-making, or control action that could be performed by humans but is actually performed by machine.'[8]

These two definitions emphasise the transfer of functions from a human operator (in the case of this paper the soldier) to a machine. Automatic is defined as '(of a

---

6 Collins *Dictionary of the English Language* edited by Patrick Hanks (London: Collins, 1982) 1492.

7 *Ibid*, …, 98.

8 Barbara Adams and Lora Bruyn, *Trust in Automated Systems Literature Review* (Toronto: Department of National Defence, Defence Research and Development Canada, CR-2003-096) 11,

http://pubs.drdc.gc.ca/PDFS/unc59/p520342.pdf Internet accessed 31 March 2008.

firearm) self loading and able to fire continuously as well as happening without conscious thought'[9]. An automatic weapon therefore transfers some of the function previously preformed by a soldier (such as pulling the trigger on a machine gun) to the machine itself. This differs from an autonomous system which performs without soldier decision-making input: it goes further than a functional approach to include system operation and management. Automation offers the potential to augment (ie to add a new capability or function) rather than purely enhance (ie improve upon) the abilities of a soldier. It can enable activities to be conducted which soldiers cannot perform. Such autonomous systems are also not necessarily unitary entities. For example, a military aircraft will contain many different levels of automation in its inter-related subsystems[10]. This paper focuses on battlefield automation (including systems that influence that environment from the air) and within the context of fully autonomous weapon systems. Autonomous systems possess self governance and independence and fully autonomous weapon systems have the ability to operate without a soldier within the decision-making process. In contrast semi autonomous systems have regulatory control exercised by the soldier (such as when to fire) although some fully autonomous systems may be capable of being switched to a semi autonomous mode. These systems can also be considered robotic in that they are 'automated machines programmed to perform specific mechanical functions'[11]. Unless stated otherwise within this paper, autonomous weapon systems are considered to be robotic in nature.

---

9 Collins *Dictionary of the English Language* …, 98.

10 Barbara Adams and Lora Bruyn, *Trust in Automated Systems Literature Review* …, 13.

11 Collins *Dictionary of the English Language* …, 1261.

This paper also considers advanced autonomous systems. These differ from autonomous systems in that they possess Artificial Intelligence (AI), a synthetic decision-making environment which has the ability to learn from experience. As these advanced systems develop and mature, the role and significance of the soldier as the 'man in the loop' could decline and eventually become redundant. It is at this stage that autonomous systems could operate without reference to soldier decision-making, although within a wider context a soldier should still decide to deploy them so ultimately a soldier is still responsible for their employment. Before this 'line in the sand' is crossed, the ethical and legal issues raised within this paper will need to have been addressed. To address them once AI technology reaches maturity is too late. Autonomous system design must be guided by ethical and legal principles and the issues debated, reviewed, discussed and resolved in order for such systems to be of future military use. Criticism has been directed towards autonomous weapon designers who have been accused of being 'solely enamoured with technology'[12]. This is perhaps an overstatement and simplification. Military R & D organisations will be aware of the necessity to understand likely battlefield doctrine, tactics, techniques and procedures (TTPs) for new technological capabilities. We have already indicated that autonomous systems can facilitate capabilities that soldiers cannot perform. Thus it is necessary to establish an iterative capability development model or process where doctrine and TTPs evolve as the new technical capabilities are tested and evaluated. This will counter critics who consider that system designers are failing to consider how such systems will be employed on the

[12] Roxana Tiron, *Lack of Autonomy Hampering Progress of Battlefield Robots* (National Defense: Arlington Volume 87, Issue 594, May 2003) 35.

battlefield, how they will be supported and sustained and how they will be integrated within the tactical environment[13].

Currently automation ranges from relatively simple mechanical automatic weapons through to complex digitisation and data management systems with significant reliance on the soldier to make decisions: whether to pull the trigger of an automatic rifle or to accept a recommendation offered by a command and control decision-making tool. Maintaining the 'soldier in the loop' in the decision making process makes such systems semi-autonomous: somewhere in the system there is human control. A fully autonomous system is capable of making a decision (or taking an action) without reference to human interaction. As science and technology mature and the products of military research and development move from science fiction to science fact such automation has the potential to become common on the battlefield of tomorrow. The majority of military systems still require some level of human 'supervision' or decision-making. The need for such supervision and oversight will undoubtedly decrease as technology, specifically computing power, sensory systems and AI, improves. Nevertheless, soldiers are likely to remain involved in the tasking process for these systems (the issue of advanced autonomous systems tasking other similar systems falls outside the scope of this paper).

---

13 *Ibid*, …, 35.

1.3 Rationale for autonomous system proliferation

It is likely that much of the technology on the battlefield of tomorrow will be either directly or indirectly developed by the US. It is therefore worthwhile to briefly highlight why the US is so reliant now, and for the foreseeable future, on battlefield technology. The US has an aversion to casualties, rooted within American culture and history, specifically following from Vietnam, and, more recently Somalia[14]. Consequently it has been argued that US political and military leadership has become casualty-phobic: the aversion to casualties is so strong 'as to elevate the safety of American troops above the missions they are assigned to accomplish'[15]. Others comment, also using the example of the US in Vietnam, that 'as the war progressed, the political need to control casualties served to increase the importance of technology'[16]. It is therefore perhaps unsurprising to note observations such as by 'utilizing technology to take a US service member out of harm's way, it is worth every penny' and that 'armed robots can be used as a force multiplier to augment an already significant force in the battle space'[17]. Within this context a force multiplier could involve a single autonomous system conducting a number of functions that soldiers may not be capable of completing. John Ellis, in his analysis of the machine gun, provides further historical perspective.

---

14 Jeffery Record, *Collapsed Countries, Casualty Dread, and the New American Way of War* (Parameters: US Army War College Quarterly, Summer 2002) 2.

15 *Ibid*, …, 2 – 3.

16 Warren Chin, *Technology, Industry and War, 1945 to 1991*, Chapter 2 within *War in the Age of Technology: myriad faces of modern armed conflict,* edited by Geoffrey Jensen and Andrew Wiest (New York: New York University Press, 2001) 52.

17 Stew Magnuson, *RoBo Soldiers* (Arlington: National Defence, Volume 92, Issue 646, September 2007) 1.

Referring to the firepower of the Gatling Gun and it's ability to make 'one soldier as effective as one hundred', he notes that a consequence of its use could have been to supersede the necessity of large armies and in doing so reduce exposure of soldiers to battle and disease[18]. Further historical rationale behind the development of automatic weapons (and as a precursor to autonomous systems) provided by Ellis was that a key driver in developing machine guns in the US in the mid nineteenth century was because of the shortage of manpower. These issues may continue to influence, to a lesser or greater degree, future weapon system designers.

1.4 Potential military advantages of autonomous battlefield systems

Autonomous systems promise highly significant enhancements to the military environment. Their ability to conduct activities which soldiers cannot perform has already been introduced. Ronald Arkin[19] suggests a number of areas where the potential utility of autonomous systems on the battlefield can be demonstrated. He proposes that such systems will have the potential to be much better equipped with battlefield observation sensors than soldiers currently possess (ie the limited spectral and spatial resolution of the human eye). This approach is contingent upon a design philosophy that prefers to equip robots rather than soldiers. Soldiers could be augmented by systems that match the sensory performance level of machines, although areas such as man machine interface, portable power sources and load carriage could complicate the process.

---

[18] John Ellis, *The Social History of the Machine Gun* …, 21.

[19] Ronald Arkin, *Governing Lethal behaviour: embedding ethics in a Hybrid Deliberative/Reactive Robot Architecture* (Georgia Institute of Technology, Technical Report GIT-GVU-07-11, 2007) 6.

Autonomous systems could be capable of integrating information from a wider range of sources before responding far faster with either lethal or non lethal force than a soldier. These systems may be capable of exploiting more of the electro-magnetic spectrum in order to enhance their ability to gather information and data throughout the wide range of battlefield environmental conditions (in particular with the capability to operate 24/7 and throughout adverse weather). They could be designed to perform beyond the normal unassisted physical abilities of the soldier, in particular with respect to strength and endurance. They should be capable of operating in harsher environments where soldiers cannot function, perhaps due to unacceptable levels of risk, such as contaminated areas or where smart munitions have been deployed[20]. If smaller than the soldier they could potentially be more survivable by presenting a reduced target to the enemy. Furthermore, such systems have the potential to be immune to the inherent human characteristic of self preservation and near misses or suppressive fire will not affect performance like they would on soldiers[21]. Arkin considers that they could be employed in a 'self-sacrificing manner', therefore avoiding soldiers being exposed to life threatening and dangerous situations[22]. However, if robotic systems are developed which possess their own thought and logic processes (but not necessarily emotion) then this characteristic could become challenging, since they could possess a drive towards self preservation. This aspect of self preservation may not necessarily be the same as that for soldiers. With soldiers the fear response often governs behaviour (particularly within the

---

[20] Thomas Cowan, *A Theoretical, Legal and Ethical Impact of Robots on Warfare* (Carlisle Barracks, Pennsylvania: US Army War College Research Project: Mar 2007) 3.

[21] *Ibid*, …, 3.

[22] Ronald Arkin, *Governing Lethal behaviour: embedding ethics in a Hybrid Deliberative/Reactive Robot Architecture*, …8.

context of suppressive fire) whereas in autonomous systems behaviour may well be governed by logic: can the mission be achieved or not[23]. Whilst this places the issue into the realms of science fiction, the issue has already been broached. The science fiction writer Isaac Asimov, writing in 1950, produced within his book '*I* Robot', the Three Laws of Robotics:

> 'A robot may not injure a human being or, through inaction, allow a human being to come to harm;
>
> A robot must obey orders given it by human beings where such orders would not conflict with the First law, and;
>
> A robot must protect its own existence as long as such protection does not conflict with the first or Second Law'[24].

Such a code could well form part of a future ethical framework. Autonomous systems also have the potential to be designed 'without emotions that cloud their judgment or result in anger and frustration with ongoing battlefield events.'[25] Whilst this may be correct in understanding soldiers' emotions, conflicting rules and logic within autonomous systems could disrupt their algorithms and thus resemble emotions from a

---

[23] http://blog.wired.com/defense/2008/01pentagon-research.html

[24] Isaac Asimov, *I Robot*, (New York: Doubleday and Company, 1950) 43.

[25] Ronald Arkin, *Governing Lethal behaviour: embedding ethics in a Hybrid Deliberative/Reactive Robot Architecture* … 6.

human perspective. Conversely, some emotion could be important in influencing subjective judgment on a suitable course of action to be followed. The following provides an example where soldier decision-making can become removed from stress and emotion. US Army Predator unmanned aerial vehicles (UAVs) targeting insurgent activity within Iraq can be flown by operators located within a low stress air-conditioned office environment from within the US. These operators are detached from some key contributors of battlefield stress such as fear of injury or death, sleep deprivation (due to prolonged periods of operation), the physical stress of combat and the confusion of a firefight. Removed from these external stresses, the operator could arguably make a better decision than a soldier subjected to a range of additional stresses. However, devoid of emotive content, a purely technological decision could also be devoid of ethical considerations as well.

Autonomous systems could potentially operate more ethically than soldiers. This is suggested by Arkin who considers that the process of making an ethical decision is one of logic based upon specific sets of criteria and values. Consequently, such decision-making could be based upon 'mathematical functional notations'[26]. In order for this to work, the logic, rules and criteria will need to be consistent and robust enough to handle any situation. Is it likely that such rules can be crafted in a manner that will handle ethical dilemmas? These ethical issues are considered in the next chapter.

