Scienza e Pace

Science & Peace

ISSN 2039-1749 Vol. XI, No 2 (2020)

To kill or not to kill?

Do autonomous weapons respect IHL when taking life and death decisions over humans?

Marina Del Greco

Online Journal of the "Sciences for Peace"

Interdisciplinary Centre - University of Pisa

This paper has been refereed through double-blind peer review

Received: 18 October 2019
Accepted: 15 November 2020

To cite this article:

Del Greco, M. (2020) *"To kill or not to kill? Do autonomous weapons respect IHL when taking life and death decisions over humans?"*, *Scienza e Pace*, XI (2), pp. 37-57.

To kill or not to kill?

Do autonomous weapons respect IHL when taking life and death decisions over humans?

di **Marina Del Greco** *

Abstract

Since 2016, a Group of Governmental Experts on emerging technologies have been meeting in Geneva, to discuss the latest developments in the field of autonomous weapons. The peculiarity of these armaments is that not only they move around, without the assistance of a remote pilot, but that they can identify targets and even engage them, using lethal force, without the need to receive inputs, nor authorisations, from human operators. The possibility that machines take alone life and death decisions over humans raises several concerns, related in particular to their capacity to abide by International Humanitarian Law (IHL). Indeed, the application of IHL norms entails the ability to carry out complex evaluations, based on shifting factors, that appear too complex for robots, considering current technological advancements. Nevertheless, on the other hand, some characteristics of autonomous weapons make them the best option to minimise collateral damages. These contradictions originated a heated debate, which is still ongoing, regarding the advisability of a ban on autonomous weapons. This study aims at identifying the core issues linked with the employment of autonomous weapons, from an IHL perspective and at analysing the feasibility of possible solutions.

Key words: IHL, International Humanitarian Law; Autonomous weapons; Killer robots; LAWs

Abstract

Ogni anno, un gruppo di esperti di tecnologie emergenti si riunisce a Ginevra, per discutere degli ultimi sviluppi nel campo delle armi autonome. La peculiarità di tali armamenti è che, non solo sono in grado di spostarsi senza la necessità di essere pilotati da remoto, ma anche che possono identificare ed attaccare degli obiettivi, senza bisogno di ricevere alcun'autorizzazione da parte di un operatore umano. Chiaramente, il fatto che dei robot possano prendere autonomamente decisioni di vita o di morte su degli esseri umani pone una serie di interrogativi, relativi soprattutto alla loro compatibilità con il Diritto Internazionale Umanitario. Difatti, le norme in questione impongono di effettuare una lunga serie di valutazioni, basate su fattori mutevoli, che appaiono troppo complesse per essere effettuate da delle macchine. D'altro canto, alcune delle caratteristiche distintive delle armi autonome le rendono lo

* **Marina Del Greco** holds a Master degree in Law. This paper has been developed during the Master in Human Rights and Conflict Management the author attended in 2018-19 at the Scuola Superiore Sant'Anna, Pisa. Email: marina.dg1992@gmail.com

strumento più efficace per minimizzare le probabilità di causare danni collaterali. Da tali contraddizioni è scaturito un acceso dibattito, ancora in corso, circa la possibilità di bandire l'utilizzo dei suddetti armamenti. Il presente studio si propone di identificare, dal punto di vista del diritto umanitario, i principali problemi connessi con l'utilizzo delle armi autonome, analizzando anche la fattibilità di eventuali soluzioni.

Parole chiave: Diritto Internazionale Umanitario; Armi autonome; Principio di distinzione; compatibilità

1. Introduction

Since when primitive humans started hunting or fighting the first battles against each other, the possession of weapons capable of hitting a target from afar, which reduced the risks for the ones employing them, represented a significant advantage. For this reason, humankind has invested its resources in the development of spears, bows, catapults, rifles, missiles etc. Nowadays, this trend has not changed. However, a significant part of States' military budget is now allocated also to the education and the lifelong learning necessities of their soldiers. Consequently, if in the past ground troops were expendables, modern militaries are highly specialised professionals, who benefitted from a costly formation, at the expense of their State. As a result, they became a precious investment for their governments, something to protect from danger, whenever possible.[1]

However, the paradoxical idea of keeping soldiers away from dangers clashes with the need to defeat elusive enemies, like modern terrorist organisations. In fact, to achieve this goal, it is fundamental to rely on a large quantity of high-quality intelligence data. Such information, however, can be acquired exclusively by getting really close to the enemy, after long, risky, monitoring operations. Eventually, this loop was broken by the invention of autonomous weapons. As it will be subsequently explained in more details, these weapons are essentially robots capable of performing reconnaissance missions, defensive tasks and even offensive operations, without the presence of a human being on board and even without the need to be piloted from afar.

