ICRC

Convention on Certain Conventional Weapons (CCW)
Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)
11-15 April 2016, Geneva

Views of the International Committee of the Red Cross (ICRC) on autonomous weapon system

11 April 2016

Introduction

Views on autonomous weapon systems, including those of the International Committee of the Red Cross (ICRC), continue to evolve as a better understanding is gained of current and potential technological capabilities, the military purpose of autonomy in weapon systems, and the resulting questions for compliance with international humanitarian law (IHL) and ethical acceptability.

Expert discussions of the last three years in the framework of the Convention on Certain Conventional Weapons (CCW) and in meetings convened by the ICRC and other organisations have been crucial to enhancing this understanding.

The ICRC held a second expert meeting on 'Autonomous weapon systems: Implications of increasing autonomy in the critical functions of weapons' from 15-16 March 2016 in Versoix, Switzerland.[1] Representatives of 20 States together with individual experts and representations of the United Nations and civil society organisations participated in the meeting. The ICRC will soon publish a summary report of the meeting.

In the meantime, as a contribution to ongoing discussions in the CCW, this paper highlights some of the key issues on autonomous weapon systems from the perspective of the ICRC, and in the light of discussions at its recent expert meeting.

1. Definitions

The ICRC has defined autonomous weapon systems as: "Any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention."

The advantage of this broad definition, which encompasses some existing weapon systems, is that it enables real-world consideration of weapons technology to assess what may make certain existing weapon systems acceptable – legally and ethically – and which emerging technology developments may raise concerns under international humanitarian law (IHL) and under the principles of humanity and the dictates of the public conscience.

Some prefer a narrow definition more closely linked to identifying the type of weapon systems of greatest concern. In addition, some narrow definitions distinguish between "highly automated" and "autonomous" weapons. However, many experts emphasise that there is not such a clear difference from a technical perspective, and the core legal and ethical questions remain the same.

---

[1] For a report of the first meeting see: ICRC (2014) *Autonomous weapon systems technical, military, legal and humanitarian aspects*, Report of an Expert Meeting held 26-28 March 2014 (published November 2014), https://www.icrc.org/en/download/file/1707/4221-002-autonomous-weapons-systems-full-report.pdf

In the view of the ICRC it is useful to start with a broad scope since lessons from experiences with existing weapon systems with autonomy in their critical functions can inform current discussions on emerging technology.

2. Autonomy in existing weapons

There are weapon systems in use today which can select and attack targets without human intervention. After activation by a human operator it is the weapon system, through its sensors and computer programming, which selects a target and launches an attack.

One broad group are **anti-material defensive weapons** used to protect vehicles, facilities or areas from incoming attacks with missiles, rockets, mortars or other projectiles. These include missile and rocket defence weapons and vehicle "active protection" weapons. The ability to effectively control these weapons and the use of force seems to require certain operational constraints including: limits on the task carried out (i.e. a single function to defend against incoming projectiles); limits on the targets (i.e. primarily objects and vehicles); controls over the operational environment (e.g. limitations on the geographical area and time frame of autonomous operation); and procedures for human intervention to deactivate the weapon, i.e. to cease its operation.

Some **offensive weapon systems**, including certain **missiles and torpedoes** also have a level of autonomy in selecting and attacking targets after launch. Many of these weapons are fired into a particular target area, after which on-board sensors and programming take over to autonomously select and attack a specific target object or person within that area. Some have more freedom of action in time and/or space, and therefore greater autonomy. These include, in particular, **loitering munitions** that search for targets over a wide geographical area for long time periods, and **encapsulated torpedo weapons** that remain stationary underwater over long time periods but can carry out attacks autonomously.

The trend in the development of missiles appears to be increasing autonomy with respect to movement in time and space. Indications are that future developments could also include increasing adaptability of these weapon systems to their environment. The ability to effectively control these weapons and the use of force may depend on a number of factors (as with defensive systems discussed above) including: limits on the task carried out; the ability of the system to discriminate targets; controls over the operational environment, such as limitations in time and space; and the ability for humans to communicate with the weapon system, for example to deactivate it. The latter is particularly difficult for underwater systems.

There are other **anti-personnel weapons** that may be capable of autonomously selecting and attacking targets, such as so called "sentry" weapons used to defend facilities and borders. However, the systems in use today apparently remain under remote control for initiation of attacks.

In sum, autonomy for selecting and attacking targets in existing weapon systems is limited by the operational parameters described above. Moreover, the technical characteristics and performance of existing weapon systems, combined with the operational parameters of their use, provide a certain degree of predictability of the outcomes of using these weapon systems. This predictability may be lost as autonomous weapon systems are used for more complex tasks or deployed in more dynamic environments than has been the case until now.