[26] Ronald Arkin, *Governing Lethal behaviour: embedding ethics in a Hybrid Deliberative/Reactive Robot Architecture* … 14.

Autonomous systems, by replacing soldiers on the future battlefield, will remove potential risk to those soldiers not deployed onto the battlefield. We have already seen that this is typically cited as the overriding reason why the technology is being pursued by the US Army and other military forces. As the country possessing the largest military R and D budget[27], the US focus on technology to reduce exposure of soldiers to battlefield risks and therefore the potential of becoming casualties cannot be overlooked.

If autonomous weapons systems are fully autonomous, then they are unlikely to need to accommodate onboard soldiers. Consequently designers could become 'unrestrained by conventional design parameters … such as shock and vibration, survivability and risk factors.'[28] Noel Sharkey, referring to US military progress (specifically with the Marine Corps) on battlefield autonomous vehicles, considers that such systems will be 'cheap to manufacture, require fewer support personnel and … can perform better in more complex missions.'[29] Such systems could also be cheaper to manufacture if the process itself is automated (ie robots building robots). Training requirements could also be reduced. Once a single autonomous system is trained and ready for operational deployment (an achievement which admittedly may prove to be

27 According to US Government Accountability Office figures, the 2007 Research and Development budget is $73.2 Bn covering over 1000 projects, taken from:

Sandra Erwin, *Defense department should refocus technology spending, experts warn*, (Arlington National: Defence, November 2007).

28 Roxana Tiron, *Lack of Autonomy Hampering Progress of Battlefield Robots* …, 35.

29 Noel Sharkey, *Automated Killers and the Computing Profession* (New York: Computer, Volume 40, Issue 11, November 2007) 124.

costly in time and resources) further identical systems could be made ready by simply uploading the previously tested algorithms.

Thomas Cowan offers a further benefit of autonomous systems, linked with demographics. He considers both increasing personnel costs and shrinking populations from which to draw soldiers will make it more challenging to deploy soldiers onto the future battlefield[30]. Personnel costs can include training, pay and pensions. As life spans increase the provision of military pensions is likely to create a significant financial overhead. For example, a soldier joining an army at 20, who serves for a further 20 years and lives to 80 will draw a military pension for twice as long as he served. Edward Luttwak provides some historical context to the issue of demographics raised by Cowan. He remarks that the loss of a son in combat from amongst preindustrial and early industrial societies, where families had many children (and where losing some to disease was normal) was 'fundamentally less unacceptable than for today's families, with their one, two or at most three children.'[31] Luttwak goes further by stating that in low-birth rate, post industrial societies there is a refusal to sanction casualties of any avoidable combat[32]. Consequently the deployment of autonomous systems could be highly appealing to both military and political leaders. The following section considers employment opportunities for these systems and provides an assessment on the possibility of these capabilities becoming achievable.

---

30 Thomas Cowan*, A Theoretical, Legal and Ethical Impact of Robots on Warfare*, …, 3.

31 Edward Luttwak, *Toward Post-Heroic Warfare,* (Foreign Affairs, Volume 74, Number 3, May/June 1995) 115.

32 *Ibid*, …, 115.

1.5 Employment opportunities on the battlefield

The following generic areas are widely referred to within literature as being applicable for the use of semi through fully autonomous systems; reconnaissance (including nuclear, chemical and biological detection), surveillance, target identification and engagement (such as through automated UAVs and land based robotic systems), de-mining (including tactical countermine and post conflict environments), explosive ordnance demolition (EOD), decontamination, logistics (such as the transport and asset management of combat supplies) and perimeter security and access control. Distinguishing between near term employment (over the next 5 years), mid term (5 to 15 year) and the future (beyond 15 years) will assist in identifying the increasing urgency to address ethical and legal issues as the systems become increasingly autonomous. Ultimately, it could be possible that autonomous systems will become sufficiently advanced so as to be able to make human-like decisions (including life and death decisions) in dynamic and changing situations. If this point is reached then such systems could be employed within any battlefield environment or role unless, perhaps, there were unaddressed ethical and legal issues.

Within the near term, numerous weapon systems possessing various levels of autonomy already exist including: 'cruise missiles, torpedoes, submersibles, robots for urban reconnaissance, uninhabited aerial vehicles (unarmed UAVs) and uninhabited combat aerial vehicles (armed UAVs).'[33] None of these strictly fall within the earlier

---

[33] Robert Sparrow, *Killer Robots*, …, 62.

definition of fully autonomous. Nevertheless, weapon systems such as UAVs possessing the ability to take off, fly to an objective and subsequently return with minimal human control are already common. Their proliferation and importance in supporting military operations will continue to develop. The GLOBAL HAWK UAV, for example, provides a key surveillance capability for the US military. Other smaller but equally capable systems are widely employed for both target surveillance and target designation[34] and have been extensively used in recent military operations in Kosovo, Afghanistan and Iraq.

South Korea has developed an 'intelligent surveillance and security guard robot' designed to replace soldiers and overcome difficulties of fatigue and operating within severe weather conditions. The manufacturers consider that the system ensures that 'the perfect guarding operation is achieved.'[35] In Gaza the Israeli Defence Force (IDF) is also deploying an integrated border guard system. Both systems are capable of delivering lethal and non lethal effects in manual, semi automatic and fully autonomous modes. Target identification for the latter is based purely on motion sensors with no form of system discrimination. Engagement areas are designated 'kill zones' and marked accordingly on the ground in order to deter entry[36].

---

34 Target designation involves remotely identifying and then tracking or marking a target (such as by a laser) enabling the target to be engaged by another weapon system such as precision guided munitions.

35 Known as the Intelligent Surveillance & Security Guard Robot. Further details given at the following link,

www.samsungtechwin.com/product/features/dep/SSsystem_e/SSsystem.html Internet accessed 19 March 2008.

36 Ronald Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture,* …, 6.

Robotic weapon systems were deployed by the US Army to Iraq in June 2007[37]. These systems, designated special weapons observation remote reconnaissance direct action system (SWORDS), are armed with a light machine gun and designed for use in high risk combat missions, seeking out and neutralising insurgent snipers before the deployment of foot patrols. They are not truly autonomous since they are remotely controlled by a soldier. However, they form a significant part of the US Army's Future Combat System's (FCS) concept project. The FCS R & D package is funded in excess of $14bn US dollars (2006 figures) and includes the US Army Defence Advanced Research Projects Agency (DARPA). A key programme principle is the integration of the 'most advanced battlefield technologies currently available, as well as those likely to become available within the next two decades.'[38] The robotics focus within this programme is threefold;

> 'to develop perception technologies enabling robots to see and understand the environment, to develop intelligent control architectures allowing autonomous planning and execution in tactical environments and to develop human-machine interfaces capable of effectively tasking robots while minimising operator workload'[39].

---

37 Stew Magnuson, *RoBo Soldiers*, …, 3.

38 Robert Sparrow, *Killer Robots*, …, 64.

39 Ian Kemp, *The American Future Combat System*, (Zurich: Armada International, Issue 2, 2006) 1.

The development of technologies enabling target discrimination is a key component of the DARPA research. The need to discriminate between targets is an essential component of the ethical issues considered in the next chapter. If this technology matures, perhaps within the next 15 years, DARPA projects such as the Armed Robotic Vehicle (ARV) with both reconnaissance and assault variants and the Multifunction Utility/Logistics and Equipment (MULE) vehicle with potential utility for countermine operations may be commonly deployed on the battlefield[40]. The Intelligent Munitions System (IMS) also within the DARPA programme is an unattended lethal or non-lethal 'mine' intended to 'provide offensive battlespace shaping and defensive force protection capabilities.'[41] This again requires an ability to discriminate between targets. Furthermore, military exoskeleton research, which has the potential to be incorporated within autonomous systems, is likely to be fully mature within the next 10 years[42].

Beyond 15 years the critical technical challenge is likely to be with the development of AI. Through the development of AI technology soldier dependent functions (such as mission planning, target distinction and engagement decisions) could ultimately be replaced by autonomous systems. A recent study of internet experts

---

40 *Ibid*, …, 6 and 7.

41 *Ibid*, …, 10.

42 Noah Shachtman, *Marines give us Exoskeletons, Self Aware Robots*, Dangeroom Wired Blog Network, November 2007,

http://blog.wired.com/defense/2007/11/marines-heart-e.html#more Internet accessed 19 March 2008.

considers that AI technology will be suitably mature by 2050 at the very latest[43]. The science writer Ray Kurzweil is more optimistic. A strong advocate of AI, he states through his website that both hardware and software to achieve human level intelligence, including the ability to match human emotions, will be available by 2029[44].

The findings above provide insight as to where the future is likely to lead with the development and employment of autonomous battlefield weapons systems. Whilst there seems to be little doubt on the impact that technology is having on the development of military systems, there is some concern over the success rate of these research projects. The failure rate of DARPA projects is in the region of 85%[45]. However, assuming that autonomous weapon systems will feature within the 15% of successful projects, it is these systems that could generate some of the difficult issues highlighted within this paper.

1.6 Autonomous systems and future doctrine

This section considers how autonomous systems relate to future military concepts and doctrine. Such systems should be developed within the overarching vision as to how military activity is to be conducted. In other words, they need to comply with future doctrine, although doctrine is unlikely to be changed until a concept and the underlying

---

43 Janna Anderson and Lee Rainie, *"The Future of the Internet II",* (Pew/Internet and American Life Project, Sep 2006),

http://www.pewinternet.org Internet accessed 20 November 2007.

44 http://www.KurzweilAI.net Internet accessed 31 March 2008.

45 Bruce Falconer, *DARPA's New SuperSoldiers*, (Washington: National Journal, Volume 35, Issue 45, November 2003) un-numbered article.

technology is thoroughly tested and demonstrated. The US Army, possessing the world's largest R and D budget, again provides a useful starting point. Chapter 4 of the US Army's Field Manual 1 (FM1), *'The Army Our Army at War Relevant and Ready Today and Tomorrow'* is devoted to the future operating environmental challenges. It highlights four operating environments that will present themselves as conflict situations for the Army of Tomorrow (AoT): traditional, irregular, catastrophic and disruptive[46].

The traditional challenge will demand an extension of the mastery of conducting major combat operations, countering conventional threats whilst at the same time preparing for those of tomorrow. Irregular challenges will be presented by state and non-state sources using unconventional methods and often manifesting it through 'small wars'. Catastrophic challenges will arise from terrorists and rogue states threatening the use of weapons of mass destruction (WMD) or other means of causing disastrous effects. Disruptive challenges will arise from competitors developing, possessing, and employing breakthrough technologies to gain an advantage in a particular operational domain. FM1 highlights a number of technologies that will be applicable to all of these future operating environments. Within these capabilities are advanced information systems supporting, amongst other capabilities, surveillance and reconnaissance, precision weaponry and force protection. Many of these capabilities could be met by autonomous systems.

---

[46] Headquarters Department of the US Army, *Field Manual 1, The Army Our Army at War Relevant and Ready Today and Tomorrow*, (Washington DC: Department of the US Army, 14 June 2005) 31.