The introduction of this new technology raised several ethical, political and legal concerns in the international community. This study is focused on the latter and, more specifically, on autonomous weapons' compliance with International Humanitarian Law (IHL). Following a preliminary clarification of the terminology used in this field, the main features of autonomous weapons will be defined. The study will then proceed with a brief description of illustrative kinds of autonomous weapons, to better comprehend how these systems work in practice. Subsequently, the main IHL principles will be briefly recalled, highlighting

1 This tendency was reinforced, during the nineties, by the Gulf war, which spread an expectation for short and almost casualties-free wars among westerns (Birkeland 2018, 2).

the aspects that are more likely to be endangered by the employment of these robots. Eventually, the opportunity to ban autonomous weapons will be evaluated, basing on the profiles that present criticalities and on possible ways to overcome them.

2. What makes a weapon "autonomous"?

Colloquially, the words "unmanned" and "autonomous" are sometimes used interchangeably, although the two concepts present important differences. Similarly, the entire category of unmanned weapons seems to comprehend only drones, such as the famous "Predator" or "Reaper". Consequently, before diving into the legal and ethical considerations related to the employment of lethal autonomous weapons (LAWs), it appears fundamental to proceed with a preliminary clarification of the terminology that will be used in this paper.

2.1 The Concept of man "in, on and out" of the loop

The notion of man "in, on and out" of the loop of the decisions related to targeting and attacking allows scientists to classify robotic systems into three groups, according to their degree of independence from human control.

The first category includes weapons that need the presence of a man "in the loop", meaning that they require a real-time authorisation, from a human being, to perform at least some of their tasks (typically the offensive ones). Drones and unmanned aerial vehicles in general are part of this category. In fact, they can be controlled remotely, or pre-programmed, to perform different activities (e.g. locating a target, flying over it, acquiring images etc.). However, they cannot fire a missile without a real-time input from a human, who "pushes the button", authorising the attack.

The situation is slightly different in relation to the robots that belong to the "man on the loop" class. These weapons can theoretically perform all their tasks independently, including the engagement of targets, without the need to receive any inputs. Nevertheless, they can be overridden in any moment by a human operator; hence, they are also referred to as "human-supervised" weapons. Finally, the armed systems designed to execute all their tasks without human inputs, nor supervision, compose the "man out of the loop" category. (Melzer 2013, 6).

2.2 Automaticity vs. autonomy

Robotic systems that are part of the "man out of the loop" group can be further subdivided into automated weapons and fully autonomous ones. Although the two terms are often confused and improperly used as synonyms, the adjective "automated" comes from the ancient Greek term "automa". An *automa* was a machine that, thanks to hidden mechanisms, seemed capable to move, as if it was "alive". The notion of autonomy, instead, comes from the words “auto” i.e. “self” and “nomos” that means “law”, which together entail something’s capacity to create its own regulations and act according to those.

Therefore, while the concept of “automaticity” consists in the ability to perceive the surrounding environment and carry out specific tasks, as predetermined responses to the inputs received from the setting, the notion of “autonomy” is more advanced. Fully autonomous weapons can, in fact, face also unforeseen circumstances, as they are capable of “learning” and adjusting their goals and reactions also to unexpected situations. As a consequence, both automatic and autonomous systems can move, identify targets and attack them, without human authorisation. However, only fully autonomous weapons can carry out these tasks in uncircumscribed environments. (Birkeland 2018)

Accordingly, it is fair to say that all autonomous entities are also automatic, but not vice versa, as the idea of autonomy encompasses the one of automaticity, requiring in addition also the ability to respond to unpredictable events. Essentially, autonomous systems can be seen as an evolution of “simple” unmanned weapons: revolutionary warlike tools capable of taking life and death decisions over humans, alone. Indeed, this aspect is the main reason why autonomous systems’ compliance with IHL entails several issues.

3. Examples of autonomous weapons

Following these preliminary clarifications, to fully understand the concerns raised by the employment of autonomous weapons, it is necessary to better define their characteristics and especially their functioning. During an armed conflict, these robotic systems can have countless applications: from mere surveillance and reconnaissance missions, to site-defence systems, offensive artillery, anti-personnel weapons etc. However, for the purpose of this study, it will be sufficient to briefly illustrate how three different kinds of autonomous weapons operate.

3.1 Offensive systems: the Harpy

The Harpy is a peculiar kind of drone specifically designed by Israel to destroy enemy’s anti-aircraft defences. The weapon can stay in the sky for up to 9 hours, at a maximal altitude of 4.500 meters, flying in a pre-determined zone, called “loitering area”. From there, the drone scans the territory underneath, looking for enemy radars’ emitters. Once one is detected, the Harpy engages it, without the need to receive any authorisation, nor other inputs, from a human operator. The attack consists in vertically diving onto the target, with an accuracy of 1 meter, in order to strike it with a 16 kg warhead.

