Professionals in Humanitarian Assistance and Protection

Autonomous Weapons – Potential advantages for the respect of international humanitarian law

Marco Sassòli

2 March 2013

Autonomous weapons are able to decide whether, against whom, and how to apply deadly force based on a pre-established computer program. "Intelligent weapons", in addition, may even be able to learn from past experience. These weapons do not base the use of force on ad hoc human decisions. I argue that autonomous weapons may bring advantages concerning the possibility of respecting international humanitarian law (IHL) – if it is technically possible to make them at least as accurate as the average soldier in terms of distinction, proportionality, and precautions.

Technological possibilities and IHL requirements

It is obviously not up to me as a lawyer to determine whether autonomous weapons that attain this standard of distinction, proportionality, and precautions currently exist or whether they can be developed in the future. Therefore, I first have to clarify that if it is not technically feasible to respect certain requirements of IHL with automated weapons, this would not be a sufficient reason for abandoning those requirements. The use of an autonomous weapon would then simply be unlawful.

This bold statement of illegality is however subject to two nuances:

1. First, both artillery and missiles are not as able as a sniper to interrupt an attack at the last moment based on changing circumstances. Nevertheless, no one claims that such weapons are inherently unlawful.
2. Second, it may well be that the average soldier is better able to respect IHL in certain respects than an automated weapon, while such a weapon is better in other respects. Even if an overall assessment is admissible, it must in my view be made concerning every given attack and may not be made in abstracto, concerning the weapon as such (although the weapon as such may be unlawful if it is always less accurate than a weapon delivered by a human being).

The ease of using force and making war

Some argue that with automated weapons it is easier to make war and to use force beyond a state's borders. But this has been true for many weapons and technologies – it was true for new weapons in the Middle Ages, and it was true when the first artillery, airplanes, and modern navies were developed. Compared with person-to-person fighting, all of these technologies made it easier to wage war. This is an issue concerning the permissibility of war (jus ad bellum) and a disarmament issue. It implies that robots, too, are subject to the general problem of disarmament.

It may well be that the (possibility of) secrecy around the use of automated weapons and the resulting difficulties of attribution make the implementation of state responsibility and of international criminal responsibility for an act of aggression more difficult. On the other hand, the fact that computer systems register everything makes an enquiry regarding criminal accountability easier, at least when undertaken by the party using the automated weapon.

In addition, there may be a psychological problem, but I am unable to judge whether it is real. One may argue that someone who builds an automated weapon and programs it, and who may be the last human being in the loop, in New Jersey, without even knowing where the weapon is to be used, feels a greater distance, feels less responsible, and will adopt more easily a computer game mentality than a soldier who is actually on the battlefield among the human beings he or she will kill. As far as I know, there is no scientific evidence for this effect or for the opposite. However, in my experience during the conflicts in the former Yugoslavia, I have seen people who killed other people face-to-face, with no more inhibition than my son playing a computer game. Moreover, I am not so sure that a programmer in New Jersey, when adequately trained and supervised and subject to appropriate accountability systems, will see the world simply as a computer game.

Many may, however, find the very idea that a robot may kill a human horrible. But frankly, is it not as horrible that one human being deliberately kills another human being without being immediately threatened by that human being? This is war. War is horrible.

Are autonomous weapons inherently unfair?

Many also consider automated weapons profoundly unfair, because most often only one party will have access to such technology, while the other side has to fight using actual human beings who will be killed. Be that as it may, but for a long time war has not been fair. The idea that war is waged by two knights fighting each other, while civilians stand by wondering who will win, belongs to the past. No one suggests that a party may not use their air force or navy if the enemy has no air force, navy, or anti-aircraft weaponry.

Contemporary reality also shows that the technologically weaker side may prevail over the stronger belligerent and impose its political will. Some claim that the weaker side tries to compensate for their lack of technology through changing the rules according to which the war is fought. It is, however, difficult to argue that, according to IHL, this risk means that a party is not allowed to use technology not available to the enemy.

Autonomous weapons will never leave you

The only valid argument of principle against automated weapons (but which is situated outside of IHL) is in in my view one that was recently put forth by Human Rights Watch[1]. Even the most ruthless dictator may be abandoned by the human beings fighting for him or her. One set of reasons for deserting is that the fight appears to be unjust, hopeless, or too inhumane, which means a dictator needs to exercise a certain minimum level of caution in their actions. On the other hand, no one has to fear that his automated weapons will abandon him or her.