The Canadian Forces Land Operations 2021[47] provides a force employment concept that is broadly shared by leading Allied Western military powers, including the US and the UK. In a future security environment exhibiting volatility and uncertainty and the likely demise of (although never removed) inter-state conflict, it is envisioned that future threats will become more diverse with a greater emphasis on highly adaptive, technology enabled adversaries operating within an asymmetric environment. Within this more complex environment the AoT will be required to operate effectively across the 'physical, moral and informational planes of any conflict'[48]. This conflict will span the continuum of operations from peacetime military engagement through to major combat operations. Through the development of the five operational functions of act, sense, command, shield and sustain, the AoT must be able to successfully engage in the continuum of operations and autonomous systems need to have utility within these functions. The AoT concludes that autonomous systems will be necessary given the advantages they provide within the increasingly complex future security environment.

1.7 Criticism of the current approach to technology and autonomous systems

The previous section demonstrates how technology and autonomous systems can support future military doctrine. This approach in using technology in the future does have its critics. This final section considers concerns over the potential utility of such systems and the potential for reliance on technology to resolve conflict. Military

---

[47] Directorate of Land Concepts and Design, *Land Operations 2021: Adaptive Dispersed Operations. The Force Employment Concept for Canada's Army of Tomorrow*, (Kingston, Ontario, 2007).

[48] *Ibid*, …, 6.

historians such as Record and Mahnken refer to the change of (US) emphasis in future conflict from interstate to failed states. Record considers that the war in Afghanistan exemplifies a 'growing American propensity for intervention in failed states' adding that there are 'no qualified conventional military adversaries left.'[49] Mahnken, writing before the 2003 Gulf War, notes the growing reliance by the US Army on high technology weapon systems and, referring in particular to the 1991 Gulf War, the 'absolute dominance in technology in recent wars.'[50] However, with the greater likelihood of conflict involving failed and failing states in the future, he remarks that a sole focus on technology should not be adopted; technology alone may not guarantee future military success. Whilst technology may well assist in achieving tactical victories it may be insufficient to ensure strategic success. Instead, in order to be successful he considers a military capability will need to; improve cooperation with local officials, law enforcement (police) and covert operations; develop less of a reliance on fire power intensive strategies which are counter productive in conflicts requiring a maintenance of public support; and for the development of skills suited to the longer term nation building rather than being directed at the shorter term decisive battle.

The military historian Van Crefeld considers that future wars will be focused on low intensity conflicts rather than conventional wars. Although writing in 1991 before the second Gulf War[51], he cites Vietnam, Nicaragua, Lebanon and Afghanistan as low

---

49 Jeffery Record, *Collapsed Countries, Casualty Dread, and the New American Way of War*, …, 2.

50 Thomas G Mahnken, *Democracies and Small Wars, edited by Efraim Inbar*, (London: Frank Carr, 2003) 78.

51 The first Gulf War is considered the Iran/Iraq conflict in the 1980s.

intensity conflicts characterised by enemy forces intermingled with the civilian population and widely dispersed[52]. He considers that high technological weapon systems, both current and future, are not sufficiently capable (especially in target recognition) to make 'much of an impression on a dispersed enemy, or (one) intermingled with friendly forces or indistinguishable from the civilian environment.'[53] John Gentry also criticises the US reliance on technology by commenting that it contributes nothing (certainly initially) to complex civil-military operations like those conducted in Haiti, Somalia, Bosnia and Kosovo[54]. However, Van Crefeld does not fully discount military technology of the future, but considers that the focus will move from expensive weapon systems to 'cheap gadgets capable of being manufactured in large numbers.'[55] This could be seen today through the widespread use of Improvised Explosive Devices (IEDs) by insurgents in Iraq and Afghanistan.

Although beyond the scope of this paper, this perceived over reliance on future technology (and therefore autonomous systems) to resolve conflict, emphasizes the need for a more strategic approach to conflict resolution, which will include not just the military but the socio-political aspects of conflict. Nevertheless, whatever the scale of conflict, there will be a high value placed upon the lives of soldiers (certainly within Western culture). Consequently the demand for technologies to reduce the risk to them will be strong and as this chapter has shown, autonomous systems have the potential to

---

52 Martin Van Crefeld, *The Transformation of War*, (New York: The Free Press, 1991) 208.

53 *Ibid*, …, 205.

54 John Gentry, *Doomed to Fail: America's Blind faith in Military Technology*, (Parameters: Volume 32, Winter 2002/2003) 91.

55 Martin Van Crefeld, *The Transformation of War*, …, 209.

reduce these risks. Therefore, the development of deployable autonomous systems capable of sensing the environment, integrating data and information, formulating a decision and then acting upon that decision is essential. Some of the challenging issues in fielding these systems on the future battlefield are discussed in the following chapters.

Chapter 2 - The Ethical Debate

This chapter considers the ethical issues surrounding the employment of autonomous weapon systems on the future battlefield. Included is an assessment of some of the key principles of '*Jus in Bello'*, just and fair conduct in war, from Just War theory. Particular emphasis is given to discrimination, responsibility, necessity and proportionality. These key ethical principles should be considered within the early stages of system design. With a soldier within the decision-making process, responsibility is clearly defined: the action either sits with the individual making the decision or with the authorising authority who gave the soldier the order to act or carry out the specific mission or task. Philosophically, there may be little differentiation between an authority giving an order to a soldier or to an autonomous system. Responsibility still rests with the authorising authority to ensure that the solider or autonomous system is properly trained, tested and or evaluated and suitably equipped to conduct the specific mission or task. However, this chapter demonstrates that discrimination and responsibility present significant challenges for autonomous systems. Although such principles of '*jus in bellum'* do not necessarily form a check list of essential issues that must all be addressed in their entirety, failure to consider any one of them could, potentially, place the employment of that autonomous system at risk by failing to meet accepted ethical norms. This chapter also considers the potential impact of autonomous systems on how war is waged. Reducing risk (not just by the removal of soldiers from the battlefield but also in the minds of decision-makers) could potentially increase the likelihood of waging war. If so then this is contrary to '*jus ad bellum'* and the principles of going to war, specifically rightful intention and as a means of last resort, in that such systems make the decision to wage war easier to make[56]. Such unintended consequences, which may not necessarily be fully understood, should at least be acknowledged as a risk factor. Risk can then be weighed against the threat and against the risk of not having these systems. The need for such work seems compelling: although the amount of technological research on autonomous systems is progressing at break-neck speed, evaluation of the impact on ethical issues is not keeping pace[57]. Sparrow specifically comments that as AI technologies improve 'it seems inevitable that they will be used more and more in warfare'[58]. Consequently he argues that the ethical issues surrounding the employment of these systems need to be addressed as a matter of urgency. There have even been suggestions that there should be a ban on autonomous weapon research until international agreement is reached on how ethical issues should be addressed and taken forward[59]. The recent Royal United Services Institute (RUSI) for Defence and Security Studies sponsored forum on The Ethics of Autonomous Military systems may assist in this process: significant ethical issues raised at this forum are discussed within this chapter[60].

---

56 Peter Asaro, *How Just could a Robot war be?,* Transcript from the 5th European Computing and Philosophy Conference, 2007.

57 Thomas Cowan, *A Theoretical, Legal and Ethical Impact of Robots on Warfare*, …, 13.

58 Robert Sparrow, *Killer Robots*, …, 62.

59 Tom Simonite, *Robot arms race underway expert warns*, NewScientistTech, February 2008, 1, http://technology.newscientist.com/article/dn 13382 Internet accessed 10 March 2008.

60 The Ethics of Autonomous Military Systems was a one day symposium sponsored by RUSI and held at Whitehall, London on 27 February 2008. The aim of the symposium was to examine the ethics of the use of autonomous military systems.

2.1 Standard of performance of autonomous systems

Any autonomous system should be able to perform to a standard which is, at the very least, as good as that conducted by the soldier. This is relatively simple when judging against characteristics such as strength or stamina. Both can be easily measured and comparisons readily made. For example, a recent loading trial of a US 'exoskeleton' robotic system demonstrated that the augmented soldier was capable of lifting a specified weight over a measured height up to 10 times as fast as an unassisted soldier[61]. Whilst not truly autonomous this demonstrates a future capability that such systems could possess. Comparison of autonomous system performance requires a system or process to compare against. When no such comparison is possible, setting the conditions for such a system to achieve as good as or better than human activity becomes much more challenging. Indeed autonomous systems are likely to enable tasks to be completed that have never been conducted before. Whilst comparative assessment is therefore unlikely it is still possible to evaluate legal and ethical consequences of capabilities that soldiers cannot perform.[62] The principles of *Jus in Bello* still apply.

---

[61] SARCOS, under a US military research programme, has designed a military exoskeleton which enhances both strength and stamina of the soldier,

http://blog.wired.com/defense/2007/11/video-fix-super.html Internet accessed 21 December 2007

[62] Comments provided by Regan Reshke, Scientific Advisor DGLCD, Director Science and Technology Land – 7, Defence R&D Canada, Kingston.

2.2 Discrimination

Autonomous systems deployed upon the battlefield will need to be capable of effective target discrimination[63] in order to be considered ethical, although discrimination is but one of several '*jus in bello'* principles against which such systems will be judged. They will need to identify whether the target is friendly or hostile, combatant or non-combatant. For example, discriminating between a soldier in camouflage uniform lying in a prone position with a weapon ready to shoot and a child dressed in khaki playing with sticks; or discriminating between an enemy leaning over and arming an explosive device by the side of a road as against a farmer going about his business in a field. These two simple examples, which are relatively easy to differentiate if a soldier is within the decision-making process, become much more challenging for autonomous systems. For the latter, an accurate assessment relies on technology to detect, define, resolve and discriminate between what is and what is not a suitable target. The outcome of these events will be a lethal or non lethal response (such as an audible warning or the taking of a decision) depending upon the construct of the system. The environment, within which these systems will need to discriminate effectively, will be complex and unpredictable. The 'number of (unfamiliar) circumstances that could occur simultaneously are vast and could result in chaotic robotic behaviour with deadly results'[64]. The technology required to deliver the necessary level of discrimination is considered by some as the greatest task that needs addressing. In particular, the 'rudimentary identification systems currently in

---

63 Discrimination is defined in the Collins English Dictionary as the 'recognition of the difference between one thing and another'.

64 Noel Sharkey *Automated Killers and the Computing Profession*, …, 122.

existence today, specifically identification friend or foe [IFF] systems, only form an elementary basis for further advanced systems.'[65] Decisions on the application of lethal force are currently made by humans and there is concern that this should never be delegated to machines. In particular, concern exists that autonomous systems will never be able to discriminate between targets and subsequently be capable of deciding for themselves when to pull the trigger[66].

Nevertheless, as indicated in the previous chapter, there seems to be optimism regarding the ability of technology to support effective deployment of autonomous systems: it seems the issue is not if the technology can be created but, rather, when. Although the reality of fully autonomous weaponised unmanned ground vehicles which can 'distinguish enemy combatants … and attack them without direct human control' remains decades away, many analysts are convinced that the technology will arrive[67]. The US DoD, for example, has a target for deployment of autonomous vehicle systems by 2015[68] requiring one third of its operational ground vehicles to be unmanned[69]. However, concern has been raised over the potential ability of future technology to

---

[65] Ronald Arkin, *Governing Lethal Behaviour; Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*, …, 11.

[66] Tom Simonite, *Robot arms race underway expert warns*, …, 1.