Besides the ability to autonomously select and engage its objectives, what distinguishes the Harpy from a missile, as well as from any other weapon, is the ability to react if the radar is suddenly shut down. In fact, in such a circumstance, the drone reacts automatically, by aborting the mission and flying back to the loitering area. There, it remains a constant threat for the enemy, until when a radar is activated again and, consequently, autonomously re-engaged (IAI Website).

Lastly, it is worth mentioning that, although the Harpy might be mistaken for an advanced guided munition, i.e. "*explosive projectiles that can actively correct for initial-aiming or subsequent errors by homing in on their targets or aim-points after being fired, released or launched*",[2] this weapon it is indeed an autonomous one and not a simple "smart bomb". In fact, before the mission, the human operator identifies only the Harpy's loitering area, but not its final target.

3.2 Defence systems: the Iron dome, the Phalanx and APS

Lately, as a response to the spread of unmanned aerial vehicles, also the demand for anti-drone technology experienced a significant growth. Modern defence systems are equipped with radars that enable them to identify approaching objects and, hence, potential threats. When something is detected, the first thing that the defence system needs to do is to establish whether the object is a threat or not.

To do so, it analyses its shape, speed and trajectory. To avoid incidents, defence systems usually have access to databases with information on civilian flights and feature special software capable of identifying pre-determined kinds of flying objects, labelling them as "not-to-attack". Anything that does not fall in one of these two categories (i.e. civilian aircrafts or other friendly items) is considered dangerous and automatically attacked. If multiple incoming threats are detected, the systems carry out an assessment and decide which target is to be prioritised. Usually, the one closer to the area they are supposed to protect is also the first engaged (Boulanin et al. 2017).

Normally, these defence systems operate autonomously, although they could also be set to require the authorisation of a human operator. Nevertheless, considering the need to take action as soon as an aerial threat is spotted, for the defence to be effective, these systems operate almost exclusively in fully autonomous mode. The potential anticipated collateral harm is allegedly justified with the need to prevent the greater damages that an air strike could inflict (Melzer 2013, 12).

One of the most famous defence systems is the Israeli "Iron Dome": a short-range counter-rocket, artillery and mortar (C-RAM) shield, capable of shooting down helicopters, aircrafts and drones (Ioanes 2019). The system was first launched in 2011 and it originally consisted of 5 gun-batteries; however, over the years, it has increased its components, doubling them. Every battery is made of three parts: a radar that identifies and tracks approaching objects, a control system that verifies whether they represent a threat or not and a missiles-launching station.

2 Boulanin et al. (2017), 47.

Once the control system detects a danger, it calculates how to intercept it with a missile and launches the counteroffensive (Jewish Virtual Library 2019).

Another system, very similar to this one, but specifically designed to be employed on war ships, is the American "Phalanx". It works in the same way the Iron Dome does, but to save space on the ship, it combines the radar, the control centre and the artillery in a single unit, instead of having three separate elements. [3] The Phalanx can detect a potential threat from a distance of 18,5 km. To evaluate whether it is dangerous or not, a software tracks it, assessing whether the object is approaching or moving away from the ship, as only incoming items must be considered. Then, the Phalanx proceeds to calculate the exact trajectory of the potential ammunition, to verify if it will actually hit the ship. Eventually, the computer determines the speed of the object, as the Phalanx only engages items with a speed ranging between a minimum and a maximum level (manually set by the operators). If the flying object satisfies all these criteria, the computer identifies it as a target, keeps tracking its path and assigns it a certain priority (in case of multiple threats). When the target is 3,7 km away, the Phalanx engages it and keeps shooting at it, until the object changes its flying path or disappears from the radar (Stoner 2017). While such a system seems infallible, there have been several reports of malfunctioning and collateral damages, including Phalanxes shooting at friendly ships' ammunitions (Miller 1989; Press 2017).

Lastly, it is important to underline that not all defence systems resemble the ones described above. In fact, there are also much smaller devices, designed to be mounted on vehicles, called "Active Protection Systems" (APSs). The task of these systems is to protect tanks from missiles and rockets, reducing the need for thick and bulky armours and hence improving vehicles' manoeuvrability. APSs operate in the same manner the Iron Dome and the Phalanx do. They feature a radar and infra-red or ultraviolet sensors to detect incoming threats, a control system to identify them and a various kinds of ammunitions to neutralise incoming threats. In particular, APSs may fire projectiles against approaching objects, or be equipped with special devices capable of disrupting the navigation systems of missiles and other ammunitions. In both cases, the aim is to prevent enemy weaponry from hitting the intended targets (Boulanin et al. 2017, 41).

