EXPERT MEETING

AUTONOMOUS WEAPON SYSTEMS

TECHNICAL, MILITARY, LEGAL AND HUMANITARIAN ASPECTS

GENEVA, SWITZERLAND
26 TO 28 MARCH 2014

ICRC

REPORT

International Committee of the Red Cross
19, avenue de la Paix
1202 Geneva, Switzerland
T +41 22 734 60 01 F +41 22 733 20 57
E-mail: shop@icrc.org www.icrc.org

AUTONOMOUS WEAPON SYSTEMS:

TECHNICAL, MILITARY, LEGAL AND HUMANITARIAN ASPECTS

EXPERT MEETING

GENEVA, SWITZERLAND
26 to 28 MARCH 2014

CONTENTS

INTRODUCTION AND STRUCTURE OF THE REPORT

The International Committee of the Red Cross (ICRC) first raised concerns about autonomous weapon systems in its 2011 report, *International Humanitarian Law and the challenges of contemporary armed conflicts*, calling on States to carefully consider the fundamental legal, ethical and societal issues raised by these weapons before developing and deploying them.[1]

From 26 to 28 March 2014, the ICRC convened an international expert meeting entitled *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*. It brought together government experts from 21 States and 13 individual experts with a wide range of legal, technical, operational, and ethical expertise. The aim was to gain a better understanding of the issues raised by autonomous weapon systems and to share perspectives.

In May 2014, at a Meeting of Experts convened by the High Contracting Parties to the Convention on Certain Conventional Weapons (CCW), the ICRC presented a summary report of its expert meeting as a resource for States.[2] The summary report is included here together with additional material prepared by the ICRC and independent experts.

There is no internationally agreed definition of an autonomous weapon system. However, for the purposes of the ICRC's meeting, 'autonomous weapon systems' were defined as weapons that can independently select and attack targets. These are weapon systems with autonomy in the 'critical functions' of acquiring, tracking, selecting and attacking targets.

This report is divided into three sections:

Section 1 is a summary report of the expert meeting, which was prepared by the ICRC under its sole responsibility and previously published in May 2014.

Section 2 comprises summaries of selected presentations given by independent experts at the meeting, and provided under their own responsibility.

Section 3 is an edited version of the background paper prepared by the ICRC and circulated to participants in advance of the expert meeting in March 2014. It does not necessarily represent institutional positions of the ICRC.

The expert meeting agenda and the list of participants are provided in **Annexes 1** and **2**.

[1] ICRC (2011) *International Humanitarian Law and the challenges of contemporary armed conflicts*. Report for the 31st International Conference of the Red Cross and Red Crescent, Geneva, 28 November to 1 December 2011, pp. 39-40: https://www.icrc.org/eng/assets/files/red-cross-crescent-movement/31st-international-conference/31-int-conference-ihl-challenges-report-11-5-1-2-en.pdf

[2] *Summary report of the ICRC meeting on autonomous weapon systems*, 26-28 March 2014, first published 9 May 2014: https://www.icrc.org/eng/assets/files/2014/expert-meeting-autonomous-weapons-icrc-report-2014-05-09.pdf

PART I: SUMMARY REPORT BY THE INTERNATIONAL COMMITTEE OF THE RED CROSS[1]

Expert meeting on *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*, 26-28 March 2014, Geneva, Switzerland.

MEETING HIGHLIGHTS

The aim of the ICRC's expert meeting was to gain a better understanding of the issues raised by autonomous weapon systems and to share perspectives among government representatives, independent experts and the ICRC. The meeting brought together representatives from 21 States and 13 independent experts. Some of the key points made by speakers and participants at the meeting are provided below although they do not necessarily reflect a convergence of views.

There is no internationally agreed definition of autonomous weapon systems. For the purposes of the meeting, 'autonomous weapon systems' were defined as weapons that can independently select and attack targets, i.e. with autonomy in the 'critical functions' of acquiring, tracking, selecting and attacking targets.

There has been rapid progress in civilian robotics in the past decade, but existing autonomous robotic systems have some key limitations: they are not capable of complex decision-making and reasoning performed by humans; they have little capacity to perceive their environment or to adapt to unexpected changes; and they are therefore incapable of operating outside simple environments. Increased autonomy in robotic systems will be accompanied by greater unpredictability in the way they will operate.

Military interest in increasing autonomy of weapon systems is driven by the potential for greater military capability while reducing risks to the armed forces of the user, as well as reduced operating costs, personnel requirements, and reliance on communications links. However, current limitations in civilian autonomous systems apply equally to military applications including weapon systems.

Weapon systems with significant autonomy in the critical functions of selecting and attacking targets are already in use. Today these weapons tend to be highly constrained in the tasks they carry out (e.g. defensive rather than offensive operations), in the types of targets they can attack (e.g. vehicles and objects rather than personnel) and in the contexts in which they are used (e.g. simple, static, predictable environments rather than complex, dynamic, unpredictable environments). Closer examination of these existing weapon systems may provide insights into what level of autonomy would be considered acceptable and what level of human control would be considered appropriate.

Autonomous weapon systems that are highly sophisticated and programmed to independently determine their own actions, make complex decisions and adapt to their environment (referred to by some as "fully autonomous weapon systems" with "artificial intelligence") do not yet exist. While there are different views on whether future technology might one day achieve such high levels of autonomy, it is notable that today machines are very good at quantitative analysis, repetitive actions and sorting data, whereas humans outperform machines in qualitative judgement and reasoning.

[1] First published 9 May 2014: http://www.icrc.org/eng/resources/documents/report/05-13-autonomous-weapons-report.htm

There is recognition of the importance of maintaining human control over selecting and attacking targets, although there is less clarity on what would constitute 'meaningful human control'. Some suggest that 'fully autonomous' weapon systems, by definition operating without human supervision, may be useful in very limited circumstances in high-intensity conflicts. However, autonomous weapon systems operating under human supervision are likely to be of greater military utility due to the military requirement for systematic control over the use of force.

Two States – the United States and the United Kingdom – have developed publicly available national policies on autonomous weapon systems. The US policy states that "autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force." The UK policy is that the "autonomous release of weapons" will not be permitted and that "…operation of weapon systems will always be under human control." Other States have either not yet developed their policy or have not discussed it openly.

There is no doubt that the development and use of autonomous weapon systems in armed conflict is governed by international humanitarian law (IHL), including the obligation to undertake legal reviews in the study, development, acquisition or adoption of new weapons. As with any new weapon, the legality of autonomous weapon systems must be assessed based on their design-dependent effects and their intended use. However, it is not clear how such weapons could be adequately tested given the absence of standard methods for testing and evaluating autonomous systems.

There is acknowledgement that programming a machine to undertake the qualitative judgements required to apply the IHL rules of distinction, proportionality and precautions in attack, particularly in complex and dynamic conflict environments, would be extremely challenging. It is clear that the development of software that would be capable of carrying out such qualitative judgements is not possible with current technology, and is unlikely to be possible in the foreseeable future. Some have nevertheless argued that weapon systems with autonomy in critical functions can comply with IHL when performing simple tasks in predictable environments, as is the case with some existing weapon systems. Others argue that it would be difficult to ensure that these systems are solely used within such constraints.

There are different views on the adequacy of IHL to regulate the development and use of autonomous weapon systems. Some take the view that existing law is sufficient. Others argue that an explicit ban on autonomous weapon systems is necessary, or the development of a legal norm requiring, and defining, 'meaningful human control'.

States, military commanders, manufacturers and programmers may be held accountable for unlawful 'acts' of autonomous weapon systems under a number of distinct legal regimes: State responsibility for violations of IHL and international human rights law; international criminal law; manufacturers or product liability; and corporate criminal liability. The lack of control over and unpredictability of autonomous weapon systems could make it difficult to find individuals involved in the programming and deployment of the weapon criminally liable for war crimes, as they may not have the knowledge or intent required for such a finding. On this basis, several speakers and participants expressed concern about a potential 'accountability gap.'

Some suggest that there may be a duty to develop new technology if it might reduce the impact of armed conflict on one's own forces and on civilians. Others argue it is more likely that autonomous weapon systems will have limited capabilities to comply with IHL, and that many of the perceived advantages could be achieved using weapon systems that are remotely operated under direct human control.

Even if autonomous weapon systems could be used in compliance with IHL rules, ethical and moral challenges need to be considered carefully. There is the question of whether the principles of humanity and the dictates of public conscience allow life and death decisions to be taken by a machine with little or no human control. It is argued that the manner in which people are killed matters, even if they are lawful targets. Some emphasize that respecting the human right to dignity means that killing capacity cannot be delegated to a machine; rather, the decision to take someone's life must remain with humans.

SUMMARY REPORT

1. BACKGROUND

The aim of the ICRC's expert meeting was to gain a better understanding of the range of issues raised by autonomous weapon systems and to share perspectives among government representatives, independent experts and the ICRC. It brought together 21 States[2] and 13 independent experts, including roboticists, jurists, ethicists, and representatives from the United Nations and non-governmental organizations. The meeting was held under the Chatham House Rule.

The ICRC first raised its concerns about autonomous weapon systems in a 2011 report, *International Humanitarian Law and the challenges of contemporary armed conflicts*,[3] calling on States to carefully consider the fundamental legal, ethical and societal issues raised by these weapons before developing and deploying them.

In preparation for the expert meeting, the ICRC reviewed available information on autonomous weapon systems and, in a background document, highlighted questions relating to: autonomy in existing weapon systems; interest in increased autonomy; compatibility with international humanitarian law (IHL); and ethical and societal concerns.

It is clear that some weapons with significant degrees of autonomy in selecting and attacking targets are already in use today, although they are used in limited circumstances. They tend to be operated in fixed positions (rather than mobile), used primarily in unpopulated and relatively simple and predictable environments, and deployed against military objects (as opposed to directly against personnel). However, there is also continued interest in increasing overall autonomy of existing weapon platforms, in particular mobile unmanned systems that operate in the air, on the ground, or at sea.

There is no internationally agreed definition of an autonomous weapon system. For the purposes of the meeting, 'autonomous weapon systems' were defined as weapons that can independently select and attack targets. These are weapon systems with autonomy in the critical functions of acquiring, tracking, selecting and attacking targets.

Discussions at the meeting were rich and wide-ranging, covering the following topics:

- Civilian robotics and developments in autonomous systems
- Military robotics and drivers for development of autonomous weapon systems
- Autonomy in existing weapon systems
- Research and development of new autonomous weapon systems
- Military utility of autonomous weapon systems in armed conflict
- Current policy on autonomous weapon systems
- Autonomous weapon systems under international humanitarian law
- Accountability for use of autonomous weapon systems
- Ethical issues raised by autonomous weapon systems

A summary of presentations and discussions is provided in Section 2. This summary is provided under the sole responsibility of the ICRC. It is not intended to be exhaustive but rather it reflects the key points made by speakers and participants. Where agreement or disagreement on certain points is indicated in the text, it reflects only a sense of the views among those who spoke.

[2] Algeria, Brazil, China, Colombia, France, Germany, India, Israel, Japan, Kenya, Mexico, Norway, Pakistan, Qatar, the Republic of Korea, the Russian Federation, Saudi Arabia, South Africa, Switzerland, the United Kingdom and the United States.
[3] ICRC (2011) *International Humanitarian Law and the challenges of contemporary armed conflicts*. Report for the 31st International Conference of the Red Cross and Red Crescent, Geneva, 28 November to 1 December 2011.

2. SUMMARY OF PRESENTATIONS AND DISCUSSIONS

2.1 Civilian robotics and developments in autonomous systems

The speaker in the first session described the rapid progress in civilian robotics in the past decade, including the development of systems with autonomous functions, such as autonomous vacuum cleaners, underwater robots used to map the seabed, and soon cars that may be able to drive autonomously.

Using examples such as autonomous cars and humanoid robots, the speaker explained the main characteristics and limitations of current autonomous robotic systems:

- They are best at performing simple tasks, and are not capable of the complex reasoning or judgement carried out by humans;
- They are best at carrying out single rather than multiple tasks;
- They have little capability to perceive their environment, and are consequently most capable in simple, predictable environments;
- They have limited adaptability to unexpected changes in their environment;
- They are unreliable in performing their assigned task and generally cannot devise an alternative strategy to recover from a failure;
- They can be slow at performing the assigned task.

Looking to the future, the speaker explained that autonomous robotic systems will gradually become more sophisticated with advances in computation techniques and sensor quality. However, there are fundamental technical challenges to address before they may become more versatile (e.g. performing multiple tasks), more adaptable (i.e. to unpredictable external environments), and capable of carrying out complex tasks that require reasoning and judgement.

During discussions the speaker explained that as robotic systems are given greater decision-making power (and therefore more autonomy) they become more unpredictable. While robotic systems performing repetitive actions according to specific rules may be more predictable, with increasing autonomy – and less strictly defined rules – there will be increasing uncertainty about how the system will operate.

Regarding public acceptance of robotic systems, the speaker emphasized there will be demand for high reliability because humans are much less forgiving of machines in making mistakes than we are of ourselves. Therefore autonomous robotic systems would be expected to outperform humans.

One participant noted that the pace of development in robotics is rapid and that the core technical challenges are being addressed by researchers. It was added that, while complex reasoning is beyond the capability of current technology, existing robotic systems are already able to outperform humans on certain tasks. The speaker suggested that this type of high performance relies on the task being very well well-defined and information about the environment (or context) pre-programmed, adding that existing robotic systems are not able to adapt to unexpected changes in the environment.

There was also a discussion among participants about the capabilities of machines to recognize objects and individuals, or even to determine human intentions. While current visual recognition technology is becoming more sophisticated, it remains unreliable. However, there were diverse views on where technology development may lead in this area.

Overall, the speaker noted that current technological limits mean it is most likely that human-robot interaction will be preferred over independent action of robots. This might be seen as ‘supervised autonomy’ where decisions requiring intelligence – and the ability to carry out complex reasoning and judgement – are retained by humans.

2.2 Military robotics and drivers for development of autonomous weapon systems

The speaker made a distinction between automatic systems and autonomous systems explaining that the former operate with pre-programmed instructions to carry out a specific task, whereas the latter act dynamically to decide if, when, and how to carry out a task. Automatic systems therefore act based on deterministic (rule-based) instructions whereas autonomous systems act on stochastic (probability-based) reasoning, which introduces uncertainty. However, the speaker emphasized that future military systems would likely be hybrids of automatic and autonomous systems.

The speaker went on to emphasize three main drivers for military interest in increased overall autonomy for weapons platforms, which are linked to the advantages of unmanned weapon systems in general. First is the potential for reduced operating costs and personnel requirements. Second is the potential for increased safety in operating these platforms (compared to manned systems). And third is the potential for increased military capability by using one weapons platform to perform all functions – from identifying through to attacking a target.

Other drivers of autonomy in weapon systems mentioned during discussions included the potential for: force multiplication (i.e. greater military capability with fewer personnel); removal of risks to one's own forces; and decreased reliance on communications links. However, a participant noted that many of these advantages may still be possible while retaining remote control of the critical functions of selecting and attacking targets

The speaker noted that some functions, such as 'autopilot' in military and civilian aircraft, have been autonomous for many years. For other functions, such as target selection and attack, direct human control is maintained for the vast majority of weapon systems today.

The speaker highlighted several limitations in the current technology of autonomous systems that are particularly relevant for military applications such as weapon systems. Firstly, current autonomous systems are 'brittle' (not adaptable and easily break down), which makes them unreliable. Secondly, existing autonomous systems still rely heavily on human input for many functions in order to correct mistakes. Thirdly, there is a lack of standard methodologies to test and validate autonomous systems. Finally, and perhaps the greatest barrier to development of autonomous weapon systems in particular, is the limited ability of autonomous robotic systems to perceive the environment in which they operate.

During discussions, speakers and participants referred to the concept of 'fully autonomous weapon systems' meaning highly sophisticated weapon systems with 'artificial intelligence' that are programmed to independently determine their own actions, make complex decisions and adapt to their environment. These do not yet exist and there was a certain divide between those optimistic about the future development of underlying technology, who suggested that 'fully autonomous systems' are inevitable and may one day be more capable than humans at complex tasks, and those who emphasised the current limits of foreseeable technology, arguing that there is a need to focus attention on managing the relationship between humans and machines to ensure that humans remain in control of robotic systems. In response to the question of whether autonomous humanoid robots – with comparable decision-making capabilities to humans – might be developed by the military, the speaker said that it is not likely even in the long term.

However, the speaker did note that it would be possible to develop a weapon system today with full autonomy in selecting and attacking targets provided the developer or user was prepared to accept a high failure and accident rate. Therefore the likelihood of these weapons being used will also depend on what is considered acceptable by the user.

The speaker also emphasized that the civilian commercial market is the driving force for development of autonomous systems in general and that, once the technology has been developed for other purposes, it may be relatively easy to then weaponize a commercially developed system.

2.3 Autonomy in existing weapon systems

Speakers in this session explained that there are already weapon systems in use that have autonomy in their 'critical functions' of selecting and attacking targets. Noting that there are no internationally agreed definitions of autonomous weapon systems, one speaker highlighted the US Department of Defense policy, which divides autonomous weapons into three types according to the level of autonomy and the level of human control:

- *Autonomous weapon system (also referred to as human 'out-of-the-loop')*: "A weapon system that, once activated, can select and engage targets without further intervention by a human operator."[4] Examples include some 'loitering' munitions that, once launched, search for and attack their intended targets (e.g. radar installations) over a specified area and without any further human intervention, or weapon systems that autonomously use electronic 'jamming' to disrupt communications.

- *Supervised autonomous weapon system (also referred to as human 'on-the-loop')*: "An autonomous weapon system that is designed to provide human operators with the ability to intervene and terminate engagements, including in the event of a weapon system failure, before unacceptable levels of damage occur."[5] Examples include defensive weapon systems used to attack incoming missile or rocket attacks. They independently select and attack targets according to their pre-programming. However, a human retains supervision of the weapon operation and can override the system if necessary within a limited time-period.

- *Semi-autonomous weapon system (also referred to as human 'in-the-loop')*: "A weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator."[6] Examples include 'homing' munitions that, once launched to a particular target location, search for and attack pre-programmed categories of targets (e.g. tanks) within the area.

The speaker identified three main considerations for assessing the implications of autonomy in a given weapon system: the task the weapon system is carrying out; the level of complexity of the weapon system, and the level of human control or supervision of the weapon system. The speaker added that critical functions of some weapons systems have been automated for many years and that a weapon system does not necessarily need to be highly complex for it to be autonomous.

The speakers in this session emphasized that autonomous weapon systems in use today – 'autonomous' or 'supervised autonomous' according to the definitions provided – are constrained in several respects: first, they are limited in the tasks they are used for (e.g. defensive roles against rocket attacks, or offensive roles against specific military installations such as radar); second, they are limited in the types of targets they attack (e.g. primarily vehicles or objects rather than personnel), and third, they are used in limited contexts (e.g. relatively simple and predictable environments such as at sea or on land outside populated areas). However, both speakers noted that there are some existing anti-personnel weapon systems that have autonomous modes, such as so called 'sentry weapons'.

[4] US Department of Defense (2012) *Autonomy in Weapon Systems, Directive 3000.09,* 21 November 2012, Glossary, Part II Definitions.
[5] *Ibid.*
[6] *Ibid.*

There was a discussion among participants that identified a number of different factors that are taken into consideration by the military in determining both the desirability of autonomy selecting and attacking targets, and the acceptability of autonomy for a given weapon system.

Major factors affecting the desirability for autonomy in existing weapons include: the military capability advantage provided by autonomy in selecting and attacking targets; the necessity of this autonomy for the particular task (e.g. the desirability for the weapon system to act faster than humans); and the reliability or susceptibility of communications links.

The assessment of how much autonomy is considered acceptable in existing weapons is influenced by a number of different factors including:

- The type of task the weapon is being used for (e.g. offensive or defensive);
- The type of target (e.g. objects or personnel);
- The type of force (e.g. non-kinetic, such as electronic 'jamming', or kinetic force);
- The context in which the weapon is used (e.g. simple or 'cluttered' environments);
- The ease of target discrimination in the particular context;
- The way in which humans interact with, and oversee, the weapon system;
- The 'freedom' of the weapon to move in space (e.g. fixed or mobile; and narrow or wide geographical area);
- The time frame of action of the weapon (i.e. attacks only at a specific point in time or attacks over a longer period of time); and
- The predictability, reliability, and therefore trust in the operation of the weapon system.

A participant emphasized that there is a need to look more closely at autonomy in existing weapons to learn lessons about the rationale for autonomy in selecting and attacking targets and the constraints placed on the operation of these weapons. This may provide useful insights into what level of autonomy would be considered acceptable and what level of human control would be considered appropriate.

2.4 Research and development of new autonomous weapon systems

As all speakers explained during this session, while some existing weapon systems have autonomous features of selecting and attacking targets, there is military interest in increased autonomous functioning for the expanding range of unmanned air, ground and maritime weapons platforms.

One speaker emphasized that much of the focus to date has been on increasing autonomy in 'non-critical functions', such as navigation (e.g. autopilot, take-off and landing, route planning) and other on-board systems, such as sensor control. Nevertheless, the speaker noted that there has been work undertaken on automating some elements of the targeting process, such as image processing, image classification, tracking, and weapon trajectory planning.

Another speaker explained that some new weapons and prototypes under development have been promoted by manufacturers, or suggested by developers, as having autonomous features of target selection and attack. As all speakers noted, these include air weapon platforms that search for potential targets within an area, underwater systems that can search for and attack ships, and ground systems that have autonomous modes for selecting and attacking targets (e.g. so called 'sentry weapons').

During discussions one speaker noted that it is difficult to gain a fuller understanding of the degree of interest in autonomy for 'critical functions' of selecting and attacking targets

because there is little information available on weapons development due to the confidentially and classification associated with these activities.

Two speakers emphasized general limitations of autonomous robotic systems that affect their suitability for weapon systems in particular: their limited ability to carry out complex decision-making; their lack of reliability and predictability; their difficultly in operating outside simple environments; and the difficulty in testing autonomous systems due to their unpredictability. Acknowledging current limitations, one speaker suggested that future technology developments over the longer term may yet enable development of autonomous weapon systems that can perform as well as, or better than, humans.

One speaker highlighted the limitations of existing vision systems developed for automatic target recognition, which are unsophisticated and can only operate in simple, low-clutter environments. Another speaker explained that these systems are limited both by their ability to use information gathered in making judgements and by the capability of their sensors to collect information. Whereas humans use multiple sensory inputs to inform decision-making, automated targeting systems may rely on one or two – such as video and acoustic detection. However, another speaker noted there are also some types of sensors where machines can offer sensing capabilities that humans do not possess, for example infra-red cameras.

As regards reliability, one speaker noted that failures or errors in autonomous weapon systems could arise from many sources including: difficulties with human-machine interaction, malfunctions, hardware and software errors, cyber-attacks or sabotage during development, and interference such as 'jamming' or 'spoofing'. Another speaker explained that a problem with human-machine interaction can be various biases, such as automation bias (i.e. too much trust in a machine) or confirmation and belief bias (i.e. tendency to trust information that confirms existing information or beliefs).

There was agreement among speakers and participants that autonomous weapon systems programmed to independently determine their own actions, make complex decisions and adapt to their environment (referred to by some as "fully autonomous weapon systems" with "artificial intelligence") are not conceivable with today's technology. However, there were different views on whether future technology might one day achieve such high levels of autonomy. One speaker highlighted the general differences between human and machine (computer) capabilities; it is notable that machines are very good at quantitative analysis, repetitive actions and sorting data, whereas humans outperform machines at qualitative judgement, reasoning and recognizing patterns.

Another speaker said that autonomy in various functions of unmanned weapons platforms will increase in the future but that this could actually lead to the need for more human supervision due to the increased unpredictability that comes with increased autonomy. Therefore it is likely that partnerships between humans and machines would be necessary rather than full autonomy for weapon systems.

One speaker argued that 'fully autonomous weapon systems' may still be of utility in narrow circumstances where they might be able to perform in a more conservative – or less risk-averse – way than humans. During discussions a participant highlighted the potential for 'function creep' or 'mission creep' where an autonomous weapon system designed for a specific limited context is then used in wider contexts, or where an autonomous system developed and used for a non-weaponized function is later weaponized. Another speaker also raised the risks associated with proliferation of autonomous weapon systems, including the potential for unpredictable interactions if these weapon systems were ever deployed against each other.

2.5 Military utility of autonomous weapon systems in armed conflict

Views on the military utility of autonomous weapon systems varied according to different perspectives of what is considered within the scope of a discussion about autonomous weapon systems. Some participants focused solely on 'fully autonomous weapon systems' that do not yet exist, while others included weapon systems already in use that have autonomy in selecting and attacking targets.

One speaker explained that a weapon system with full autonomy in target selection and attack potentially offers increased capabilities in force protection, particularly in situations where time is limited, and it further removes the risks for the user of the weapon system and their soldiers. It has been suggested that autonomous weapon systems may offer savings in personnel and associated costs, however the speaker suggested this may not be the case since these weapons are likely to have high procurement and maintenance costs. Another speaker emphasized the potential utility of these weapon systems for 'dull, dirty, dangerous and deep' – so called '4D' – missions.

One speaker explained that a 'fully autonomous weapon system' should be understood as a weapon system that, once programmed by humans, is given a mission task in a generic way and then operates without further intervention. Such a weapon system, by definition, would not be supervised. The speaker discussed the military utility of 'fully autonomous weapon systems' based on the central assumption that these future systems would be capable of complying with IHL. However, during discussions a participant noted that the lack of supervision and the inherent unpredictability of a 'fully autonomous weapon system' raise questions as to whether there could ever be full confidence that it would comply with IHL in all circumstances.

One speaker suggested that 'fully autonomous weapon systems' may not be useful in low-intensity conflicts but they could find a role in high-intensity conflicts against military objects, and in very limited circumstances. These situations might include time-critical defensive situations, particularly those where the tempo of operations and time pressure for a response is high.

Both speakers noted that the operating environment would also be an important factor, since identification of legitimate targets may be easier in some contexts, e.g. at sea or in unpopulated areas on land, than in others, e.g. populated urban areas. The speakers noted that use in complex environments against personnel would be problematic, as the weapon system would need to make very fine judgements such as recognizing a soldier who is injured or surrendering, and determining whether a civilian is directly participating in hostilities. One speaker noted that use in populated areas would also be problematic from the perspective of gaining support of the local population during counter-insurgency type operations. Other difficulties could arise in the use of autonomous weapon systems by coalitions of different countries since they may have different policies and rules of engagement.

One speaker noted that the role of the weapon system – defensive or offensive – and the type of target – military object (so called 'anti-materiel') or combatant (i.e. anti-personnel) may also be key factors affecting their utility. Based on examples of current weapon systems, defensive anti-materiel autonomous weapon systems might be seen as more acceptable, and therefore of more utility, than offensive weapon systems targeting personnel.

Another speaker explained that, with an increased number of armed robotic systems in use, it is possible that in the future autonomous weapon systems could be used alongside soldiers, or in attacks against other autonomous weapon systems, with unpredictable results. More broadly the speaker expressed concerns that autonomous weapon systems could risk

making conflict more likely by lowering the threshold for the use of force since they could provide opportunity to attack without risks to the users.

During discussions a participant expressed concern that autonomous weapon systems that are not capable of complying with IHL might be deployed despite their limitations, or used in environments that they are not equipped to operate in. A participant also said that the use of autonomous weapon systems might provoke strong reactions by the side being targeted, since the acceptability of attacks carried out against humans by autonomous robots might be considered differently to those carried out with existing means.

During presentations and discussions there was recognition of the importance of retaining human control over selecting and attacking targets but less clarity on what would constitute 'meaningful human control'. One speaker explained that the military requirement for systematic control of the use of force would mean that autonomous weapon systems under supervision are likely to be of greater military utility. A participant raised questions about the meaningfulness of human supervision if the time window for human intervention is extremely short.

Nevertheless, one speaker noted that it is still possible that 'fully autonomous weapon systems', operating without human supervision, may be of military value in critical situations – such as responding to an overwhelming attack, or where a mission is critical but communications links are not available or 'jammed' – provided that the user is confident that the autonomous weapon system would perform better than humans in the same situation.

2.6 Current policy on autonomous weapon systems

Two States – the United States and the United Kingdom – are known to have developed national policy on autonomous weapon systems, and representatives of these countries presented their respective policies at the meeting. Other States have either not yet fully developed their policy or have not discussed it openly. However they were encouraged to do so by some participants during discussions.

United Kingdom

The speakers explained that the UK policy is based on a distinction between automated weapon systems and 'fully autonomous weapon systems'. Under UK definitions an automated or automatic system is "…programmed to logically follow a pre-defined set of rules with predictable outcomes" whereas an autonomous system is "…capable of understanding higher level intent and direction."[7] An autonomous weapon system would be capable of understanding and perceiving its environment, and deciding a course of action from a number of alternatives without depending on human oversight and control. The UK understanding is that the overall activity of such a system would be predictable but individual actions may not be.

The speakers noted that current UK policy is that the 'autonomous release of weapons' will not be permitted and that "…operation of weapon systems will always be under human control".[8] As a matter of policy, the UK is committed to using remotely piloted rather than highly automated systems as an absolute guarantee of oversight and authority for weapons release.

[7] UK Ministry of Defence, Development, Concepts and Doctrine Centre (2011) *Joint Doctrine Publication 0-01.1: UK Supplement to the NATO Terminology Database,* September 2011, p. A-2.
[8] UK Ministry of Defence (2013) *Written Evidence from the Ministry of Defence submitted to the House of Commons Defence Committee inquiry 'Remote Control: Remotely Piloted Air Systems - current and future UK use'*, September 2013, p3.

The speakers added that the UK government has previously stated to the UK parliament that "no planned offensive systems are to have the capability to prosecute targets without involving a human."[9] They explained that for existing automated weapon systems this human control could be seen as the human setting the pre-programmed parameters of the weapon system's operation.