[67] Preston Lerner, *Robots go to War*, (Popular Science, 2005),

http://www.popsci.com/popsci/printerfriendly/technology/generaltechnology/f59ca60b2e948010vgnvcm1000004eecbccdrcrd.html Internet accessed 12 November 2007.

[68] Cited within the article *Robot Wars*, Economist.com, 7 June 2007,

http://www.economist.com/science/tq/displaystory.cfm?story_id=9249201. Internet accessed 21 December 2007.

[69] Thomas Cowan, *A Theoretical, Legal and Ethical Impact of Robots on Warfare*, …, 1.

deliver fully autonomous systems. Even simple decision-making requires 'complex machinery … comprising of optical analysers, motion guidance systems, simulations, databases … conflict-resolvers and many others.'[70] Sensors may be unable to 'identify human motives, measure human emotions, quantify the coherence of human organsiations or assess the importance of the data they gather.'[71] Certainly there have been setbacks with the development of AI. This has been attributed by some to inflated expectations, partly fuelled by science fiction writers and films (such as *2001: A Space Odyssey)*, failing to materialize[72] [73]. Others consider that R&D programmes working to relatively short timeframes (within the next 5 years) are based upon a 'mythical view of AI[74]'.

A review of some online sources indicates a range of global research into AI and robotics with much of this development attributed to the exponential growth in computer power[75]. Such a pace of change is likely to continue to increase[76]. AI has been shown to out perform man in some narrow expert fields such as in certain medical diagnostic

---

70 Steven Pinker, *How the Mind Works,* (New York: Norton, 1999), 19.

71 John Gentry, *Doomed to fail: America's blind faith in Military Technology*, …, 91.

72 Heather Havenstein, *Spring comes to AI Winter*, (Computerworld, February 2005), 1, http://www.computerworld.com/printthis/2005/0,4814,99691,00.html Internet accessed on 28 February 2008.

73 Raymond Kurzweil, *The Age of Intelligent Machines,* (Massachusetts Institute of Technology: Dai Nippon Printing Company, 1990) 9.

74 Tom Simonite, *Robot arms race underway expert warns*, …, 1.

75 http://www.transit-port.net/AI.CogSci.Robotics/cogsci.html Internet accessed 28 February 2008.

76 Raymond Kurzweil, *The Age of Intelligent Machines,* …, 9.

areas[77]. Whilst true AI[78] remains elusive, this does not mean that the goal is impossible to reach. There does seem to be a growing body of scientific research that points to its possibility if not its inevitability. Consequently, it is important to consider ethical and legal implications in the use of advanced AI autonomous systems today in order to that their development can be shaped in a controllable rather than reactionary manner.

2.3 Proportionality and necessity

Justice in the conduct of war requires that actions are carried out with both proportionality and necessity. Proportionality requires that acts of war should not yield damage disproportionate to the ends that justify their use. Noncombatant harm is justifiable only when it is truly collateral: it is indirect and unintended[79]. It requires a balance between military advantage and damage to non combatants. With these aspects being compared to 'oranges and apples' there is considerable difficulty in their assessment[80]. Whilst there may be little difficulty in identifying when the disproportion is great enough (ie where one trumps the other), difficulty lies with ascertaining a set of

---

77 BBC News, *PC beats doctor in scan tests*, (February 2008), http://news.bbc.co.uk/2/hi/science/nature/7257730.stm Internet accessed 10 March 2008.

78 True AI is more commonly referred to in AI research as "General Artificial Intelligence".

79 Ronald Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberate/Reactive Robot Architecture*, …, 2.

80 Kenneth Anderson, *Law of War and Just War Theory Blog: the Ethics of Robot Soldiers*, July 2007, http://kennethandersonlawodwar.blogspot.com/2007/07/ethics-of-robot-soldiers-html Internet accessed 12 December 2007.

decision rules which provides a value to assess one against the other[81]. The challenge this creates for a soldier is the same as that faced by designers of autonomous systems, with the latter attempting to 'reproduce a moral calculus at the machine level'[82].

System designers could err on the side of caution, since ultimately an autonomous system is at stake and not a soldier's life. Consequently there could be a limit placed upon the destructive power of the autonomous system. By limiting destructive power proportionality issues could be reduced although there would need to be a trade off against mission effectiveness. In other words, a system limited in design with non lethal weapons will not be able to accomplish a mission that may require lethal force. However, having a range of weapon options (both lethal and non lethal) could not only enable a system to select the most appropriate effect required to achieve mission success but also provide options to apply minimum force. Key is whether or not a system can be designed in such a manner that it 'knows' when it is appropriate to use lethal or non lethal means.

Necessity within the realms of '*Jus in Bello'* demands that no more destruction is permitted than is strictly necessary in order to achieve a military objective. Autonomous systems will therefore need to discriminate between target sets, assess any threat that the target may present and then carry out a suitably measured response. The level of response will depend upon the two previous conditions being met. Regarding weapon

---

81 *Ibid*, ..., un-numbered article.

82 *Ibid*, …, un-numbered article.

performance, autonomous systems should be capable of greater consistency than that achieved by soldiers. For example, accuracy and consistency results from weapon firing within the SWORDS programme demonstrate 100% performance levels, far exceeding those achieved by soldiers[83]. Key will be selection of the most suitable level of response. If equipped with different weapon types, then logical selection (such as demanded by a mobility kill rather than target destruction) of the most suitable weapon, delivered in a more accurate and consistent manner, should lead to effective mission success. If these areas are addressed then the delivery of a necessary level of minimum force should be readily achieved by autonomous systems.

2.4 Responsibility

Jus in Bello requires that when conducting combat there must be responsibility for actions conducted. This is particularly important when killing enemy combatants. Robert Sparrow notes if war is to be governed by morality at all, then 'someone should accept responsibility, or be capable of being held responsible, for the decision to take human life.'[84] Others remark that an essential component of war is the relationship between the soldier and his enemy. How soldiers treat their enemies is the basis of the ethics of war and that such ethics 'requires us to take responsibility for our actions.'[85] In other words, soldiers are responsible for their actions toward other soldiers. Therefore,

---

[83] Thomas Cowan, *A Theoretical, Legal and Ethical Impact of Robots on Warfare*, …, 7.

[84] Robert Sparrow, *Killer Robots,* …, 8.

[85] Christopher Coker, *Waging War without Warriors: the Changing Culture of Military Conflict*, (London: Lynne Rienner Publishers, IISS Studies in International Security, 2002) 181.

not only is there the need to operate within a set of rules in order to show proper respect in the taking of human life, there remains the need to ensure there is the ability to accept responsibility for the outcome of any action. For example if there is failure within an autonomous targeting system and a non combatant target is killed or destroyed. The inability to readily identify responsibility for an event will ensure that prosecution of those responsible will be impossible. This is particularly significant when seeking culpability for war crimes. Consequently, the need to adhere to the principle of responsibility reflects a respect of human life and the ethical taking of it.

Currently the majority of autonomous systems operate on a ‘fire and forget’ principle: the decision to operate the system is made by a human somewhere along the target engagement process and a weapon is typically launched against a known target or a target area. The weapon determines the suitable trajectory in order to ensure successful impact with the target. Target identification is limited, for example, to whether the target (such as an armoured vehicle) is located at a specified point or area. There is little or no discrimination between friendly and enemy and combatant and non-combatant. Reliance on the soldier provides a readily identifiable individual who can be judged responsible for an action. This becomes more important if an innocent civilian is accidentally targeted and killed on the battlefield. The soldier is seen to have failed in selecting the correct target and is held responsible for these actions. Failure could also be considered if excessive collateral damage is created because of the proximity of civilians to a legitimate military target. The meaning of failure should also be clarified: whilst there is likely to be ethical failure as a result of the killing of civilians, the mission could be considered tactically successful if a legitimate military target was destroyed. Either way, explanation for erroneous actions will need to be given. If the actions are supported by the principles of Just War then the soldier is afforded a degree of protection. If the actions fall outside these principles, then culpability and ultimately punishment will follow.

Not withstanding the requirement within just war for responsibility for actions, autonomous systems could assist with understanding how such errors could occur. If designed to record every parameter and condition of an engagement they could provide better evidence with regards to a failure than the subjective observations of a soldier under stress or duress. However, as discussed later, such systems may well be able to be designed to ensure that such failure does not happen in the first place. Let us now consider the potential impact of more advanced AI autonomous systems.

2.5 The potential impact of artificial intelligence

If advances in technology continue, AI systems may not only be able to determine a course of action but also be able to act upon that decision once made. Such systems will therefore be able to make decisions about the nature of the target; not just what it is, but what it is doing. Is it, for example, presenting itself as a threat, going to the assistance of a fallen comrade or offering itself for surrender? They could also judge whether a command was ethical or not. If the latter, conflict could arise between achieving the mission and carrying out an unethical action. These decision-making processes will form part of the system design and most likely follow a set of logic rules. Isaac Asimov's Laws of Robotics, with the emphasis on not harming humans, could provide a base from which to develop such rules. Those that support this approach stress that these rules are 'meant to ensure robots would not harm people in any circumstance.'[86] However these rules would seem to be incompatible with those that govern decisions on killing soldiers (in other words the ability to make life or death decisions). Therefore, if autonomous systems are to be deployed on missions that may result in death, system design needs to include rules and laws to ensure casualties are not 'unethically killed.'[87]

Advanced AI systems could become capable of learning from experience. Whilst this does not necessarily imply that existing operating rules or laws will be adjusted, it could create a degree of randomness in the decision-making process and therefore result in a degree of unpredictability. If this follows, Anderson notes there could be difficulty in holding anyone other than the system responsible for the actions undertaken. However, the command authority deploying such systems should still be considered to be ultimately responsible. In just the same manner as ensuring soldiers are prepared for battle, the command authority should endeavour to test, evaluate and certify autonomous systems by subjecting them to expected operating conditions. Judgement will then be required on the risk of using the systems or not. If the level of risk is considered

---

86 Kenneth Anderson, *Law of War and Just War Theory Blog: the Ethics of Robot Soldiers*, July 2007,

http://kennethandersonlawodwar.blogspot.com/2007/07/ethics-of-robot-soldiers-html Internet accessed 12 December 2007.

87 *Ibid*.

unacceptable then perhaps this draws into question whether such systems should be given advanced AI, capable of independent thought and logic, in the first place. If such systems can choose their own outcomes then perhaps they have to be held responsible for their actions. However, there may be limited value in holding advanced autonomous systems responsible unless they have the ability to feel emotions and therefore 'suffer' from punishment. This perspective of seeking revenge through punishment could be considered unique to humans and, therefore, instead of exerting punishment a simpler approach could be to re-programme the system. Let us consider further the issue of responsibility.

2.6 Responsibility for the actions of autonomous battlefield systems

Responsibility for the actions of autonomous systems can be considered under the headings of programmer/designer, authorising officer (commanding officer, CO) and of the autonomous system itself[88]. These can be expanded to consider a causal chain stretching from the manufacturer, programmer, designer, department of defence to the officers in charge of the operation[89]. Noel Sharkey remarks that 'the military needs clear guidelines to ensure that all those involved in the deployment of lethal robots have a chain of responsibility.'[90] Consequently, where such systems detect, discriminate and respond, there needs to be an acknowledged level of responsibility or culpability if there

---

[88] Robert Sparrow, *Killer Robots*, …, 8.