3. Emerging technology and future autonomous weapons

Although it is difficult to foresee the future development of autonomous weapon systems, it is clear there are a number of **military drivers for increased autonomy**, including enabling: increased mobility of robotic/unmanned weapon systems or platforms; operation of these systems in "communications denied" environments; shorter decision-making times between identifying and attacking a target; increased performance over remotely operated systems; and operation of increased numbers of robotic/unmanned weapon systems by fewer operators. Interest in autonomous weapon systems could also develop in different ways among non-State armed groups.

The general trend in civilian robotics is towards **supervised autonomy**, where robotic systems are increasingly autonomous while human operators retain oversight and often the ability to intervene.

The degree of autonomy in a robotic system is related to the level of human intervention in its operation, both in terms of the degree of human intervention and the stages at which the intervention is made. Even so, machines can and do effectively take decisions that have been delegated to them by humans through their computer programming, and without the need to be "conscious" or to have human-like levels of intelligence.

There are a number of developments that might make increasingly autonomous weapon systems become **less predictable**. These include: increased **mobility**, meaning the weapon system would encounter more varied environments over greater time periods; increased **adaptability**, such as systems that set their own goals or change their functioning in response to the environment (e.g. a system that defends itself against an attack) or even incorporate learning algorithms; and increased **interaction of multiple weapon systems** in self-organising swarms. In addition to decreasing the predictability of the weapon system, these developments could raise related problems for the validity of testing to ensure reliability.

For example, autonomous weapon systems that could set their own goals, or even "learn" and adapt their functioning, would by their nature be unpredictable. Highly mobile autonomous weapon systems would also create problems of predictability with respect to the target of specific attacks, especially if a system moved over a wide area or carried out multiple attacks.

4. Legal and ethical implications of increasing autonomy in weapon systems

It is clear that IHL rules on the conduct of hostilities are addressed to the parties to an armed conflict, more specifically to the human combatants and fighters, who are responsible for respecting them, and will be held accountable for violations. These obligations cannot be transferred to a machine. Still, in practical terms, the question remains, what limits are needed on autonomy in weapon systems to ensure compliance with IHL.

Control exercised by human beings can take various forms and operate at different stages of the "life cycle" of an autonomous weapon system, including: 1) the development of the weapon system, including its programming; 2) the deployment and use of the weapon system, including the decision by the commander or operator to use or activate the weapon system; and 3) the operation of the weapon system during which it selects and attacks targets.

It is clear that human control is exerted in the development and deployment stages of the weapon's "life cycle". It is in stage three, however, when the weapon is in operation – when it autonomously selects and attacks the target(s) – that the important question arises as to **whether human control in the first two stages is sufficient to overcome minimal or no human control at this last stage**, from a legal, ethical and military-operational standpoint (see also section 5 below).

The assessment of whether an autonomous weapon system can be used in compliance with IHL may depend on **the specific technical characteristics and performance of the weapon system and the intended and expected circumstances of its use**. Certain technical characteristics and their interaction with different operational parameters could significantly affect this assessment, including:

- The **task** the weapon system carries out;
- The **type of target** the weapon system attacks;
- The **environment** in which the weapon system operates;
- The **movement** of weapon system in space;
- The **time-frame** of operation of the weapon system;
- The **adaptability** of the weapon system, i.e. its ability to adapt its behaviour to changes in its environment, to determine its own functions and to set its own goals;
- Degree of **reliability** of the weapon system, i.e. robustness to failures and vulnerability to malfunction or hacking; and
- Potential for **human supervision** and intervention to deactivate the weapon system.

The combination of these technical characteristics and performance of the weapon system with the operational parameters of its use are critical to determining the foreseeable effects of the weapon – in other words, the **predictability** of the outcomes of using the weapon – and therefore in determining whether it can be used in conformity with IHL rules.

Indeed, deploying a weapon system whose effects are wholly or partially unpredictable would create a significant risk that IHL will not be respected. The risks may be too high to allow use of the weapon, or else mitigating the risks may require limiting or even obviating the weapons' autonomy. In this respect, the last factor in the list above – human intervention – could be considered as a **risk mitigation** factor. The level of risk resulting from a decrease in predictability of the weapon system may be influenced by a number of factors, such as the environment in which the weapon is used.