From a UK legal perspective, the speakers explained that all weapons developed or acquired are subject to legal review in accordance with Article 36 of Additional Protocol I. Such legal reviews incorporate an assessment of the compatibility of the weapon with the core rules of IHL as well as an assessment of whether the weapon is likely to be affected by the current and future trends in the development of IHL. The UK considers the existing provisions of international law sufficient to regulate the use of autonomous weapon systems.

United States

The speaker explained that US policy on autonomy in weapon systems is found in Department of Defense Directive 3000.09 of November 2012. It covers manned and unmanned platforms, as well as guided munitions, and excludes mines, cyber weapons, and manually guided munitions.

The speaker stated that the policy was developed in order to reduce risks associated with autonomy in weapon systems and specifically it "establishes guidelines designed to minimize the probability and consequences of failures in autonomous and semi-autonomous weapon systems that could lead to unintended engagements",[10] with the recognition that no policy can completely eliminate the possibility of such failures. The policy states that "autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force."[11]

The speaker noted that the policy does not further define what is considered an appropriate level of human judgement. Such an assessment may be different for different weapon systems depending on the operating environment and the type of force used. The speaker explained that factors in determining levels of autonomy in weapon systems include: the capability of the weapon system of carrying out a military mission or task; the robustness of the system against failures and enemy hacking; a design that ensures human judgement is retained for appropriate decisions; and the capability of the system to be used in compliance with IHL, as determined by legal review.

The US policy recognized the increased risks associated with reduced human control, i.e. moving from human 'in-the-loop' through human 'on-the loop' to human 'out of the loop'. The speaker noted that while weapon systems may become more capable with increased autonomy, they may become less predictable due to an increased ability to define their own actions. US policy is broad in that it covers existing and potential future weapons that have some autonomy in selecting and attacking targets. In this sense it covers the full range of weapon systems with autonomy in selecting and attacking targets.

The policy sets out three types of autonomous weapon systems and associated constraints. A 'semi-autonomous weapon system' (see Section 2.3 for the US definition) is considered acceptable for lethal offensive and defensive applications, and current examples include homing munitions, unmanned aircraft with GPS-guided bombs, and intercontinental ballistic missiles.

An 'autonomous weapon system' (see Section 2.3 for the US definition) is considered acceptable for some non-lethal applications – such as electronic jamming of materiel targets

[9] *Ibid.*
[10] US Department of Defense (2012) *Autonomy in Weapon Systems, op. cit.,* para 1(b)
[11] *Ibid*, para 4(a).

– due to the type of force and the type of target, which is seen to present lower risks. Under US policy, the speaker explained that any future development of offensive autonomous weapon systems employing lethal force would require specific additional review and approval before development and again before fielding.

Under the policy a sub-category of an 'autonomous weapon system' is a 'supervised autonomous weapon system' (see Section 2.3 for the US definition), which is considered acceptable for lethal operations against vehicle and materiel targets but in local defensive operations only. Current examples include ship defence systems and land-based air and missile defence systems. Development of an offensive supervised autonomous weapon system, or one used defensively to target humans, would require specific additional review and approval before development and again before fielding.

Wider discussions

Discussions on current policy illustrated some differences in approach and in the scope of weapons under consideration. Some participants noted that the US policy is designed to cover autonomy in existing and future weapon systems, whereas the UK policy is solely focused on potential future 'fully autonomous weapon systems'.

A participant highlighted the difficulties associated with carrying out legal reviews of autonomous weapon systems due to challenges with testing. One speaker noted that realistic testing is a challenge for any weapon system and simulations can be used. However, the speaker acknowledged that verifying and validating complex software systems, as might be incorporated in an autonomous weapon system, is a very difficult process.

While there was broad agreement among speakers and participants of the need to retain human control over the use of force, several participants highlighted a lack of clarity over what constitutes 'appropriate' or 'meaningful' human control over weapon systems that independently select and attack targets.

2.7 Autonomous weapon systems under international humanitarian law

There was no doubt that the development and use of autonomous weapon systems in armed conflict is governed IHL, including the obligation to undertake legal reviews in the study, development, acquisition or adoption of new weapons, as required by Article 36 of Additional Protocol I to the Geneva Conventions (API) and implemented by some States not party to API.

In considering the capabilities that a 'fully autonomous weapon system' might need to be able to comply with IHL, several speakers emphasized that qualitative decision-making is typically required when applying the IHL rules of distinction, proportionality and precautions in attack. For instance, the IHL rule of distinction requires that attacks only be directed at combatants and military objectives. Civilians are protected from direct attack, unless and for such time as they are directly participating in hostilities. Military objectives are defined as "those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."[12] In this regard, one speaker emphasized that determining who and what can be attacked under IHL, and under what circumstances and using which means, is therefore context-dependent.

[12] *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflict* (*Additional Protocol I* or AP I) (adopted on 8 June 1977, entered into force on 7 December 1978), art 52(2).

The rule of proportionality, according to which incidental casualties and damages can be lawful if they are not excessive in relation to the concrete and direct military advantage anticipated, is said to be among the most complex to interpret and apply under IHL, as it requires a case-by-case qualitative judgement, in often rapidly changing circumstances. In addition, IHL requires parties to armed conflicts to take constant care to spare the civilian population, civilians and civilian objects. This obligation underlies the rule of precautions in attack, which also requires making a number of qualitative evaluations to avoid or in any event minimize incidental harm to civilians and civilian objects.

Legal reviews of new weapons

Undertaking legal reviews of autonomous weapon systems raises a number of challenges. Firstly, the timing of the reviews is important. Article 36 refers to an obligation to determine the legality of new weapons in the study, development, acquisition or adoption of new weapons. Two speakers emphasized that legal reviews should be carried out throughout the development process, and not just when the weapon is ready for procurement. One speaker highlighted the fine line between research and development and suggested that the obligation to undertake a legal review does not apply to open-ended research, but it does apply as soon as such research is carried out for a specific weapon programme. Already at this early stage, there is an interest in ensuring that the weapon complies with the law, before further resources are invested into its development.

Regarding the content of legal reviews, speakers queried how weapons with varying degrees of unpredictability could be tested. It was emphasized that current testing and evaluation procedures have limitations and there are no standard methods for testing autonomous systems. Although testing autonomous weapon systems may be affected by limited weapons budgets, States are obliged to test new weapons to verify their performance, and must find ways of ensuring that the testing process is effective. One participant noted that States could exchange experiences on development and use of weapons, and that cooperation in testing would be advantageous. Another participant made the point that, as with the development of other weapons, the legality of autonomous weapon systems must be assessed based on their design-dependent effects and their intended use.

Speakers and participants expressed different views regarding the relevance of the Martens Clause to legal reviews of new weapons. Some were of the opinion that States were under an obligation to assess whether a new weapon complies with the principles of humanity and the dictates of public conscience. Others were of the view that the Martens Clause is not a criterion in its own right; rather, it operates as a reminder that even if new technologies are not covered by particular treaty law, other international norms nevertheless apply to them.

Challenges in complying with targeting rules under IHL

All of the speakers acknowledged the complexity of the assessments and judgements involved in applying the IHL rules of distinction, proportionality and precautions in attack, especially in dynamic conflict environments. These assessments and judgements appear to be uniquely human (some referred to "subjective" appreciation), and would seem extremely challenging to program into an autonomous weapon system. Current technology, including heat sensors, visual sensors capable of detecting military uniforms or weapons, and sensors that detect incoming fire would not be capable of independently making the nuanced distinctions required by the principle of distinction, including distinguishing persons that are *hors de combat* from combatants, and civilians from those who are directly participating in hostilities. It is clear that the development of software that would be capable of carrying out such qualitative judgments is not possible with current technology. Some speakers even found it difficult to imagine a day when technology could make this possible.

One speaker made the point that an evaluation of military advantage (under the rule of distinction for the purpose of determining whether an object is a military objective, and under the rule of proportionality to determine whether the incidental harm would be excessive in relation to the concrete and direct military advantage anticipated) requires not only an ability to perceive and analyse the immediate circumstances, but also requires knowledge of the broader context of the conflict. Assuming that an autonomous weapon system is incapable of this, a human would have to be in constant communication with the system, to input information relevant to this broader assessment. On the other hand, there may be ways of updating the information database of the machine so that it is aware of the real-time military advantage associated with attacking the category of objective in question.

Under the obligation to cancel or suspend an attack if it becomes apparent that the attack is indiscriminate or disproportionate, one speaker noted that an autonomous weapon system would need to be capable of quickly perceiving and analysing changes in the environment, and adapting its operations accordingly. Again, this represents a significant programming challenge.

In contrast, a participant noted that weapon systems that perform simple tasks in predictable environments could be easier to develop. When operating within such limits, autonomous weapon systems may be capable of complying with IHL. In response, speakers and participants acknowledged the difficulty in enforcing such restrictions, particularly regarding use by non-State armed groups.

Working on the assumption that technology may one day be capable of complying with IHL rules without human intervention, two speakers pointed out the potential advantages of autonomous weapon systems. In particular, autonomous weapon systems would not be affected by fear, hatred or other emotions. Autonomous weapon systems may also be able to take additional precautionary measures because they would not be concerned about their own 'safety'. Finally, autonomous weapon systems may allow for greater transparency than humans, as they could be equipped with audiovisual recording devices and would not be 'motivated' to conceal information. In response, several participants made the point that many of these perceived advantages could also be achieved using weapon systems that are remotely operated under direct human control.

One speaker argued that predictability of the autonomous weapon system's compliance with IHL is vital; if it is not possible to guarantee that the weapon system will comply with IHL in all circumstances then it would not be lawful.

Adequacy of international humanitarian law

Speakers and participants expressed different views regarding the adequacy of IHL to regulate the development and use of autonomous weapon systems. Some were of the view that existing law is sufficient, although additional guidance on testing and legal reviews of autonomous weapon systems would be beneficial. Others expressed the view that an explicit ban on autonomous weapon systems is necessary, or development of a legal norm requiring, and defining, 'meaningful human control.'

2.8 Accountability for the use of autonomous weapon systems

The discussion on accountability for serious IHL violations committed by autonomous weapon systems raised a number of issues, including concern about a possible 'accountability gap' or 'accountability confusion.' Some suggested that such an accountability gap would render the machines unlawful. Others were of the view that a gap will never exist as there will always be a human involved in the decision to deploy an autonomous weapon system to whom responsibility could be attributed. However, it is unclear how responsibility

could be attributed in relation to 'acts' of autonomous machines that are unpredictable. How can a human be held responsible for a weapon system over which they have no control? In addition, error and malfunction, as well as deliberate programming of an autonomous weapon system to violate IHL, would require that responsibility is apportioned to persons involved in various stages, ranging from programming and manufacturing through to the decision to deploy the weapon system.

Speakers and participants raised a number of potential legal frameworks through which States, individuals, manufacturers and programmers could be held accountable, including the law of State responsibility, individual criminal responsibility, manufacturer's liability (for example, negligence or breach of contract), as well as corporate criminal liability (if an accepted concept under domestic law).

Many speakers and participants favoured the law of State responsibility as an appropriate legal framework for accountability for serious violations of IHL. One speaker suggested that states could and should be held liable if a legal review of an autonomous weapon system is inadequate, leading to a serious violation of IHL that could have been prevented through better testing and review of the weapon system. In this respect, views were expressed regarding the need to develop more precise regulations for testing and review of such weapons.

Speakers and participants also discussed international criminal law, although questions were raised regarding difficulties in proving knowledge or intention (required for a finding of criminal liability) when the weapon system is operating autonomously, or in cases of error or malfunction. One participant suggested that a programmer that intentionally programs an autonomous weapon system to commit war crimes could be held accountable. It was argued that, even if the programming occurred in peacetime, the programmer could be held liable for committing or being an accessory to a war crime if the autonomous weapon system carried out the act in an armed conflict. However, it would be challenging to identify a specific individual in the complex development and manufacturing chain, and very challenging to prove.

Another speaker highlighted the importance of accountability under international human rights law, including the right to life and human dignity, which, according to some experts, would apply even in armed conflict, though possibly subject to restrictions on their extra-territorial application.

An important question arising from the discussion is whether an autonomous weapon system that is capable of independently determining its actions and making complex decisions would be held to the same standard as humans in complying with IHL. Several speakers and participants suggested that machines should be held to a higher standard of performance than humans, partly because the public would be even less tolerant of war crimes committed by autonomous weapon systems than if they were committed by humans.

2.9 Ethical issues and the dictates of public conscience

Even if autonomous weapon systems could be used in such a way as to comply with IHL rules, there are ethical and moral challenges that need to be considered carefully. There is the related question of whether the principles of humanity and the dictates of public conscience (the Martens Clause) allow life and death decisions to be taken by a machine with little or no human control.

One speaker made the point that although moral sentiment and ethical judgement are not specified in the law and should not be confused with the law, these ethical elements are often used as a basis for formulating legal rules. For example, it was argued that moral

judgement underlies the determination of whether a weapon is of a nature to cause superfluous injury. Likewise, the Martens Clause embodies a moral framework whereby in the absence of a necessity to kill, lethal force should not be used even against lawful targets. In addition, it was argued that IHL rules governing the conduct of hostilities appeal specifically to humans exercising human judgement.

The speaker also pointed out that it matters *how* people are killed, even if they are lawful targets. According to one participant this is particularly true from the perspective of the affected community, which may be more aggrieved if the individual is killed by a machine – especially if there is an 'accountability gap' – than if lethal force is applied by a human. If someone is killed by a machine, this may also lead to a sense of injustice.

From an ethical perspective, one speaker asked what the consequences will be if we override the right to life through a piece of software? With increasing "dehumanization of warfare" we may lose responsibility and moral accountability, as well as our ability to define human dignity. The speaker emphasized that this is irresponsible, since morality requires meaningful human supervision of decisions to take life. In this regard, international human rights law also provides a moral framework; respecting the human right to dignity means that we do not delegate killing capacity to a machine, rather, the decision to take someone's life must remain with humans. A participant argued that moral responsibility relating to use of an autonomous weapon system will always remain with the last human in the chain of command.

At the same time, one participant stressed that we may have a duty to explore new technology if there is a chance it might reduce the impact of armed conflict on one's own forces and on civilians. Some other participants shared this view, noting the responsibility of States to explore ways of reducing risks to one's own forces.

In response, a speaker noted that a utilitarian approach must involve an assessment of both the possible humanitarian benefits of developing autonomous weapon systems and the potential risks, as well as the likelihood of these benefits and risks. Given the lack of evidence to indicate that autonomous robotic systems will ever be able to undertake complex reasoning and nuanced judgements, it will more likely be the case that autonomous weapon systems will have limited capabilities and would be unable to comply with IHL. The speaker also raised concerns about proliferation of autonomous weapon systems and its impact on the escalation of conflict.

The discussion also addressed the question of an ethical charter, with one participant referring to national discussions aimed at developing an ethical charter for programmers and manufacturers of civilian robots. One participant also noted the diverse ethical frameworks amongst States and suggested that there may be divergence between States on whether or not autonomous weapon systems are acceptable from an ethical standpoint.

Finally, a speaker suggested that human control and human decision-making are implicitly and explicitly required by international human rights law and international humanitarian law. As such, it was argued that there is a need to develop a legal norm requiring, and defining, 'meaningful human control' of weapon systems, and that further discussions on this issue are vital.

PART II: SELECTED PRESENTATIONS

Speaker's summary

CIVILIAN ROBOTICS AND DEVELOPMENTS IN AUTONOMOUS SYSTEMS

Dr Ludovic Righetti, Max Planck Institute for Intelligent Systems, Germany

This presentation provided an overview of current developments in civilian robotics and autonomous systems. The main objective of the presentation was to help answer several questions: “What are autonomous robotic systems and what are their current capabilities and uses?” and “What are the technological limits of current autonomous systems and foreseeable developments in this technology?”

What are autonomous robotic systems? Our notions of autonomous robots are certainly biased due to representations in science fiction and in movies. As a result we tend to attribute human capabilities, such as intelligence or cognitive reasoning, to robots. However, it is important to realize that, to date, there is no system with such capabilities. Over the past several decades there have been impressive advances in artificial intelligence, machine learning and robotics research but the science necessary to create machines with cognitive or ‘intelligent’ capabilities does not yet exist. Despite our hopes as researchers, it is not even clear how such a goal might be achieved. The most impressive examples of ‘intelligent’ machines, such as computer programs playing chess or Jeopardy, or personal voice recognition assistants such as Siri, are only able to perform the task for which they were programmed (e.g. chess programs cannot play a different game) in a carefully controlled environment (e.g. voice recognition programs do not like French accents in English) and, most importantly, they do not act in the real world. Therefore, it seems reasonable to argue that a machine with reasoning capabilities comparable to a human is not likely to exist in the foreseeable future. By extension, notions of conscious machines (e.g. the singularity) are more closely related to science fiction fantasies than tangible possibilities supported by scientific facts.

More realistically, we might consider a robot that can perform a certain number of complicated tasks without human intervention to be autonomous. For example, self-driving cars do not require human intervention on the road, and can therefore be considered autonomous. However, it is not easy to define what exactly constitutes an autonomous system because many different types of robots exist for many different applications. For example, the Roomba robot, a small vacuum-cleaning robot on wheels, is a perfectly autonomous machine. Once activated it will eventually sweep the apartment and even find its recharging deck when its battery level gets low. However, the robot can only accomplish one task. It behaves in a manner pre-programmed by the engineers, has a limited scope of action, and cannot understand its surroundings or make complex decisions.

Several characteristics of autonomy in robotic systems are particularly relevant for consideration of potential military applications. What is the level of autonomy of the system and what is the required precision of the human command required to activate the system? (e.g. a remote-controlled car which receives a constant stream of precise commands vs a self-driving car which only receives a desired address). What are the latencies in the human intervention, i.e. how much time does the human have available to give a command to the robot or to intervene in its current behaviour? How much adaptability does the robot have?

i.e. how much variation or how many unknowns in the environment can be tolerated while still ensuring good performance? How versatile is the robot? i.e. how many tasks can the robot perform? Can it learn new tasks for which it was not programmed? For each of these questions there is a continuum of possibilities. In order to understand the current capabilities of (partially) autonomous robotic systems and their main limitations, it is useful to consider the following important achievements in different areas of robotics.

There are already underwater and flying robots that can navigate autonomously. More recently, self-driving cars have also been developed. These cars are supposed to be able to drive completely autonomously, stopping at red lights, obeying traffic rules and avoiding other cars or pedestrians. While these achievements are truly impressive, there are important limitations. Perhaps the most important one is that, for all these machines, their understanding of complex and ever changing environments is very limited. In order to address this issue, the algorithms developed for these systems require a large amount of information about the environment they will operate in prior to being put to use. For example, self-driving cars generally need to know the type of crossings and traffic signs they might encounter at a particular location in advance in order to simplify computer vision algorithms. It is very likely that a car designed to drive in California would not perform so well in Switzerland and would most likely create a lot of problems in Great Britain, where people drive on the other side of the road.

In the past two decades, there has been important progress in legged locomotion. There are now quadruped robots that are able to walk on complicated terrain and recover from strong, unexpected external pushes. These machines can also reliably walk on terrain that they do not know beforehand, crossing small unknown obstacles. Biped robots, such as humanoids, are able to walk, and even run in some cases, and they can also recover from unexpected external pushes. However, biped locomotion is very unstable and these robots are not able to walk reliably on entirely unknown and non-flat terrain. In all cases of legged locomotion, robots are able to adapt to a certain level of uncertainty but they are not able to adapt to all types of uncertainty in the environment and so the robustness of the locomotion remains an issue. These robots also lack the ability to understand their environment. Moreover, due to the large number of degrees of freedom (i.e. their articulations), it is still very difficult to conceive algorithms that can compute the necessary motions for a robot to cross arbitrary obstacles or react to any type of external variation. It is fair to say that, in most cases, they perform best in highly controlled environments.

Aside from navigation tasks, another important challenge for robotics research is manipulation (i.e. how to grasp objects and use them to perform certain tasks). Nowadays robots are able to perform relatively complicated tasks, such as opening doors, using drills or cooking simple recipes in partially unknown environments. However, in all these cases the environments are still very much controlled. Moreover, it is difficult to guarantee that the robots will succeed in achieving the tasks all the time. In most complex manipulation scenarios, failure to achieve the task is still relatively high. Manipulation is particularly difficult because it requires automatic planning of complex sequences of actions that will lead to successful achievement of the task, as well as reasoning about the properties of the object in order to understand how they can be used. Moreover, this needs to be done in a constantly changing environment, for example in the kitchen of a restaurant where humans are also working. The number of possible actions is so high that current algorithms are not able to reason in such a general setting. This is one reason why successful autonomous robotic applications still require controlled environments. They help reduce the amount of possible actions and engineers can program pre-defined actions before the execution of the tasks. Another issue arises from the inability of robots to understand complex and changing environments. For example, the recognition of objects in a cluttered environment remains an important challenge. In addition, the robustness of programmed behaviour is another challenge, i.e. where the robot needs to make additional decisions if something does not work as planned.

A trend in robotics that seems promising for concrete applications is the use of supervised autonomy. In this case, instead of allowing complete autonomy for the robot, a human operator stays 'in the loop' to provide all the important cognitive abilities that the robot lacks. The DARPA Robotics Challenge, initiated in 2013, exemplifies this concept. It is a competition between several top research laboratories in the world where the goal is to develop robots that can reliably perform tasks in a disaster or emergency response scenario. The robots are required to traverse difficult terrain, open doors, climb up ladders and use tools. In all these scenarios, a human operator is allowed to remotely provide instructions to the robots, for example to help detect objects, select where to step on the ground or decide which action to take next. The complexity of the tasks carried out illustrates clearly the potential utility of such robots while the level of required remote assistance by human operators underlines the fundamental limitations in the development of fully autonomous systems.

There are two types of challenge that need to be addressed if autonomous, intelligent systems are to be created. Technological limitations, such as computational power, actuation and sensor quality and density, are limitations that will likely be overcome, or made less severe, in the near future given sufficient time and financial investment. These advances will lead to: better performances for self-driving cars; increased agility for walking machines; and higher dexterity in robotic manipulation. On the other hand, there are scientific challenges that we do not yet know how to solve. These include, for example: creating algorithms that can understand the world at a human level or reason about complicated tasks during manipulation; and creating versatile machines that can adapt to arbitrary environments. It is impossible to predict when or if these challenges will be solved. Therefore, despite tremendous progress in robotics in recent decades with constant improvement in the skills of robots in carrying out different tasks, there are still fundamental and difficult obstacles to developing robotic systems with true autonomy.

Speaker's summary

AUTONOMOUS WEAPONS AND HUMAN SUPERVISORY CONTROL

Professor Noel Sharkey, University of Sheffield, UK

Those who support the development of autonomous weapons often make the error of believing that it will provide their State with an asymmetric advantage and no one else will have the technology to keep up. However, history has shown us that new weapons technology proliferates rapidly. The automation of warfare is no exception. Once many States have autonomous weapons there will be rapid developments of counter weapons and counter-counter weapons. This leaves us with an uncertain future. We cannot know how such weapons will interact with one another except that it will be unpredictable.

A number of states have been discussing the development and use of autonomous weapon systems for more than a decade. These are weapons that once activated, select targets and engage them with violent force without further intervention by human operators. However, such systems pose considerable challenges for international humanitarian law (IHL), in particular to the principles of distinction, proportionality and precaution.

Although there has been considerable testing of autonomous combat platforms, none has yet been fielded.[1] The minimum requirements for fully autonomous weapons to comply with IHL are that they can:

- distinguish between military and non-military persons and objects
- determine the legitimacy of targets
- make proportionality decisions
- adapt to changing circumstances
- handle unanticipated actions of an adaptive enemy
- deal with other autonomous systems controlled by unknown combat algorithms.

The state of the art in computing machinery is unlikely to meet all of these requirements within the foreseeable future. Although computers are better at some tasks than humans, humans are better at some tasks than computers (see Table 1 for examples). Military control and humanitarian impacts are best served by playing to the strengths of both.

Currently ‘automatic target recognition methods’, despite decades of research, only work in low cluttered environments and with military objects such as tanks in the desert and ships at sea. The methods are unreliable with medium to high-clutter environments and are not used. This is unlikely to change significantly in the near to medium term future although improvements are expected in the longer term.

Distinguishing between combatants and civilians and others who are *hors de combat* is considerably more difficult. Sensing and vision processors will improve in the longer term future. But methods to determine the legitimacy of targets are not even in the pipeline yet.

[1] There are a number of weapons systems that Sense and React to Military Objects (SARMO) for protection against fast incoming munitions such as mortar shells and missiles (e.g. C-RAM, Phalanx, Mantis). These are not fully autonomous in that they are programmed to automatically perform a small set of defined actions repeatedly. They are used in highly structured and predictable environments that are relatively uncluttered with a very low risk of civilian harm. They are fixed base and have constant vigilant human evaluation and monitoring for rapid shutdown – US Department of Defense uses the term ‘supervised autonomy.’ These may be acceptable when used against military objects but caution must be exercised about expanding their role.

Table 1: The differing skills of computers and humans

Computers	Humans
calculating numbers	deliberative reasoning
searching large data sets	perceiving patterns
responding quickly to control tasks	meta-cognition (thinking about thinking)
simultaneous repetitive routine tasks	reasoning inductively
carrying out multiple complex tasks	applying diverse experience to novel tasks
sorting data	exercising meaningful judgement

Many targeting decisions are subjective in nature. Decisions such as the proportionate use of force require the deliberative reasoning of an experienced human commander who must balance civilian lives and property against direct military advantage. A human can even reason about their reasons for choices before making a decision (meta-cognition). These are not strengths of computing.

There are a very large, perhaps infinite, number of novel and unanticipated circumstances that can occur in warfare, and this is where humans score higher than computers. Autonomous weapons may catch out a clever enemy to begin with, but they will soon adapt their methods and 'game' the technology to make it strike unintended targets.

Computers are also susceptible to a number of potential problems that make them unpredictable: human error, human-machine interaction failures, malfunctions, communications degradation, software coding errors, enemy cyber-attacks, infiltration into the industrial supply chain, jamming, spoofing, decoys, and other enemy countermeasures or actions. Again, a human in the control loop can determine that a system is displaying aberrant behaviour and take appropriate action.

Some States already understand the unpredictability of autonomous weapons and propose to keep a person in the control loop. The US Department of Defense issued the first policy guidelines on autonomous weapons: "Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgement over the use of force."[2] On 26 March 2013, the Parliamentary Under Secretary of State, Lord Astor of Hever, replying to questioning in a House of Lords debate, acknowledged that fully autonomous systems might not be predictable and stated, "Let us be absolutely clear that the operation of weapons systems will always be under human control."[3]

However, the question remains as to what is meant by appropriate levels of human control or judgement. Humans need to exercise meaningful control[4] over weapons systems to counter many of the problems that arise from automation. The control of weapons mediated by computer programs raises its own problems. Perhaps the most important of these is the delicate human-computer balancing act. Because humans sometimes fail at some tasks, it does not mean that machines can do them any better. It can simply mean that humans are being asked to perform in a mode of operation that is not well suited to human psychology. This needs to be part of the equation of ensuring efficient and meaningful human supervisory control of weapons.

[2] US Department of Defense, *Autonomy in Weapon Systems*, Directive 3000.09, November 21 2012.
[3] UK House of Lords Hansard, 26 March 2013, http://www.publications.parliament.uk/pa/ld201213/ldhansrd/text/130326-0001.htm#st_14
[4] Article 36, *Structuring debate on autonomous weapons systems*. Memorandum for delegates to the Convention on Certain Conventional Weapons, November 2013.

Sharkey[5] has proposed a reframing of autonomy in terms of 5 levels of human supervisory control of weapons (rather than levels of autonomy). This clarifies the role of the human, makes the chain of command transparent and allows for clearer accountability.

Table 2: Levels of human supervisory control of weapons

1. human deliberates about a target before initiating any attack
2. program provides a list of targets and human chooses which to attack
3. program selects target and human must approve before attack
4. program selects target and human has restricted time to veto
5. program selects target and initiates attack without human involvement

Levels 1 and 2 are acceptable given the adequate consideration of the decision-making environment. Level 3 could be acceptable given the adequate time for deliberation. Levels 4 and 5 pose unacceptable risks of mishap**.**

Conclusions

IHL compliance with autonomous weapon systems cannot be guaranteed for the foreseeable future. Although we can expect considerable improvements in some facilities, other facilities may not be possible into the foreseeable future.

The predictability of fully autonomous weapon systems to perform mission requirements cannot be guaranteed. Testing such systems for unanticipated circumstances is not viable.

The unpredictability of autonomous weapons in unanticipated circumstances makes weapons reviews extremely difficult or even impossible to guarantee IHL compliance.

The combined strengths of humans and computers operating together with the human in charge of targeting decisions makes better military sense and will maintain greater humanitarian impact providing that the human reasoning process is taken into account.

[5] Sharkey, N, "Towards a principle for the human supervisory control of robot weapons," Special Issue on "Investigating the Relationship between Future Technologies, Self and Society," *Politica & Società,* No. 2, May-August 2014.

Speaker's summary

ETHICAL RESTRAINT OF LETHAL AUTONOMOUS ROBOTIC SYSTEMS: REQUIREMENTS, RESEARCH, AND IMPLICATIONS[6]

Professor Ronald Arkin, Georgia Institute of Technology, USA

Robotic systems are now widely present in the modern battlefield, providing intelligence gathering, surveillance, reconnaissance, and target acquisition, designation and engagement capabilities. Limited autonomy is also present or under development in many systems as well, ranging from the Phalanx system "capable of autonomously performing its own search, detect, evaluation, track, engage and kill assessment functions"[7], fire-and-forget munitions, loitering torpedoes, and intelligent antisubmarine or anti-tank mines among numerous other examples. Continued advances in autonomy will result in changes involving tactics, precision, and just perhaps, if done correctly, a reduction in atrocities, as outlined in research conducted at the Georgia Tech Mobile Robot Laboratory (GT-MRL).[8] This paper asserts that it may be possible to ultimately create intelligent autonomous robotic military systems that are capable of reducing civilian casualties and property damage when compared to the performance of soldiers. Thus, it is a contention that calling for an outright ban on this technology is premature, as some groups already are doing.[9] Nonetheless, if this technology is to be deployed, then restricted, careful and graded introduction into the battlefield of lethal autonomous systems must be standard policy as opposed to haphazard deployments, which I believe is consistent with existing international humanitarian law (IHL).