Robots are not addressees of the law

When trying to apply the rules of IHL, there are some preliminary issues to clarify. Only human beings are subject to legal rules, and only human beings are obliged to follow them. In the case of automated weapons, IHL is therefore applicable to those who devise them, produce them, program them, and decide on their use. No matter how far we go into the future and regardless of in what ways artificial intelligence will work, there will always be a human being involved, at least at the machine's conception. A human being will decide that this machine will be created and then create the machine. Even if one day robots construct robots, it is still a human who has constructed the original robot. This human being is bound by the law. The machine is not bound by the law.

This may raise problems concerning the temporal field of application of IHL, since the last human intervention may happen before an armed conflict exists. Nevertheless, the program instructing the automated weapon when to use lethal force must already comply with IHL. At the same time, the temporal issue may also raise difficult legal questions in terms of criminal accountability, as war crimes can only be committed in armed conflicts.

Advantages of not being human

The main advantage of automated weapons – from an IHL compliance perspective – is that only human beings can be inhumane and only human beings can deliberately choose not to comply with the rules they were instructed to follow. As soon as robots have artificial intelligence, it will be necessary to make sure that such intelligence is not used – as human intelligence is sometimes used – to circumvent the rules or to decide from a utilitarian perspective that non-respect of their IHL instructions is preferred, as it makes the achievement of the main objective of overcoming the enemy easier.

If this risk can be avoided, a robot cannot hate, cannot fear, and has no survival instinct. While a human often kills to avoid being killed, a robot can wait with the use of force until the last moment when it is established that the target and the attack are legitimate. However, this advantage would be diminished if the enemy prepares against such robots and then those who produce the robots program them to shoot before they are destroyed by anti-robot weapons. We would then be back to what we have now with human beings – only it would be automatic. Nevertheless, for automated weapons, many more precautionary measures are feasible than for human beings and must therefore be taken under IHL.

Fundamental IHL questions rendered more acute

As always, the question arises as to which rules apply. I think robots simply raise anew – only more acutely – fundamental questions of humanitarian law on which there are already ongoing general debates. The most elemental question that comes to mind is the definition of armed conflict itself since outside an armed conflict, robots could only be used if they were able to arrest a person rather than use (lethal) force. As we know that there is no unique definition of an armed conflict, the questions are rather what is an international armed conflict and what is a non-international armed conflict.

What is the lower threshold of violence between a state and a non-state actor (or between non-state actors) that makes it an armed conflict? This is not a specific question for robots, and even where automated weapons are used, the answer must be given and will perforce be given by a human being. But the answer becomes even more important when automated weapons are used.

Many other questions must find an answer before an automated weapon can be programmed, for example:

- What exactly constitutes direct participation in hostilities?
- What is the relationship between international human rights law and international humanitarian law?
- What is the geographical scope of application of IHL and what is the scope of the battlefield?

Automated weapons raise this latter question more acutely, but legally, the considerations must be the same as for an aerial bombardment: may a belligerent attack a target which would be a legitimate target under IHL far away from the actual fighting, restrained only by the rules of IHL? Or does in such a place IHL not apply at all? Or does international human rights law prevail as the lex specialis?

Discrimination and distinction

The main question remains a technical one: is it possible or will it one day be possible to develop a program which enables a robot to distinguish not too stupidly between on the one hand legitimate targets, military objectives, combatants, and civilians directly participating in hostilities, and on the other, civilians, civilian objects, and specially protected objects such as cultural property, medical units, and objects indispensable for the survival of the civilian population? Human beings may make a lot of mistakes, even more mistakes than many machines have technical failures. Nevertheless, there are many elements which enable a human being to understand what is a legitimate target or not and those factors must be reproduced in a computer program. For the programmer who has to translate the ICRC guidance on direct participation in hostilities into a computer program, this will be incredibly challenging. It is, however, necessary to overcome this challenge if autonomous weapons are to be used.

