ICRC

Report of the ICRC Expert Meeting on ‘Autonomous weapon systems: technical, military, legal and humanitarian aspects’, 26-28 March 2014, Geneva

9 May 2014

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.

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 emphasise 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[1] and 13 independent experts, including roboticists, jurists, ethicists, and representatives from the United Nations and non-governmental organisations. 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',[2] 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. A more detailed meeting report will be published later in 2014.

[1] 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.

[2] 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 emphasised 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 recognise 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 emphasised that future military systems would likely be hybrids of automatic and autonomous systems.

The speaker went on to emphasise 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 emphasised 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 weaponise 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 Defence 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-the-loop’)*: “A weapon system that, once activated, can select and engage targets without further intervention by a human operator.”[3] 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.”[4] 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.”[5] 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 emphasised 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

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

areas). However, both speakers noted that there are some existing anti-personnel weapon systems that have autonomous modes, such as so called ‘sentry weapons’.

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 emphasised 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 emphasised 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 emphasised 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 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 recognising 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-weaponised function is later weaponised. 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 emphasised 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 recognising 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".[6] 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".[7] 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.

[6] 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.

[7] 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."[8] 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 weapons 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",[9] 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."[10]

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 recognises 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.

[8] *Ibid.*

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

[10] *Ibid*, para 4(a).

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 – 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 emphasised 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."[11] In this regard, one

[11] *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).

speaker emphasised that determining who and what can be attacked under IHL, and under what circumstances and using which means, is therefore context-dependent.

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 emphasised 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 program. 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 emphasised 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 programme 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 assumptions 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 audio visual 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, manufacturers 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 judgment 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 judgment.

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 emphasised 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.