Multiple potential benefits of intelligent war machines have already been declared by the military, including: a reduction in friendly casualties; force multiplication; expanding the battlespace; extending the soldier's reach; the ability to respond faster given the pressure of an ever increasing battlefield tempo; and greater precision due to persistent stare (constant video surveillance that enables more time for decision-making and more eyes on target). This argues for the inevitability of development and deployment of lethal autonomous systems from a military efficiency and economic standpoint, unless limited by IHL.

It must be noted that past and present trends in human behaviour in the battlefield regarding adhering to legal and ethical requirements are questionable at best. Unfortunately, humanity has a rather dismal record in ethical behaviour in the battlefield. Potential explanations for the persistence of war crimes include:[10] high friendly losses leading to a tendency to seek revenge; high turnover in the chain of command leading to weakened leadership; dehumanization of the enemy through the use of derogatory names and epithets; poorly trained or inexperienced troops; no clearly defined enemy; unclear orders where intent of the order may be interpreted incorrectly as unlawful; youth and immaturity of troops; external pressure, e.g. for a need to produce a high body count of the enemy; and pleasure from the power of killing or an overwhelming sense of frustration. There is clearly room for improvement and autonomous systems may help address some of these problems.

---

[6] This summary is an abridged version of Arkin, R C, "Lethal Autonomous Systems and the Plight of the Non-combatant", *AISB Quarterly*, No. 137, July 2013

[7] US Navy, "Phalanx Close-in Weapons Systems", United States Navy Factfile, http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=800&ct=2 (accessed 8/2008)

[8] Arkin, R C, *Governing Lethal Behavior in Autonomous Robots*, Chapman-Hall, 2009.

[9] Notably Human Rights Watch, International Committee on Robot Arms Control (ICRAC) and Article 36.

[10] Boothby, B (Ed.), *Law of War Workshop Deskbook*, International and Operational Law Department, Judge Advocate General's School, June 2000; Danyluk, S, "Preventing Atrocities", *Marine Corps Gazette*, Vol. 8, No. 4, June 2000, pp. 36-38; Parks, W H , "Crimes in Hostilities. Part I", *Marine Corps Gazette*, August 1976; Parks, W H, "Crimes in Hostilities. Conclusion", Marine Corps Gazette, September 1976; Slim, H, *Killing Civilians: Method, Madness, and Morality in War*, Columbia University Press, New York, 2008.

Robotics technology, suitably deployed, may assist with the plight of the innocent non-combatant caught in the battlefield. If used without suitable precautions, however, it could potentially exacerbate the already existing violations by human soldiers. While I have the utmost respect for our young men and women soldiers, modern warfare puts them in situations in which no human being was ever designed to function. In such conditions, expecting strict adherence to the laws of war seems unreasonable and unattainable by a significant number of soldiers.[11] Battlefield atrocities have been present since the beginnings of warfare, and despite the growth of IHL over the last 150 years or so, these tendencies persist and are well documented,[12] even more so in the days of CNN and the internet. The dangers of abuse of unmanned robotic systems in war, such as the Predator and Reaper drones, are well documented; they occur even when a human operator is directly in charge.[13]

Given this, questions then arise regarding if and how these new robotic systems can conform as well as, or better than, our soldiers with respect to adherence to existing IHL. If achievable, this would result in a reduction in collateral damage, i.e. non-combatant casualties and damage to civilian property, which translates into saving innocent lives. If achievable this could result in a moral requirement necessitating the use of these systems. Research conducted in our laboratory[14] focuses on this issue directly from a design perspective. No claim is made our research provides a fieldable solution to the problem, far from it. Rather these are baby-steps towards achieving such a goal, including the development of a prototype proof-of-concept system tested in simulation. Indeed, there may be far better approaches than the one we currently employ, if the research community can focus on the plight of the non-combatant and how technology may possibly ameliorate the situation.

As robots are already faster, stronger, and in certain cases (e.g. Deep Blue, Watson) smarter than humans, is it really that difficult to believe that ultimately they will be able to treat us more humanely on the battlefield than we do each other, given the persistent existence of atrocious behaviour by a significant subset of soldiers? Is there any cause for optimism that this form of technology can lead to a reduction in non-combatant deaths and casualties? I believe so, for the following reasons:

- The ability to act conservatively, i.e. they do not need to protect themselves in cases of low certainty of target identification. Autonomous armed robotic vehicles do not need to have self-preservation as a foremost drive, if at all. They can be used in a self-sacrificing manner if needed and appropriate without reservation by a commanding officer. There is no need for a 'shoot first, ask-questions later' approach, but rather a 'first-do-no-harm' strategy can be adopted instead. They can truly assume risk on behalf of the non-combatant, something that soldiers are schooled in, but which some have difficulty achieving in practice.

- The eventual development and use of a broad range of robotic sensors better equipped for battlefield observations than humans currently possess. This includes ongoing technological advances in electro-optics, synthetic aperture or wall-penetrating radars, acoustics, and seismic sensing, to name but a few. There is reason to believe in the future that robotic systems will be able to pierce the fog of war more effectively than humans ever could.

[11] US Surgeon General's Office, Mental Health Advisory Team (MHAT) IV Operation Iraqi Freedom 05-07, Final Report, 17 November 2006.
[12] For a more detailed description of these abhorrent tendencies of humanity discussed in this context, see Arkin, R C, "The Case for Ethical Autonomy in Unmanned Systems", *Journal of Military Ethics*, 9:4, pp. 332-341, 2010.
[13] Adams, J, "US defends unmanned drone attacks after harsh UN Report", *Christian Science Monitor*, June 5, 2010; Filkins, D, "Operators of Drones are Faulted in Afghan Deaths", *New York Times*, May 29, 2010; Sullivan, R, "Drone Crew Blamed in Afghan Civilian Deaths", *Associated Press*, May 5, 2010.
[14] For more information see Arkin, R C, *Governing Lethal Behavior in Autonomous Systems,* Taylor and Francis, 2009.

- Unmanned robotic systems can be designed without emotions that cloud their judgement or result in anger and frustration with ongoing battlefield events. In addition, "Fear and hysteria are always latent in combat, often real, and they press us toward fearful measures and criminal behaviour."[15] Autonomous agents need not suffer similarly.

- Avoidance of the human psychological problem of 'scenario fulfilment' is possible. This phenomenon leads to distortion or neglect of contradictory information in stressful situations, where humans use new incoming information in ways that only fit their pre-existing belief patterns. Robots need not be vulnerable to such patterns of premature cognitive closure. Such failings are believed to have led to the downing of an Iranian airliner by the USS Vincennes in 1988.[16]

- Intelligent electronic systems can integrate more information from more sources far faster before responding with lethal force than a human possibly could in real time. These data can arise from multiple remote sensors and intelligence (including human) sources, as part of the US Army's network-centric warfare concept and the concurrent development of the Global Information Grid. "Military systems (including weapons) now on the horizon will be too fast, too small, too numerous and will create an environment too complex for humans to direct."[17]

- When working in a team of combined human soldiers and autonomous systems as an organic asset, they have the potential capability of independently and objectively monitoring ethical behaviour in the battlefield by all parties, providing evidence and reporting infractions that might be observed. This presence alone might possibly lead to a reduction in human ethical infractions.

But there are many counterarguments as well. These include the challenge of establishing responsibility for war crimes involving autonomous weaponry, the potential lowering of the threshold for entry into war, the military's possible reluctance of giving robots the right to refuse an order, proliferation, effects on squad cohesion, the winning of hearts and minds, cyber security, proliferation, and mission creep.

There are good answers to these concerns I believe, and are discussed elsewhere in my writings.[18] If the baseline criteria becomes outperforming humans in the battlefield with respect to adherence to IHL (without mission performance erosion), I consider this to be ultimately attainable, especially under situational conditions where bounded morality (narrow, highly situation-specific conditions) applies,[19] but not soon and not easily. The full moral faculties of humans need not be reproduced to attain to this standard. There are profound technological challenges to be resolved, such as effective *in situ* target discrimination and recognition of the status of those otherwise *hors de combat*, among many others. But if a war-fighting robot can eventually exceed human performance with respect to IHL adherence, that then equates to a saving of non-combatant lives, and thus is a humanitarian effort. Indeed if this is achievable, there may even exist a moral imperative for its use, due to a resulting reduction in collateral damage, similar to the moral imperative Human Rights Watch has stated with respect to precision-guided munitions when used in urban settings.[20] This seems contradictory to their call for an outright ban on lethal autonomous robots[21] before determining via research if indeed better protection for non-combatants could be afforded.

---

[15] Walzer, M*, Just and Unjust Wars,* 4th ed., Basic Books, 1977.
[16] Sagan, S, "Rules of Engagement", in *Avoiding War: Problems of Crisis Management* (Ed. George, A), Westview Press, 1991.
[17] Adams, T, "Future Warfare and the Decline of Human Decisionmaking", in *Parameters*, US Army War College Quarterly, Winter 2001-02, pp. 57-71.
[18] E.g. Arkin, R C, *op. cit.,* 2009.
[19] Wallach, W and Allen, C, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, 2010.
[20] Human Rights Watch, "International Humanitarian Law Issues in the Possible U.S. Invasion of Iraq", *Lancet*, 20 February 2003.
[21] Human Rights Watch, "Losing Humanity: The Case Against Killer Robots", 19 November 2012.

How can we meaningfully reduce human atrocities on the modern battlefield? Why is there persistent failure and perennial commission of war crimes despite efforts to eliminate them through legislation and advances in training? Can technology help solve this problem? I believe that simply being human is the weakest point in the kill chain, i.e. our biology works against us in complying with IHL. Also the oft-repeated statement that "war is an inherently human endeavour" misses the point, as then atrocities are also an inherently human endeavour, and to eliminate them we need perhaps to look to other forms of intelligent autonomous decision-making in the conduct of war. Battlefield tempo is now outpacing the soldier's ability to be able to make sound rational decisions in the heat of combat. Nonetheless, I must make clear the obvious statement that peace is unequivocally preferable to warfare in all cases, so this argument only applies when human restraint fails once again, leading us back to the battlefield.

While we must not let fear and ignorance rule our decisions regarding policy towards these new weapons systems, we nonetheless must proceed cautiously and judiciously. It is true that this emerging technology can lead us into many different futures, some dystopian. It is crucially important that we not rush headlong into the design, development, and deployment of these systems without thoroughly examining their consequences on all parties: friendly forces, enemy combatants, civilians, and society in general. This can only be done through reasoned discussion of the issues associated with this new technology. Toward that end, I support the call for a moratorium to ensure that such technology meets international standards before being considered for deployment as exemplified by the recent report from the United Nations Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions.[22] In addition, the United States Department of Defense has recently issued a directive[23] restricting the development and deployment of certain classes of lethal robots, which appears tantamount to a quasi-moratorium.

The advent of these systems, if done properly, could possibly yield a greater adherence to the laws of war by robotic systems than from using soldiers of flesh and blood alone. While I am not averse to the outright banning of lethal autonomous systems in the battlefield, if these systems were properly inculcated with a moral ability to adhere to the laws of war and rules of engagement – while ensuring that they are used in narrow bounded military situations as adjuncts to soldiers – I believe they could outperform soldiers with respect to conformance with IHL. The end product then could be, despite the fact that these systems could not ever be expected to be perfectly ethical, a saving of non-combatant lives and property when compared to the behaviour of soldiers.

We must continue to examine the development and deployment of lethal autonomous systems in forums such as the United Nations and the International Committee of the Red Cross (ICRC) to ensure that the internationally agreed upon standards regarding the way in which war is waged are adhered to as this technology proceeds forward. If we ignore this, we do so at our own peril. A call for a ban on these autonomous systems may have as much success as trying to ban artillery, cruise missiles or aircraft bombing and other forms of standoff weaponry. A better strategy perhaps is to try and control its uses and deployments, which existing IHL appears at least at first glance to adequately cover, rather than a call for an outright ban, which seems unenforceable even if enacted.

Until it can be shown that existing IHL is inadequate to cover this new technology, only then should such action be taken to restructure or expand the law. This may be the case, but unfounded pathos-driven arguments based on horror and Hollywood in the face of potential reductions of civilian casualties seem at best counterproductive. These systems counter intuitively could make warfare safer in the long run for innocents in the battlespace, if coupled with the use of bounded morality, narrow situational use, and careful graded introduction.

[22] Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary, and Arbitrary Execution*, United Nations Human Rights Council, 23rd Session, 9 April 2013.
[23] US Department of Defense, *Directive 3000.09, Autonomy in Weapons Systems*, 21 November 2012.

I believe, however, that we can aid the plight of non-combatants through the judicious deployment of these robotic systems, if done carefully and thoughtfully, particularly in those combat situations where fighters have a greater tendency or opportunity to stray outside IHL. But what must be stated is that a careful examination of the use of these systems must be undertaken now to guide their development and deployment, which many of us believe is inevitable given the ever increasing tempo of the battlefield as a result of ongoing technological advances. It is unacceptable to be 'one war behind' in the formulation of law and policy regarding this revolution in military affairs that is already well underway. The status quo with respect to human battlefield atrocities is unacceptable and emerging technology in its manifold forms must be used to ameliorate the plight of the non-combatant.

Speaker's summary

RESEARCH AND DEVELOPMENT OF AUTONOMOUS 'DECISION MAKING' SYSTEMS

Dr Darren Ansell, University of Central Lancashire, UK

Robotic technology has been around for many years but its ability to appear 'human-like' in its behaviours has, to date, been limited. Some of the best examples of human-like behaviour in robots relate to mobility functions such as balance, movement, climbing and walking or even running. Even with these highly advanced systems, it is still easy for a novice human observer to see the limitations in performance of these systems.

When you consider that a human being uses its senses to take in data about the world, thinks about that data in order to reason, make plans, make decisions and act on those decisions, this is a complex chain of steps that are equally difficult to reproduce in a computer system, implemented in software.

If we take each of the steps in turn from the viewpoint of a computerized robotic system (the robot), the robot must be able to use sensors to sense and receive information about its environment. In some cases the raw performance of sensors for capture of electromagnetic data outperforms human beings ability. Take modern digital cameras for example with extremely high many-mega pixel arrays. These systems can capture a scene with enormous resolution and focus beyond a human's ability. These systems are not infallible though. Errors can be introduced, external interference can add 'artefacts' to imagery and equipment can and does fail. Deliberate external means of disrupting or spoofing the sensing action can also take place.

In any event, having sensed or after having received data from another source, the robot must then make sense of it. This task of extracting information from the data is complex and challenging. As a human being we can look at the scene in front of our eyes and immediately recognize objects (we are in fact 'classifying' objects into their specific types, e.g. chair, table, window, telephone, etc.). It is very challenging for a machine to perform this function accurately. Often image-processing techniques have to be used which may, for example, attempt to classify an object by matching it against a database of many thousands of similar images (looking for correlation). Usually the output from these software programs is a classification accompanied with some form of confidence or error rating. For example the software may report that it is 90% certain that it is observing a blue car.

Given the uncertainties that will arise, the robotic system must then piece together the individual classifications in order to make a diagnosis of what is actually happening. These next steps are where the machine infers knowledge and what is often described in the computer science world as 'creating beliefs.' By combining smaller pieces of information, the machine will create many beliefs and of course, as these beliefs are built upon uncertain source data, they too will be uncertain. A machine may for example, monitor its fuel gauge and the rate at which fuel is being consumed over time. If the fuel level falls quickly, the machine could create a belief that it has a fuel leak. However, if the system that was measuring the amount of fuel remaining was inaccurate, then that belief may be untrue. One possible problem here is that the human designer may fail to incorporate the machines' ability to include a vital piece of information that would have a major impact on the belief being accurate.

The next step a robotic system must take is to plan based upon these classifications and beliefs. The plan may be a simple one such as move left, speed up or slow down. It may be a short-term plan (fractions of a second long), or a longer-term plan that lasts several hours or even days.

The ability of a machine to create a plan varies enormously depending on the action that is to be taken, and what source information is needed to form a feasible (and safe) plan. Errors can be introduced during the creation of the plan, through poor design in the planning software. A machine must make an assessment of how good a proposed future plan is, in order to be able to choose whether or not to action the plan.

Take for example a car satellite navigation system. It may propose a route plan for you to follow based upon its limited knowledge of the road network. If that navigation system does not have access to traffic or road works information, its assessment of the suitability of that route will be incorrect. The machine needs a means of assessing a plan, before acting upon it, in order to predict how good it will be. The fidelity and accuracy with which the plan can be assessed will limit how good the robot performs in the real world.

In some robotic systems, the planning functions generate many alternative plans, and the system must choose between them. This selection process is a form of machine decision making, where some software function must be able to look at the performance predictions for several plans and select one. The selection process too, can be another source of error as the machine may make its selection without full knowledge of, for example, laws and regulations, or fully take into account how the uncertainties have propagated through the various processing stages, compounding the errors. It is common practice for the machine to interact with a human being at this stage, in order to seek authorization to commit to some plans, but this human interaction is only possible when communication links are working.

If the human designer fails to adequately constrain both the planning process and the decision-making process, the system could then commit an unsafe or even unlawful act. Many of the better performing software techniques that are often used for ‘classification’, ‘planning’ and ‘decision making’ in the computer science domain are often unpredictable, and may even use random processes in order to function. In safety-involved or safety-critical systems, such as flight control or weapons release, these software functions do not currently have a route to system certification with regulatory authorities. All aircraft safety-critical software in operation today is entirely predictable (deterministic), i.e. given a certain set of inputs, it will always produce the same output. The same cannot be said for many of the advanced computer science techniques that are enhancing more general robotic technology today.

To achieve the complex sensing and computer-processing stages described above, both the sensor technology and processing technology need to be carried on the host platform. On larger platforms with adequate space and power supplies, this can usually be achieved. On smaller platforms such as small drones and even the micro air vehicles, this is not a feasible option as size, weight and power provision are very limited.

For machines to make reliable decisions consistently in a wide range of scenarios, free from human interaction, there needs to be major leaps in computer science and autonomous systems technology. Today’s systems can only demonstrate reliable, consistent, trusted performance when placed in known environments which are predictable and well understood.

Speaker's summary

CAN AUTONOMOUS WEAPON SYSTEMS RESPECT THE PRINCIPLES OF DISTINCTION, PROPORTIONALITY AND PRECAUTION?

Professor Marco Sassòli, University of Geneva, Switzerland

1. A first preliminary technical assumption: it is possible to keep artificial intelligence under control.

I assume even autonomous weapon systems with artificial intelligence, though capable of learning, cannot do what the human beings who created them do not want them to do – or that it is at least possible to limit their autonomy in this regard. This must be done because they are not addressees of the law. This also implies that it must be always possible to predict what they do; otherwise humans cannot remain responsible for their conduct and only human beings are addressees of international humanitarian law (IHL).

2. A second preliminary technical assumption: it is not technically impossible to develop robots that are as able as the average soldier to make distinctions.

Furthermore, I cannot exclude that it may one day be possible to construct autonomous weapon systems which are capable of perceiving the information necessary to comply with IHL – this appears to me to be the main challenge – and then to apply IHL to that information. For the time being and pending evidence of revolutionary technical developments, it may be wise to limit the use of autonomous weapons to situations in which no proportionality assessment is needed and where the enemy consists of declared hostile forces in high intensity conflicts (although even then they must be able to sense who surrenders or is otherwise *hors de combat* before they may be deployed). I guess it will still take some time before they can be used in counterinsurgency operations.

3. A legal preliminary question: must targeting decisions involve subjective judgements?

As for the law, my understanding is that IHL on targeting does not require subjective value judgements, which machines are unable to make, but depends on an objective assessment of facts. Several authors and military manuals mention that its application involves a subjective determination. The question is, however, whether this is simply a description of the unfortunate reality, while the determination should ideally be as objective as possible, or whether this is a normative proposition and the determination should be subjective. For the application of the proportionality principle, for instance, I think that both for human operators and for autonomous weapons it would be desirable if a formula for such a calculation, together with indicators of the elements that should/should not be taken into account, could be agreed upon. Obviously the determination must be made on a case-by-case basis (and modelling and determining indicators for the infinite variety of possible situations will be a perhaps insurmountable difficulty for producers of genuinely autonomous weapons), but I do not see why it should be “subjective.”

4. Advantages of autonomous weapon systems

If my two technical assumptions and my understanding of IHL are correct, an attack executed by autonomous weapons would have many advantages in terms of distinction, proportionality and precautions over an attack directly executed by human beings. Only human beings can be inhuman and only human beings can deliberately choose not to

comply with the rules they were instructed to follow. To me, it seems easier to expect (and to ensure) a person who devises and constructs an autonomous weapon in a peaceful workplace to comply with IHL than a soldier in the midst of a battlefield or in a hostile environment. A robot cannot hate, cannot fear, cannot be hungry or tired and has no survival instinct. The robot can delay the use of force until the last, most appropriate moment, when it has been established that the target and the attack are legitimate. Robots do not rape. They can sense more information simultaneously and process it faster than a human being can. As the weapons actually delivering kinetic force become increasingly quicker and more complex, it may be that humans become simply too overwhelmed by information and decisions that must be taken to direct them.

The development of autonomous weapons may even lead, because of programming needs, to a clarification of many rules that have so far remained vague and whose protective utility depends upon subjective value judgments. Most arguments of principle against autonomous weapons either do not withstand comparison with other, alternative, means and methods of warfare or they are based upon an erroneous understanding of IHL. There are nevertheless some challenges when applying existing IHL to autonomous weapons, which necessitate agreement on the proper interpretation of IHL by every State using them and between States.

5. Difficulties resulting from the temporal field of application of IHL

Agreement has to be found on the temporal field of application of IHL (beyond Article 36 of Protocol I) to conduct in peacetime which may produce results during armed conflict. I would suggest that IHL applies to all conduct of a State aimed at having effects during an armed conflict. Anyway, a State using a weapon which was programmed in peacetime not to comply with IHL has not taken, as soon as a conflict starts, all feasible precautionary measures to avoid incidental civilian losses. For criminal responsibility, the issue is trickier.

6. Difficulties to apply the proportionality principle

In my view, the greatest difficulty for an autonomous weapon system to apply the proportionality principle is not linked to the evaluation of the risks for civilians and civilian objects but to the evaluation of the military advantage anticipated. While I could imagine a robot sensing the necessary information to evaluate risks for civilians and even to proceed to the necessary evaluation if objective formulas are adopted, the ‘concrete and direct military advantage anticipated’ resulting from an attack against a legitimate target constantly changes according to the plans of the commander and the development of military operations on both sides. Except where no, or clearly negligible, effects upon civilians can be anticipated, a machine, even if perfectly programmed, could therefore not be left alone in applying the proportionality principle, but must be constantly updated about military operations and plans. This is in my view the most serious IHL argument against the even theoretical possibility of deploying genuinely autonomous weapons that remain fully autonomous over considerable periods of time.

7. Difficulties and opportunities in taking feasible precautions

The feasibility of precautions must be understood to refer to what would be feasible for human beings using the machine, not to the options available to the machine. An autonomous weapon could be a means to render certain precautions feasible which would not be so for a soldier. In my view, a consolidated assessment of advantages and disadvantages is admissible on whether an autonomous weapon is as good as an average soldier in respecting IHL, but such an assessment must be made for every attack. For this, parameters must be fixed for comparison with the performance of human beings in attacks. An autonomous weapon would therefore have to make such an determination in relation to the specific circumstances of each attack, and indicate, if necessary, that it cannot execute

that attack (but that, for example, it has to be executed by a human being). States developing, producing and deploying autonomous weapon systems must – and will in their own interest – also take measures to avoid the enemy tampering with the system, using it against them and their civilians.

Important precautions, such as the obligations to verify the nature of the target and the legality of the attack, to choose means and methods avoiding or minimizing incidental effects on civilians, and to respect the proportionality principle are addressed only to "those who plan or decide upon an attack." Some wonder whether or not this means a human being must plan and decide. In my view, all rules of IHL are addressed only to human beings. This does not, however, preclude the human planners and decision makers from being temporally and geographically removed from the attack, as long as they define the parameters according to which the robot attacks, make sure that it complies with them, and that the robot has the necessary information to apply such parameters.

A particularly tricky issue is determining what the obligation to interrupt an attack implies, when it becomes apparent that it is unlawful, when autonomous weapons are used in terms of their sensing capability and ability to change behaviour. In my view this obligation implies that an autonomous weapon system must be constructed to be as good as a human being to perceive changes in the environment.

Speaker's summary

INCREASINGLY AUTONOMOUS WEAPON SYSTEMS: ACCOUNTABILITY AND RESPONSIBILITY

Professor Christof Heyns, University of Pretoria, South Africa

What happens when things go wrong with increasingly autonomous weapon systems? This may happen in the context of armed conflict (where international humanitarian law, or IHL, determines the targeting rules) when someone who has surrendered is shot, or there could be an excessive number of civilian casualties. Likewise, in a law enforcement context (where international human rights law, or IHRL, sets the standards) there could be excessive use of force by the police.

Traditionally, where humans would take the decision to use force, this could lead to prosecutions, disciplinary action or the need to pay compensation. The question arises, however, of what happens where humans do not exercise meaningful human control over the use of force during armed conflict or law enforcement, but delegate it to computers.

The underlying assumption of this question is that autonomous weapon systems are not illegal weapons – that they may be used under certain circumstances. There is of course a view according to which they are illegal weapons under existing law and/or should be declared as such by new law. This is based on arguments, for example, that their use cannot meet the requirements of IHL that protect the lives of civilians (such as distinction and proportionality) in the case of armed conflict, or that they cannot meet the requirements of IHRL that protect those against whom force may be used in the context of law enforcement (such as necessity and proportionality).

It has also been argued that delegating (or 'outsourcing') decisions over life and death to machines is inherently wrong, whether this is done in conformity with the formal requirements of IHL or IHRL, or not. What is at stake here is not just the protection of the lives of those mentioned, but also the human dignity of anyone at the receiving end of the use of such autonomous force (including combatants and suspected perpetrators who may otherwise lawfully be targeted). To use such weapons under any circumstances would be illegal and any use should lead to accountability. These weapons should also be banned formally because they violate the public conscience.

However, assuming they are not illegal weapons and may be used under certain circumstances – or that there is a grey area where it is not clear whether they are illegal weapons because, for example, we are dealing with 'increasingly' autonomous weapons –- at one point or another things may go wrong if they are used. There may be a malfunction; the machines may learn things they were not supposed to learn; there could be other unexpected results. Normally humans are held accountable on the basis of the control they exercised in making decisions, but humans are by definition out of the loop where machines are used that take autonomous, and in many cases unpredictable, decisions. It clearly makes no sense to punish a machine for its autonomous decisions. The question arises whether there will be an accountability vacuum.

This will not be acceptable because it will mean that the underlying values – the protection of humanitarian values and the rights to life and dignity – are in effect rendered without protection. War crimes are not crimes if there cannot be prosecution. It is, for example, a component of the rights to life and dignity that in cases of violation there will be accountability. If such an accountability vacuum is a necessary component of the use of

autonomous weapon systems, this will be a further reason why they should be regarded as illegal weapons, in addition to the targeting considerations mentioned above.

The obvious response in assessing accountability is to shift attention from the immediate loop of the targeting decision to what may be called the 'wider loop.' Autonomous weapon systems work on the basis of programs that are manufactured, acquired and activated by human beings. The very definition of an autonomous weapon system is that, once activated, it can select and engage targets without further human intervention. There will always be a human in the 'wider loop' who has activated (and manufactured, etc.) the system.

However – and this is the main point I want to make – the extent to which those people can be held responsible for the actions of autonomous weapon systems is far from clear. The scenario where any such person would say 'the machine did it' is easy to imagine. The argument would be that the machine took its own decisions, which are unpredictable, not because computers act randomly, but because the environments in which they operate are so complex that all possible interactions between the system and the surroundings cannot be foreseen. Even if accountability can in theory be assigned by law, in practice those who activate autonomous weapon systems may find a lot of sympathy from judges and others who have to assess their conduct. The danger of an accountability gap, in law or in practice, remains.

Accountability requires, as a starting point, knowledge of the facts by those applying the norms. In this regard, autonomous weapon systems may in fact offer some advantages over systems where humans take the decisions. Every move of such high-tech equipment is certain to be monitored and recorded, and will be available to those in charge. However, the additional question is to what extent do both the law and legal processes allow humans in the 'wider loop' to be held responsible?

One question in this regard is: **who** can potentially be held accountable?

Clearly, if someone anywhere in the 'wider loop' acts in malice and, for example, programs a machine to cause a disaster or violate standing orders, that person can be held accountable. Nevertheless, what about other cases, where things go wrong in the course of 'normal, authorized usage'?

One possibility is the individual commander or operator who deploys the system. *Mens rea* is required and it is not clear to what extent such a person, if he or she is given the authority to use autonomous weapon systems, can be expected to 'foresee the unforeseeable' if the system is used within the confines provided. Can they really be expected to understand the details of the technology? Those accused under such circumstances are very likely to point a finger to another and say "I was authorized; I have very little knowledge of these systems and was relying on those with better knowledge to evaluate the risks." Command responsibility may appear analogous, but in its current form it deals with responsibility for the actions of humans over whom a commander exercises control.

States are responsible for wrongful acts under international law that are attributable to them, but this obviously does not extend to criminal responsibility. They may be held accountable for human rights violations and may be required to cease unlawful actions and required to pay compensation. So far, however, this has not often happened for IHL violations. However, given the role of States in deciding which weapons to acquire, and the obligation of weapons review, this form of accountability could play a potentially important role.