Predicting the outcome of using autonomous weapon systems may become increasingly difficult as the weapon systems become more complex or are given more freedom of action in their operations. For example, the legal assessment of an autonomous weapon system that carries out a single task against a limited type of target in a simple (uncluttered) environment, and that is stationary and limited in the duration of its operation (e.g. some existing missile and rocket defence systems) may conclude that there is an acceptable level of predictability, allowing for responsibility and accountability of the human operator. However, the conclusion may be very different regarding an autonomous weapon system that carries out multiple tasks (or is adaptable) against different types of targets in a complex (cluttered) environment, and that is mobile over a wide area and/or operating for a long duration.

For the purposes of an assessment under IHL there is no legal distinction between an offensive attack and a defensive one, as they both constitute attacks under the law. Nevertheless, the distinction between defensive and offensive weapon systems may be of greater relevance from a military-operational or ethical perspective.

The obligation to carry out **legal reviews of new weapons** under article 36 of Additional Protocol I to the Geneva Conventions is important to ensure that a State's armed forces are capable of conducting hostilities in accordance with its international obligations. The above challenges for IHL compliance will need to be carefully considered by States when carrying out legal reviews of any autonomous weapon system they develop or acquire. As with all weapons, the lawfulness of a weapon with autonomy in its critical functions depends on its specific characteristics, and whether, given those characteristics, it can be employed in conformity with the rules of IHL in all of the circumstances in which it is intended and expected to be used. The ability to carry out such a review entails fully understanding the weapon's capabilities and foreseeing its effects, notably through testing. Yet foreseeing such effects may become increasingly difficult if autonomous weapon systems were to become more complex or to be given more freedom of action in their operations, and therefore become less predictable.

Questions arise as to how IHL's "targeting rules" (e.g. the rules of proportionality and precautions in attack) are considered in reviewing weapons. Where it is the weapon itself that takes on the targeting functions, the legal review would demand a very high level of confidence that the weapon is capable of carrying out those functions in compliance with IHL.

An additional challenge for reviewing the legality of an autonomous weapon system is the absence of standard methods and protocols for testing and evaluation to assess the performance of these weapons, and the possible risks associated with their use. Questions arise regarding: How is the reliability (e.g. risk of malfunction or vulnerability to cyber-attack) and predictability of the weapon tested? What level of reliability and predictability are considered to be necessary? The legal review procedure faces these and other practical challenges to assess whether an autonomous weapon system will perform as anticipated in the intended or expected circumstances of use.

Although there are different views on the adequacy of national legal reviews of new weapons for ensuring IHL compliance of autonomous weapon systems, especially given the low level of implementation among States, this mechanism remains a critical measure for States to ensure respect for IHL. In any case, efforts to strengthen national legal review processes should be seen as complementary and mutually reinforcing of CCW discussions at the international level.

Some have raised concerns that use of autonomous weapon systems may lead to an "**accountability gap**" in case of violations of IHL. Others are of the view that no such gap would ever exist as there will always be a human involved in the decision to deploy the weapon to whom responsibility could be attributed.

Under **IHL and international criminal law**, the limits to control over, or the unpredictability of, an autonomous weapon system could make it difficult to find individuals involved in the programming and deployment of the weapon liable for serious violations of IHL. They may not have the knowledge or intent required for such a finding, owing to the fact that the machine can select and attack targets independently. Programmers might not have knowledge of the concrete situations in which at a later stage the weapon system might be deployed and in which IHL violations could occur. On the other hand, a programmer who intentionally programmes an autonomous weapon to commit war crimes would certainly be criminally liable. Likewise, a commander would be liable for deciding to use an autonomous weapon system in an unlawful manner, for example deploying in a populated area an anti-personnel autonomous weapon that is incapable of distinguishing civilians from combatants. In addition, a commander who knowingly decides to deploy an autonomous weapon whose performance and effects he/she cannot predict may be held criminally responsible for any serious violations of IHL that ensue, to the extent that his/her decision to deploy the weapon is deemed reckless under the circumstances. Overall, as long as there will be a human involved in the decision to deploy the weapon to whom responsibility could be attributed, there might not be an accountability gap.

Under the **law of State responsibility** a State could be held liable for violations of IHL caused by the use of any autonomous weapon system. Indeed under general international law governing the responsibility of States, they would be held responsible for internationally wrongful acts, such as violations of IHL committed by their armed forces using autonomous weapon systems. A State would also be responsible if it were to use an autonomous weapon system that it has not, or has inadequately, been tested or reviewed prior to deployment.