[89] Noel Sharkey, *Automated Killers and the Computing Profession*, …, 122.

[90] *Ibid*, …, 124.

is system failure. Equally important is the need for minimum levels of system validation and verification standards and procedures. Sharkey further states that ‘within [the] chain of responsibility mishaps can occur, such as justifiable failure in adverse conditions, faulty or partial sensor readings or the ubiquitous friendly fire.’[91] Mishaps can also occur when this causal chain comprises only of soldiers. The issue then becomes, which chain is likely to minimise or avoid such mishaps. We have previously noted areas of weakness for soldiers where autonomous systems could potentially perform better. If technology continues to develop then the balance of which causal chain possesses the least likelihood of making errors could swing towards autonomous systems.

Autonomous system failure could result from a specific design fault, such as failure in a mechanical or electronic logic circuit board process. Responsibility can be apportioned and, as necessary, negligence determined. Conversely a designer may be aware of system limitations and make these clear to the user. If so responsibility could shift from the designer to the command authority deploying the system onto the battlefield. Alternatively the system could be employed in a manner that was beyond that predicted by the designer. Again responsibility shifts to the command authority.

The command authority must understand the implications of any known system limitations and be able assess risk in using them. This is particularly important with ability to self select targets. The need to test and validate in as real situations as possible therefore becomes an essential requirement. The use of AI systems could place commanders at risk of being held responsible for the actions of machines whose

---

[91] *Ibid*, …, 122.

decisions are not directly under the control of these commanders. Commanders may be unable to predict actions of such systems and exert control over the outcome of events. Sparrow considers that it would be unfair for a commander to be accused in such a manner remarking 'if machines chose their own targets how can COs be accused of being responsible?'[92] However, this differs little with soldiers who themselves are able to act with a degree of independence in accordance with the situation, mission requirements and mission command and within rules of engagement. Is it really going to be possible to determine which is less predictable, the soldier or the machine? Perhaps part of the answer lies in the extent of commander's diligence in pre-deployment training and validation. If this is thoroughly conducted areas of risk should be identified enabling commanders to assess whether these systems should be deployed for a mission or not.

When considering whether autonomous weapon systems could be held responsible for the consequences of their actions, the issue of punishment for wrongful actions should be included. We have already noted that advanced AI could possess abilities at least equal and more likely superior to those of soldiers (since they would not, for example, be subject to the fallibility of human fear, anger and fatigue). How could these systems be punished? A simple answer may be destruction, but there may be little value in doing this. Rather, corrective design and testing could take place and if the problem could not be fixed, then the systems would need to be taken out of service or given alternative tasks. Stopping the purchase of more systems (and terminating support contracts) could provide sufficient punishment for a contractor or system designer. There may be little need to be concerned whether an autonomous system is punished at all if

[92] Robert Sparrow, *Killer Robots,* …, 10.

they are incapable of both understanding the implications of actions and of linking punishment with any crime created.

A simple solution addressing responsibility could be to place a soldier within the system, removing the ‘autonomous’ nature. Whilst this could be acceptable initially, this would negate some key system advantages, in particular reducing the risk of soldier casualties by removing them from the battlefield and by relying on soldiers to handle the increasing amount of data and information ever more quickly. For example, time sensitive targeting, which relies on quick decision-making involving a range of information to exploit a target opportunity, could be better performed by autonomous systems. In such scenarios, maintaining a ‘soldier in the loop’ may not therefore be the best solution to apportioning responsibility. This does not necessarily imply that there will no longer be a place for the soldier on the future battlefield. Having acknowledged soldier weaknesses, design philosophy for autonomous systems could be best directed at maximising and exploiting the potential strengths of such systems. This could see a future battlefield with a mix of both soldiers and autonomous systems, complementing each other in the areas that they do best.

It seems that identifying where responsibility lies for the actions performed by autonomous systems presents significant challenges for future weapon designers and to the military chain of command destined to deploy them. As systems become more complex and autonomous they will present greater difficulty in apportioning responsibility to those who design them or order their use on the battlefield. Future commanders could be held responsible for actions conducted by autonomous systems that could be neither predicted nor controlled[93]. This would seem to suggest that no amount of testing, validation and training would be sufficient to render these systems suitable for fully autonomous life and death decisions. Future battlefield situations may therefore prove to be so complex that autonomous systems are unable to cope and actions could therefore become unpredictable. If presented with such a situation the system could be designed to shut down, but only if this action did not place soldiers lives at risk (an aspect of Asimov's robot laws).

Key to just war is the requirement for someone to be held responsible for deaths that occur in the course of war. As a 'necessary condition for fighting a just war … someone can be justly held responsible for deaths that occur in the course of the war, and as this condition cannot be met in relation to deaths caused by autonomous weapon systems it would therefore be unethical to deploy such systems in warfare'[94]. Difficulty exists in allowing machines to be given the ability to take responsibility for killing. Irrespective of the future capabilities of these systems, responsibility for taking life is a human action that cannot be delegated. This approach is supported by the view that 'since technology is both morally blind and irresponsible, machines cannot take over responsibility.'[95] If decision-making is transferred to machines, it could distance soldiers from the consequences of any actions. In other words, handing over responsibility to

---

93 Robert Sparrow, *Killer Robots*, …, 13.

94 *Ibid*, …, 62.

95 Christopher Coker, *Waging War without Warriors: the Changing Culture of Military Conflict*, …, 182.

machines and therefore removing soldiers from decision-making would make the battlefield a less humane environment: there will be relatively fewer soldiers on a future battlefield relying on the actions of autonomous weapon systems. This seems to imply that future systems will never be as capable as the soldier in making ethical decisions. Only future technology will be able to prove or disprove this.

We have seen that advances in AI development could, some time in the future, enable the 'Holy Grail' of ethical decision-making to be reached. Some believe that advanced AI systems will be able to match or improve upon human ethical decision-making since 'ethics is a cognitive pursuit where correct answers can be arrived at by reasoning and balancing of evidence and that such a process could be more accurately answered by advanced AI than by humans.'[96] In other words questions about ethics, which would depend upon reasoning and weighting of evidence, could be more accurately answered by advanced AI than by humans. This is supported by Arkin who believes such rules can be used to create an artificial conscience within an autonomous system, capable of acting more humanely than a soldier[97]. These comments are, of course, merely an unproven hypothesis and do not reflect current reality. Even if autonomous systems are unable to be designed with this level of AI, they could still be deployed on the future battlefield. Whilst the system may not be able to be held responsible for its actions, the authorising officer within the causal chain who made the

[96] Nick Bostrom, *Ethical Issues in Advanced Artificial Intelligence*, (Oxford: Oxford University Philosophy Faculty, UK, 2007) 3.

[97] Ronald Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberate/Reactive Robot Architecture*, …, 14.

decision to deploy the system (with a knowledge and understanding of systems risks and limitations) can be.

2.7 Autonomous weapon systems and the potential effect on the honour of soldiers

The deployment of autonomous weapons could impact upon the sense of honour that many soldiers possess. Here honour is considered as respect, regards and esteem both amongst soldiers and between soldiers and their enemies[98]. A historical perspective is provided by Christopher Coker who remarks that warriors (soldiers) derived value from war, and that it was (and remains) an 'intensely human activity … where warriors not only respected themselves but also were respected by their enemies.'[99] The deployment of autonomous systems could result in the soldier becoming 'disengaged, emotionally and morally, from the battlefield, with no contact and no experience gained.'[100] This could have a negative effect as to how a soldier feels about him or herself: in other words a lowering of self value and esteem. In extremis, soldiers could become transformed into technicians where there is no place for emotion, fear, courage or even endurance. This perspective is important since it appears to directly contradict some of the potential advantages offered by autonomous systems which have already been discussed. Indeed Coker acknowledges that autonomous systems can reduce stress, battle fatigue and the frequency of post traumatic stress disorder (PTSD). Nevertheless,

---

98 Collins Dictionary of the English Language, …, 704.

99 Christopher Coker, *Waging War without Warriors: the Changing Culture of Military Conflict*, …, 160.

100 *Ibid*, …, 161.

removing these risks to soldiers could separate them both emotionally and psychologically from the battlefield. Consequently, there would be less scope for soldiers to 'overcome, benefit and enhance themselves through the experience of battle.'[101] Others note that, irrespective of whether soldiers are completely or only partly removed from the battlefield as a consequence of the deployment of robotic weapon systems, the effect will be the same: 'war ceases being personal, it quits being conducted at the human level; it becomes more of a video game than a deadly act.'[102] Van Crefeld further remarks that where there is no risk to the soldier on the battlefield the activities are conducted without honour[103]. This does seem to imply that a soldiers' perception of importance and honour is related to the level of exposure to risk. It also seems to overlook the fact that throughout history soldiers have sought to implement new technologies in an attempt to not only limit risk to them but also to maximize the threat to others. Nevertheless, this questions whether there is honour in killing without risk and whether the use of autonomous weapons would be a fair means of war fighting. The dilemma is that whilst autonomous weapons have great potential to reduce risk to soldiers, if these weapons kill enemy combatants this could be considered as being an unfair and dishonourable way to conduct war. But if military and political leaders are concerned over casualty rates, the reduction of risk associated with autonomous weapons could dominate any debate on their use.

---

101 *Ibid*, …, 172.

102 Thomas Cowan*, A Theoretical, Legal and Ethical Impact of Robots on Warfare*, …, 6.

103 Martin Van Crefeld, *The Transformation of War*, …, 217.

Future battlefield scenarios could involve autonomous weapons being deployed against forces, such as insurgents, who do not have reciprocal capabilities; a scenario of machine versus man. Whilst issues of risk and honour, already discussed could apply, other issues could also arise. The insurgents could be faced with a significant 'overmatch' in capability as a result of the deployment of these autonomous weapons. It could force insurgents to change tactics, especially if the weapons were perceived impossible to defeat. Insurgent violence could be exported from the area of conflict and directed elsewhere. There could be greater emphasis on achieving a spectacular event such as that achieved in the destruction of the World Trade Centre on 11 September 2001. Arguably, technology overmatch exists today with US military superiority against insurgents in both Iraq and Afghanistan. However, such potential implications brought about by the use of autonomous weapons should be understood by future political leaders and military commanders. This would assist in mitigating unintended consequences of their use. If the risk of creating these consequences is considered by commanders to be greater than the benefits of deploying the systems then they should not be deployed.

2.8 The potential effect of autonomous battlefield systems on the likelihood to wage war

The final section within this chapter considers whether the ability of a nation to deploy fully autonomous weapon systems could impact upon the likelihood of waging war. If these systems reduce casualties then commanders, military or political, may be more compelled to use them. Consequently this could result in an increased likelihood in waging war. In other words, the availability of fully autonomous systems reducing the risk to soldiers could result in a lowering of the 'threshold' of a nation to undertake war. Commenting on US concerns over casualties, Sharkey notes that a reduction in casualties as a direct consequence of deploying autonomous weapons would provide fewer disincentives for starting wars. Without a visible display of body bags coming back to the US, there would be less care about action abroad[104]. However, within 'jus ad bellum' just war principles of going to war, if war becomes more likely through the use of autonomous systems then they could be considered unethical. This perspective is not supported by Arkin who remarks that this ethical dilemma is common with any significant technological advance in weaponry[105]. Nevertheless, if autonomous weapons make the likelihood of conducting war greater then they risk being considered unethical.