What about the programmers? In this regard, one of the problems is that each one of them is normally involved in developing one aspect of the programme only, which may be used for a range of purposes, also non-military ones. Manufacturers and suppliers have a more

comprehensive role and product liability, and even corporate criminal responsibility, may in some cases play a role, but so far it is largely untested in this domain.

It should be noted, however, that assigning responsibility to non-human entities such as States and corporations entails a very different form of responsibility compared to the case where humans are held accountable. The underlying goals and thus also the effects of accountability which apply where humans are brought to book – which may for example include retribution – may not be the same where non-human entities are at stake.

Moreover, if responsibility becomes too diffused, and is shared by a host of human and non-human entities, it may lose its practical effect. If everyone is responsible, no one is responsible.

It could also be asked against what benchmarks the performance of autonomous weapon systems should be measured, if they were used? Is it sufficient if they do a little bit better than humans in terms of targeting, as some have suggested?

It seems clear that at least marginally enhanced performance will be required. Article 57 of Additional Protocol I (API) to the Geneva Conventions requires the minimization of civilian casualties in attack. If autonomous weapon systems cannot do as well as humans, they should not be deployed. Likewise, IHRL requires all force used in law enforcement to be graduated.

I would argue that while autonomous weapon systems cannot be required to be perfect, they will in practice be held to standards that are significantly higher than those posed for humans. As a matter of law we hold medical specialists – and it seems to me technology in general – to higher standards. As a matter of fact there is likely to be significant outrage when machines kill civilians. Moreover, the potential impact of increased autonomous killing on human dignity demands that there will have to be very good reasons for their deployment.

Responsibility is thus uncertain. The question must be asked: what can be done to limit the situation where the issue arises and to ensure that where it does, accountability is effective?

There are a number of dangers that must be guarded against:

- Automation bias – the tendency of humans to defer to computers, including in contexts where computers are ill-suited to take decisions, for example where value judgments are at stake.

- The tendency to use weapons outside their specified boundaries. The recent experience with drones has shown that unmanned systems in general can easily be deployed in areas without a nexus to armed conflict, while the more permissive targeting rules of IHL are invoked to justify their use in ways that are impermissible in terms of IHRL.

To conclude: without clear accountability for their use and inherent risks, it will be irresponsible and, I would argue, unlawful to use autonomous weapon systems. Accountability follows control. As a result there should be a positive duty on all States to ensure that meaningful human control is exercised over each attack or use of force. It should be a priority for the international community to develop guidelines on what is meant by meaningful human control, and how to enforce it in practice.

It is only realistic, however, to recognize that the question of exactly what constitutes meaningful human control will be a contested issue. Even if weapons that do not comply with these standards are banned, there will always be a grey area, and the question will arise as to how to deal with responsibility in the case of weapons with increasing autonomy that approach the level of autonomous weapon systems.

This brings to the fore the importance of weapons review, in order to ensure that accountability can take place within a well-regulated environment. In this context I want to emphasize the importance of the Article 36 API weapons review and the need to exercise additional care as far as autonomous weapon systems are concerned – this could take the form of placing limitations on the circumstances under which they may be used, such as time and space requirements or limitations, provisions on the training and licensing of operators, and regulations on the circumstances under which increasingly autonomous weapons may be traded. It may, furthermore, become necessary to introduce an equivalent of the Article 36 review procedure at the domestic level, to cover autonomous weapon systems that are used in law enforcement.

In addition to the above, there is also the question of accountability of States to the international community. This requires transparency. If the international institutions that are watchdogs in this area give an unqualified go-ahead for the use of autonomous weapon systems, States will see this as a green light to use them in an increased range of situations – allowing States and individuals alike to point the finger when things go wrong and say that the humanitarian and human rights watchdogs of the world considered the matter – at meetings like these – and did not find any problem with the use of these weapons. In considering responsibility for the use of autonomous weapon systems, we also need to consider our own responsibility as those who influence and take decisions on these matters.

Speaker's summary

ETHICAL ISSUES RAISED BY AUTONOMOUS WEAPON SYSTEMS

Dr Peter Asaro, The New School, USA

My presentation on the ethical issues surrounding autonomous weapon systems sought to refocus and reframe the central moral questions, and their relation to the central legal questions. In part I aim to move the discussion to the role of moral reasoning in targeting and firing decisions, and in part I aim to understand the role of morality within the law, and how morality might shape the development of new laws.

Much of the legal discussion on autonomous weapon systems has focused on whether their use is permissible in international humanitarian law (IHL). The debate has focused on whether these systems can meet the requirements of distinction and proportionality, and whether they might undermine human legal responsibility. From the perspective of existing law, it becomes a technological question of whether such systems might be developed which could be sufficiently discriminate and proportionate in their application of violent force. If so, then these systems would be legal. If not, then these systems would be prohibited.

It is important to note the difference between law and morality for two reasons. First, while the law may tell you that a certain act is permissible under the law, it does not tell you whether that act is moral. While IHL may tell us that it is lawful to kill an enemy combatant, this does not necessarily mean we should kill a particular enemy combatant in a particular situation. Just because you can do something does not mean that you should. Second, if we conclude there is a need for new law or regulation, the source of that law ought to be based in firm moral foundations. This might, in part, involve the elaboration or clarification of existing law, but the development of new law should also have moral guidance. The Martens Clause serves not only to extend protections which may not be specifically defined under IHL, but also points to the “principles of humanity and the dictates of the public conscience” as a valid source for new IHL.

Where does the law run out? What are the situations that might arise through the use of autonomous weapon systems where the law is indifferent to immorality, or the systematic violation of human rights and dignity? We are thus confronted with the question of whether there might be autonomous weapon systems which could pass Article 36 review, and conform to existing IHL in some uses, that would actually be problematic in their general use and widespread adoption? What would make their use problematic, and how might we create regulations to avoid those problems?

This, I believe, is the point at which we must move beyond the legal question to the moral question. It is necessary to avoid legalism here – the view that morality and legality are equivalent. If we are to extend the existing body of IHL, the best place to start is with moral reflection. As a global community, we may hold different moral values and theories, but historically we have been able to come to broad-based agreement on certain moral issues. The UN Declaration of Human Rights (1948) is a significant example. It is valuable to consider autonomous weapon systems from a variety of moral theories and perspectives. I believe that taking such an approach leads us to a convergence of various views and perspectives on the conclusion that the best option for regulating autonomous weapon systems is a prohibition on their use.

So what are the moral questions that arise beyond the existing law with regard to permitting autonomous weapon systems to kill human beings? Insofar as we define autonomous

weapon systems as systems which are capable of selecting targets and directing violent force against those targets without meaningful human control, then autonomous weapon systems are significantly different than other weapons. While there are weapons in use in which particular objects and people are not consciously targeted by the operator of the weapon, including artillery. However, there is still a human who makes a targeting decision, with an understanding of the weapon and its potential impact.

In giving over the responsibility to make targeting decisions to machines, we fundamentally change the nature of the moral considerations involved in the use of violent force. While it is claimed that a human will write the computer program the autonomous weapon system will follow, there is no way for a programmer to anticipate every situation and circumstance of the use of force, or the moral values and military necessity for the use of force in those future instances. At best, such a program might approximate the choices made by humans.

To be clear, a programmed system would not be conducting the moral reasoning of human beings, at least for the foreseeable future. Moral reasoning requires an ability to view a situation from multiple and conflicting perspectives, weigh incomparable values against each other, including the significance and value of human life, and to choose a course of action that one can take responsibility for. While moral reasoning is a challenge for humans, and they often fall short, it is impossible for algorithms.

This is why I have previously argued[24] that the requirements in IHL, including Article 57 of Additional Protocol I (AP I), which requires commanders to take all reasonable precautions to protect civilians, also implies that they do this sort of moral reasoning about the potential harm to civilians and weigh this against the military necessity of a given attack, for each and every attack. This constitutes an implicit requirement for moral reasoning about the principles of distinction, proportionality and military necessity, before an attack. Of course, the authors of AP I did not envision a machine or algorithm making such a decision. Given that this technological possibility is now before us we ought to consider whether it is indeed acceptable or not, and, if not, how we might ensure that there is meaningful human control over each and every use of violent force.

In practice, people tend to employ a variety of moral frameworks in making moral decisions and resolving moral dilemmas. In the Western philosophical tradition, the major moral theories are utilitarianism, Kantian and rights-based theories, sentimentalism, and virtue ethics. While I do not believe that any one of these alone is the 'correct' theory, they each capture a compelling element of human moral reasoning. Moreover, I believe an analysis of the question of the morality of autonomous weapon systems according to each theory points to the need and desirability of a prohibition on their use.

We have heard repeatedly that the utilitarian analysis might actually support the development and use of autonomous weapon systems, insofar as they might reduce civilian casualties and violations of IHL. While this might be possible in theory, in order to make the utilitarian moral argument it is necessary to show not only the possible benefit of the technology, but also the probability of that outcome. In other words, if we permit the unregulated development of autonomous weapon systems, what is the likelihood that we will only have autonomous weapon systems which meet Article 36 reviews, and which actually achieve better protection of civilians? Moreover, how would we know this, or measure it? And can we rely upon the existing Article 36 review process – given its largely non-specific character and the fact that formal review processes are only employed by a handful of countries – as sufficient to guarantee that the widespread development and global proliferation of autonomous weapon system technologies will be exclusively or predominately to the benefit of civilians. I believe that this outcome is unlikely, but it is ultimately an empirical question. To

[24] **Error! Main Document Only.**Asaro, P, "On Banning Autonomous Lethal Systems: Human Rights, Automation and the Dehumanizing of Lethal Decision-making," Special Issue on New Technologies and Warfare, *International Review of the Red Cross,* 94(886) Summer 2012, pp. 687-709.

the degree that we are uncertain about the consequences of developing this technology, and in the light of the clear potential negative outcomes of military and political instability, lack of legal accountability, risks of unintentional attacks and hacking, the burden of proof is on the proponents of these systems to establish processes that would guarantee this positive outcome. And even then, this would still not address the other moral aspects of the issue.

From the Kantian or rights-based perspective, the analysis focuses on the human right to life, and the conditions under which that right might be deprived. Under IHL, it is lawful to kill any and all enemy combatants, except for those who are *hors de combat*. However, it does not follow that you *should* kill someone just because it is legal to do so. This goes beyond questions of sympathy, empathy or mercy. The military objective of an attack may not require the killing of an individual enemy combatant. From a moral perspective, it is preferable to spare that individual's life while achieving the military objective. More fundamentally, the question of whether it is morally and legally justifiable to deprive a human of their right to life requires another rational or moral agent to make that determination. That is, we cannot accept that an algorithmic system, which lacks situational, social, cultural and moral understanding, could be capable of determining whether its rules for determining who is a 'lawful target' or 'enemy combatant' provide sufficient justification for the use of violent force.

In this sense, there is no way for an autonomous weapon system, without meaningful human control, to ensure that the killing it does is not arbitrary. As such, any killing done without meaningful human control would be, by definition, arbitrary. And while that might be lawful under existing IHL, it might be fundamentally immoral, and a threat to human rights.

Some have characterized the moral objection to autonomous weapon systems as the 'ugh' factor. In moral theory we call this sentimentalism – the idea that we have moral sentiments and feelings which we might articulate as rules, but which nonetheless drive our decisions even when unarticulated. It is an empirical question whether the 'public conscience' as pointed to by the Martens Clause actually contains this particular sentiment. Do the majority of people in the world actually feel that it is wrong to be killed by a machine? If this is true, then it would seem to be a strong basis for extending the law with a new regulation.

It is also possible to articulate these not-yet-articulated sentiments. Indeed, if we look historically at the concept of 'superfluous injury' we find something similar. Though it was not yet explicit in IHL, there was a broadly felt sentiment that certain forms of injury were morally bad and militarily unnecessary. The authors of IHL sought to articulate this sentiment and developed the concept of superfluous injury to do so. I believe that we can do something similar by articulating the concept of meaningful human control in a manner that captures this moral sentiment in the public conscience.

According to virtue theory, there are a set of virtues which we should seek to exhibit and live up to in our actions. When these virtues come into conflict with each other, such as loyalty and honesty, we demonstrate our moral character in choosing one over the other. In the case of autonomous weapon systems, there is a real danger that by removing human control over targeting decisions and the use of force, we are similarly removing the moral responsibility of the operators of those systems. As such, we are precluding operators from exercising their moral character and warrior virtues. Similarly, as automated decision-making moves from tactical decisions to strategic or even political decisions, we risk alienating military commanders and military leaders from their leadership virtues and human responsibility over the course of human events more generally.

Finally, there are questions of human dignity involved in considering what it means to allow machines to take human lives without meaningful human control. On the one hand, the law already acknowledges that it matters how and why people are killed, and not simply whether or not they are killed. Whether civilian casualties are merely tragic accidents or war crimes depends on *mens rea*, and the intentions of those who order attacks and carry them out.

More generally, we have a different sense of the justice of a given death based on the reasons behind it. We can better understand a death when we can understand the reasons and circumstances behind it. In other words, how we kill, even in war, does matter.

Moreover, what does it mean for the value of human life if we allow automated machines to take human lives? As we give over the decisions of life and death to technological systems, are we diminishing the value of human life? Slavery and torture are evil and unjust not just because of the immediate suffering of the individual subjected to them, but of the collective diminishing of human value that they represent. So too, we may decide that it diminishes human dignity and the value of human life by giving over the authority to take human lives to automatic machines.

Speaker's summary

AUTONOMOUS WEAPON SYSTEMS AND ETHICS

Dr Peter Lee, University of Portsmouth, UK

This presentation covered four areas: (1) The Perception Problem; (2) An ethical framework; (3) Autonomous weapons, politics and war; and (4) The moral calculus of oversight and accountability (past, present and future).

1. The Perception Problem

As of March 2014, such a thing as a fully autonomous, cognisant, self-reasoning weapon system does not exist (simple systems like Improvised Explosive Devices are excluded from this discussion, even though they may act autonomously – or at least automatically). While scientific, technological and theoretical advances bring such a possibility closer, attitudes to – and perceptions of – possible autonomous weapons are already being shaped by perceptions of existing remotely piloted drones and their current military application.

To illustrate, the expert meeting participants were asked the question, "According to the Bureau of Investigative Journalism (a journalism organisation critical of drones), how many civilians were killed by CIA drone strikes in Pakistan in 2013?" and given a range of possible answers from 0 to 1,000+. The entire range of potential answers was offered by different participants, with only three offering the 'correct', though disputed answer, of between 0 and 10. The *perception* of drone activities is widely disputed and understood even amongst a particularly well informed gathering of State representatives and independent experts.

Since there are currently no existing autonomous weapons, ethical considerations concerning their possible future existence is necessarily shaped by two things: perceptions of the nearest equivalents (drones, currently remotely piloted), and the influence of science fiction and the Hollywood effect. Consequently, any ethical analysis is subject to contestation and lacking demonstrable 'facts'. Despite these challenges, however, consideration can and should be given to hypothetical possibilities.

2. An ethical framework

Peter Asaro has already set out a number of ethical approaches that can be applied to the challenge of autonomous weapons: utilitarianism, deontology, just war and more. Each contains advantages and disadvantages that do not need to be explored again at this point. I suggest that whatever moral framework, or combination of competing moral frameworks, is used needs to recognize that ethical choices are constituted in numerous overlapping and disputed discourses: technological; military; science fiction; politics and war; and truth/knowledge claims.

3. Autonomous weapons, politics and war

Technical and military situation: Put briefly, there is no current autonomous cognisant, self-reasoning technology that is weaponized for use in war. As a result, it is difficult to interrogate the ethical conceptions of proportionality and discrimination in applied force when the accuracy of discriminating between legitimate and non-legitimate targets is not known. Further, when the level of force that can potentially be applied is also not known then the question of proportionality of force is reduced to guessing or speculation.

Science fiction: From H.G. Wells's book *War of the Worlds* at the dawn of the twentieth century to the more recent *Terminator* and *I, Robot* films, science fiction not only entertains but can normalize, in some, expectations and attitudes that are not reflected in lived experience. Ethical analysis should address what is currently and foreseeably possible and try to avoid the realm of the imagined. When human lives are potentially at stake, unverified scare stories can end up having unintended consequences: if projected fears are not realized it may result in a diminution of future ethical concerns at a crucial point in the development of autonomous weaponry.

Politics and war: Autonomous weapons, like earlier weapons and current remotely piloted drones, and any asymmetric military advantage it may confer, do not inevitably lead to victory in war, counter-insurgency or anti-terrorism activities. US military asymmetric advantage did not lead to victory in Vietnam and NATO's current operations in Afghanistan do not promise long-term success despite the military advantages held. If Clausewitz's maxim that "War is a continuation of politics by other means" is applied to autonomous weapons, the claims and counter-claims to truth surrounding the technology must inevitably be applied to diplomatic, economic and other forms of engagement between contending parties, both State actors and sub-state actors.

4. The moral calculus of oversight and accountability (past, present and future)

I suggest that in the application of military force, moral responsibility is attributed in relation to an individual's freedom of thought and action, which will have implications for autonomous weapons. Consider the hierarchy of moral responsibility for the British World War II bomber offensive against Germany; current Reaper (drone) operations in Afghanistan; and potential future deployments of autonomous weapon system (other morally responsible contributors like intelligence officers, scientists and weapons manufacturers are not included for simplicity):

WWII bombing 'kill chain'	**Reaper 'kill chain'**	**Autonomous weapon 'kill chain'**
1. Winston Churchill (Prime Minister & Defence Minister) 2. Charles Portal (Chief of the Air Staff) 3. Arthur Harris (AOC-inC Bomber Command) 4. Bomber squadron commanders 5. Bomber crews	1. Prime Minister 2. Defence Minister 3. Chief of Defence Staff 4. Chief of the Air Staff 5. AOC 1 Group 6. Reaper squadron commanders 7. Reaper crews	1. Prime Minister 2. Defence Minister 3. Chief of Defence Staff 4. Chief of the Air Staff 5. AOC 1 Group 6. Autonomous weapon system squadron commanders ~~7. Autonomous weapon system~~

The circumstances of World War II prompted changes in attitudes and developments in international humanitarian law that will hopefully result in those events not being repeated. The autonomous weapon or weapon system in Column 3 cannot be an ethical actor since it does not have, and in my view will never have, the capacity for ethical calculation that human beings possess (even if that capacity is not always used when it could be). In Column 2, Reaper crew members actively make ethical decisions when striking a target and can be held accountable for their actions. If they were replaced by an autonomous system, that moral responsibility would not disappear. It would be added to the moral responsibility already existing at every level of the kill chain, military and political.

Conclusion and recommendation

Ethical assessments of autonomous weapons are currently as limited as the technological, military and political assumptions they are based upon. I recommend that ongoing ethical debate take place *as developments occur* in the coming years.

PART III: BACKGROUND PAPER BY THE INTERNATIONAL COMMITTEE OF THE RED CROSS, 17 March 2014[1]

EXECUTIVE SUMMARY

Recent technological developments and military operational demands are leading to increasing levels of autonomy in weapon systems, notably in mobile unmanned systems. For the purposes of this paper, an 'autonomous weapon system' is one that can independently select and attack targets, with or without human oversight. The term refers to weapon systems that are fitted with autonomous functions of acquiring, tracking, selecting and attacking targets ('critical functions'). The term excludes weapon systems that select and attack targets under remote control by a human operator.

Autonomous weapon systems currently in use are various fixed weapon systems in stationary roles, including ship and land-based defensive weapon systems and fixed gun systems, with different levels degrees of human oversight. Developers envisage increasing autonomy in mobile unmanned weapon systems which are already in use in a range of military operations. Overall drivers for increasing autonomy in weapon systems include decreasing the personnel requirement for operating unmanned systems, reducing their reliance on communications links, and increasing their performance and speed of decision-making.

As regards autonomy in targeting, existing weapon systems, including manned aircraft and defensive weapon systems, are fitted with rudimentary capabilities to distinguishing simple objects in 'low clutter' , relatively predictable and static environments. While automatic target recognition technology is becoming more sophisticated, there appear to remain significant challenges in developing technology that could make much finer distinctions in more complex, 'cluttered' and dynamic environments. Delegating all targeting and firing decisions to a weapon system would require a very high level of confidence that it will not make the wrong assessment, in view of the dangerous consequences of failure.

It is well accepted that new technologies of warfare must abide by existing international law, in particular IHL. The use of an autonomous weapon system would need to comply with the fundamental rules of IHL, i.e. the rules of distinction, proportionality and precautions in attack. Significant challenges lie in programming machines to distinguish objects, and even more so persons, in particular to distinguish civilians from combatants and persons *hors de combat* from active combatants. Assessments required by the rules of proportionality and precautions in attack are also highly context-dependent and require weighing up many qualitative variables in an unpredictable battlefield. Programming an autonomous weapon system to respect these rules in a dynamic environment, notably where combatants and military objectives are comingled with civilians and civilian objects (a defining feature of contemporary armed conflicts), would appear to remain a formidable challenge.

In addition, there are many questions regarding how to ensure accountability for acts performed by autonomous weapons that amount to violations of IHL, be it through individual criminal responsibility or State responsibility.

[1] **This is an edited version of the background paper circulated to participants in advance of the ICRC's expert meeting**. It was drafted by Neil Davison, Science Adviser, Nathalie Weizmann, Legal Adviser, and Isabel Robinson, Legal Attaché from the Arms Unit in the Legal Division at the ICRC.

These challenges underscore the crucial importance of carrying out rigorous legal reviews of new autonomous weapon systems or of modifications of existing weapon systems involving autonomy in critical functions, to determine whether their use may violate international law in some or all circumstances. Such reviews would need to be based on thorough testing and evaluation of the weapon's functions and capabilities. However, questions about the degree of predictability and reliability of autonomous weapon systems could affect the ability to carry out effective testing and an accurate legal review.

Even if technology could one day allow an autonomous weapon system to be fully compliant with IHL in a dynamic environment, there remain some fundamental questions: Would the dictates of public conscience be prepared to yield to a machine the decision to take human life on the battlefield? And if it is agreed that some human control or oversight is required in such life and death decisions, what kind and degree of human control would be meaningful? These fundamental questions, as well as the above-mentioned technical, military and legal issues, call for thorough discussions at national and international levels.

INTRODUCTION

This background paper was prepared as an aid for discussions at the ICRC's expert meeting on *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects*, held from 26 to 28 March 2014. It does not necessarily represent institutional positions of the ICRC.

The paper is organised in three parts:
- Part A: Autonomy in weapon systems
- Part B: Applying international humanitarian law; and
- Part C: Ethical and societal concerns, and the dictates of public conscience

PART A: AUTONOMY IN WEAPON SYSTEMS[2]

1. The rise of robotic weapon systems

During the past 15 years there has been a dramatic increase in the use of robotic systems by military forces, in particular various unmanned systems that operate in the air, on land, and on – or under – the sea.[3] Although the gradual increase in sophistication of military machinery and in physical distance of military personnel from the battlefield is a process as old as war itself, recent developments in robotics and computing combined with military operational demands are influencing the development and use of novel robotic systems with increasing levels of autonomy.

For example, in 2001 the United States military had around 50 unmanned air systems. In 2013, some twelve years later, it had around 8,000 unmanned air systems and 12,000 unmanned ground systems.[4] In the air at least, these systems are expected to take over most of the functions currently carried out by manned aircraft since they offer the potential for increased capability and 'persistence' with reduced risks to military personnel at potentially lower financial cost in the future.[5]

Increasingly these unmanned systems are being adapted, designed and used to deliver weapons, therefore becoming weapon systems.[6] While only a few countries are known to have used unmanned air systems to deliver weapons, estimates of the number of countries to have developed these systems range from 50 to over 80.[7] A significant number of these countries now possess unmanned air systems adapted or designed to deliver weapons. Other robotic weapon systems, used at fixed positions and generally not described as

[2] Part A of this paper relies primarily on sources from the United States and the United Kingdom. This is a reflection of the current availability of public information on the topic and not a preference for sources from particular countries.
[3] P. W. Singer, *Wired for War* (New York: Penguin, 2009,); P. Rogers, 'Unmanned Air Systems: The Future of Air & Sea Power?' *Institut Français des Relations Internationales (IFRI) Focus Stratégique,* No. 49 (January 2014); J. Gertler, *U.S. Unmanned Aerial Systems,* Congressional Research Service (January 2012), p. 3; US Department of Defense, *Unmanned Systems Integrated Roadmap FY2013-2038* (2013), p. 19, http://www.defense.gov/pubs/DOD-USRM-2013.pdf.
[4] P. W. Singer, 'The Predator Comes Home: A Primer on Domestic Drones, their Huge Business Opportunities, and their Deep Political, Moral, and Legal Challenges,' *Brookings*, 8 March 2013, http://www.brookings.edu/research/papers/2013/03/08-drones-singer; US Department of Defense, Defense Science Board, *Task Force Report: The Role of Autonomy in DoD Systems* (19 July 2012), p. 78, http://www.acq.osd.mil/dsb/reports/AutonomyReport.pdf.
[5] UK Ministry of Defence, Development, Concepts and Doctrine Centre*, The UK Approach to Unmanned Aircraft Systems, Joint Doctrine Note 2/11* (30 March 2011) paras 102 -103, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/33711/20110505JDN_211_UAS_v2U.pdf.
[6] US Department of Defense, *Unmanned Systems Integrated Roadmap 2013, supra* note 3, p. 73.
[7] S. Joshi and A. Stein, 'Emerging Drone Nations,' *Survival*, Vol. 55:5 (2013), pp. 53–78; R. O'Gormann and C. Abbott, *Remote control war: Unmanned combat air vehicles in China, India, Israel, Iran, Russia and Turkey*, Open Briefing (20 September 2013), p. 2, http://issuu.com/openbriefing/docs/remote_control_war; G. Taylor, 'U.S. intelligence warily watches for threats to U.S. now that 87 nations possess drones,' *The Washington Times*, 10 November 2013, http://www.washingtontimes.com/news/2013/nov/10/skys-the-limit-for-wide-wild-world-of-drones/?page=all.

'unmanned', have been in use for many years as defensive systems for ships or ground installations.

Various weaponized unmanned systems are already deployed in the air, on the ground and at sea and the range of their envisioned operations is broad. For example, in the air, it is foreseen that unmanned air systems may be used for air-to-air combat, electronic warfare and suppression of air defences, in addition to their current use for targeted strikes. Unmanned ground systems may be used for armed reconnaissance and combat operations, as well as in law enforcement-type operations. At sea unmanned underwater and surface vehicles may be used to lay and destroy mines, as well as other armed operations.[8] The development and acquisition of weaponized unmanned systems are likely to expand globally in the coming years.

2. The evolution towards autonomy

Closely linked to the adoption of more robotic systems are their increasing levels of autonomy.[9] Autonomy in unmanned systems, for example, is seen by some as critical for future conflicts.[10] Indeed in 2010 the US Air Force identified the potential offered by autonomous military systems as the "single greatest theme to emerge" from their study of new technologies, concluding that autonomy could offer "potentially enormous increases in its capabilities."[11]

Most armed unmanned air systems, for example, are already highly automated in different functions – such as take-off, landing and navigation – which enable the weapon system to be "operated rather than continuously piloted."[12] This level of autonomy is expected to increase in order to reduce the workload for the operator and increase the time available for them to focus on decision-making.[13]

Such automatic systems still require a certain level of human control when changes are needed in order for them to adapt to external circumstances in their environment. However, one focus of research and development is in autonomous unmanned systems that can make decisions and react to circumstances without human intervention.[14] Indeed the ultimate aim of some research is for operators to give an unmanned system its objectives and for the system to then adapt autonomously to fulfil these, even perhaps through autonomous collaboration and data sharing among multiple unmanned systems.[15]

Research on autonomous systems in general encompasses developments in information and communications technology, machine learning, artificial intelligence, and cognitive and behavioural sciences, which are active fields in the commercial and academic sectors, and so access to emerging technologies is widespread.[16] However, it is not at all clear whether such technologies are likely to be sufficiently sophisticated to enable autonomous functioning in highly complex decision-making, such as the process of selecting and attacking targets with weapon systems.

---

[8] US Department of Defense, *Unmanned Systems Integrated Roadmap 2013, supra* note 3, p. 24.
[9] Ibid., pp.15-16; US Department of Defense, Defense Science Board, *Task Force Report, supra* note 4; UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, pp. 5-4.
[10] US Department of Defense, *Unmanned Systems Integrated Roadmap 2013, supra* note 3, p. 67.
[11] US Air Force, *Report on Technology Horizons: A Vision for Air Force Science & Technology During 2010-2030*: *Volume 1* (15 May 2010), p. 105, file://gva.icrc.priv/DfsRoot/Users/IROB/My%20Documents/Downloads/ADA525912.pdf.
[12] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, pp. 6-7.
[13] Ibid.
[14] US Department of Defense, *Unmanned Systems Integrated Roadmap 2013, supra* note 3, p. 68.
[15] US Air Force, *Report on Technology Horizons, supra* note 11, p. 23.
[16] US Department of Defense, *Unmanned Systems Integrated Roadmap 2013, supra* note 3, p. 13.

For now, when unmanned systems are used to deliver weapons, the decision to fire a weapon, and the specific target to fire at, are still taken by a person and not a machine.[17] However, many of the functions leading up to that point are already automated and delegated to sensors and computer systems. For example, some armed unmanned air systems have automated systems for take-off and landing, navigation, response to a lost communication link, and to a lesser degree, elements of target acquisition.[18] In theory, it would only be one step further to allow a mobile unmanned system to fire a weapon without human intervention, but one that would entail a technological leap in capability while simultaneously raising significant legal and ethical questions.[19]

In fact there are some robotic fixed weapon systems, as opposed to mobile unmanned systems, for which this step has already been taken. A number of countries have for many years operated weapon systems that have autonomous modes that are used to defend ships or ground installations from air or sea-borne threats such as rockets, mortars and aircraft. (See section 4.1 for further details). These types of weapon system can select and engage targets without human intervention although, when in such a mode, there is the capability for humans to intervene to override the system.