Autonomous weapon systems also raise **ethical** concerns that deserve careful consideration. The fundamental question at the heart of concerns, and irrespective of whether they can be used in compliance with IHL, is whether **the principles of humanity and the dictates of public conscience** would allow machines to make life-and-death decisions in armed conflict without human involvement. The debates of recent years among States, experts, civil society and the public have shown that there is a sense of deep discomfort with the idea of any weapon system that places the use of force beyond human control.

The question remains, however, what degree of human control is required, and in which circumstances, in light of ethical considerations? Is it sufficient for a human being to program an autonomous weapon system according to certain parameters and then make the decision to deploy it in a particular context? Or is it necessary that a human being bring his or her judgment to bear also on each individual attack? If the weapon autonomously uses force against a human target, what ethical considerations would this entail?

5. Human control

The notion of human control has become the **overarching issue** in the debates on autonomous weapon systems. There is broad agreement that human control over weapon systems and the use of force must be retained, although less clarity on whether this is for legal, ethical, military operational, and/or policy reasons, and what makes it “meaningful”, “appropriate” or “effective”.

From the ICRC’s perspective, a focus on the role of the human in the targeting process and the human-machine interface could provide a fruitful avenue for increasing understanding of concerns that may be raised by autonomous weapon systems, rather than a purely technical focus on the ‘level of autonomy’ of weapon systems.

A certain level of human control over attacks is inherent in, and required to ensure compliance with, the IHL rules of distinction, proportionality and precautions in attack. Considering more closely what these requirements are could help determine the boundaries of what is acceptable under IHL with respect to autonomy in the critical functions of selecting and attacking targets.

There are already a number of considerations that have been suggested, which provide avenues for future work to establish these requirements, including:

- **Predictability** of the weapon system in its intended or expected circumstances of use;
- **Reliability** of the weapon system in its intended or expected circumstances of use;

- **Human intervention** in the functioning of the weapon system during its development, deployment and use;
- Knowledge and accurate **information** about the functioning of the weapon system and the context of its intended or expected use; and
- **Accountability** for the functioning of the weapon system following its use.

Many of the technical characteristics and operational parameters that are relevant to assessing compliance with IHL (see section 4) are also important factors for determining the requisite human control over the use of force, as well as relevant for military-operational or ethical considerations. For example, human control over existing autonomous weapon systems (see section 2) is largely governed by technical and operational constraints on the functioning of the system (e.g. limited tasks and targets, limits in space and time, physical controls over the environment, and human supervision and ability to deactivate).

The **military** have a clear interest in maintaining human control of weapon systems, both to ensure compliance with IHL and to ensure that the commander has control over a given military operation. Rules of engagement associated with particular weapon systems are a primary way in which operational control is exerted over weapon systems and the use of force, and are therefore an important element in ensuring human control over the weapon system and the use of force.

Deeper consideration of the elements constituting human control and understanding **human-machine interaction** is needed to help determine the boundaries necessary to ensure that human control is maintained over weapon systems and the use of force. As outlined above, there are already concerns that autonomous weapon systems which could adapt or change their functioning, and those that 'hunt' for targets over wide areas could raise serious questions about human control and IHL compliance due to the lack of predictability on when and where the use of force and specific attacks would take place.

6. The way forward

There are at least **three broad approaches** that States could take to address the legal and ethical questions raised by autonomous weapon systems. The first regards strengthening national mechanisms for legal review and implementation of IHL to ensure any new weapons, including autonomous weapon systems, can be used in compliance with IHL.

The second is for States to develop a definition of "lethal autonomous weapon systems" in terms of the weapon systems that may be problematic from a legal and/or ethical perspective with a view to establishing specific limits on autonomy in weapon systems.

The third approach is for States to develop the parameters of human control in light of the specific requirements under IHL and ethical considerations (principles of humanity and the dictates of public conscience), thereby establishing specific limits on autonomy in weapon systems.

Both the second and third approaches recognise that international consideration is needed of the limits of autonomy in weapons systems to ensure legal compliance and ethical acceptability. Ultimately these two approaches might lead to the same end point in terms of identifying weapon systems requiring possible regulation or prohibition.

From the ICRC's perspective, the third approach focussing on human control and the human-machine interaction could be an effective way forward for the CCW, with efforts to strengthen national legal reviews to be pursued in parallel as a mutually reinforcing initiative. The framework of human control provides a useful baseline from which common understandings can be developed among States, and through which boundaries or limits on autonomy in weapon systems can be established. This is consistent with the broad agreement among States, experts and other stakeholders that there is a need to maintain human control over weapon systems and the use of force in view of legal obligations, military operational requirements, and ethical considerations.

+ + +