A further perspective on the threshold of going to war can also be presented. Widespread development of armies equipped with autonomous systems could generate a similar situation to that of the Cold War. This would see military stand off resulting from the risk of mutually assured destruction between the two belligerent nations. Therefore, instead of lowering the threshold of waging war, it is raised. Either way these examples indicate wide ranging impact that autonomous weapons could indirectly create and an issue military commanders and political leadership will need to acknowledge, understand and consider.

---

[104] Noel Sharkey, *Automated Killers and the Computing Profession* , …, 122.

[105] Ronald Arkin, *Governing Lethal Behaviour; Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*, …, 10.

This chapter has shown that autonomous systems will need to be compliant with a number of ethical principles of Just War to have military utility on the future battlefield. In principle this should be no different in how other military activities are ethically judged. Discrimination and responsibility provide significant challenges for both autonomous system designers and command authorities. Achieving effective discrimination within intensive combat environments must not be underestimated. If the risk of these systems failing to meet these ethical norms is assessed to be too great and outweighs the benefits of deploying the systems in the first place, then they should not be used in an autonomous mode and perhaps they should not be used at all. Placing a soldier in the decision-making loop negates some of the benefits that autonomous systems potentially offer, especially in reducing the risk to soldiers on the future battlefield. However, responsibility is easily identified. Responsibility for the actions of fully autonomous weapon systems is more challenging to apportion. However, much relies on the control of humans, specifically with the system designers. These systems should be what we design them to be, even with fully autonomous AI systems that can make and act upon their own decisions. Such decision-making must be conducted within rules and laws which are discussed in the next chapter. Responsibility therefore will rest with commanders who authorise the systems to be deployed. Thorough system validation and assessment will assist these commanders in judging risk of deploying autonomous systems. This will also assist in identifying unforeseen consequences. If such systems are considered unethical in either their construct (therefore actions) or by consequences (such as lowering the threshold to wage war), then they will be judged as having failed in reaching ethical norms. If so, their use on the future battlefield should be seriously questioned.

Chapter 3 – The Legal Debate

This chapter considers some of the fundamental legal challenges that should be both considered and addressed before advanced battlefield autonomous systems are deployed onto the battlefield. If such systems perform illegally, sanctions such as 'unilateral or multilateral retaliatory force and punishment, criminal prosecution, and the force of world opinion'[106] could follow. This chapter provides a brief historical background behind the legal environment within which military action is conducted today. It highlights both laws of war (LOW) and rules of engagement (RoE) which together form the legal framework for military conflict. It assesses key legal areas likely to directly affect autonomous systems, in particular the need to minimise collateral damage. How and why soldiers fail to act within laws of war is discussed. This leads to an assessment of whether autonomous systems have the potential to improve upon these failings. Consequently, this chapter considers whether autonomous systems can potentially operate more legally than a soldier. The Ottawa Treaty on the use of anti-personnel mines is used as an example where legal (and ethical) shortcomings resulted in a comprehensive (albeit not unanimous) world wide ban. Failure to address similar concerns with autonomous weapon systems could result in a similar ban.

---

106 Anthony Hartle, *Moral Issues in Military Decision Making*, (Kansas: University Press of Kansas, 2004), 106 – 107.

3.1 Today's legal military environment

The codification of the LOW or the Laws of Armed conflict (LOAC) has been developed over centuries of military action. Such LOAC, embedded within protocols such as the Geneva Convention and RoE, prescribe 'what is and what is not acceptable in the battlefield in both a global (standing RoE) and local (Supplementary RoE) context'[107]. The US military defines LOAC as 'that part of international law that regulates the conduct of armed hostilities' and RoE as 'directives issued by competent military authority that delineate the circumstances and limitations under which … forces will initiate and/or continue combat engagement with other forces encountered.'[108] The UK joint guide to RoE remarks that '[such RoE] are used … for all operational tasks … in peacetime, in periods of tension or limited conflict and war.'[109] Such rules provide guidance to military commanders at all levels on the political policy to be pursued and the 'parameters within which armed forces may operate'[110]. Arkin remarks that the development of LOAC is an evolutionary process, affected by the development of technology. Consequently he considers that lethal autonomous systems must not only

---

107 Ronald Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture, …*, 2.

108 US Department of Defence, *Joint Publication 1-02, Dictionary of Military and Associated Terms*, (DOD: 2007), 2.

109 Royal Naval Staff College, *The Tri-Staff College Guide to Rules of Engagement*, (Greenwich: UK MoD, DR 3A-2, 1996), 1.

110 *Ibid*, …, 1.

comply with current rules but that 'they must be able to adapt to the new policies and regulations as they are formulated by international law'[111].

The laws of war existing today are based upon instructions drawn up by the American Francis Lieber in 1863 resulting in the Lieber Code. Such a code was a 'manual for combatants concerning the conduct of war that attempted to codify customary law and standards applying to all nations.'[112] It formed the basis for future Hague Conventions which are mainly concerned with permissible means of fighting. These rules were augmented by those presented by Henry Dunant following his experiences on the battlefield of Solferino[113]. His efforts to mitigate the suffering of combatants should war arise eventually led to the first Geneva Convention: such conventions concern the treatment of the victims of war[114]. In addition to subsequent Geneva and Hague Conventions, Protocols and Regulations, judgments of national and international courts (such as the Permanent Court of International Justice at The Hague) and international tribunals (such as the Nuremburg Tribunal) have established principles and precedents that contribute to the laws of war. For these rules and regulations to be of value, states need to be signatories to them and consequently abide by them. The legal issues surrounding the use of autonomous weapons fall into two overarching groups: signatory nations legally employing such weapons, and the impact of non signatory states

111 Ronald Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture,* …, 3.

112 Anthony Hartle, *Moral Issues in Military Decision Making,* …, 105.

113 The Solferino battle was conducted during 1859 between Austria and France.

114 Anthony Hartle, *Moral Issues in Military Decision Making,* …, 107.

and non states (including insurgent organisations) using such weapons. These are considered next.

3.2 Potential legal areas of risk with autonomous weapons systems

The need for discrimination was considered in the previous chapter and presents an extremely challenging technical problem to resolve. Whilst we have seen that some believe this is in fact too challenging, let us assume that a technical solution will be produced enabling accurate target detection, location, identification and classification[115]. Autonomous systems will nevertheless have a range of battlefield limitations which will need to be addressed in order for them to be deployed within a suitably robust legal environment. Arkin provides a summary of limitations regarding the potential use of lethal autonomous systems. These include the acceptance of surrendering combatants and the humane treatment of prisoners of war (POW), the use of proportional force in a conflict, the protection of both combatants and non-combatants from unnecessary suffering and the avoidance of collateral damage (to both people and property)[116]. In addition autonomous systems must prohibit attacks on personnel operating under Red Cross/Crescent flags, white flags or neutral states, ensure the avoidance of torture and the mutilation of corpses. In order to address these areas, autonomous weapon systems must be able to: identify legitimate legal targets; assess whether the target is a threat (and not,

---

[115] Following target definition principles given in US DoD, *Joint Publication 3-60, Joint Targeting* (US DoD, April 2007), I – 1.

[116] Ronald Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture,* …, 3.

for example, surrendering); consider options and then select a suitable course of action (such as the use of lethal or non lethal means); consider risks and consequences of this course of action (such as the need to provide a warning of imminent action or the risk of collateral damage); and on completion of the mission assess effectiveness and the need for follow-on action.

If autonomous systems become widely available on the commercial defence market then signatory states could be faced with an opponent using similar autonomous systems but operating them with unrestricted rules of engagement; in effect causing signatory states to 'fight with one hand tied behind the back'. This situation of operating against an opponent who does not demonstrate a similar legal (and ethical) 'modus operandi' is not new, such as the actions of Taliban insurgents in Afghanistan. Nevertheless, the threat of lethal autonomous weapons being deployed against signatory countries by organsiations not adhering to the same legal rules could be difficult to resolve. One approach could be to ensure that within the design stage a legal architecture is developed to ensure non legal actions are not permissible. This is not just a challenging technical issue for future system designers to resolve, but it would inevitably be vulnerable to tampering in order to remove or reduce the effect of any inbuilt system legal constraints. How much trust can be placed on designers to produce a future failsafe system which will ensure such weapons comply with extant law? Such failsafe systems may be difficult to achieve especially acknowledging the frequent success of computer hackers breaking into a wide range of commercial and military security systems.

3.3 Why do soldiers legally fail on the battlefield?

Advantages that autonomous systems potentially offer have already been discussed. In particular, these systems could potentially operate in a more ethical and legal manner than soldiers. Before considering this let us first identify some of the reasons why soldiers may act unlawfully on the battlefield. Simply put 'to err is human' but behind this statement is a wide range of contributing factors. For example, soldiers may not fully understand the law and the implications of their actions. This could be through ignorance, forgetfulness, lack of understanding, by being poorly briefed or even by not being briefed at all. Unlawful targeting of non combatants and civilians could occur through poor target identification or through deliberate choice. Whilst the former situation could involve soldiers receiving some legal support if they could justify their actions using extant RoE, the latter would not. Combat stress, fatigue and hunger can all have a detrimental effect on soldier's performance. Furthermore, the loss of comrades in combat could result in acts of revenge. This could be compounded by poor leadership, poor training and inexperienced soldiers[117]. All these factors can contribute to soldiers acting unlawfully.

3.4 Can autonomous weapons systems operate more legally than the soldier?

It may be possible that autonomous weapon systems could be designed to perform more consistently within the LOW than soldiers. These systems could be designed to be

---

[117] Ronald Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture,* …, 8.

less susceptible to many of the soldier areas of failing discussed here. They will not become tired or emotional or hungry (acknowledging the requirement for replenishment and maintenance). Autonomous systems could be employed in a sacrificial mode, avoiding placing soldiers within life threatening and dangerous situations. This could become an issue with advanced AI systems thinking for themselves (or having the ability to learn from experience). Nevertheless, if suitable protocols and logic processes are included within the design stage then such difficulties could be addressed. This concept could lead to the development of 'mathematical methods enabling the development of autonomous system architectures capable of supporting ethical behaviour regarding the application of lethality in war'[118]. The difficultly will be to put theory into reality: described in the previous chapter, whether this can be achieved is a considerable challenge and represents a serious risk if it cannot be.

3.5 The risk of collateral damage: the use of close defence autonomous gun systems

The risk of causing collateral damage will be a key consideration to deploy future autonomous weapon systems. These difficulties exist today. The example considered here concerns the deployment of close in gun systems designed to neutralise incoming artillery and mortar rounds. These autonomous weapon systems, such as the Phalanx and 'Goalkeeper' Gatling gun'[119], are common on naval platforms and provide a final layer

---

[118] Ronald Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture,* …, 14.

[119] The Gatling Gun design incorporates a number of barrels rotated in sequence allowing for a high cyclic rate of fire without excessive barrel wear and barrel temperature increase.

of defence against incoming antiship missiles and low flying aircraft. They rely on a high cyclic rate of fire to deliver a 'wall of steel' which destroys the incoming threat. Capable of operating within less than 1 km of the ship, the response required from target identification, threat assessment, ballistic computation for impact and when to fire demands an autonomous system capable of operating faster than a human. Consequently, in a high threat environment, such systems are likely to be switched to autonomous mode. Although the ballistic range of the munitions exceeds several km, the relatively benign environment of the ocean can allow the operation of such systems with minimal risk of collateral damage.