These fixed weapon systems are currently used in fairly narrow roles and circumstances, for example operations against specific targets in so called low clutter environments where the terrain is simple. Mobile unmanned weapon systems, on the other hand, feature in a wider range of military operations. However, while developers envisage increasing autonomy for the latter, the challenge of autonomy in 'cluttered', complex environments is far greater.[20]

3. Defining autonomous weapon systems

3.1 Robots and autonomy

It is clear that the current discussion of autonomy in weapons systems primarily relates to robotic systems. **Robots**[21] are generally understood to be machines that follow a sense-think-act paradigm; they gather information with sensors, process this information with a computer or other information processing capability, and then use actuators to interact with the physical world.[22] Patrick Lin *et al* provide a useful summary of what might be considered as robotic military systems:

> "Most robots are and will be mobile, such as vehicles, but this is not an essential feature; however, some degree of mobility is required, e.g., a fixed sentry robot with swivelling turrets or a stationary industrial robot with movable arms. Most do not and will not carry human operators, but this too is not an essential feature; the distinction becomes even more blurred as robotic features are integrated with the body. Robots can be operated semi- or fully-autonomously but cannot depend entirely on human control: for instance, tele-operated drones such as the Air Force's Predator unmanned aerial vehicle would qualify as robots to the extent that they make some

[17] Ibid., p. 24.
[18] P. W. Singer, 'The Predator Comes Home,' *supra* note 4; UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, p. 2-3.
[19] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, p. 5-4.
[20] P. W. Singer, *Wired for War, supra* note 3, pp. 126-128.
[21] '**Robot**: machine capable of carrying out a complex series of actions automatically; especially one programmable by a computer,' *Oxford English Dictionary.*
[22] P. Lin, P. G. Bekey and K. Abney, *Autonomous Military Robotics: Risk, Ethics, and Design*, Ethics + Emerging Sciences Group at California Polytechnic State University (2008), p. 4, http://ethics.calpoly.edu/ONR_report.pdf; J.L. Chameau, W. Ballhaus and H. Lin (eds), *Emerging and Readily Available Technologies and National Security — A Framework for Addressing Ethical, Legal, and Societal Issues*, National Research Council, US National Academies (Washington: The National Academies Press, 2014), pp. 3-1 and 3-2.

> decisions on their own, such as navigation, but a child's toy car tethered to a remote control is not a robot since its control depends entirely on the operator."[23]

Autonomy in machines can be understood as the capacity of a robot, following activation, to operate without any external control in some or all areas of its operation for extended periods of time.

The literature on military robotic systems and autonomy tends to distinguish three main categories of control and automation: (1) **remote controlled** or tele-operated systems, which are controlled directly by a remote operator; (2) **automated** systems (also called **semi-autonomous**), which can act independently of external control but only according to a pre-defined set of programmed rules; and (3) **autonomous** systems, which can act without external control and define their own actions albeit within the broad constraints or bounds of their programming and software.

In general, automated or semi-autonomous systems tend to be distinguished from autonomous systems by virtue of their level of adaptability and decision-making in relation to their external environment. Semi-autonomous systems are pre-programmed to carry out actions with little adaptability to their external environment.[24] Truly autonomous systems would be able to make decisions that define their actions and adapt to their environment based on pre-programmed rules or boundaries.[25] This does not mean that an autonomous system has fully independent thought and action since it always operates within the human-designed limits of its software algorithm.[26]

Autonomy is sometimes discussed as a continuum; from remote controlled to automated and then to autonomous.[27] However, this may not be a useful distinction when considering systems, including weapon systems, which may incorporate both remote controlled and autonomous operation for different functions.[28] For example, although some armed unmanned air systems are often described as 'remote controlled' when in fact many of their functions are certainly automated and may even have a certain level of autonomy – such as in take-off and landing, navigation, pre-planned response to a specific event, and even some aspects of target acquisition.[29]

Therefore, for a discussion of autonomous weapon systems, it may be useful to focus on **autonomy in critical functions** rather than autonomy in the overall weapon system. Here the key factor will be the level of autonomy in functions required to select and attack targets (i.e. critical functions), namely the process of target acquisition, tracking, selection, and attack by a given weapon system. Indeed, as discussed in parts B and C of this paper, autonomy in these critical functions raises questions about the capability of using them in accordance with international humanitarian law (IHL)[30] and raises concerns about the moral acceptability of allowing machines to identify and use force against targets without human involvement.

Following this logic, a weapon system that is remotely controlled by (a) person(s) for all aspects of the targeting and firing process would be excluded from a discussion of

---

[23] P. Lin, et al., *Autonomous Military Robotics, supra* note 22, p. 4.
[24] US Department of Defense, *Unmanned Systems Integrated Roadmap 2013, supra* note 3, p. 66.
[25] US Department of Defense, Defense Science Board, *Task Force Report, supra* note 4, p. 1; US Department of Defense, *Unmanned Systems Integrated Roadmap 2013, supra* note 3, pp. 66-67.
[26] US Department of Defense, Defense Science Board, *Task Force Report, supra* note 4, pp. 1 and 21.
[27] US Department of Defense, *Unmanned Systems Integrated Road Map FY2011-2036*, (2011), p. 44, http://www.acq.osd.mil/sts/docs/Unmanned%20Systems%20Integrated%20Roadmap%20FY2011-2036.pdf, p. 46, describing four levels of autonomy.
[28] US Department of Defense, Defense Science Board, *Task Force Report, supra* note 4, p. 4.
[29] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, p. 2-3.
[30] This is without prejudice to issues and concerns that may arise under other bodies of law.

autonomous weapon systems from an IHL perspective. Conversely, if a weapon system has, or is developed to have, autonomy in the targeting and firing process then it should be included in the discussion.

3.2 Existing definitions of autonomous weapon systems

There is no agreed definition of an autonomous weapon system although various definitions proposed by different countries and organisations share similar themes. The US Department of Defence provides three:

> **Autonomous weapon system**: "A weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation."[31]
>
> **Human supervised autonomous weapon system**: "An autonomous weapon system that is designed to provide human operators with the ability to intervene and terminate engagements, including in the event of a weapon system failure, before unacceptable levels of damage occur."[32]
>
> **Semi-autonomous weapon system**: "A weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator."[33]

The UN Special Rapporteur's report to the Human Rights Council on autonomous weapon systems – or 'Lethal Autonomous Robots' – provides another:

> **Autonomous weapon system**: "Lethal Autonomous Robotics (LARs) refers to robotic weapon systems that, once activated, can select and engage targets without further intervention by a human operator. The important element is that the robot has an autonomous 'choice' regarding selection of a target and the use of lethal force."[34]

Human Rights Watch has also provided some definitions according to the level of human input and supervision in selecting and attacking targets:

> **Human-in-the-Loop Weapons**: "Robots that can select targets and deliver force only with a human command";
> **Human-on-the-Loop Weapons**: "Robots that can select targets and deliver force under the oversight of a human operator who can override the robots' actions";
> **Human-out-of-the-Loop Weapons**: "Robots that are capable of selecting targets and delivering force without any human input or interaction."[35]

---

[31] US Department of Defense, *Autonomy in Weapon Systems, Directive 3000.09* (21 November 2012), http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf.

[32] Ibid.

[33] Ibid., p. 14. Semi-autonomous weapon system includes: (1) 'Semi-autonomous weapon systems that employ autonomy for engagement-related functions including, but not limited to, acquiring, tracking, and identifying potential targets; cueing potential targets to human operators; prioritizing selected targets; timing of when to fire; or providing terminal guidance to home in on selected targets, provided that human control is retained over the decision to select individual targets and specific target groups for engagement;' and (2) 'Fire and forget" or lock-on-after-launch homing munitions that rely on TTPs [tactics, techniques, and procedures] to maximize the probability that the only targets within the seeker's acquisition basket when the seeker activates are those individual targets or specific target groups that have been selected by a human operator.'

[34] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns*. UN General Assembly, A/HRC/23/47 (9 April 2013), para. 38, http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf.

[35] B. Docherty, *Losing Humanity: The Case Against Killer Robots,* Human Rights Watch (November 2012), p. 2, http://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf.

> **Autonomous weapon system**: "The term 'fully autonomous weapon' refers to both out-of-the-loop weapons and those that allow a human on the loop, but that are effectively out-of-the-loop weapons because the supervision is so limited."[36]

The ICRC has also raised some general definitions in its 2011 report on IHL and challenges in contemporary armed conflicts:

> **Automated weapon system**: "An automated weapon or weapons system is one that is able to function in a self-contained and independent manner although its employment may initially be deployed or directed by a human operator. … Although deployed by humans, such systems will independently verify or detect a particular type of target object and then fire or detonate."[37]
>
> **Autonomous weapon system**: "An autonomous weapon system is one that can learn or adapt its functioning in response to changing circumstances in the environment in which it is deployed. A truly autonomous system would have artificial intelligence that would have to be capable of implementing IHL." [38]

Common to all the above definitions is the inclusion of weapon systems that can independently select and attack targets, with or without human oversight. This includes both weapon systems that can adapt to changing circumstances and 'choose' their targets and weapon systems that have pre-defined constraints on their operation and potential targets or target groups. However, the distinction between autonomous and automated (weapon systems is not always clear since both have the capacity to independently select and attack targets within the bounds of their human-determined programming. The difference appears only to be the degree of 'freedom' with which the weapon system can select and attack different targets.

Also common to all these definitions is the exclusion of weapon systems that select and attack targets only under remote control by a human operator. This would exclude current armed unmanned air systems (i.e. 'drones') since targeting and firing is carried out remotely by a human operator. However, it should be noted that if existing remote controlled weapon systems have, or are developed to have, the capability to independently select and/or attack targets (with or without human supervision) then they would become *de facto* semi-autonomous or autonomous weapon systems.

As mentioned, there is not always a clear line between 'automated' and 'autonomous' weapon systems, and some of the questions and issues raised by autonomous weapons – including legal questions – are also raised by automated weapon systems.[39] For the purposes of this paper, the term 'autonomous weapon systems' refers to weapon systems for which critical functions (i.e. acquiring, tracking, selecting and attacking targets) are autonomous.

---

[36] Ibid., p. 2.
[37] ICRC, *International Humanitarian Law and the challenges of contemporary armed conflicts*, Official working document of the 31st International Conference of the Red Cross and Red Crescent (28 November - 1 December 2011), p. 39, http://www.icrc.org/eng/assets/files/red-cross-crescent-movement/31st-international-conference/31-int-conference-ihl-challenges-report-11-5-1-2-en.pdf.
[38] Ibid. See also: ICRC, *Autonomous weapons: States must address major humanitarian, ethical challenges, FAQ* (9 February 2013), http://www.icrc.org/eng/resources/documents/faq/q-and-a-autonomous-weapons.htm.
[39] ICRC, *International Humanitarian Law and the challenges of contemporary armed conflicts* (2011), *supra* note 37, pp. 39-40.

4. Autonomy in existing weapon systems

This section examines levels of autonomy in weapon systems already in use as well as some that are under development. It is not intended to be exhaustive but rather gives an impression of relevant weapon systems operating in fixed positions as well as mobile unmanned systems on the ground, in the air and at sea.

Some examples of relevant weapon systems are provided in associated footnotes. However it should be noted that there is relatively little openly available information with which to assess the technical characteristics and mechanism of operation for these weapon systems, including the degree of autonomy used in selecting and attacking targets.

4.1 Fixed weapon systems

Current weapon systems with the highest degree of autonomy are various fixed weapon systems in stationary roles as opposed to mobile unmanned systems. These include ship and land-based defensive weapon systems and fixed gun systems (sometimes referred to as sentry guns) with different levels of human oversight.

Many countries currently use autonomous weapon systems for defence of ships or ground installations against rockets, mortars, missiles, aircraft and high-speed boats.[40] Many of these weapons have autonomous modes that carry out target selection and attack automatically under the overall supervision of human operators. According to the US Department of Defence definition, they are classed as "human supervised autonomous weapons"[41] and, for example, they are not permitted to select humans as targets, although they can attack manned vehicles (such as aircraft or small boats).[42]

Fixed anti-personnel weapons designed to guard specific sites are apparently being developed to have increasing levels of autonomy. Some have automatic modes whereby it is claimed they can automatically select and attack targets that have been detected by on-board sensors without further human intervention.[43] When used in these modes they might

[40] *Examples include:*

- **Aegis Weapon System**; a ship-based system combining radar to automatically detect and track targets with various missile and gun systems, http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=200&ct=2.
- **Patriot** surface-to-air missile system; a missile defence system that automatically detects, and tracks targets before firing interceptor missiles, http://www.raytheon.com/capabilities/products/patriot/.
- **Phalanx Close-in Weapon System**; a ship-based 20 mm gun system that autonomously detects, tracks and attacks targets. A related ship-based weapon system, the **SeaRAM Close-in Weapons System**, combines the Phalanx's technology with an 11 tube missile launcher. A variant in the very early stages of testing is the **Laser Weapon System**, which combines the Phalanx with a high-energy laser weapon, http://www.navy.mil/navydata/fact_print.asp?cid=2100&tid=487&ct=2&page=1; http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=456&ct=2.
- Other 'Close-in Weapon Systems' include the **Type 730** and **Type 1030 Close-in Weapon Systems** and the **Goalkeeper Close-in Weapon System**, https://www.thalesgroup.com/en/content/goalkeeper-close-weapon-system.
- **Counter Rocket, Artillery, and Mortar System**; a land-based fixed weapon system that employs the same technology as the Phalanx Close-in Weapon System to target and attack rockets, artillery and mortars, http://www.msl.army.mil/Pages/C-RAM/cram.html.
- **Iron Dome**; a ground based air defence system which automatically selects targets and fires interceptor missiles, http://www.rafael.co.il/Marketing/186-1530-en/Marketing.aspx.
- **NBC MANTIS** (Modular, Automatic and Network-capable Targeting and Interception System); an automated ground based air defence system using 35 mm guns to automatically target rocket, artillery and mortars, http://www.rheinmetall-defence.com/en/rheinmetall_defence/public_relations/news/archive_2012/aktuellesdetailansicht_4_2560.php.

[41] US Department of Defense, *Directive 3000.09, supra* note 31; US Department of Defense, *Unmanned Systems Integrated Roadmap 2013, supra* note 3, p. 24.

[42] US Department of Defense, *Directive 3000.09, supra* note 31: 'Human-supervised autonomous weapon systems may be used to select and engage targets, with the exception of selecting humans as targets, for local defense to intercept attempted time-critical or saturation attacks for: (a) Static defense of manned installations; (b) Onboard defense of manned platforms.'

[43] *Examples include:*

- **Samsung SGR-A1**; a sentry gun system fitted with a 5.56 mm machine gun and a grenade launcher. Apparently it has two modes, both of which feature some level of autonomous target selection and attack. In a semi-automatic mode the weapon

also be classed as "human supervised autonomous weapon systems" or simply "autonomous weapon systems."

4.2 Ground weapon systems

The two main potential military uses of unmanned ground systems are 1) for accessing areas that are inaccessible or too dangerous for humans; and 2) for use as weapon systems. When used as the latter their perceived advantages include providing significant force multiplication capability.[44]

Various unmanned ground systems have been fitted with weapons to enable, at minimum, remote operation but also potentially a certain level of autonomy.[45] Some of the these robotic ground systems are already used for other purposes – such as bomb disposal – some are purpose designed as weapon systems, and others combine vehicles with mounted fixed weapon systems (see section 4.1). Two main foci of current efforts to increase autonomy in unmanned ground vehicles are on improving their navigation over complex terrain and their ability to fire weapons within the rules of engagement.[46] The latter is seen as requiring significant advances in the development of autonomous systems and human-machine interaction in general.

In addition to overall force multiplication,[47] the deployment of autonomous ground combat systems to fight against enemy ground combat systems is seen as a potential future scenario in the long term, depending on advances in technology.[48]

---

system identifies a target and sends a signal back to a remote operator to authorise the attack. In automatic mode the weapon automatically selects and attacks targets, http://www.stripes.com/machine-gun-toting-robots-deployed-on-dmz-1.110809.

- **DoDamm aEgis** and **Super aEgis**; sentry gun systems fitted with various rifle, machine gun, grenade launcher and missile weapons. According to the manufacturer the weapon systems all have autonomous detection and tracking of targets and the modes for manual and autonomous firing, http://www.dodaam.com/eng/sub2/menu2.php.
- **Sentry-Tech Stationary Remote Controlled Weapon Station**; a weapon system with machine guns and grenade launcher that is operated by remote control, although there have been reports of aims to develop autonomous functions, http://www.rafael.co.il/Marketing/396-1687-en/Marketing.aspx.

[44] US Department of Defense, Defense Science Board, *Task Force Report, supra* note 4, p. 92.

[45] *Examples include:*

- **Guardium**; a 4-wheeled unmanned ground combat system. The manufacturer claims that it navigates autonomously and can be fitted with various remotely operated weapon systems but that a 'fully-autonomous' version is planned, http://g-nius.co.il/unmanned-ground-systems/guardium-mk-iii.html. A related system is the **Avantguard**, http://g-nius.co.il/unmanned-ground-systems/avantguard.html.
- **Athena**; an unmanned combat vehicle operating on land and water that is fitted with the aEgis fixed weapon system (see note 93), http://www.dodaam.com/eng/sub2/menu2_1_6.php.
- **Unmanned Ground Combat Vehicle PerceptOR Integration** (or 'Crusher'); a ground system with autonomous capabilities that will be able to carry an 8,000 lb payload, including weapons, http://www.cmu.edu/cmnews/extra/060428_crusher.html.
- **Talon SWORDS**; a ground robot that can be fitted with a variety of weapons including a rifle, machine guns, grenade and rocket launchers, which are operated by remote control, http://www.qinetiq.com/media/news/releases/Pages/talon-robots-demo-swords-at-dsei.aspx.
- **Modular Advanced Armed Robotic System**; a ground robot with a machine gun and a 40 mm grenade launcher on a rotating turret, which are operated by remote control, https://www.qinetiq-na.com/products/unmanned-systems/maars/.
- Various armed ground robots tested with 7.62 machine guns in late 2013 by the US Army, including: **Modular Advanced Armed Robotic System** (as above); **710 Warrior**; **Remote Armed Maneuver Platform**; **Common Remotely Operated Weapon Station**; **Protector**; **Mobile Armed Dismount Support System**; **Carry-all Modular Equipment Landrover**; and **Atlas**, a humanoid robot, http://www.armytimes.com/article/20131012/NEWS/310140003/UGV-models-face-off-over-firepower-load-carrying.
- Other ground robots tested with weapons, including: **Packbot**, which is normally used for bomb and improvised explosive device (IED) disposal but has also been tested with shotguns, Taser electric-shock weapons, and claymore mines, http://news.cnet.com/8301-10784_3-9736833-7.html; and **Andros**, which is also used in bomb disposal but can be fitted with a shotgun or a Taser electric-shock weapon, http://www.northropgrumman.com/Capabilities/Remotec/Applications/Pages/Swat.aspx.

[46] US Department of Defense, *Unmanned Systems Integrated Roadmap 2013, supra* note 3, pp. 87, 94.

[47] P. McLeary, 'US army Studying Replacing Thousands of Grunts with Robots,' *Defense News,* 20 January 2014, http://www.defensenews.com/article/20140120/DEFREG02/301200035/US-Army-Studying-Replacing-Thousands-Grunts-Robots.

4.3 Air weapon systems

a) Unmanned air systems

During the past 10-15 years some unmanned air systems have been adapted and used to fire weapons.[49] There are now a large number of unmanned air systems capable of being armed, which have been acquired or are under development. It is estimated that around 20 countries have developed or acquired this capability although only a few have so far used them to fire weapons in armed conflict. [50]

While these weapon systems as a whole are becoming more autonomous with the automation of functions such as take-off and landing, and navigation, as far as is known the decision to select and attack targets is retained by a human operator instructing the weapon system by remote control. However, it appears that these weapon systems do use some level of automation for aspects of target acquisition and tracking.[51]

Many existing unmanned air systems were primarily developed for reconnaissance and intelligence purposes and later adapted to carry weapons and carry out attacks. However a new generation of these weapon systems are being designed as combat systems, and may also make use of more automated and autonomous features.[52]

b) Missiles, homing munitions, sensor-fused munitions, and loitering munitions

Another broad category of weapons that already feature a level of autonomy in target selection and attack are various munitions with active guidance systems for target selection and attack. They generally do not require additional external guidance once fired and are sometimes referred to as 'fire and forget' munitions.

Missiles by definition all have some form of on-board guidance system but for this discussion we will include only those designed to have some capability to independently determine the target after firing.[53] Generally these types of missiles fly to a pre-programmed location after which they use in-built sensor and information processing capabilities, such as active radar and millimetre wave radar, to determine their target. Some self-destruct or deactivate if a suitable target is not found.

---

[48] US Department of Defense, Defense Science Board, *Task Force Report, supra* note 4, p. 94.

[49] *Examples include:*

- **MQ1-Predator**; a medium-altitude, long-endurance unmanned air system carrying two Hellfire air-to-surface anti-tank missiles, http://www.af.mil/AboutUs/FactSheets/Display/tabid/224/Article/104469/mq-1b-predator.aspx.
- **MQ-9 Reaper**; a medium to high-altitude, long endurance unmanned air system, which is larger than the Predator and carries more weapons; up to 16 Hellfire missiles or a mixture of 500 lb and 250 lb laser guided bombs, http://www.af.mil/AboutUs/FactSheets/Display/tabid/224/Article/104470/mq-9-reaper.aspx.

[50] M. Daly (ed.), *IHS Jane's. All the World's Aircraft: Unmanned 2012-2013* (IHS Global Limited, 2012); R. O'Gormann and C. Abbott, *Remote control war, supra* note 7.

[51] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, p. 2-3; P. W. Singer, 'The Predator Comes Home,' *supra* note 4.

[52] *Examples include:*

- **X-47B**; a developmental unmanned combat air system, which is currently undergoing testing by the US Navy. It can take off, fly, and land autonomously while overseen by an operator with a computer, http://www.nytimes.com/2013/05/13/opinion/drones-and-the-rivalry-between-the-us-and-china.html
- **Taranis**; a developmental unmanned combat air system undergoing testing by the UK Air Force, which will have autonomous flight capability and perhaps increased automation in its targeting systems, http://www.ft.com/cms/s/0/0ef5939e-cf57-11e2-be7b-00144feab7de.html#axzz2vHo7UNUn

[53] *Examples include:*

- **Air launched missiles** such as: AIM-120 Advanced Medium-Range Air-to-Air Missile; AGM-114 Hellfire Longbow air-to-surface missile; Brimstone air-to-surface missile; and R-77 air-to-air missile.
- **Cruise missiles** such as: AGM-158 Joint Air-to-Surface Standoff Missile; BrahMos; and Tactical Tomahawk.
- **Anti-ship missiles** such as: Harpoon; Naval Strike Missile; and YJ-82 (or C-802) anti-ship missile.
- **Portable missiles** such as FGM-148 Javelin anti-tank missile.

Also included are other types of homing munitions and sensor fused munitions that employ on-board systems for target selection and attack.[54] Excluded from this discussion, however, are munitions that use guidance systems (e.g. laser, GPS) to attack only specific pre-programmed and pre-selected specific targets, since they do not have the ability to actively select an alternative target.[55]

Loitering munitions are another category of munitions with autonomy in target selection and attack.[56] Here the lines between unmanned combat air vehicles and missiles become increasingly blurred. Certain loitering munitions are essentially unmanned air systems that integrate a weapon as part of their construction. They are expendable weapon systems rather than acting as a platform from which to launch a weapon. These types of weapons have been in development for some years but only relatively few weapon systems have come into use.[57]

4.4 Maritime weapon systems

Unmanned maritime systems of various sizes and functions are also being developed as weapons platforms. There are two main types: unmanned surface vehicles,[58] whose operations include anti-submarine warfare and surface warfare; and unmanned underwater vehicles,[59] which may be used for anti-submarine warfare, laying mines and other types of attack.[60]

Autonomous capability for unmanned underwater vehicles is of particular interest due to the difficulties of communication underwater and the size of potential operating areas. These vehicles may operate without human interaction for many days.[61] Unmanned underwater vehicles with a level of autonomous function are already used to detect mines, map

[54] *Examples include:*
- **Artillery projectiles** such as: BONUS 155 mm; M982 Excalibur 155 mm; and SMArt 155.
- **Other munitions** such as: BLU 10, an air dropped sensor fused munition; and CBU-97 Sensor Fuzed Weapon, a cluster bomb.

[55] Even this distinction may not be strictly accurate since munitions that use a heat-seeking signature to determine their target may have a crude ability to select between different types of targets.

[56] *Examples include:*
- **Harpy**; an armed unmanned air vehicle incorporating an explosive warhead. It is an anti-radiation weapon that employs active guidance to autonomously detect electromagnetic emissions from radar equipment before attacking, http://www.iai.co.il/2013/16143-16153-en/IAI.aspx.
- **Harop**; a longer range version of Harpy.
- **Low-Cost Autonomous Attack System**; a now cancelled developmental system using 'swarms' of multiple loitering munitions to autonomously search for and attack targets over a wide area.

[57] Defense Update, 'Loitering Autonomous Weapons,' *Defense Update*, January 2007, http://defense-update.com/features/du-1-07/armedUAVs_8.htm

[58] *Examples include:*
- **Protector**, a unmanned surface vehicle in use by several countries, which can be fitted with a **mini-Typhoon** remote controlled weapon system combining a machine gun and 40 mm grenade launcher. The **Typhoon** and **min-Typhoon** are also used on manned ships and are similar to ship-board fixed weapon systems (see section 4.1 of this paper). According to the manufacturer they can be operated remotely or in human supervised autonomous mode, where they can carry out automatic target selection and attack, http://www.rafael.co.il/marketing/SIP_STORAGE/FILES/1/941.pdf.

[59] *Examples include:*
- **Anti-Submarine Warfare Continuous Trail Unmanned Vessel**, a developmental underwater system designed to autonomously track submarine targets, http://www.darpa.mil/Our_Work/TTO/Programs/Anti-Submarine_Warfare_%28ASW%29_Continuous_Trail_Unmanned_Vessel_%28ACTUV%29.aspx.

[60] US Department of Defense, Defense Science Board, *Task Force Report, supra* note 4, p. 85; A. Martin, 'U.S. Expands Use Of Underwater Unmanned Vehicles,' *National Defense*, April 2012, http://www.nationaldefensemagazine.org/archive/2012/April/Pages/USExpandsUseOfUnderwaterUnmannedVehicles.aspx.

[61] US Department of Defense, Defense Science Board, *Task Force Report, supra* note 4, p. 86.

oceanography and for various commercial purposes.[62] However, weaponized operations are envisaged.[63]

5. Drivers for autonomy

Armed robotic systems, particularly unmanned systems, offer a number of advantages to the users including: force multiplication; reduced risk to military personnel; increased capability over a wider area and deeper in the adversaries' territory; increased persistence in the battlefield; and all this at potentially lower cost.[64]

Increasing autonomy of these systems in general could enhance these advantages. But there are some specific drivers for increasing autonomy in order to reduce the requirement for human control.[65] These include: decreasing the personnel requirement for operating unmanned systems; reducing their reliance on communications links; and increasing their performance and speed of decision-making.[66] These are drivers for increasing autonomy in the weapon system as a whole but they may also be relevant to increasing autonomy in the critical features of targeting and firing.

5.1 Decreased personnel requirements

One driver for increased autonomy in unmanned air systems is as a means to decrease the number of personnel needed to operate them.[67] Autonomous functions might reduce the personnel requirements in a number of ways, such as enabling one person to control multiple unmanned systems and automating the processing and analysis of the data they collect.[68] With increased autonomy, unmanned ground systems might also be used to substitute or expand ground forces (i.e. force multiplication).[69] Overall reductions in personnel requirements may also have the added attraction of reducing costs.

If increasing autonomy in various functions decreases the overall workload for the operator then some argue this will increase the time available for the operator to focus on supervision and critical decisions such as firing of weapons. However, if a single person is operating multiple armed unmanned systems then this reduced workload benefit could be lost or weakened since their time and attention would necessarily be shared among multiple weapon systems.

5.2 Reduced reliance on communications links

Another driver for overall autonomy is to reduce the dependence on high-speed communication links used to operate robotic systems.[70] Autonomous systems would be able

[62] W. Connors, 'Underwater Drones Are Multiplying Fast,' *Wall Street Journal*, 24 June 2013, http://online.wsj.com/news/articles/SB10001424127887324183204578565460623922952.

[63] E. Whitman, 'Unmanned Underwater Vehicles: Beneath the Wave of the Future,' *US Navy*, http://www.navy.mil/navydata/cno/n87/usw/issue_15/wave.html; N. Hopkins, 'Ministry of Defence plans new wave of unmanned marine drones,' *The Guardian*, 2 August 2012, http://www.theguardian.com/world/2012/aug/02/ministry-defence-plans-unmanned-marine-drones.

[64] G. Marchant, B. Allenby, R. Arkin, E. Barrett, J. Borenstein, L. Gaudet, O. Kittrie, P. Lin, G. Lucas, R. O'Meara, and J. Silbermann, 'International Governance of Autonomous Military Robots,' *Columbia Science and Technology Law Review*, Vol. XII (2011), pp. 272-315.

[65] US Department of Defense, *Unmanned Systems Integrated Road Map 2011, supra* note 27, p. 3.

[66] Ibid., p. vi; UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, pp. 5-10.

[67] US Department of Defense, *Unmanned Systems Integrated Road Map 2011, supra* note 27, p. 44.

[68] Ibid., p. 27; US Department of Defense, *Unmanned Systems Integrated Roadmap 2013*, p. 19, *supra* note 3; J. Gertler, *U.S. Unmanned Aerial Systems, supra* note 3, p. 3.