Nor is it simply an issue of distinguishing between combatants and civilians in a Cold War-like situation. In today's conflicts, civilians often directly participate in hostilities, making the distinction more difficult for human soldiers, but also for automated weapons. An advantage of automated weapons is that they cannot be told, as we were so often told by military colleagues when drafting the ICRC Direct Participation Guidance, that it all depends on the given situation. To write a computer program, it must be clarified which factors guide the process of distinction and how those factors can be determined. In addition, computer programs cannot be instructed to apply unrealistic criteria, such as determining the intent of an enemy.

It may be particularly difficult to automatize the indicators which convince a human being that a certain person belongs to a category or has a conduct which makes that person a legitimate target. It will be equally difficult to formalize factors which convince a human being that he or she must interrupt an attack because the target is not lawful. To have a machine take such decisions autonomously may be even more difficult because nothing hinders the enemy feigning those indicators which make the robot believe that it is not confronted with a legitimate target. If the enemy artificially fulfills the indicators which make a robot decide that it may not attack under IHL, the fascinating question arises whether a machine can be "led to believe" something, or whether it is possible to "invite the confidence" of a machine – two elements of an act of perfidy, prohibited under IHL.

Proportionality

On proportionality, the need to translate principles into a computer program for automated weapons may be an opportunity to enhance the observance of IHL. The proportionality principle codified in Article 51 (5) (b) of Protocol I2, prohibits attacks, even if directed at a military objective, if they "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". Despite the qualifications of the military advantage, it remains very difficult to compare a military advantage with civilian losses, and the comparison remains dependent on inevitable subjective value judgments. This is especially the case if there is uncertainty regarding the advantage and the effect on the civilian population.

It might, however, be possible that military and humanitarian experts work to identify indicators and criteria to evaluate the proportionality and to make the subjective judgments regarding proportionality slightly more objective. Until now, such suggestions were rejected by military lawyers who insist that it all depends on the circumstances and on good faith. This would need to change with autonomous weapons, since a machine needs clear criteria and a formula to calculate proportionality.

Feasibility of precautions

Finally, under IHL precautions have to be taken only if they are feasible. Naturally, the feasibility has to be measured in regard to the possibilities of those who plan and decide upon an attack or who execute it, not according to the feasibility for a certain kind of machine. This being said, automated weapons are able to take additional precautions, because the human life of the pilot or weapons operator is not at risk.

The feasibility of precautions evolves through experience. When precautions taken in the past proved to have been unsuccessful, that may imply the need to take measures (and belligerents have in my view an obligation to foresee pertinent procedures) to avoid such incidents in the future. It is essential to make sure that autonomous weapons can be recalled and reprogrammed quickly and that human beings monitor the development of that intelligence.

Concluding remarks

My basic conclusion is that the advantages and disadvantages of autonomous weapons under IHL all depend on the technicians working on them and on what they are able to produce. I do not think there is any rule of principle of IHL which prohibits the use of automated weapons. All depends on how able they are to respect the rules of IHL. For the time being, and pending evidence of revolutionary technical developments, it may be wise to limit the use of automated weapons to situations in which no proportionality assessment is needed and to fighting declared hostile forces in conflicts of a high level of intensity. For IHL to be respected, it seems it will still be some time until autonomous weapons can be used in counterinsurgency operations.

About the author

Marco Sassòli, is professor of international law and Director of the Department of international law and international organization at the University of Geneva. From 2001-2003, Marco Sassòli has been professor of international law at the Université du Québec à Montreal, Canada, where he remains associate professor. He is also associate professor at the University de Laval, Canada and chairs the board of Geneva Call, an NGO engaging armed groups to respect international humanitarian norms.

Marco Sassòli graduated as doctor of laws at the University of Basel (Switzerland) and was admitted to the Swiss bar. He has worked from 1985-1997 for the International Committee of the Red Cross at the headquarters, inter alia as deputy head of its legal division, and in the field, inter alia as head of the ICRC delegations in Jordan and Syria and as protection coordinator for the former Yugoslavia. During a sabbatical leave in 2011, he joined again the ICRC, as legal adviser to its delegation in Islamabad. He has also served as executive secretary of the International Commission of Jurists and as registrar at the Swiss Supreme Court.

Marco Sassòli has published on international humanitarian law, human rights law, international criminal law, and the sources of international law and state responsibility.

Notes

[1] Human Rights Watch, *Losing Humanity: The Case against Killer Robots*, p. 38.
[2] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977.