Commanders involved in counterinsurgency (COIN) operations in Iraq and Afghanistan, have demanded a means to counter insurgent artillery and mortar attacks on static security force (SF) bases. A number of countries (notably US, UK and Canada) assessed that a Gatling gun system[120], adapted from in service naval versions outlined above, could provide an effective solution to counter the threat. Whilst the US and UK deployed systems into theatre, Canada assessed the legal and ethical difficulties surrounding potential collateral damage as too significant; the high rate of fire and the down range risk of munitions landing at the end of their trajectory were of considerable concern. A cyclic rate of 3000 rounds per minute, firing 20mm ammunition, results in 50 rounds being fired every second[121]. Whilst timed fuses detonate before ground impact,

---

[120] The gun is the General Dynamics/Raytheon Phalanx Mark 15 close in weapon system, and comprises of 6 barrels firing 20mm ammunition at a cyclic rate of 3000 rounds per minute to an effective range of 1500 meters,

www.armedforces.co.uk/navy/listings/0014.html Internet accessed 12 March 2008.

[121] A standard engagement can last up to 5 seconds.

there is a considerable 'beaten zone' where unexploded rounds or their fragments fall. Consequently, the risk of collateral damage can be considerable. This can, to a limited degree, be countered by the imposition of physical hard stops both in azimuth and elevation. Constraints in azimuth control the direction the gun is facing while constraints in elevation limit both the minimum and maximum effective range of the system[122]. However, these limitations reduce the effectiveness of the gun system to counter a threat which could be delivered from any direction. Furthermore, insurgents deliberately site firing platforms amongst the civilian population further increasing the potential risk for collateral damage. Given that the effect of an engagement in the defence of a Coalition installation must not be detrimental to the support of the people which the Coalition was attempting to protect, the Canadian military chose not to deploy the system. Acknowledging similar concerns, the US and UK deployed comparable systems into theatre. However, they also identified a lack of trust of soldiers in the gun systems when they were switched into autonomous mode. Commanders were not comfortable with the concept of having a fully autonomous system operating. As such, imposed constraints limited the full potential of these systems. Ultimately commanders will need to make a balanced assessment between the benefits that the system delivers against the risk of collateral damage and legal consequences which could result.

---

[122] Information provided by Major JJ Shamehorn Royal Canadian Artillery and Battery Commander of 128 Field Battery.

3.6 What lessons can be learnt from the Ottawa Treaty banning anti-personnel mines?

The Ottawa Treaty on the ban of the military use of anti personnel mines provides a useful example of potential legal and ethical difficulties faced with the development of autonomous weapon systems. Whilst it is accepted that comparing anti personnel mines to lethal autonomous weapon systems may be seen as tenuous, these mines 'identify' a target (through external stimulus) and subsequently initiate (ideally) a non lethal[123] response in an autonomous manner. Furthermore, some consider the law banning antipersonnel mines is 'immensely relevant' to the theorizing on the use of robotic technologies in future war[124].

The ban in using anti personnel mines was instigated by the late Diana, Princess of Wales in early 1997. Her focus was primarily on the dehabilitating injuries caused by these mines and injuries caused to children who picked them up. Visiting Angola and Bosnia she used the media to highlight the non discriminatory manner of the mines and therefore their illegal and unethical nature. The International Campaign to Ban Landmines resulted in countries ratifying and signing what became the Ottawa Treaty which banned the use, production, trade and stockpiling of antipersonnel mines. Within the preamble of the treaty signatories to the ban are;

---

[123] A non lethal effect commits additional soldiers to evacuate the injured soldier, requires medical facilities for treatment and can contribute to a reduction in morale and fighting spirit.

[124] Peter Asaro, '*Robots and Responsibility from a Legal Perspective*, (Umea University, Research Paper, January 2007) 2.

> "determined to put an end to the suffering and casualties caused by anti personnel mines, that kill or maim hundreds of people … mostly innocent and defenceless civilians and especially children … and have other severe consequences for years after emplacement"[125].

As of the end of January 2008, the number of States party to the treaty 'had reached 156 countries, including mine affected countries and as well as former users and producers of the weapon.'[126] Whilst it is not the purpose of this paper to review the treaty in detail, it is important to note that even when presented with what seems to be a compelling argument, not all countries have signed up to adhere to the ban. Notable exceptions include the US which considers that 'military capabilities provided by landmines remain necessary for the US military to protect … forces and save lives.'[127] However, the US differentiates between the persistent and non persistent nature of mines and is 'committed to eliminate persistent landmines *of all types* [author's emphasis] from its arsenal.'[128] China, Russia and India have also not signed up to the ban, demonstrating

---

125 International Campaign to Ban Landmines, *"Convention on the prohibition of the use, stockpiling, production and transfer of anti-personnel mines and on the their destruction",* 1,

http://www.icbl.org/layout/set/print/content/view/full/6942. Internet accessed 28 January 2008.

126 International Campaign to Ban Landmines, "*The Mine Ban Treaty: A success in progress",*

http://www.icbl.org/layout/set/print/content/view/full/22687. Internet accessed 28 January 2008.

127 US Department of State, "*New United States policy on landmines: reducing humanitarian risk and saving lives of United States Soldiers*", Washington DC, Bureau of Political-Military Affairs, February 2004, 1,

http://www.state.gov/t/pm/rls/fs/30044.htm Internet accessed 29 January 2008.

128 *Ibid*, …, 1.

that countries can possess differing interpretations and approaches towards the same situation.

3.7 Ensuring autonomous systems act legally

Just as we have seen the need for autonomous systems to comply with accepted ethical norms so to must they act legally when conducting missions and tasks. Conceptually these systems will need to be capable of following logical processes which refer to laws and rules. These processes will then create singular or multiple courses of action (CoA) upon which a decision on the most suitable will need to be made. Arkin notes the difficulty in creating such systems. In particular he highlights: that laws and rules are almost always provided in a conceptual, abstract level; that they may be subject to interpretation; that they may have different meanings in different contexts and that rules could conflict with each other[129]. In connection with the use of lethal force, he suggests that the key overarching design requirement for an autonomous system is that it 'does not engage a target until obligated to do so, consistent with the situation and there exists no conflict with law of war and rules of engagement'.

Following from these system design challenges, a military version of Alan Turing's 'Turing Test' has been suggested as a tool to judge whether autonomous

---

[129] Ronald Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture,* …, 39.

systems could be legally deployed[130]. This theoretical test assesses the capability of a machine to demonstrate intelligence[131]. In doing so, a machine would need to communicate, reason, possess knowledge (such as a database) along with the ability to learn. Turing's test (originally considered a game) involves human judgement, achieved through conversation, differentiating between man and machine. Turing proposed that 'a computer that can regularly … fool a discerning judge in this game would be intelligent, a computer that thinks, beyond any reasonable doubt[132]'. Failing the test may not necessarily prove anything since advanced AI systems could, like humans, decline to prove themselves or choose to fail. Passing the test could, however, be used to confirm intelligence[133]. In passing a military Turing test an autonomous system should be no worse than a soldier at taking decisions, specifically as to what constitutes a legitimate target. Taking into account continuing growth of computing power, some consider that Turing capable computers could be manufactured around 2020[134]. Whilst we have seen that expectations of AI computers have not met reality, the Turing test does possess potential utility in being used to judge the legal deployability of autonomous systems.

---

130 New Scientist Technology Blog site, *Military Turing test would make war robots legal*, 2,

http://www.newsientist.com/blog/technology/2008/02/military-turing-test-would-make-war-robots-legal Internet accessed 10 March 2008.

131 Alan Turing, *Computing Machinery and Intelligence*, Mind, Volume LIX, Number 236, October 1950, 3, reproduced on the following internet site,

http://www.abelard.org/turpap/turpap/html Internet accessed 10 March 2008.

132 Daniel Dennett, *Can Machines Think?* within *Part One: The Roots of Artificial Intelligence* by Raymond Kurzweil, (Massachusetts Institute of Technology: Dai Nippon Printing Company, 1990) 48.

133 *Ibid*, …, 49.

134 Raymond Kurzweil, *The Age of Intelligent Machines*, …, 10.

System evaluation and validation may also identify areas of risk and guide commanders to assess whether an autonomous system is suitable for a specific task or mission. It may therefore be possible to take precautions during mission planning in order for such systems to be used within the law. This could result in limiting the type of mission or situation that autonomous systems could be exposed to in order to reduce or remove the risk of breaking any law[135]. Greatest risk will lie with lethal systems. Asaro considers that this is no different than any other system engineering activity in that it is necessary to take proper care in imagining, assessing and mitigating the potential risks of all [new] technology[136]. As technology matures, the risks associated with deploying autonomous systems will change. Hence the need for a continual development process in order to adapt to advances in technology. Whatever the outcome, the utility of autonomous systems on the future battlefield will be judged against ethical norms and legal compliance. Failure in either of these areas will result in failure in the system itself.

---

[135] Tom Simonite, Robot arms race underway expert warns, …, 2.

[136] Peter Asaro, '*Robots and Responsibility from a Legal Perspective*, …, 4.

Conclusion

As scientific research and technological developments continue to advance it seems possible, if not inevitable, that fully autonomous weapon systems on the future battlefield will become a reality. Such systems will be complex, capable of sensing the environment, integrating data and information from a plethora a sensors, formulating options, deciding upon a specific course of action and then acting upon that decision. Furthermore, through the development of AI, they will be capable of learning from experience and making judgements and decisions based upon embedded principles, rules and laws. Whatever configuration that these future systems eventually comprise, they must satisfy a number of ethical and legal norms. Only if judged capable of meeting these norms should the deployment of these systems follow.

Autonomous systems, possessing lethal and non lethal capabilities could be employed in reconnaissance, surveillance, target engagement, tactical and post conflict demining, explosive ordnance disposal, decontamination, logistic support, perimeter security and access control. They will not only have the potential to enhance the activities of soldiers by improving upon their performance, but also augment activities by enabling missions to be completed by soldiers who could not otherwise accomplish them on their own.

Autonomous systems will not be affected by a number of shortcomings which influence the performance of soldiers on the battlefield. Whilst acknowledging the requirement for replenishment and maintenance, they will not be prone to certain psychological and emotional conditions such as battlefield stress, fear and duress. They will be less affected by environmental conditions and be capable of far greater endurance and strength. Systems designed with lethal weapons potentially offer far greater consistency of weapon performance. Autonomous systems could be employed in a sacrificial mode, avoiding placing soldiers within life threatening and dangerous situations. By reducing the exposure of soldiers to risk, they have the potential to reduce casualty rates. With western societal sensitivity to casualty rates, this is arguably the overriding reason for developing autonomous battlefield systems.

Ethical issues can be considered within the framework of 'Jus in Bello' and the conduct of just war. Autonomous systems should be assessed against ethical principles of discrimination, necessity, responsibility and proportionality. The technical challenges in meeting these principles must not be underestimated. In particular, discriminating between what is and what is not a legitimate target is likely to be one of the greatest challenges for system designers. Nevertheless, discrimination will be critical in order for future systems to perform both ethically and within current and future law. Moreover there is the need to have a defined chain of responsibility to ensure that any actions can, if necessary, be brought to account. Human commanders will inevitably feature predominantly within this causal chain. Even with systems that can make their own decisions, responsibility should lie with leaders or commanders (either military or political) authorising their deployment. How these advanced systems operate is subject to human design. It is, therefore, our responsibility to ensure that they are developed within a decision-making framework which is ethically and legally robust. Future commanders who will have autonomous system capabilities at their disposal will also need to ensure that meticulous system evaluation and validation is conducted using likely scenarios in order to highlight areas of risk and concern. Human judgement will then be required to balance potential risk against benefit offered.