[69] US Department of Defense, *Unmanned Systems Integrated Roadmap 2013, supra* note 3, p. 68.

[70] Ibid., p. 75; UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, pp. 3-13.

to continue with their operations if communication links were degraded[71] and they would enable operations in areas where communication is very limited or not possible, such as in caves and under water.[72]

Wireless communications are also susceptible to intentional disruption such as hacking, 'jamming' and 'spoofing', which could render unmanned systems inoperable, for example, if they require remote operation for most functions. In 2012 researchers demonstrated the latter technique using a 'fake' GPS communications signal to re-direct the path of an unmanned air system.[73] More sophisticated hacking of could enable complete takeover of their operation, including potential release of weapons.

Even where communications links are not disrupted, there is a problem of limited bandwidth to transmit sensor data and video feeds back to operators. As unmanned systems become more numerous the stresses on available bandwidth are only likely to increase.[74]

Autonomy again is an attractive way of mitigating these problems. However, any such autonomous weapon systems would effectively be beyond human oversight unless communication links were re-established for critical decisions such as selecting targets and firing weapons.

5.3 Increased performance and speed of decision-making

The limits of human capabilities in areas such as quantitative information analysis, speed of decision-making, and reaction time provide some attraction to increased autonomy in military systems, including weapon systems. It has even been suggested that the increasing speed, complexity, and information overload of warfare may become too difficult for humans to direct.[75] Interest stems from the potential for autonomous systems to perform faster than humans in certain tasks, in particular rapid decision-making and reaction to situations, which might be translated into a significant capability advantage over adversaries.[76]

A critical caveat, however, is that while machines may perform better and more quickly than humans at some quantitative tasks, there remains a significant technical barrier to their outperforming humans in many qualitative tasks requiring sophisticated human judgement, decision-making and reasoning.

Some observers have noted that other 'human factors', such as the tendency to make mistakes over time, the susceptibility to fatigue and low morale, and the potential for cognition and perception to be impaired by environmental circumstances, are also drivers for the pursuit of autonomous weapon systems.[77] Another perspective is that overall systems for human supervision of autonomous weapon systems need to take into account the limits of human performance.[78] In any case, for the time being the advantages of machine performance over humans may be most applicable to simple repetitive tasks rather than

[71] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, pp. 5-10.
[72] US Department of Defense, *Unmanned Systems Integrated Road Map 2011, supra* note 27, p. 45.
[73] P. W. Singer, 'The Predator Comes Home,' *supra* note 4; BBC News, 'Researchers use spoofing to 'hack' into a flying drone,' 19 June 2012, http://www.bbc.co.uk/news/technology-18643134.
[74] A. Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Surrey, Burlington: Ashgate Publishing, 2009), pp. 37-39.
[75] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, pp. 5-10; UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Global Strategic Trends – Out to 2040* (2010, 4th edition), p. 148, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/33717/GST4_v9_Feb10.pdf.
[76] G. Marchant, et al, 'International Governance of Autonomous Military Robots,' *supra* note 64; US Air Force, *Report on Technology Horizons, supra* note 11, p. 59.
[77] P. Lin, et al, *Autonomous Military Robotics, supra* note 22, p. 1; A. Krishnan, *Killer Robots, supra* note 74, pp. 39-42.
[78] M. Cummings, S. Bruni and P. Mitchell, 'Human Supervisory Control Challenges in Network-Centric Operations,' *Reviews of Human Factors and Ergonomics*, Vol. 6:1 (2010), pp. 34-78.

tasks requiring complex reasoning and decision-making, such as the decision to select and attack targets with a weapon system.

6. Barriers to autonomy

6.1 Autonomy in targeting and firing

Particularly relevant to the legal and ethical discussion about autonomous weapon systems is the degree of autonomy in critical functions of a particular weapon system. While autonomous navigation of an unmanned weapon system may raise less significant issues as it is not directly related to the process of firing weapons, a critical aspect for any analysis will be the degree of autonomy in the targeting and firing process, including acquiring, tracking, selecting and attacking targets.

Automated target recognition systems have been used with existing weapon systems, including manned aircraft and defensive weapon systems. However, current capabilities are rudimentary and limited even in distinguishing simple objects (e.g. a tank) in simple, low clutter environments (e.g. a field).[79]

If automatic target recognition technology were to become more sophisticated and verifiable then it might lead to increases in the overall levels of autonomy in the targeting and firing process for various weapon systems. However, there would be formidable obstacles in developing technology that could make the much finer distinctions that would be required under IHL for any operations in more complex, cluttered environments. It remains questionable whether such technical developments are conceivable.[80]

6.2 Predictability and reliability

Perhaps the greatest technical barrier to the adoption of autonomous systems in general will be ensuring that they function as intended, something that takes on particular significance with any weaponized systems where failures are likely to have dangerous consequences.[81] Since autonomous systems are adaptable (within programmed boundaries) they are necessarily unpredictable. And due to the sheer number of possible situations an autonomous weapon system might be faced with, it is only possible to test an insignificant fraction of these. With existing methods, therefore, it is not possible to verify or certify the operation of autonomous systems for all but the simplest of applications.[82] In addition, testing itself may raise safety issues.[83]

Added to the problem of unpredictability is the inherent problem of reliability present with any complex system. For autonomous weapon systems, failures might result for a wide variety of reasons including: human error, breakdown in human-machine interaction, malfunctions, degraded communications, software failures, cyber-attacks, jamming and spoofing, among others.[84]

---

[79] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, p. 6-1; J.L. Chameau, et al, *Emerging and Readily Available Technologies, supra* note 22, p. 3-2.
[80] N. Sharkey, 'Towards a principle for the human supervisory control of robot weapons,' *Politica & Società* (forthcoming, 2014).
[81] J.L. Chameau, et al, *Emerging and Readily Available Technologies, supra* note 22, p. 3-3; C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 98.
[82] US Air Force, *Report on Technology Horizons, supra* note 11, p. 105.
[83] US Department of Defense, Defense Science Board, *Task Force Report, supra* note 4, p. 64.
[84] US Department of Defense, *Directive 3000.09, supra* note 31; J.L. Chameau, et al, *Emerging and Readily Available Technologies, supra* note 22, p. 3-2.

Problems with predictability and reliability are only likely to increase as autonomous systems become more complex.[85] As a result there will always be a level of uncertainty about the way an autonomous weapon system will interact with the external environment.[86] This raises the related question of whether predictability will be sufficiently high – and uncertainty sufficiently low – to enable an accurate legal review of autonomous weapon systems.

6.3 Malfunctions, accidents and vulnerabilities

Interrelated with questions of predictability and reliability are the risks associated with malfunctions, accidents and vulnerabilities in the use of autonomous weapon systems.[87] Past accidents related to autonomous elements of target selection and firing in existing weapon systems raise concerns over potential risks with weapons that have increasing levels of autonomy.

Examples of accidents include the shooting down of two US fighter jets in 2003, which were mistakenly identified as incoming rockets by a US Patriot air defence system in Iraq. Military researchers analysing these incidents identified two main causes: 'undisciplined automation' whereby too many functions of weapon system are automated without full consideration for how operators can effectively monitor the process and override the system if necessary; and 'automation bias' whereby operators place too much trust in the automated capabilities of the weapon system.[88]

A major problem with attaining effective human oversight of autonomous weapon systems is that there may be insufficient time for a person to take the decision to intervene and then override. One argument for greater autonomy in some weapon systems, therefore, is to remove this difficulty for human interaction with machines within such short time periods. However, delegating all targeting and firing decisions to a weapon system requires a very high level of confidence that it will not make the wrong assessment. And so the counter argument is that the focus needs to be on improving human-machine interaction and supervision.

More recently, in 2007, during a training exercise by the South African National Defence Force a 35 mm anti-aircraft weapon malfunctioned in its automatic mode killing 9 soldiers and seriously injuring 14 others.[89] An enquiry into the incident blamed a mechanical failure[90] but nevertheless it illustrates the dangers when there are malfunctions with weapons that have some level of automation in targeting and firing.

Malfunctions have also been reported with armed unmanned ground robots that were deployed by the US military but not yet used in combat. The Talon SWORDS ground robot, which can be fitted with various weapons including a rifle, machine guns, grenade and rocket launchers,[91] was deployed in Iraq by the US Army in 2007.[92] Operated by remote control,

---

[85] US Department of Defense, *Unmanned Systems Integrated Road Map 2011, supra* note 27, p 50; P. Lin, et al, *Autonomous Military Robotics, supra* note 22, p. 8.
[86] US Department of Defense, Defense Science Board, *Task Force Report, supra* note 4, p. 63.
[87] Ibid., p. 7-8.
[88] J. Hawley, 'Not by widgets alone,' *Armed Forces Journal*, 1 February 2011, http://www.armedforcesjournal.com/not-by-widgets-alone/.
[89] N. Shactman, 'Robot Cannon Kills 9, Wounds 14,' *Wired.com*, 18 October 2007, http://www.wired.com/dangerroom/2007/10/robot-cannon-ki/; N. Shactman, 'Inside the Robo-Cannon Rampage (Updated),' *Wired.com*, 19 October 2007, http://www.wired.com/dangerroom/2007/10/inside-the-robo/.
[90] G. Hosken, 'Army blames gun's maker for Lohatla,' *iol news*, 26 January 2008, http://www.iol.co.za/news/south-africa/army-blames-gun-s-maker-for-lohatla-1.387027#prof.
[91] QinetiQ, 'QinetiQ weaponises its Talon® robots for demo of SWORDS at DSEi,' 11 September 2007, http://www.qinetiq.com/media/news/releases/Pages/talon-robots-demo-swords-at-dsei.aspx.
[92] N. Shactman, 'First Armed Robots on Patrol in Iraq (Updated),' *Wired.com*, 2 August 2007, http://www.wired.com/dangerroom/2007/08/httpwwwnational/.

these robotic ground systems were placed in fixed positions, in contrast to original plans for them to be mobile, and apparently did not fire their weapons.[93] Different reports in 2008 put these limitations down to concerns about reliability of the robots and previous malfunctions during testing.[94]

In addition to malfunctions and accidents, further concerns arise from the potential for intentional interference with autonomous weapon systems. One of the major challenges in the development of software for autonomous systems in general will be protecting it from cyber-attacks both during development and during operations.[95] If an autonomous weapon system were to be hacked and diverted from its normal functioning then the potential consequences could be disastrous.[96]

7. Wider considerations and risks

The potential barriers to development and use of autonomous weapon systems discussed above are only some of the technical issues. There are significant legal, ethical and societal questions that also need to be addressed.[97] These are explored in Parts B and C of this paper.

Some experts have raised the implications for international security and strategic stability should autonomous weapon systems be further developed and integrated into the military structures of States, or even those of non-State armed groups, but these issues are beyond the scope of this paper.

[93] S. Magnusun, 'Future of Armed Ground Robots in Combat Still Debated,' *National Defense*, 15 August 2013, www.nationaldefensemagazine.org/blog/Lists/Posts/Post.aspx?ID=1236; E. Sofge, 'America's Robot Army: Are Unmanned Fighters Ready for Combat?' *Popular Mechanics*, 18 December 2009, http://www.popularmechanics.com/technology/military/robots/4252643.

[94] Popular Mechanics, 'The Inside Story of the SWORDS Armed Robot "Pullout" in Iraq: Update,' *Popular Mechanics,* 1 October 2009, http://www.popularmechanics.com/technology/gadgets/4258963; S. Weinberger, 'Armed Robots Still in Iraq, But Grounded (Updated),' *Wired.com*, 15 April 2008, http://www.wired.com/dangerroom/2008/04/armed-robots-st/

[95] US Department of Defense, Defense Science Board, *Task Force Report, supra* note 4, p. 11.

[96] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 98; B. Docherty, *Losing Humanity, supra* note 35, p. 6.

[97] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, pp. 5-1 to 5-12.

PART B: APPLYING INTERNATIONAL HUMANITARIAN LAW

It is well accepted that new technologies of warfare must abide by existing international law, in particular IHL – also known as the law of armed conflict. Whether or not a new weapon (including an autonomous weapon system) is capable of use in accordance with IHL is determined by assessing the weapon's foreseeable effects based on its design, and its foreseeable use in normal or expected circumstances. The use of such a weapon system would need to comply with the fundamental rules of IHL, i.e. the rules of distinction, proportionality and precautions in attack. In a dynamic environment, respect for these rules largely involves the capacity to exercise qualitative judgments. Based on what is known of current and emerging autonomous technologies, discussed in Part A, achieving such capacity would appear to pose significant challenges. In any case, States would need to carry out rigorous legal reviews of such weapon systems before they are used. If the use of such a weapon system were to result in a violation of IHL, this raises the question of who would be legally responsible for the performance of the weapon.

As discussed in Part A, for the purposes of this paper, an autonomous weapon system is one that can independently select and attack targets. The term refers to weapon systems that are fitted with autonomous features of acquiring, tracking, selecting and attacking targets (critical functions). Parts B and C of this paper will use the term 'autonomous weapon system' to imply autonomy in such functions.

1. Legal reviews of new weapons[98]

It is important to carry out rigorous legal reviews of new technologies of warfare to ensure that they may be used lawfully.

1.1 The requirement to carry out a legal review of new weapons

Article 36 of Additional Protocol I to the Geneva Conventions of 1949 (AP I) states:

> In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.[99]

The aim of Article 36 is to prevent the use of weapons that would violate international law in all circumstances and to impose restrictions on the use of weapons that would violate international law in some circumstances, by determining their lawfulness before they are developed, acquired or otherwise incorporated into a State's arsenal.

All States have an interest in assessing the legality of new weapons, regardless of whether they are party to AP I. A State's faithful and responsible application of its international law obligations would require it to ensure that the new weapons, means and methods of warfare it develops or acquires will not violate its legal obligations when used. [100]

---

[98] This section is based in part on the ICRC's *Guide to the Legal Review of New Weapons, Means and Methods of Warfare* (Geneva: ICRC, 2006), p. 4, http://www.icrc.org/eng/assets/files/other/icrc_002_0902.pdf.

[99] While there is disagreement on whether this obligation restates customary law, M. Schmitt argues that 'the obligation to conduct legal reviews of new means of warfare before their use is generally considered – and correctly so – reflective of customary international law.' M. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics,' *Harvard National Security Journal: Features Online* (2013), p. 28.

[100] See ICRC's *Guide to the Legal Review of New Weapons, Means and Methods of Warfare*, *supra* note 98, p. 4.

Today at least nine States are known to have formal mechanisms or procedures to review the legality of new weapons. The two States that are known to have policies that address autonomous weapon systems – the US and the UK – recognize in their policies the requirement to carry out legal reviews of such weapons.[101]

1.2 Scope of the legal review of new weapons

The legal review applies to weapons and the ways in which they are used, bearing in mind that a weapon cannot be assessed in isolation from its expected method of use. This seems clear from the text of Article 36, which requires the determination of whether the employment of the weapon would "in some or all circumstances" be prohibited by IHL. The use that is made of a weapon can be unlawful in itself, or it can be unlawful only under certain conditions. The legal review should therefore cover, *inter alia*, a weapon that a State develops or acquires, modifications to an existing weapon in a manner that alters its function, as well as the ways in which the weapon is foreseen to be used pursuant to military doctrine, tactics, rules of engagement, operating procedures and counter-measures.[102]

In determining the legality of a new weapon, the reviewing authority must apply existing international law rules which bind the State – be they treaty-based or customary. Article 36 of AP I refers in particular to the Protocol and to "any other rule of international law applicable" to the State. This includes treaty and customary prohibitions and restrictions on specific weapons, as well as the general IHL rules applicable to all weapons, means and methods of warfare. General rules include those aimed at protecting civilians from the indiscriminate effects of weapons and combatants from superfluous injury and unnecessary suffering. As noted in the ICRC's Commentary on the Additional Protocols to the Geneva Conventions, a State need only determine "whether the employment of a weapon *for its normal or expected use* would be prohibited under some or all circumstances. A State is not required to foresee or analyse all possible misuses of a weapon, for almost any weapon can be misused in a way that would be prohibited."[103]

The acceptability of autonomous weapon systems should also be examined according to the principles of humanity and the dictates of public conscience. This refers to the 'Martens Clause', which Article 1(2) of API formulates as follows:

> "In cases not covered by this Protocol or by any other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from dictates of public conscience."[104]

In its advisory opinion on the *Legality of the Threat or Use of Nuclear Weapons*, the International Court of Justice (ICJ) stated that the Martens Clause had "proved to be an effective means of addressing rapid evolution of military technology."[105] The Court also found that the Martens clause represents customary international law. A weapon that is not covered by existing rules of IHL would be considered contrary to the Martens Clause if it is

[101] See Enclosure 3 setting out Guidelines for Review of Certain Autonomous or Semi-Autonomous Weapon Systems, US Department of Defense, *Directive 3000.09, supra* note 31; UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, para. 503.
[102] ICRC's *Guide to the Legal Review of New Weapons, Means and Methods of Warfare*, *supra* note 98, section 1.1.
[103] ICRC, *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (The Netherlands: Martinus Nijhoff Publishers, 1987), para. 1469 (emphasis added).
[104] The Martens Clause was first articulated in the preamble to the 1899 Hague Convention (II), and subsequently appeared in the preamble to the 1907 Hague Convention (IV). It also appears in Article 1(2) of AP I.
[105] ICJ, *Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion,* ICJ Reports 1996, para. 78, http://www.icj-cij.org/docket/files/95/7495.pdf.

determined per se to contravene the principles of humanity or the dictates of public conscience.[106] However, some dispute this interpretation.[107]

In terms of when the assessment should take place, Article 36 of AP I requires it at the stage of study, development, acquisition or adoption. Practically speaking, for a State that produces weapons, this would mean carrying out a review at the conception/design stage and thereafter at different stages of its technological development and before entering the production phase. For a State that is procuring weapons for the first time, the review should be carried out when the weapon is being studied for purchase and before a purchase agreement is entered into. For a State adopting a technical modification or a field modification to an existing weapon, the review should be carried out at the earliest possible stage.[108]

1.3 Legal reviews of autonomous weapon systems

In light of the above requirements, a legal review would apply to new weapon systems that are wholly or partially fitted with autonomous features, as well as to existing weapon systems that are fitted with new autonomous features. In assessing the legality of such autonomous weapon systems, the reviewers should also look at the normal or expected circumstances of their use. This requires foreseeing how the weapon will perform in the environment in which it is intended to be deployed, based on the weapon's design and how it actually functions. If the reviewers find that the autonomous weapon system could be used lawfully only in limited circumstances, these limits must be incorporated in the instructions and rules of engagement applying to the weapon. The permitted circumstances of use may in some cases be so limited or complex that it may be more appropriate to prohibit the weapon's use altogether.

Waxman and Anderson have recommended careful and continuous development of internal norms, principles and practices for the design and implementation of autonomous weapon systems and a clear articulation of the legal and moral principles by which autonomous weapon systems should be judged.[109] Along these lines, United Kingdom Joint Doctrine Note 2/11 states, "if we wish to allow systems to make independent decisions without human intervention, some considerable work will be required to show how such systems will operate legally."[110] However, the Joint Doctrine clearly states that the UK Ministry of Defence "currently has no intention to develop systems that operate without human intervention in the weapon command and control chain, but it is looking to increase levels of automation where this will make systems more effective."[111] Likewise, the US Department of Defense policy on autonomous weapon systems requires a human role when lethal force is used: "Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force."[112]

As seen in Part A (section 6.2) the unpredictability of more complex autonomous weapon systems might make it very difficult in practice to effectively review their legality. In addition, there is controversy today over whether an autonomous weapon system would pass such legal review or not. Human Rights Watch has concluded that autonomous weapon systems would be "incapable of abiding by key principles of international humanitarian law"[113] and has recommended that "any review of fully autonomous weapons should recognize that for many people these weapons are unacceptable under the principles laid out in the Martens

[106] ICRC's *Guide to the Legal Review of New Weapons, Means and Methods of Warfare*, *supra* note 98, p. 7.
[107] See Part C (section 2) of this background paper.
[108] Ibid., section 2.3.1.
[109] M. Waxman and K. Anderson, 'Law And Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can' (2013), p. 13, http://ssrn.com/abstract=2250126, p.10.
[110] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, para. 503.
[111] Ibid., para 508.
[112] US Department of Defense, *Directive 3000.09, supra* note 31, pp. 2-3.
[113] B. Docherty, *Losing Humanity, supra* note 35, p. 30.

Clause."[114] In contrast, Schmitt has argued, "while it is true that some autonomous weapon systems might violate international humanitarian law norms, it is categorically not the case that all such systems will do so. Instead, and as with most other weapon systems, their lawfulness as such, as well as the lawfulness of their use, must be judged on a case-by-case basis."[115]

Sassòli has expressed the concern that, "politically, there is a risk that once the technology has been developed at great expense, vested interests will make it nearly impossible to conclude that the result is unlawful. The solution may be to accompany the development process with constant reviews."[116] In this respect, the U.S.'s Department of Defense Directive 3000.09 states that autonomous weapon systems must be subjected to two legal reviews: a preliminary legal review before a decision to enter into formal development, and another legal review before fielding.[117]

2. The fundamental rules of IHL in the conduct of hostilities, and programming challenges

Among the principal concerns about autonomous weapon systems is the question of whether they can be programmed to comply with the fundamental rules of IHL in the conduct of hostilities, namely the rules of distinction, proportionality, and precautions in attacks. Some systems might be able to comply with these rules in environments where there are few or no civilians, where their functions would pose little or no risk to civilians, or where they would be meant for 'machine-on-machine' operations.[118]

While it is more readily acknowledged that a machine can be programmed to carry out quantitative evaluations, it remains difficult, today, to encode qualitative judgements into a machine. The fundamental challenge in applying IHL rules on the conduct of hostilities lies in the fact that both quantitative and qualitative judgements would be made *de facto* by the machine based on the algorithms it is given.[119]

2.1 The rule of distinction

Article 48 of AP I describes the fundamental rule of distinction as follows:

> "In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives."[120]

[114] Ibid., p. 36. See also Part C (2) of this background paper.
[115] M. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics', *supra* note 99, p. 8.
[116] M. Sassòli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified,' manuscript to be published in U.S. Naval War College, International Law Studies, Vol. 90 (2014).
[117] See Enclosure 3 setting out Guidelines for Review of Certain Autonomous or Semi-Autonomous Weapon Systems, US Department of Defense, *Directive 3000.09, supra* note 31*,* p. 7.
[118] M. Waxman and K. Anderson, 'Law And Ethics for Autonomous Weapon Systems,' *supra* note 109, p. 13.
[119] See M. Wagner, 'Autonomy in the Battlespace,' in D. Saxon (ed), *International Humanitarian Law and the Changing Technology of* War (The Netherlands: Martinus Nijhoff Publishers, 2013), p. 120. According to UK Joint Doctrine Note 2/11, 'meeting the requirement for proportionality and distinction would be particularly problematic, as both of these areas are likely to contain elements of ambiguity requiring sophisticated judgement. Such problems are particularly difficult for a machine to solve and would likely require some form of artificial intelligence to be successful.' UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, para. 508*.*
[120] This rule also exists under customary IHL in both international and non-international armed conflicts. See ICRC, *Customary International Humanitarian Law*, Volume I: Rules (Cambridge: Cambridge University Press, 2005), Rule 1, p. 3.

The rule is made operational through two other fundamental provisions prohibiting attacks on civilians and civilian objects (in Articles 51(2) and 52(1) AP I respectively).[121] In addition, Article 50(1) states that, "in case of doubt whether a person is a civilian, that person shall be considered to be a civilian."[122] Article 52(3) AP I states a similar rule for civilian objects.[123]

As discussed in Part A, current fixed autonomous weapon systems used in narrow roles and operating in relatively static, low clutter environments can be programmed to distinguish simple objects. In particular, "established technology… enables sensors to detect and recognise pre-determined categories of military equipment, such as artillery pieces, tanks, armoured personnel carriers, anti-aircraft batteries and so on. (…)"[124] Indeed, it seems that today a number of weapon systems are capable of determining the military nature of a target, based on quantitative data.[125]

Over time, programming might evolve and allow for more complex reasoning, including the capacity to make qualitative judgments. Autonomous weapon systems, it has been submitted, may even become good substitutes for humans, even if only in restricted contexts.[126] However, it remains an open question whether such technical developments are conceivable.[127]

However, as mentioned above in Part A (section 6.1) current autonomous target recognition capabilities are rudimentary and limited even in distinguishing simple objects in non-complex, low clutter environments.[128] According to UK Joint Doctrine Note 2/11, "for operating environments with easily distinguished targets in low clutter environments, a degree of autonomous operation is probably achievable now and data from programmes such as Brimstone and ALARM, for example, would have direct read-across. However, this is unlikely to be of much help to unmanned systems that we expect will have to operate in the future cluttered urban and littoral environments on long endurance missions."[129] As the UK Doctrine suggests, applying the rule of distinction in more complex environments would require a qualitative assessment, thus making it far more challenging for autonomous weapon systems to be IHL-compliant in such environments.

a) Distinguishing civilian objects from military objectives

Article 52(1) of AP I protects civilian objects from attacks or reprisals, and defines them as all objects which are not military objectives as defined in paragraph 2, which reads:

> "Attacks shall be limited strictly to military objectives. In so far as objects are concerned, military objectives are limited to those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."

The definition of a military objective is context-dependent. Certain objects will meet the definition in virtually any armed conflict (e.g. tanks, combat aircraft, military bases). As

[121] These rules also exist under customary IHL in both international and non-international armed conflicts. Ibid., Rules 1 and 7.
[122] Regarding the customary law equivalent, some states include this rule in their military manuals, whereas other states have expressed reservations. Ibid., p. 24.
[123] The customary nature of these two rules is not accepted by all States. Ibid., pp. 35–36.
[124] W. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' in D. Saxon (ed.), *International Humanitarian Law and the Changing Technology of* War (The Netherlands: Martinus Nijhoff Publishers, 2013), p. 55.
[125] See M. Wagner, 'Autonomy in the Battlespace,' *supra* note 119, p. 113.
[126] M. Waxman and K. Anderson, 'Law And Ethics for Autonomous Weapon Systems,' *supra* note 109, p. 12.
[127] N. Sharkey, 'Towards a principle for the human supervisory control of robot weapons,' *supra* note 80.
[128] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, p. 6-1; J.L. Chameau, et al, *Emerging and Readily Available Technologies, supra* note 22, p. 3-2.
[129] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, para. 508.

discussed in Part A, with existing technological capabilities, it could be possible to programme the rudimentary characteristics of such objects (e.g. shape, dimension, location of fixed objects) into an autonomous weapon, making it capable of identifying a target by matching the target's characteristics with those of its programme.[130] This would be a mechanical exercise based on quantitative data, and its reliability would depend on a predictable, low clutter environment.

Conversely, objects which are *a priori* civilian objects (e.g. hospitals, schools, apartment buildings) may become military objectives if the criteria of the above-quoted definition of a 'military objective' have been met. These criteria rely on a number of factors, the assessment of which would pose a significant challenge for an autonomous weapon system, especially in a dynamic environment. Indeed, determining both an 'effective contribution to military action' and a 'definite military advantage' requires assessing contextual elements that vary with the circumstances. As such, this exercise would involve a qualitative or subjective judgment.

In particular, an attack on the object in question must bring about 'a definite military advantage' which must be offered 'in the circumstances ruling at the time'. If the destruction of a given object does not yet offer, or no longer offers, a definite military advantage, the object does not constitute a military objective and must not be attacked. In a dynamic environment, this would require the attacker, or the autonomous weapon in use, to constantly interpret the information before it and reassess the situation and thus the military advantage to be gained.

Akerson argues that the definition of military objective is "expressed in general, subjective terms for precisely the reason that it cannot be articulated with any more precision without reference to the context in which the commander must apply it. The paradigm is thus unsuitable for a computer algorithm for two reasons: it cannot be expressed with precision and its value can only be determined in the context of application."[131]

b) Distinguishing civilians / persons hors de combat from combatants

Developing the capability of an autonomous weapon to distinguish persons (as opposed to objects, discussed above) in accordance with the rules stemming from the principle of distinction poses significant challenges, notably in two scenarios: distinguishing civilians from combatants or other fighters, and distinguishing persons that are *hors de combat* from combatants.

Civilians vs. combatants

Under the rule of distinction, attacks must only be directed at combatants.[132] Civilians are protected from deliberate attack, unless and for such time as they are directly participating in hostilities.[133]

[130] See M. Wagner, 'Autonomy in the Battlespace,' *supra* note 119, p. 113.
[131] D. Akerson, 'The Illegality of Offensive Lethal Autonomy' in D. Saxon (ed.), *International Humanitarian Law and the Changing Technology of* War (The Netherlands: Martinus Nijhoff Publishers, 2013), p. 79.
[132] The term 'combatant' here is used in its generic sense, meaning a person who does not enjoy the protection against attack accorded to civilians, but does not imply a right to combatant status or prisoner of war status. In this regard, the text occasionally uses the term 'fighter' interchangeably with 'combatant'.
[133] *Geneva Conventions of 12 August 1949*, adopted on 12 August 1949, entered into force on 21 October 1950, Article 3 common to four Geneva Conventions; *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflict*, adopted on 8 June 1977, entered into force on 7 December 1978 (*Additional Protocol I* or AP I), Articles 51(2) and (3); *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflict*, adopted on 8 June 1977, entered into force on 7 December 1978 (*Additional Protocol II* or AP II), Article 13.

Applying this rule is not so straightforward and would pose particular challenges for the programmer of an autonomous weapon system. For example, in a 'classic' armed conflict involving uniformed combatants, an autonomous weapon system would need to be capable of distinguishing an armed and uniformed soldier from an armed and uniformed civilian such as a police office or a hunter.[134]

Moreover, in contemporary armed conflicts, with the shift of military operations into civilian population centres, with civilians increasingly becoming involved in the hostilities (both on the side of States and organized armed groups) and with such fighters often not wearing distinctive uniforms, there are increasing difficulties in distinguishing between lawful targets and persons protected from attack.

Such difficulties would pose significant challenges in terms of programming an autonomous weapon system to be IHL-compliant in a populated environment. One key challenge would be ensuring that the autonomous weapon can accurately distinguish a civilian directly participating in hostilities from one who is not. In order to qualify as 'direct participation in hostilities',[135] a civilian's specific act must meet the following three cumulative criteria: 1) the act must be likely to adversely affect the military operations or military capacity of a party to an armed conflict or, alternatively, to inflict death, injury or destruction on persons or objects protected against direct attack; 2) there must be a direct causal link between the act and the harm likely to result either from that act, or from a coordinated military operation of which that act constitutes an integral part; and 3) the act must be specifically designed to directly cause the required threshold of harm in support of a party to the conflict and to the detriment of another. Measures in preparation of a specific act of direct participation in hostilities, and the deployment to and the return from the location of the act also form an integral part of that act. When civilians cease their direct participation in hostilities, they regain full civilian protection against direct attack.

Programming these criteria into a machine would appear a formidable task because of the qualitative analyses required by each, such as the assessment of the likely adverse effects of an act, and whether the individual is acting in support of a party to the conflict. This also involves interpreting an individual's intentions. These criteria are challenging for humans to apply, let alone machines, in view of the limits of current and foreseeable technologies.[136]

Persons hors de combat

Challenges would also arise in ensuring the autonomous weapon is capable of distinguishing an active combatant from one who is *hors de combat*. Article 41 of AP I prohibits attacks against persons who are *hors de combat* in the following terms:

> A person is 'hors de combat' if:
>
> (a) he is in the power of an adverse Party;
>
> (b) he clearly expresses an intention to surrender; or
>
> (c) he has been rendered unconscious or is otherwise incapacitated by wounds or sickness, and therefore is incapable of defending himself;

[134] To the extent that these armed civilians are not directly participating in the hostilities. See next footnote and accompanying text.
[135] See the ICRC's *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law* (Geneva: ICRC, 2009), which presents the ICRC's recommendations on how the notion of 'direct participation in hostilities' should be interpreted in contemporary armed conflicts. http://www.icrc.org/eng/assets/files/other/irrc-872-reports-documents.pdf.
[136] M. Wagner, 'Autonomy in the Battlespace,' *supra* note 119, p. 114.

> provided that in any of these cases he abstains from any hostile act and does not attempt to escape.[137]

The ICRC Commentary indicates that a defining feature of each of these elements is the fact that the person is 'defenceless', whether or not the person has laid down arms.[138] Recognizing whether a person is *hors de combat* requires interpretation of a person's intentions and behaviour in the given circumstances.

Illustrative of the programming challenge presented by this rule is the difficulty in detecting a person's willingness to surrender. There is no general agreement on the precise requirements to surrender, although many States and authors refer to the classical examples of raising hands, throwing away one's weapon or waving a white flag. The assessment of whether a person has surrendered requires detecting the individual's intention to surrender. This depends heavily on information reasonably available to commanders and others responsible for deciding upon attacks at the time they take their action.[139]

In relation to the difficulties in interpreting intent to surrender, Boothby argues that "there are legal implications flowing from the apparent fact that decision-making which is difficult for a pilot [of manned aircraft] or operator [of a remotely piloted aircraft] becomes unlikely bordering on impossible for autonomous weapons."[140] In this case, "it is the weapon system itself that presents the problem, not the circumstances. It would seem that to employ a weapon system that renders it virtually impossible to comply with the Article 41 rule would not be lawful, unless it is clear that the rule is not relevant to the circumstances of the mission that is being planned."[141]

Another equally fundamental question in relation to persons surrendering is whether it is effectively possible to surrender to a machine. The act of surrendering creates responsibilities under IHL for the party to which the combatant is surrendering (e.g. treating the wounded, protecting them from the dangers arising from the ongoing hostilities, etc.) Would it be practically possible to programme an autonomous weapon system to discharge such responsibilities? These would seem to require human involvement.

c) Doubt

As noted above, Article 50(1) API creates a presumption of civilian status in cases of doubt as to whether a person is a legitimate target[142] or as to whether a civilian object has become a military objective.[143] In such situations, a careful assessment has to be made under the conditions and restraints governing a particular situation as to whether there are sufficient indications to warrant an attack.

In cases of doubt as to the civilian status of a person, Schmitt argues, "the degree of doubt that bars attack is that which would cause a reasonable attacker in the same or similar circumstances to hesitate before attacking". He adds that developing an "algorithm that can both precisely meter doubt and reliably factor in the unique situation in which the autonomous weapon system is being operated will prove hugely challenging."[144] He nevertheless adds that algorithms that would enable an autonomous weapon system to

[137] This rule is also customary IHL in both international and non-international armed conflicts. See ICRC Customary Law Study, *supra* note 120, Rule 47.
[138] ICRC, *Commentary on the Additional Protocols, supra* note 103, para. 1630.
[139] M. Bothe, K. Partsch and W. Solf, *New Rules for Victims of Armed Conflicts: Commentary of the Two 1977 Protocols Additional to the Geneva Conventions of 1949* (The Hague: Martinus Nijhoff Publishers, 1982), p. 220.
[140] W. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' *supra* note 124, p. 59.
[141] Ibid., p. 60.
[142] Regarding the customary status of this rule, see above at note 122.
[143] See above at note 123.
[144] M. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law,' *supra* note 99, pp. 16-17.

compute doubt are possible in theory. What will remain difficult is determining the threshold of doubt at which an autonomous weapon system would need to refrain from attack.[145] A similar challenge would arise in case of a doubt relating to the status of an object.

2.2. The rule of proportionality

The rule of proportionality recognizes that civilian persons and objects may be incidentally affected by an attack that is directed at a military objective. According to the rule of proportionality, incidental civilian casualties and damages can be lawful under treaty and customary law if they are not excessive in relation to the concrete and direct military advantage anticipated and provided other rules are respected.[146]

Article 51(5) of AP I formulates the rule as follows:

> (b) an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated.

This rule is also considered to be customary law in all types of armed conflict.[147] It is said to be among the most complex to interpret and apply under IHL, as it requires a case-by-case qualitative judgement in often rapidly changing circumstances, of whether civilian loss would be excessive in relation to anticipated military advantage. Importantly, the assessment is based on information reasonably available at the time of the attack, and is not conducted *ex post facto*. Moreover, the assessment is always context specific. For example, the civilian loss resulting from an attack directed at an enemy tank that is about to fire may be acceptable in terms of the anticipated military advantage to be gained from the attack, whereas a similar number of civilian casualties may be considered excessive in an attack on a tank that is posing no immediate threat.[148] Boothby writes that these are "comparisons that challenge commanders, planners and other decision-makers" and reflect "the considerable intellectual difficulties associated with the implementation of the proportionality rule."[149]

According to the ICRC's Commentary, even if the rule is based "to some extent on a subjective evaluation, the interpretation must above all be a question of common sense and good faith for military commanders. In every attack they must carefully weigh up the humanitarian and military interests at stake."[150] The International Criminal Tribunal for the Former Yugoslavia (ICTY) has held that, "in determining whether an attack was proportionate it is necessary to examine whether a reasonably well-informed person in the circumstances of the actual perpetrator, making reasonable use of the information available to him or her, could have expected excessive civilian casualties to result from the attack."[151]

Programming for the rule of proportionality in attack would require attributing values to objects and persons and making calculations based on probabilities and context. The very nature and complexity of the rule could make it impossible to programme an autonomous weapon system to respect it, especially in a dynamic environment.[152] Even if one day programmers were able to achieve this, creating systems that can apply the proportionality rule in areas dense with civilian persons and objects would appear to be a long way off[153]

---

[145] Ibid., p. 17.
[146] J. Queguiner, 'Precautions under the Law Governing the Conduct of Hostilities,' *International Review of the Red Cross*, Vol. 88, No. 864 (2006), p. 794.
[147] ICRC Customary Law Study, *supra* note 120, Rule 14.
[148] See, e.g., M. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law,' *supra* note 99, p. 20.
[149] W. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' *supra* note 124, p. 56.
[150] ICRC, *Commentary on the Additional Protocols, supra* note 103, para. 2208.
[151] ICTY, *Prosecutor v. Stanislav Galić*, Case No. IT-98-29-T, Judgment, Trial Chamber (5 December 2003), para. 58.
[152] See N. Sharkey, 'Automated Killers and the Computing Profession,' *Computer*, 40:11 (2007), p. 122.
[153] M. Waxman and K. Anderson, 'Law And Ethics for Autonomous Weapon Systems,' *supra* note 109, p. 13.

because of the large number of variables that it would have to interpret in real time.[154] It seems such an assessment would require uniquely human judgment.[155]

As the rule states, a proportionality assessment requires an evaluation of both the expected incidental civilian casualties and damage and the anticipated military advantage. Some authors argue that it would be possible to programme a machine to assess the likelihood of incidental harm to civilians and damage to civilian objects near a target. Schmitt in particular points to the "collateral damage estimate methodology" or CDEM used by the US military in planning attacks, to assess factors such as a weapon's precision, its blast effect, attack tactics, the likelihood of civilian presence, and the composition of buildings. The CDEM itself "does not resolve whether a particular attack complies with the rule of proportionality", rather it is described as "a policy-related instrument used to determine the level of command at which an attack causing collateral damage must be authorized."[156] The higher the probability of incidental (collateral) damage, the higher the required level of command for approval. In Schmitt's view, "there is no question that autonomous weapon systems could be programmed to perform CDEM-like analyses to determine the likelihood of harm to civilians in the target area" and would produce results no less reliable than the CDEM, which itself is "heavily reliant on scientific algorithms". Thurnher also writes "it is conceivable that AWS could lawfully operate upon a framework of pre-programmed values. The military operator setting these values would, in essence, pre-determine what constitutes excessive collateral damage for a particular target. (…) these values would invariably need to be set at extremely conservative ends to comply with the rule."[157]

However, Schmitt acknowledges the difficulty in programming an autonomous weapon system to assess the military advantage against the incidental civilian casualties and damage to civilian objects: "Given the complexity and fluidity of the modern battle space, it is unlikely in the near future that, despite impressive advances in artificial intelligence, 'machines' will be programmable to perform robust assessments of a strike's likely military advantage."[158] Indeed, Human Rights Watch has concluded it would not be possible to duplicate the psychological processes in human judgment that are required to assess proportionality.[159] In light of today's challenges in programming for the various qualitative evaluations that this IHL rule requires, the use of autonomous weapon systems would need to be limited to cases where the risk to civilians is minimal.[160]

Schmitt does nevertheless argue that "military advantage algorithms could in theory be programmed into autonomous weapon systems" by pre-programming them to recognize a "maximum collateral damage threshold" for a given military objective like a tank, for instance. The threshold would have to be adjustable by a military operator based on changing locations, phases of operations, and other circumstances. In Schmitt's view, the requirement to assess military advantage "would likely not be able to account for all imaginable scenarios and variables that might occur during hostilities," just as is the case for a "human confronted with unexpected or confusing events when making a time sensitive decision in combat." He also reminds that under IHL the standard is reasonableness, not perfection,[161] particularly as

[154] B. Docherty, *Losing Humanity, supra* note 35, p. 32.
[155] W. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' *supra* note 124, p. 57. Boothby does, however, say that there may 'nevertheless be, perhaps restrictive, circumstances in which the use of these systems would be legitimate notwithstanding the limitations and difficulties (…) context is likely to be an important factor here,' at p. 57.
[156] M. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law,' *supra* note 99, p. 19.
[157] J. Thurnher, 'Examining Autonomous Weapon Systems form a Law of Armed Conflict Perspective,' in H. Nasu and R. McLaughlin (eds), *New Technologies and the Law of Armed Conflict* (The Netherlands, T.M.C Asser Press, 2014), p. 222.
[158] M. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law,' *supra* note 99, p. 20. See also M. Wagner, 'Autonomy in the Battlespace,' *supra* note 119, p. 121.
[159] B. Docherty, *Losing Humanity, supra* note 35, p. 33.
[160] M. Wagner, 'Autonomy in the Battlespace,' *supra* note 119, p. 122; B. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' *supra* note 124, p. 57.
[161] M. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law,' *supra* note 99, p. 21.

the proportionality assessment is based on the information reasonably available at the time of the attack. Whether the same standard or a higher one would be imposed on autonomous weapon systems remains to be seen.

An additional, important question in relation to the use of an autonomous weapon system is: when must the proportionality assessment occur to fulfil the rule? Is it sufficient that the proportionality of the attack be assessed in the planning and programming phase? Boothby has suggested that this would in principle meet the requirement of the rule if the 'planning assumption' remains reliable for the duration of the autonomous weapon system's deployment.[162] This would be most likely when the autonomous weapon system is deployed in a static, predictable, low-clutter environment. In a dynamic environment, this reliability would appear to be doubtful.

2.3 The rule of precautions in attack

In the conduct of hostilities, IHL requires the parties to armed conflicts to take constant care to spare the civilian population, civilians and civilian objects. This basic principle underlies the rule of precautions in attack, which is of customary IHL in both international and non-international armed conflict.

Listed below are some of the fundamental precautions required by the rule and set out in Article 57(2) of AP I:

> 2. With respect to attacks, the following precautions shall be taken:
>
> (a) those who plan or decide upon an attack shall:
>
> (i) do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects and are not subject to special protection but are military objectives within the meaning of paragraph 2 of Article 52 and that it is not prohibited by the provisions of this Protocol to attack them;
>
> (ii) take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects;
>
> (iii) refrain from deciding to launch any attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated;
>
> (b) an attack shall be cancelled or suspended if it becomes apparent that the objective is not a military one or is subject to special protection or that the attack may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated;

These precautions pose a number of challenges for autonomous weapon systems, discussed below.

a) Feasibility of precautions

[162] See W. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' *supra* note 124, p. 58. A similar point is made below in section 2.3 a) below on precautions in attack.

All of these obligations would apply to the use of autonomous weapon systems. It is important to remember that the feasibility of a precaution depends on the possibilities available to the party who plans, decides on and executes the attack. It does not depend on the machine's feasibility.[163] Indeed, Boothby says there is "no *carte blanche* in favour of complete autonomy", meaning that where some precautions are not feasible in relation to autonomous attack it is incorrect to conclude that they are not required in respect of these attacks. Instead, if a more conventional (i.e. non-autonomous) method of attack would permit precautions, then such precautions would be considered to be feasible and thus required.[164]

AP I does not define "feasible precautions." In Amended Protocol II to the 1980 Convention on Certain Conventional Weapons (CCW), "feasible precautions" are defined as "those precautions which are practicable or practically possible taking into account all circumstances ruling at the time, including humanitarian and military considerations."[165]

It has been argued that if, at the planning/programming stage, it is feasible to make an assessment of the potential incidental civilian casualties and damage to civilian objects that would remain reliable for the period during which the autonomous weapon system is deployed, this would in principle meet the obligation to take precautions in attack.[166] This is more likely to be feasible when the autonomous weapon is deployed in a static, predictable, low-clutter environment. Conversely, the reliability of such assessment would be doubtful where the autonomous weapon is deployed into a dynamic environment, at least based on current and foreseeable technological capabilities (discussed in Part A).

Sassòli writes that autonomous weapon systems may be able to take additional precautions because the human life of a pilot or soldier is not at risk. Moreover, they would have an advantage in that the feasibility of precautions would evolve with experience; unsuccessful precautions would lead to lessons being learned and reprogramming of the autonomous weapon systems.[167]

b) Verifying the nature of the objective

The requirement under Article 57(2)(a)(i) that those planning or deciding upon an attack do everything feasible to verify that the target is a military objective aims to ensure that operations will target strictly military objectives and thus contributes to preserving the immunity of both civilian populations and objects.[168]

According to Schmitt, this obligation "would, for example, require full use of on-board or external sensors that could boost the reliability of target identification." Thurnher adds, "the advanced recognition capabilities of AWS [autonomous weapon systems] would be sufficiently precise and reliable to fulfil this requirement. Yet at other times, (…) a force may have to augment AWS with other sensors to help validate the target."[169] Boothby has said that this rule could be complied with if the military objective being sought is susceptible to 'mechanical target recognition', meaning "technology which enables sensors to detect and recognize pre-determined categories of military equipment, such as artillery pieces, tanks, armoured personnel carriers, anti-aircraft batteries and so on."[170]

---

[163] M. Sassòli, 'Autonomous Weapons and International Humanitarian Law,' *supra* note 116.
[164] W. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' *supra* note 124, p. 61.
[165] *Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices as amended on 3 May 1996* (Amended Protocol II to the Convention on Conventional Weapons), adopted on 3 May 1996, entered into force on 3 December 1998, Article 3(10).
[166] See W. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' *supra* note 124, p. 58.
[167] Ibid., p. 31.
[168] J. Queguiner, 'Precautions,' *supra* note 146, p. 797.
[169] J. Thurnher, 'Examining Autonomous Weapon Systems,' *supra* note 157, p. 222.
[170] W. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' *supra* note 124, p. 55.

It is important to note, however, that the obligation imposed by this paragraph of Article 57 cannot be interpreted as obliging the parties to a conflict to possess modern and highly sophisticated means of reconnaissance. But it does require that the most effective and reasonably available means be used systematically in order to obtain the most reliable information possible before an attack.[171]

c) Choosing means and methods with a view to avoiding or minimizing incidental loss

The obligation to choose the methods and means of warfare likely to cause the least danger to civilian lives and to civilian objects is set out under Article 57(2)(a)(ii) and reflects customary law.[172]

Regarding the *means* of warfare, the obligation could apply to autonomous weapon systems in two distinct ways. Firstly, in terms of the decision of a commander to deploy the weapon; and secondly regarding the specific means chosen by an autonomous weapon system when it engages a target. In relation to the first, Schmitt has written, "the only situation in which an autonomous weapon system can lawfully be employed is when its use will realize military objectives that cannot be attained by other available systems that would cause less collateral damage."[173] Schmitt also argues that it is conceivable that an autonomous weapon system could be able to "achieve a military objective with less threat of collateral damage than a human controlled system."[174] Provided that it is clearly foreseeable in a given case that the deployment of an autonomous weapon would cause fewer incidental civilian casualties and less incidental damage to civilian objects compared to the use of conventional weapons, the rule on precautions in the choice of means and methods of warfare may therefore require that a commander consider using the autonomous weapon, if practicable (and subject to other considerations discussed in Part C).[175]

With regard to the second aspect of this rule, the challenge would be to programme the autonomous weapon system to be capable of making the qualitative evaluations required by this rule. It may be difficult to programme an autonomous weapon system so that it is capable of choosing the most appropriate means at its disposal. In addition to challenges arising from the means of warfare, the obligation to take precautions in the choice of *methods* of warfare also imposes restrictions on the timing, location, or even angle of an attack.[176]

d) Cancelling or suspending an attack

The obligation to cancel or suspend an attack, set out in Article 57(2)(b) of AP I (quoted above), poses a particular challenge for autonomous weapon systems in view of their persistency, i.e. the length of time between their deployment and their identification and attack of a target. This is especially true in a dynamic environment, where the circumstances around the target are likely to have changed between deployment and the attack. Instructions issued in advance of an attack can never be definite, and an autonomous weapon system would need to be designed to allow for verification of the target as required by the rule of precautions, either through programming or by a human operator.[177] This is where it has been suggested that human 'override' be built in,[178] with the risk that information

[171] J. Queguiner, 'Precautions,' *supra* note 146, pp. 797-798.
[172] ICRC Customary Law Study, *supra* note 120, Rule 17.
[173] M. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law,' *supra* note 99, p. 24.
[174] Ibid., p. 25.
[175] J. Kellenberger, 'International Humanitarian Law and New Weapon Technologies,' *ICRC*, Keynote address at 34th Round Table on Current Issues of International Humanitarian Law, San Remo (8-10 September 2011), http://www.icrc.org/eng/resources/documents/statement/new-weapon-technologies-statement-2011-09-08.htm.
[176] J. Queguiner, 'Precautions,' *supra* note 146, pp. 800 – 801.
[177] M. Sassòli, 'Autonomous Weapons and International Humanitarian Law,' *supra* note 116.
[178] J. Thurnher, 'Examining Autonomous Weapon Systems,' *supra* note 157, pp. 223-224.

processing would be so rapid and based on such large amounts of information that a human would not effectively be able to interrupt the autonomous weapon systems.[179]

One expert draws a parallel between the deployment of an autonomous weapon system and the launching of tactical cruise missiles that can have such long transit times that 'collateral circumstances' can change in the target area, in which case in his view the use of such weapon would not necessarily be unlawful "because the legitimacy of the attack decision is assessed by reference to the information reasonably available at the time that decision was made."[180] However, this analogy only holds so far: given that an autonomous weapon system would have greater persistency than a tactical cruise missile, it would be foreseeable already at the time of its planning/programming and deployment that the circumstances around the target are likely to have changed by the time the weapon finds it. Again, this is particularly relevant when the autonomous weapon is being deployed in a dynamic environment.

3. Who is responsible for acts by autonomous weapon systems that would amount to violations of international humanitarian law?

Ensuring accountability for acts of an autonomous weapon system poses some significant challenges. In Joint Doctrine Note 2/11 on the UK Approach to Unmanned Aircraft Systems, the UK has said that legal responsibility for any military activity remains with the last person to issue the command authorizing a specific activity.[181] The US has also accepted that those persons involved in operations of autonomous weapon systems could be accountable for their decisions. In its Directive 3000.09, the US Department of Defense has said that "persons who authorize the use of, direct the use of, or operate autonomous and semi-autonomous weapon systems must do so with appropriate care and in accordance with the law of war (…)."[182]

However, as mentioned above, there will always be a level of uncertainty about the way autonomous systems will interact with the external environment.[183] As the UK Joint Doctrine Note states, there is an "implicit assumption that a system will continue to behave in a predictable manner after commands are issued; clearly this becomes problematical as systems become more complex and operate for extended periods."[184]

As already mentioned, it is only possible to test the performance of an autonomous weapon system for a fraction of the situations that it might face, and the adaptable nature of an autonomous weapon system would make its performance difficult to predict. It would therefore make it challenging to effectively control the weapon system's actions[185] or hold anyone accountable for its unpredictable behaviour. In addition, it is uncertain whether commanders or operators would have the necessary knowledge or understanding to grasp how an autonomous weapon system functions.[186] In light of these challenges, who would be held responsible if an autonomous weapon system were to operate in a way that amounts to a violation of IHL? Persons who could be considered responsible include programmers,

---

[179] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para 41.
[180] W. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' *supra* note 124, p. 58.
[181] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, para. 510.
[182] US Department of Defense, *Directive 3000.09, supra* note 31, p. 3.
[183] US Air Force, *Report on Technology Horizons, supra* note 11, p. 63.
[184] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, para. 510.
[185] US Air Force, *Report on Technology Horizons, supra* note 11, p. 105.
[186] U.S. Department of Defense Directive 3000.09 says at section 4(a)(3) that '[i]n order for operators to make informed and appropriate decisions in engaging targets, the interface between people and machines for autonomous and semi-autonomous weapon systems shall: (a) Be readily understandable to trained operators. (b) Provide traceable feedback on system status. (c) Provide clear procedures for trained operators to activate and deactivate system functions.' Section 4(b) states: "Persons who authorize the use of, direct the use of, or operate autonomous and semi-autonomous weapon systems must do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement (ROE).' See US Department of Defense, *Directive 3000.09, supra* note 31, pp. 2-3.

manufacturers, officers who deploy the autonomous weapon systems, military commanders, and political leaders.[187]

3.1 Soldiers/operators and commanders

Thurnher suggests that an individual who knowingly deploys an autonomous weapon system that is incapable of distinguishing combatants from civilians into areas where civilians are expected to be located would be responsible for using the system in an unlawful manner. He adds, "a human could also be held responsible for the underlying subjective targeting decisions that laid the foundation for the ultimate strike. These actions would be measured for reasonableness." Such subjective decisions would have two components: first, the decision to deploy the autonomous weapon system, and second, the expected length of time from the launch of the system to the strike on the target. Both these aspects would need to be reasonable to avoid responsibility.[188]

Sassòli has argued that a commander's responsibility would be more akin to direct responsibility than command responsibility under international law, just as that of a soldier firing a mortar.[189] What is unclear, however, is whether the person involved in the deployment of the autonomous weapon – be it the soldier or operator deploying the weapon, or the commander ordering the deployment – would be able to sufficiently understand the programming of the weapon to fulfil the *mens rea* criterion for criminal responsibility. For Sassóli, "the operator need not understand the complex programming of the robot, but must understand the result, that is, what the robot is able and unable to do."[190]

Under IHL, a commander can be held criminally responsible for the acts of a subordinate "if they knew, or had information which should have enabled them to conclude in the circumstances at the time, that he was committing or was going to commit such a breach and if they did not take all feasible measures within their power to prevent or repress the breach."[191] This IHL rule was intended for a commander's responsibility for the actions of a human rather than the performance of a weapon system. Heyns has nevertheless proposed that "[s]ince a commander can be held accountable for an autonomous human subordinate, holding a commander accountable for an autonomous robot subordinate may appear analogous."[192] In any event, a commander could be held responsible for the decision by a human subordinate to launch an autonomous weapon system.

3.2 Programmers and manufacturers

Another possibility would be to hold programmers and manufacturers liable (either for civil damages or under criminal law).

---

[187] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 77.

[188] J. Thurnher, 'Examining Autonomous Weapon Systems,' *supra* note 157, p. 225. Thurnher states at p. 224 that 'human subjective decisions will ultimately be examined for reasonableness. Timing plays a pivotal role in any measure of reasonableness with AWS. The longer the amount of time between the last human operator input and the autonomous strike itself, the greater the risk of changes on the battlefield resulting in an unanticipated action. The greater the risk, the less reasonable the decision to deploy AWS (…). Certainly, the risk could be lowered if the autonomous weapon is capable of regularly submitting data about the environment back to the human operator who could potentially adjust the engagement criteria. This may not always be an option however.'

[189] M. Sassòli, 'Autonomous Weapons and International Humanitarian Law,' *supra* note 116, p. 324.

[190] Ibid., p. 324.

[191] *Additional Protocol I, supra* note 133, Article 86. Under customary law, the *mens rea* element is that the commander 'knew or had reason to know' that subordinates were about to commit or were committing a war crime. See ICRC Customary Law Study, *supra* note 120, Rules 152 and 153. Under the Rome Statute of the ICC, command responsibility covers situations where the commander 'either knew or, owing to the circumstances at the time, should have known' that forces were about to commit or were committing war crimes. See *Rome Statute of the International Criminal Court,* adopted on 17 July 1998, entered into force on 1 July 2002, Article 28(a)(i).

[192] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 78.

Thurnher has argued that a person who intentionally programs an autonomous weapon system to carry out acts that amount to war crimes would be liable, but he does not specify under whether this is civil or criminal liability.[193] There is the question of whether the person that has programmed the autonomous weapon before an armed conflict can be held responsible for war crimes carried out by the weapon during an armed conflict, bearing in mind that war crimes can only be committed in armed conflicts. Like Thurnher, Boothby appears to assume that the programming of an autonomous weapon system would occur specifically for each mission, and therefore during the armed conflict.[194] Sassòli suggests that a person who deliberately misprograms a machine to commit a war crime could be held responsible as an indirect perpetrator, or as a guarantor who breaches his/her obligation by failing to intervene during the armed conflict to avoid the commission of the war crime. If the operator is conscious of the limits of the weapon but uses it anyway, the programmer would be an accessory to the ensuing war crime.[195]

Regarding civil suits, according to Lin, the law governing product liability has not been tested sufficiently when it comes to robotics.[196] In addition, Heyns has proposed that it might not be equitable to put the onus of a civil suit on victims of armed conflict, as they would not necessarily have the resources and would likely be in a different country.[197] Sassòli has suggested the "drafting of specific standards of due diligence, both for manufacturers and for commanders."[198]

3.3 An accountability gap?

If finding the commander, programmer or manufacturer responsible is not a practical possibility, then it is feared that there could well be a 'responsibility gap' that would enable impunity for the use of autonomous weapon systems.[199] It has been posited that if "there is no fair and effective way to assign legal responsibility for unlawful acts committed by fully autonomous weapons, granting them complete control over targeting decisions would undermine yet another tool for promoting civilian protection."[200] Heyns has proposed that if "the nature of a weapon renders responsibility for its consequences impossible, its use should be considered unethical and unlawful as an abhorrent weapon."[201]

On the other hand, it has been proposed that accountability could be assigned in advance,[202] along with a requirement to install recording devices on the autonomous weapon systems to review footage of lethal uses.[203] The transparency that would be enabled by such an electronic trail (even after operations are carried out) would be useful before a court and could help prove the lawfulness or unlawfulness of the weapon systems' operations. This, in turn, it has been argued would reinforce the credibility of IHL.[204] Another option would be to distribute responsibility among the different actors along the chain from programming to deployment.[205] Such an approach may, however, violate the customary IHL rule stating "no

[193] J. Thurnher, 'Examining Autonomous Weapon Systems,' *supra* note 157, p. 225.
[194] W. Boothby, 'How Far Will the Law Allow Unmanned Targeting to Go?' *supra* note 124, p. 58.
[195] M. Sassòli, 'Autonomous Weapons and International Humanitarian Law,' *supra* note 116, p. 325.
[196] P. Lin, 'Introduction to Robot Ethics,' in P. Lin et al (eds) *Robot Ethics: The Ethical and Social Implications of Robotics* (USA: MIT Press, 2012), p. 8.
[197] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 79.
[198] M. Sassòli, 'Autonomous Weapons and International Humanitarian Law,' *supra* note 116, p. 325.
[199] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 77.
[200] B. Docherty, *Losing Humanity, supra* note 35, p. 42.
[201] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 80, citing: G. Verugio and K. Abney, 'Roboethics: The Applied Ethics for a New Science' in P. Lin et al (eds), *Robot Ethics: The Ethical and Social Implications of Robotics* (USA: MIT Press, 2012) and R. Sparrow, 'Killer Robots,' *Journal of Applied Philosophy,* 24:1 (2007).
[202] R. Arkin, 'The Robot didn't do it,' Position Paper for the Workshop on Anticipatory Ethics, Responsibility and Artificial Agents, University of Virginia (24-25 January 2013), p.1, http://www.cc.gatech.edu/ai/robot-lab/online-publications/position_paperv3.pdf
[203] G. Marchant, et al, 'International Governance of Autonomous Military Robots,' *supra* note 64, p. 7.
[204] See M. Sassòli, 'Autonomous Weapons and International Humanitarian Law,' *supra* note 116, p. 338.
[205] A. Krishnan, *Killer Robots, supra* note 74, p. 105.

one may be convicted of an offence except on the basis of individual criminal responsibility."[206] Put differently, no penalty can be inflicted on persons for acts that they have not personally committed.