The deployment of autonomous systems will need to comply with international and humanitarian law, uniform codes of military and the Geneva and Hague conventions. Failure to ensure such legal compliance could result in retaliatory force or sanction, criminal prosecution and the creation of negative world opinion. The Ottawa Treaty banning the use of antipersonnel mines demonstrates the ability of world opinion to be galvanised against a specific weapon system if that system is considered to be both illegal and unethical. Consequently autonomous system designers must consider these issues at the earliest stages of system development so that progress can be shaped in a controllable rather than reactionary manner.

Thorough system validation will need to identify and highlight the likelihood of unforeseen circumstances that may arise with the deployment of autonomous systems. An assessment of potential courses of action that an adversary may take when faced with autonomous systems will be necessary. Whilst acknowledging that identifying future intentions of an adversary faced with these systems could be challenging, consideration must at least be given to the potential for unintended consequences. Furthermore, autonomous systems could influence national security policies by potentially lowering the threshold above which a decision is made to go to war. If there is less risk of causalities in conducting war, then the obstacles placed upon a nation to go to war could be reduced. Alternatively, we have seen that the potential widespread deployment of autonomous systems on a future battlefield between two similar nations could create a Cold War mutually assured destruction stand-off situation and therefore raise the threshold to wage war. Either way, political leadership and military commanders need to fully understand potential consequences of developing such systems.

Ultimately the deployment of autonomous weapon systems has the potential to reduce the risk to the soldier on the future battlefield and is a compelling reason to continue to develop these systems. Key within this development process is a thorough understanding of the implications that these future systems will present to commanders, both military and political. Such understanding must be developed through open debate, review and assessment. It is more than likely that sometime in the future fully autonomous weapons will be sufficiently technologically advanced so as to be ready for deployment onto the battlefield. Some could replace the soldier whilst others could augment what soldiers can do. The range of potential tasks is considerable. How successful and widespread these systems become depends upon the human decisions and designs of today and tomorrow. Decision makers will need to ensure that whatever systems are created, they are restricted to operating ethically and legally. Failure to do so could be disastrous. A decision to deploy fully autonomous lethal systems on the future battlefield before these issues are addressed completely and effectively would be pure folly.

BIBLIOGRAPHY

Adams, Barbara and Lora Bruyn. "*Trust in Automated Systems Literature Review"* (Toronto: Department of National Defence, Defence Research and Development Canada, CR-2003-096, 2003);

http://pubs.drdc.ca/PDFS/unc59/p520342.pdf Internet accessed 31 March 2008.

Anderson, Janna and Rainie Lee. *"The Future of the Internet II"* (Pew/Internet and American Life Project, September 2006);

http://www.pewinternet.org Internet accessed 20 November 2007.

Anderson, Kenneth. "*Law of War and Just War Theory Blog: the Ethics of Robot Soldiers*" (Blogsite July 2007);

http://kennethandersonlawodwar.blogspot.com/2007/07/ethics-of-robot-soldiers-html Internet accessed 12 December 2007.

Arkin, Ronald. "*Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture"* (Georgia Institute of Technology: Technical Report GIT-GVU-07-11, 2007).

Asaro, Peter. *"How Just could a Robot war be?* Transcript from the 5th European Computing and Philosophy Conference, 2007).

Asaro, Peter. "*Robots and Responsibility from a Legal Perspective*" (Umea University: Research Paper, January 2007).

Asimov Isaac. "*I Robot"* (New York: Doubleday and Company, 1950).

BBC News, "*PC beats doctor in scan tests"* (bbc.co.uk February 2008);

http://news.bbc.co.uk/2/hi/science/nature/7257730.stm Internet accessed 10 March 2008.

Boot, Max. "*War Made New: Technology, Warfare, and the Course of history 1500 to Today*" (New York: Gotham Books, 2006).

Bostrom, Nick. "*Ethical Issues in Advanced Artificial Intelligence*" (Oxford: Oxford University Philosophy Faculty, UK, 2007).

Chin, Warren. "*Technology, Industry and War, 1945 to 1991 Chapter 2 within War in the Age of Technology: myriad faces of modern armed conflict"* edited by Geoffrey Jensen and Andrew Wiest (New York University Press: New York, 2001).

Coker, Christopher. "*Waging War without Warriors: the Changing Culture of Military Conflict*" (London: Lynne Rienner Publishers, IISS Studies in International Security, 2002).

Collins Dictionary of the English Language edited by Patrick Hanks (London, Collins, 1982).

Cowan, Thomas. *"A Theoretical, Legal and Ethical Impact of Robots on Warfare"* (Carlisle Barracks, Pennsylvania, US Army War College Research Project, Mar 2007).

Dennett, Daniel. "*Can Machines Think?"* within *Part One: The Roots of Artificial Intelligence* by Raymond Kurzweil, (Massachusetts Institute of Technology: Dai Nippon Printing Company, 1999).

Directorate of Land Concepts and Design, Canadian Department of Defence. "*Land Operations 2021: Adaptive Dispersed Operations. The Force Employment Concept for Canada's Army of Tomorrow*" (Kingston: Ontario, 2007).

Ellis, John. "*The Social History of the Machine Gun*" (London: Pimlico, 1976).

Erwin, Sandra. "*Defense department should refocus technology spending, experts warn*" (National Defence: Arlington, November 2007).

Falconer, Bruce. "*DARPA's New SuperSoldiers"* (Washington, National Journal, Volume 35, Issue 45, November 2003).

Gaffney, Steven. "*Battlefield Automation: A Luddite's View*" (US Marine Corps, 1991);

www.globalsecurity.org/military/library/report/1991/gsj.html. Internet accessed 12 December 2007.

Gentry, John. "*Doomed to Fail: America's Blind faith in Military Technology*" (Parameters: Volume 32, Winter 2002/2003).

Hartle, Anthony. "*Moral Issues in Military Decision Making*" (Kansas: University Press of Kansas, 2004).

Havenstein, Heather. "*Spring comes to AI Winter*" (Computerworld, February 2005);

http://www.computerworld.com/printthis/2005/0,4814,99691,00.html Internet accessed on 28 February 2008.

Headquarters Department of the US Army. "*Field Manual 1, The Army Our Army at War Relevant and Ready Today and Tomorrow*" (Washington, DC, 14 June 2005).

International Campaign to Ban Landmines. *"Convention on the prohibition of the use, stockpiling, production and transfer of anti-personnel mines and on the their destruction"* ICBL Webpage;

http://www.icbl.org/layout/set/print/content/view/full/6942. Internet Accessed 28 January 2008.

International Campaign to Ban Landmines. "*The Mine Ban Treaty: A success in progress"* ICBL Webpage;

http://www.icbl.org/layout/set/print/content/view/full/22687. Internet accessed 28 January 2008.

Kemp, Ian. "*The American Future Combat System*" (Zurich, Armada International, Issue 2, 2006).

Kurzweil, Raymond*. "The Age of Intelligent Machines"* (Massachusetts Institute of Technology: Dai Nippon Printing Company, 1990).

Lerner, Preston. "*Robots go to War"* (Popular Science, 2005);

http://www.popsci.com/popsci/printerfriendly/technology/generaltechnology/f59ca60b2e948010vgnvcm1000004eecbccdrcrd.html Internet accessed 12 November 2007.

Luttwak, Edward. "*Toward Post-Heroic Warfare"* (Foreign Affairs, Volume 74, Number 3, May/June 1995).

Magnuson, Stew. "*RoBo Soldiers"* (National Defence: Arlington, Volume 92, Issue 646, September 2007).

Mahnken, Thomas G. "*Democracies and Small Wars"* edited by Efraim Inbar (London, Frank Carr, 2003).

Pinker, Steven. "How the Mind Works" (New York: Norton, 1999).

Record, Jeffery. "*Collapsed Countries, Casualty Dread, and the New American Way of War*" (Parameters: US Army War College Quarterly, Summer 2002).

Reshke, Regan, "*Brave New Conflicts: Emerging Global Technologies and Trends"* (Canadian Forces Department of National Defence Research Paper, 2007);

http://lfdts.army.mil.ca/DLCD-DCSFT/pubs/jadex/paper1/JADEX-Paper-1-en.pdf Internet accessed 9 January 2008.

Royal United Services Institute (RUSI), *"The Ethics of Autonomous Systems: programme of events'* (London: February 2008);

http://www.rusi.org/events/ref:E47385996DA7D3/ Internet accessed 10 March 2008.

Shactman, Noel. *"Marines give us Exoskeletons, Self Aware Robots*, Dangeroom Wired Blog Network, November 2007;

http://blog.wired.com/defense/2007/11/marines-heart-e.html#more Internet accessed 19 March 2008.

Sharkey, Noel. "*Automated Killers and the Computing Profession"* (New York: Computer, Volume 40, Issue 11, November 2007).

Simonite, Tom. "*Robot arms race underway, expert warns"* (NewScientistTech, February 2008);

http://technology.newscientist.com/article/dn 13382 Internet accessed 10 March 2008.

Sparrow, Robert. "*Killer Robots"* (Journal of Applied Philosophy, Volume 24, Number 1, 2007).

The Economist. "*Robot Wars"* (Economist.com, 7 June 2007);

http://www.economist.com/science/tq/displaystory.cfm?story_id=9249201 Internet accessed 21 December 2007.

The New Scientist Technology Blog site. "*Military Turing Test would make war robots legal"*, 27 February 2008.

http://www.newsientist.com/blog/technology/2008/02/military-turing-test-would-make-war-robots-legal Internet accessed 10 March 2008.

The Royal Naval Staff College. "T*he Tri-Staff College Guide to Rules of Engagement*" (Greenwich: United Kingdom Ministry of Defence, DR 3A-2, 1996).

Tiron, Roxana. *"Lack of Autonomy Hampering Progress of Battlefield Robots"* (National Defense: Arlington Volume 87, Issue 594, May 2003).

Turing Alan, "*Computing Machinery and Intelligence"* (Mind, Volume LIX, Number 236, October 1950), reproduced on the following internet site;

http://www.abelard.org/turpap/turpap/html Internet accessed 10 March 2008.

US Department of Defense. "*Joint Publication 1-02, Dictionary of Military and Associated Terms"* (US Department of Defense, 2007).

US Department of Defense. "*Joint Publication 3-60, Joint Targeting"* (US Department of Defense, April 2007).

US Department of State, "*New United States policy on landmines: reducing humanitarian risk and saving lives of United States Soldiers*" (Washington DC, Bureau of Political-Military Affairs, February 2004);

http://www.state.gov/t/pm/rls/fs/30044.htm Internet accessed 29 January 2008.

Van Crefeld, Martin. "*The Transformation of War*" (New York: The Free Press, 1991).

Yudkowsky, Eliezer. *"Artificial Intelligence as a Positive and Negative Factor in Global Risk"* (California: Global Catastrophic Risks, August 2006);

http://www.singinst.org/upload/artificial-intelligence-risk.pdf Internet accessed 14 April 2008.