Waxman and Anderson have expressed some scepticism on insisting on an opportunity for criminal prosecution: "post-hoc judicial accountability in war is just one of many mechanisms for promoting and enforcing compliance with the laws of war, and its global effectiveness is far from clear."[207]

3.4 State responsibility

As mentioned above, IHL was originally designed as part of a system governing relations between States. Within this system lies the notion that every internationally wrongful act of a State entails the international responsibility of that State. A State has committed an internationally wrongful act when conduct consisting of an action or omission is attributable to the State under international law and constitutes a breach of an international obligation of the State.[208] For example, the conduct of a person exercising elements of government authority, such as a soldier or commander, would be attributable to the State.[209] According to Article 3 of Hague Convention No. IV and Article 91 of AP I to the Geneva Conventions, a party to the conflict "shall be responsible for all acts by persons forming part of its armed forces."

A State's obligations arising from its responsibility require that it cease the unlawful conduct and make full reparation, which includes restitution, compensation or satisfaction.[210] These obligations may exist towards persons or entities other than States, for example in the case of IHL violations or "other breaches of international law where the primary beneficiary of the obligation breached is not a State."[211]

Heyns has suggested that a stronger emphasis on State as opposed to individual criminal responsibility may be called for, except in respect of use by non-state armed actors.[212] Also in favour of a State responsibility approach, Waxman and Anderson have written that "it would be unfortunate to sacrifice real-world gains consisting of reduced battlefield harm through machine systems (assuming there are any such gains) simply in order to satisfy an *a priori* principle that there always be a human to hold accountable. It would be better to adapt mechanisms of collective responsibility borne by a 'side' in war (…)."[213] The need to adapt mechanisms of State responsibility seem to be borne out by the fact that it is rare to see a State's being found responsible in contentious cases with regard to serious violations of IHL.[214]

---

[206] ICRC Customary Law Study*, supra* note 120, Rule 102.
[207] M. Waxman and K. Anderson, 'Law and Ethics for Robot Soldiers', *Policy Review* No. 176 (2012), p. 7.
[208] International Law Commission, *Articles on the Responsibility of States for Internationally Wrongful Acts,* Report on the Work of its Fifty-third Session (23 April–1 June and 2 July–10 August 2001), subsequently adopted by the General Assembly by Resolution A/RES/56/83 of 12 December 2001, Official Records, Fifty-fifth Session, Supplement No. 10 (A/56/10), http://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/56/83&Lang=E, Articles 1 and 2.
[209] Ibid., Article 4.
[210] International Law Commission, *Articles on the Responsibility of States*, *supra* note 208, Articles 30 – 31, 34 – 39. Article 3 of the Hague Convention No. IV and Article 91 of AP I mention only financial compensation. See *Hague Convention (IV) respecting the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land,* adopted on 18 October 1907, entered into force on 26 January 1910; *Additional Protocol I, supra* note 133.
[211] International Law Commission, 'Commentary to the Draft Articles on Responsibility of States for International Wrongful Acts,' *Yearbook of the International Law Commission,* Vol. II, Part II (2001), p. 87, http://legal.un.org/ilc/texts/instruments/english/commentaries/9_6_2001.pdf. For more information on State responsibility for violations of IHL, see M. Sassòli, 'State Responsibility for Violations of International Humanitarian Law,' *International Review of the Red Cross*, Vol. 84, No. 846 (2002).
[212] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 81.
[213] M. Waxman and K. Anderson, 'Law And Ethics for Autonomous Weapon Systems,' *supra* note 109, p. 17.
[214] See for instance ICJ, *Military and Paramilitary Activities in and against Nicaragua* (Nicaragua v. United States of America), Merits, ICJ Reports 1986, http://www.icj-cij.org/docket/files/70/6503.pdf; ICJ, *Armed Activities on the Territory of the Congo* (DRC v. Uganda), Merits, ICJ Reports 2005, http://www.icj-cij.org/docket/files/116/10455.pdf.

PART C: ETHICAL AND SOCIETAL CONCERNS AND THE DICTATES OF PUBLIC CONSCIENCE

Concerns have been raised that autonomous weapon systems would have no 'higher purpose' sense on which to make decisions, no ability to deal with ambiguity, no empathy or compassion or any capacity to imagine or take responsibility for the consequences of their actions. On the other hand, it has been argued that because autonomous weapon systems cannot be emotive and therefore cannot hate, it is more likely than a human being to 'behave' lawfully or ethically.[215] Even so, the potential use of autonomous weapon systems evokes a very difficult question: Is the delegation to machines of life and death choices morally acceptable?

1. The role of humans in the decision to apply force, including lethal force

In a survey conducted by Arkin, engaging robotics researchers, military, policy makers and the general population, it was found that "people are clearly concerned about the potential use of lethal autonomous robots."[216] Commonly heard is a moral objection to the idea of removing all human involvement from a decision to use force. Yet, Waxman and Anderson have written that "this is a difficult argument to engage, since it stops with a moral principle that one either accepts or not", but also that "it raises a further question as to what constitutes the tipping point into impermissible autonomy, given that the automation of weapons functions is likely to occur in incremental steps."[217]

According to Heyns, it is "an underlying assumption of most legal, moral and other codes that when the decision to take life or to subject people to other grave consequences is at stake, the decision-making power should be exercised by humans."[218] This is implied by IHL treaties, the rules of which assume the conduct of human soldiers or commanders, rather than machines.[219] Likewise, for Asaro, "the very nature of IHL (…) presupposes that combatants will be human agents" in the same way that judges, prosecutors, defenders, witnesses and juries all assess "the match between an abstract set of rules and any given concrete situation."[220]

Heyns has also asked whether it would be "inherently wrong to let autonomous machines decide who and when to kill. (…) The question here is whether the deployment of LARs [lethal autonomous robots] against (…) enemy fighters, is in principle acceptable, because it entails non-human entities making the determination to use force."[221]

On the other hand, Waxman and Anderson have said, "What matters morally is the ability consistently to behave in a certain way and to a specified level of performance. The 'package' it comes in, machine or human, is not the deepest moral principle."[222]

---

[215] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, paras 520-521; D. Saxon, 'Introduction,' in D. Saxon (ed.), *International Humanitarian Law and the Changing Technology of War*, (The Netherlands: Martinus Nijhoff Publishers, 2013), p. 7.

[216] R. Arkin, *Governing Lethal Behavior in Autonomous Robots* (USA: Taylor and Francis Group, 2009), pp. 49, 52.

[217] M. Waxman and K. Anderson, 'Law and Ethics for Robot Soldiers,' *supra* note 207, p. 7.

[218] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 89.

[219] Ibid., citing *Hague Convention II with Respect to the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land*, adopted on 29 July 1899, entered into force on 4 September 1900; *Hague Convention IV, supra* note 210; *Additional Protocol I, supra* note 133, Article 1(2). While the reference to a person may make the case for a requirement of human involvement, the idea of intervention by any entity other than a human was probably not considered in the late 19th and early 20th century.

[220] P. Asaro, 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanisation of Lethal Decision-Making,' *International Review of the Red Cross*, Vol. 94, No. 886 (2013), p. 700.

[221] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 92.

[222] M. Waxman and K. Anderson, 'Law And Ethics for Autonomous Weapon Systems,' *supra* note 109, p. 16.

In Thurnher's view, "in autonomous attacks, the main targeting decisions remain subjective, and those value judgments will continue to be made exclusively by humans. However, the subjective choices may be made at an earlier stage of the targeting cycle than with the more traditional human controlled systems. Sometimes these judgment calls will be made before the AWS [autonomous weapon systems] are even launched."[223] He adds, "to comply with the law, humans will need to inject themselves at various points into the process to make the necessary subjective determinations." This can happen through programming at the design phase, before launching the autonomous weapon systems, or remotely during its mission. The human operator, therefore, "is framing the environment" in which the autonomous weapon system operates.[224]

Recently there has been discussion of a 'human-on-the-loop' model of weapon systems, entailing human supervision of one or more systems that have autonomous functions in the decision cycle to carry out lethal force.[225] According to Asaro, this would be a "middle position between the direct human control of the human-in-the-loop model and an autonomous weapon system." But in his view, "including a human in the lethal decision process is a necessary, but not a sufficient requirement. A legitimate lethal decision process must also meet requirements that the human decision-maker involved in verifying legitimate targets and initiating lethal force against them be allowed sufficient time to be deliberative, be suitably trained and well informed, and be held accountable and responsible."[226] In other words, the process should allow meaningful human control.

2. "Principles of humanity" and "dictates of public conscience" (Martens Clause)

In the absence of rules of IHL explicitly prohibiting or restricting autonomous weapon systems, the acceptability of autonomous weapon systems should be examined according to the principles of humanity and the dictates of public conscience. The terms "principles of humanity" and "dictates of public conscience" were first referred to in the Martens clause, which was included in the Preamble of the Hague Conventions II of 1899 and IV of 1907 respecting the laws and customs of war on land. It has since been introduced in the main body of AP I to the Geneva Conventions and the preamble of Additional Protocol II (AP II) to the Geneva Conventions.[227]

Article 1(2) of AP I formulates the Martens Clause as follows:

> In cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from dictates of public conscience.

The ICRC Commentary to Article 1(2) AP I states that there were two reasons why it was considered useful to include this clause in the Protocol. First, "it is not possible for any

[223] J. Thurnher, 'Examining Autonomous Weapon Systems,' *supra* note 157, p. 223.
[224] Ibid., pp. 223-224.
[225] P. Asaro, 'On Banning Autonomous Weapon Systems,' *supra* note 220, pp. 694-695.
[226] Ibid., p. 695.
[227] See, e.g., *Hague Convention II, supra* note 219; *Hague Convention IV, supra* note 210; *Geneva Protocol on the Use of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare*, adopted on 17 June 1925, entered into force on 8 February 1928; *Geneva Conventions, supra* note 133, Articles 63, 62, 142, 158(4); *Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction*, adopted on 10 April 1972, entered into force on 26 March 1975; *Additional Protocol I, supra* note 133, Article 1(2); *Additional Protocol II, supra* note 133; *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to be Excessively Injurious or to have Indiscriminate Effects*, adopted on 10 October 1980, entered into force on 2 December 1983; *Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction*, adopted on 18 September 1997, entered into force on 1 March 1999.

codification to be complete at any given moment; thus the Martens Clause prevents the assumption that anything which is not explicitly prohibited by the relevant treaties is therefore permitted. Secondly, it should be seen as a dynamic factor proclaiming the applicability of the principles mentioned regardless of subsequent developments of types of situation or technology."[228]

The ICRC Commentary to the formulation of the Martens Clause in the preamble in AP II states that since "they reflect public conscience, the principles of humanity actually constitute a universal reference point and apply independently of the Protocol."[229]

As mentioned above, the International Court of Justice (ICJ) in its opinion on the *Legality of the Threat or Use of Nuclear Weapons* affirmed the importance of the Martens Clause "whose continuing existence and applicability is not to be doubted" and stated that it "had proved to be an effective means of addressing rapid evolution of military technology."[230] The ICJ also found that the Martens Clause represents customary IHL. In the case of *Prosecutor* v. *Kupreskić*, the ICTY stated, "In the light of the way states and courts have implemented it, this Clause clearly shows that principles of international humanitarian law may emerge through a customary process under the pressure of the demands of humanity or the dictates of public conscience, even where state practice is scant or inconsistent."[231]

Notwithstanding these judicial pronouncements, the exact interpretation of the Martens clause remains subject to significant variation among experts. It is debated whether the "principles of humanity" and the "dictates of the public conscience" are separate, legally binding yardsticks against which a weapon or a certain type of behaviour could be measured in law or whether they are rather moral guidelines. Some have suggested that they correspond to a feeling by the international community that some methods or means of warfare are beyond the pale and therefore not to be tolerated. The principle of humanity and the dictates of the public conscience would act as restraining factors.

Veuthey has written, "[f]irstly, one could say that public conscience is the trigger mechanism of every codification of IHL. Secondly, public conscience is the driving force behind the implementation and enforcement of IHL. Thirdly, public conscience forms a sort of safety net for humanity for circumstances that written law has overlooked or not yet covered."[232] Schmitt has suggested that the Martens clause "applies only in the absence of treaty law. In other words, it is a failsafe mechanism meant to address lacunae in the law; it does not act as an overarching principle that must be considered in every case."[233] Boothby has also said that "it is not entirely clear whether the 'dictates' are cited by the Clause as an element of the law in their own right or whether they are another constituent element of customary law. Those who take the former view will look upon the Martens Clause as justification for the proposition that morality is a distinct element in the law of armed conflict. (…) The safer interpretation is, probably, that the dictates of the public conscience drive the evolution of custom, and perhaps of the law as a whole, by inspiring treaty negotiators."[234]

It is submitted that the principles of humanity and dictates of public conscience can play an influential role when examining the desirability of developing and deploying autonomous weapon systems. As argued above in Part B (section 1.3) on legal reviews of new weapons,

---

[228] ICRC, *Commentary on the Additional Protocols, supra* note 103, para. 55.
[229] Ibid., para 4434.
[230] *Nuclear Weapons Advisory Opinion*, *supra* note 105, para 78.
[231] ICTY, *Prosecutor* v. *Kupreškić et al*, IT-95-16-T, Judgment, Trial Chamber (14 January 2000),http://www.icty.org/x/cases/kupreskic/tjug/en/kup-tj000114e.pdf, para. 527.
[232] M. Veuthey, 'Public Conscience in International Humanitarian Law,' in D. Fleck (ed) *Crisis Management and Humanitarian Protection* (Berlin: Berliner Wissenschafts-Verlag, 2004), p. 614.
[233] M. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics', *supra* note 99, p. 32.
[234] W. Boothby, *Weapons and the Law of Armed Conflict* (Oxford: Oxford University Press, 2009), p. 14. See also Y. Dinstein, *The Conduct of Hostilities under the Law of International Armed Conflict* (Cambridge: Cambridge University Press, 2004), p. 57.

a weapon that is not explicitly covered by existing rules of IHL would be considered contrary to the Martens Clause if it is determined per se to contravene the principles of humanity or the dictates of public conscience.[235] However, the view of others is that the principles of humanity and the dictates of the public conscience would act as moral guidelines. In the end, the critical question for consideration by the international community is whether the dictates of public conscience or the principles of humanity would allow delegating life and death decisions to machines on the battlefield, in particular, 'full' autonomy in target identification and application of lethal force.

3. Asymmetry and its consequences

According to Heyns, autonomous weapon systems "present the ultimate asymmetrical situation, where deadly robots may in some cases be pitted against people on foot." They are "likely – at least initially – to shift the risk of armed conflict to the belligerents and civilians of the opposing side."[236] The UK has asked: "will future wars be fought remotely, at least initially, with little or no loss of friendly human life? Is human nature such that the next arms race will seek to pitch increasingly complex unmanned systems against other unmanned systems or humans?"[237]

On the other hand, it is also recognized that commanders have a moral responsibility to limit loss of life on both sides of a conflict. The use of certain technologies that shield belligerents from certain risks could therefore be morally justified.

Additional concerns relate to the possibility that the ability to use unmanned systems, without risk to an operator's life, can make the use of armed force more attractive. Indeed, the UK Ministry of Defence has called for debate on the implications of resorting to remote-controlled weapons in order to "ensure that we do not risk losing our controlling humanity and make war more likely."[238] This concern would equally apply to the use of autonomous weapon systems.

Asaro and Heyns have argued that "the unavailability of a legitimate human target of the LAR [lethal autonomous robot] user State on the ground may result in attacks on its civilians as the 'best available' targets and the use of LARs could thus possibly encourage retaliation, reprisals and terrorism."[239]

[235] See above Part B (section 1.2).
[236] C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 86.
[237] UK Ministry of Defence, Development, Concepts and Doctrine Centre, *Joint Doctrine Note 2/11*, *supra* note 5, para. 516.
[238] Ibid., para. 517.
[239] P. Asaro, How Just Could a Robot War Be?' in P. Brey, et al. (eds), *Current Issues in Computing and Philosophy*, (Amsterdam: IOS Press, 2008), p. 13, cited in C. Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, supra* note 34, para. 87.

ANNEX 1: EXPERT MEETING AGENDA

Autonomous weapon systems: Technical, military, legal and humanitarian aspects

Château de Penthes, Geneva, Switzerland, 26–28 March 2014

DAY ONE – 26 MARCH 2014

9:00 – 9:15 **Welcome**

Dr Philip Spoerri, Director for International Law and Cooperation, ICRC

9:15 – 9:30 **Introduction and scope of the meeting**

Dr Knut Dörmann, Head of the Legal Division, ICRC

SESSION 1 SETTING THE SCENE

Chair: Dr Knut Dörmann, Head of the Legal Division, ICRC

Session objective:
Provide an introduction to developments in robotics and autonomous systems, and the drivers for military use of these technologies in weapon systems.

9:30 – 10:45 **Part 1: Civilian robotics and developments in autonomous systems**

Dr Ludovic Righetti, Autonomous Motion Department, Max Planck Institute for Intelligent Systems, Germany

Discussion

11:15 – 12:30 **Part 2: Military robotics and drivers for development of autonomous weapon systems**

Professor Mary Cummings, Humans and Autonomy Laboratory, Duke University, USA

Discussion

SESSION 2 EXISTING AND NEW AUTONOMOUS WEAPON SYSTEMS

Chair: Ms Kathleen Lawand, Head of the Arms Unit, Legal Division, ICRC

Session Objective:
Provide an understanding of autonomy in existing weapon systems and of trends in the research and development of new autonomous weapon systems.

14:00 – 15:30 **Part 1: Autonomy in existing weapon systems**

Panel of government experts followed by opportunity for other government experts to share experience and discussion

Mr Paul Scharre, Center for a New American Security, USA

Dr Dinesh Kumar Yadvendra, Headquarters Integrated Defence Staff, Ministry of Defence, India

Discussion

16:00 – 18:00 **Part 2: Research and development of new autonomous weapon systems**

Professor Noel Sharkey, Professor of Artificial Intelligence and Robotics, University of Sheffield, UK

Professor Ronald Arkin, College of Computing, Georgia Institute of Technology, USA

Dr Darren Ansell, School of Computing, Engineering and Physical Sciences, University of Central Lancashire, UK

Discussion

DAY TWO – 27 MARCH 2014

SESSION 3 **MILITARY UTILITY AND POLICY GOVERNING AUTONOMOUS WEAPON SYSTEMS**

Chair: Dr Neil Davison, Scientific Adviser, Arms Unit, Legal Division, ICRC

Session Objective:
Provide an understanding of the military utility and roles envisaged for autonomous weapon systems, their advantages and disadvantages, and policy governing their development and use.

09:00 – 10:30 **Part 1: Military utility of autonomous weapon systems in armed conflict**

Panel of government experts followed by opportunity for other government experts to share experience and discussion

Lt Col Olivier Madiot, Arms Control Division, Ministry of Defence, France

Dr Niu Yifeng, College of Mechatronic Engineering and Automation, National University of Defence Technology, China

Discussion

11:00 – 12:30 **Part 2: Current policy on autonomous weapon systems**

Panel of government experts followed by opportunity for other government experts to share experience and discussion

Lt Col Peter Sonnex, Arms Control and Counter-proliferation Policy, Ministry of Defence, UK; and
Mr Jeremy Wilmshurst, Arms Export Policy Department, Foreign and Commonwealth Office, UK

Mr Paul Scharre, Center for a New American Security, USA

Discussion

SESSION 4 AUTONOMOUS WEAPON SYSTEMS UNDER INTERNATIONAL HUMANITARIAN LAW

Chair: Dr Knut Dörmann, Head of the Legal Division, ICRC

Session Objective:
Examine the main rules of IHL that apply in the conduct of hostilities and the challenges in programming autonomous weapons to comply with them. Other legal issues to explore include accountability for acts that would amount to IHL violations.

14:00 – 15:30 **Autonomous weapon systems under IHL: Legal review, fundamental rules of IHL and the role of the legal adviser**

Ms Nathalie Weizmann, Legal Adviser, Arms Unit, Legal Division, ICRC

Professor Marco Sassòli, Department of International Law and International Organization, University of Geneva, Switzerland

Dr William Boothby, Geneva Centre for Security Policy, Switzerland

Discussion

16:00 – 17:00 **Case studies**

With commentary from independent experts followed by discussion

17:00 – 18:00 **Accountability for use of autonomous weapon systems**

Professor Christof Heyns, Institute for International and Comparative Law in Africa, University of Pretoria, South Africa

Dr William Boothby, Geneva Centre for Security Policy, Switzerland

Discussion

DAY THREE – 28 MARCH 2014

SESSION 5	**DICTATES OF PUBLIC CONSCIENCE AND HUMAN OVERSIGHT**
	Chair: Ms Kathleen Lawand, Head of the Arms Unit, Legal Division, ICRC
	Session Objective: Explore the ethical issues raised by minimal or absent human control/oversight in target identification and use of force by autonomous weapon systems.
09:00 – 10:30	**Ethical issues raised by autonomous weapon systems**
	Dr Peter Asaro, School of Media Studies, The New School, USA
	Dr Peter Lee, Portsmouth Business School, University of Portsmouth, UK
	Discussion
11:00 – 12:00	Discussion continued
12:00 – 12:30	**Summary by ICRC and close of the meeting**

+ + +

ANNEX 2: LIST OF PARTICIPANTS

GOVERNMENT EXPERTS (* indicates speaker)

Algeria	H.E. Boudjemâa DELMI Ambassador and Permanent Representative, Permanent Mission of Algeria to the United Nations and other international organizations in Geneva
	Mr Hamza KHELIF Deputy Permanent Representative, Permanent Mission of Algeria to the United Nations and other international organizations in Geneva
Brazil	Mr Diogo COELHO Third Secretary, Ministry of External Relations of Brazil
	Mr Carlos VALLIM Jr Captain, Military Engineer, Brazilian Army, Ministry of Defence
China	Mr Hoajun JI Deputy Division Director, Ministry of Foreign Affairs
	Mr Yifeng NIU * Associate Professor, National University of Defense Technology
	Ms Mujin YU Staff Officer, Ministry of National Defence
Colombia	Mr Fernando RESTREPO PUERTA Legal Advisor, Ministry of Defense
	Ms Luz Marina URREA VANEGAS Legal Adviser, Colombian Army, Ministry of National Defence
France	Lt Col Olivier MADIOT * Adviser, Ministry of Defense
	Ms Marie-Gaelle ROBLES Adviser, Permanent Mission of France to the Conference on Disarmament
	LCL Erwan ROCHE Adviser, Permanent Mission of France to the Conference on Disarmament
Germany	Mr Peter PAUELS Military Adviser, Permanent Mission, Geneva
	Ms Pamela PREUSCHE Desk Officer, Conventional Arms Control, Disarmament and Arms Control Department, Federal Foreign Office

India	Mr Dinesh KUMAR YADVENDRA * Scientific Adviser to the Chief of Integrated Defence Staff, Headquarters Integrated Defence Staff, Ministry of Defence
	Mr Siddhartha NATH First Secretary (Disarmament), Permanent Mission of India to the United Nations, Geneva
	Ms Uma SEKHAR Counsellor, Permanent Mission of India to the United Nations, Geneva
Israel	Col Noam NEUMAN Head, International Law Department, Israel Defense Forces
Japan	Lt Col Jun KANAI Air Force Officer, Defense and International Policy Planning Division, Defense Plans and Policy Department (J-5), Joint Staff Office, Ministry of Defense
	Col Kosuke MIDORIKAWA First Secretary and Defense Attaché, Delegation of Japan to the Conference on Disarmament
Kenya	Lt Col Yvonne KIRUI Kenya Defense Forces, Ministry of Defense
Mexico	Colonel F.A.P.A. D.E.M.A. Mario Alberto ESCAMILLA PEREZ Ministry of Defence
	Ms Mariana SALAZAR Director of International Humanitarian Law, Legal Advisory, Ministry of Foreign Affairs
Norway	Ms Annette BJORSETH Legal adviser, Norwegian Ministry of Foreign Affairs
	H.E. Terje HAUGE Ambassador, Norwegian Ministry of Foreign Affairs
Pakistan	Mr Mohammad Aamir KHAN Counsellor, Permanent Mission of Pakistan, Geneva
	Mr Aamar Aftab QURESHI Deputy Permanent Representative, Permanent Mission of Pakistan, Geneva
Qatar	Mr Hassan AL-EMADI Military Cooperation, Coordination and Follow-up Department, Ministry of Defence
	Col Nasser Abdul Rahiman Moussa AL-ISSAC Military Cooperation, Coordination and Follow-up Department, Ministry of Defence

Republic of Korea	Lt Col Heonuk PARK Defense Attaché, Embassy of the Republic of Korea in Bern
	Mr Younghyo PARK Counsellor, Permanent Mission of the Republic of Korea in Geneva
Russian Federation	Mr Andrey MALOV Senior Counsellor, Permanent Mission of the Russian Federation to the United Nations and other International Organizations, Geneva
	Mr Mikhail PETROSYAN Attaché, Ministry of Foreign Affairs
	Mr Oleg POMAZUEV Ministry of Defense
Saudi Arabia	Mr Solieman ALHAMMAD Officer, Ministry of Defence
	Mr Abdullah ALQURASHI Officer, Ministry of Defence
South Africa	Ms Chantelle NAIDOO First Secretary, Permanent Mission of South Africa in Geneva
Switzerland	Major François GARRAUX Policy and Military Adviser, Arms Control and Disarmament Permanent Mission of Switzerland in Geneva
	Mr Michael SIEGRIST Legal Officer, Federal Department of Foreign Affairs
United Kingdom	Lt Col Peter SONNEX Military Policy and Technical Adviser, Arms Control and Counter-proliferation Policy, Ministry of Defence
	Mr Jeremy WILMSHURST Conventional Arms Policy Officer, Foreign & Commonwealth Office
United States	Mr Matthew McCORMACK Deputy Legal Counsel, Department of Defense
	Mr Michael MEIER Attorney-Adviser, Department of State
	Mr Paul SCHARRE * Consultant; Center for a New American Security, USA

INDIVIDUAL EXPERTS

Dr Darren ANSELL	Space and Aerospace Engineering Lead, School of Computing, Engineering and Physical Sciences, University of Central Lancashire, UK
Professor Ronald ARKIN	Regents' Professor and Associate Dean for Research & Space Planning, School of Interactive Computing, College of Computing, Georgia Institute of Technology, USA

Dr Peter ASARO	Assistant Professor, School of Media Studies, The New School, USA; Vice-Chair, International Committee for Robot Arms Control
Dr Bill BOOTHBY	Associate Fellow, Geneva Center for Security Policy, Switzerland
Professor Mary CUMMINGS	Director, Humans and Aunonomy Laboratory, Duke University, USA
Mr Steve GOOSE	Executive Director, Arms Division, Human Rights Watch, USA
Professor Christof HEYNS	Professor of Human Rights Law, Co-Director, Institute for International and Comparative Law in Africa, University of Pretoria, South Africa
Dr Peter LEE	Principal Lecturer in Military and Leadership Ethics & Assistant Director (Academic), Portsmouth Business School, University of Portsmouth, UK
Ms Hine-Wai LOOSE	Political Affairs Officer, Implementation Support Unit, Convention on Certain Conventional Weapons (CCW), United Nations Office at Geneva
Dr Ludovic RIGHETTI	Group Leader, Autonomous Motion Department, Max Planck Institute for Intelligent Systems, Germany
Professor Marco SASSÒLI	Professor of international law and Director of the Department of international law and international organization, University of Geneva, Switzerland
Professor Noel SHARKEY	Professor of Artificial Intelligence and Robotics and Professor of Public Engagement, University of Sheffield, UK
Ms Kerstin VIGNARD	Chief of Operations, United Nations Institute for Disarmament Research (UNIDIR), United Nations Office at Geneva

ICRC

Dr Yves SANDOZ	Member of the International Committee of the Red Cross
Dr Philip SPOERRI	Director of International Law and Cooperation
Dr Knut DÖRMANN	Head, Legal Division
Ms Kathleen LAWAND	Head, Arms Unit, Legal Division
Dr Neil DAVISON	Science Adviser, Arms Unit, Legal Division
Ms Nathalie WEIZMANN	Legal Adviser, Arms Unit, Legal Division
Mr Laurent GISEL	Legal Adviser, Thematic Unit, Legal Division
Ms Isabel ROBINSON	Legal Attaché, Legal Division
Ms Veronika VAJDOVA	Legal Attaché, Legal Division

MISSION

The International Committee of the Red Cross (ICRC) is an impartial, neutral and independent organization whose exclusively humanitarian mission is to protect the lives and dignity of victims of armed conflict and other situations of violence and to provide them with assistance. The ICRC also endeavours to prevent suffering by promoting and strengthening humanitarian law and universal humanitarian principles. Established in 1863, the ICRC is at the origin of the Geneva Conventions and the International Red Cross and Red Crescent Movement. It directs and coordinates the international activities conducted by the Movement in armed conflicts and other situations of violence.

ICRC

4221/002 11.2014 500