Theoretically, APSs can be overridden by the human operators present in the vehicle; however, they usually operate in fully autonomous mode, identifying and engaging targets without any kind of human authorisation. The reason is the same one explained in relation to bigger defence systems: a human input would waste too much time, nullifying the

3 More information on the functioning of this autonomous weapons are available on the website of its producing companies. General Dynamics – Ordnance and Tactical Systems, '[Phalanx CIWS](#)'. Raytheon Company, '*[Phalanx Close-In Weapon System](#)*'.

chances to intercept incoming artillery.[4] Clearly, this aspect raises concerns regarding the collateral damages that APSs could potentially inflict. In fact, while an unsupervised employment of instruments that disrupt the navigation systems of incoming missiles do not represent an issue, counteroffensive weapons could obviously make civilian casualties, in their attempt to destroy enemy ammunitions (Feickert 2016, 8).

3.3 Anti-personnel autonomous weapons: sentry robots

The last example of autonomous weapons that is necessary to analyse is constituted by sentry robots. These peculiar machines do not have the task to destroy enemy's structures, nor to protect something from incoming artillery, but are, instead, employed as anti-personnel weapons. Depending on the type, they can be fixed structures or portable gun turrets mounted on vehicles and sent on patrol. Currently, these defence systems are not widespread and, so far, only two countries use them: Israel and South Korea. The first employs a system, called Sentry Tech, on the border of the Gaza strip, while the second developed the "SGR-A1" and placed it in the demilitarized buffer zone, between itself and North Korea. Both systems are equipped with cameras, motion sensors and other tools that enable them to spot humans up to a distance of 3 km (Boulanin et al. 2017, 41). Korean robots, in particular, would be able to detect also explosive devices wore on, or even under, the clothes. When one of these systems detects a person, it alerts a human operator, who can use the robot's microphone and speakers to interact with the person and ask for identification. Depending on the answer, the operator can then decide to let the person go, or authorise the sentry robot to fire rubber bullets or grenades (Prigg 2014).

Since the Sentry Tech and the SGR-A1 clearly fall in the "man in the loop" category, it could be assumed that Israel and South Korea did not develop more advanced models. Truth is that Israel produced also an armoured unmanned vehicle, equipped with both lethal and non-lethal weapons, called "Guardium". Such robot is usually sent on roads where it is common to find improvised explosive devices (IEDs), to perform routine reconnaissance missions. However, it can also be programmed to autonomously react to a wide variety of potential scenarios, which it could encounter during patrols (Melzer 2013, 12). Similarly, after the successful employment of the SGR-A1, South Korea decided to design a fully autonomous version of its sentry robot, called "Super Aegis II". After having warned intruders to go away with a pre-recorded message, the Super Aegis can shoot them, without the need for human authorisation. Nevertheless, after several requests from its clients, the producing company introduced the possibility to insert limitations to the autonomy of

4 Most sophisticated APSs have a reaction time of 0,0005 seconds, which enables them to intercept incoming rockets launched from as close as 10 meters away. A delay of less than 0,3 seconds in their response, would make it impossible for them to intercept any anti-tank missiles launched from less than 400 metres away. (Feickert 2016).

the robot, which can now be set up to require a password to authorise every shot (Maslen et al. 2018, 40).

4. Legal Framework

After having examined the functioning of autonomous weapons, it is essential to analyse the relevant legal framework. Obviously, the most pertinent branch of law is IHL: the part of international law that regulates armed conflicts. The primary purpose of this set of norms is the protection of civilians, prisoners of war, soldiers *hors de combat* and, in general, of all the people that do not, or no longer, take part in hostilities.

IHL's most important principles are contained in the four Geneva Conventions of 1949 and in the two additional protocols of 1977, related to the protection of victims of armed conflicts. In addition to these norms and to the ones contained in other treaties, devoted to specific issues, IHL is also composed by customary law: a peculiar set of obligations that are binding on all belligerent parties, regardless of the international agreements they have ratified. The ratio of this peculiarity resides in the fact that customary law protects values that are considered particularly relevant by the international community as a whole, drawing its peremptory character, not from written conventions, but from practices consistently followed by states, from a sense of legal obligation (Guzman 2005).

Since IHL is designed to be applied during armed conflicts, the slightest variation in the meaning attributed to a norm might imply the possibility to lawfully hurt, or even kill, an individual. Hence, IHL's dispositions should ideally be clear, straightforward and phrased in a way that leaves little discretion to the interpreter. Nevertheless, there are also IHL rules formulated on purpose in general terms, to be applied also to cases not specifically regulated elsewhere. An example is the Martens clause,[5] first appeared in the preamble to the 1899 Hague Convention, then granted the status of customary law (ICJ 1966). The clause "*has proved to be an effective means of addressing the rapid evolution of military technology*",[6] as it basically states that anything not expressly forbidden is not automatically to be regarded as lawful (Singh et al. 1989). This concept implies that a specific ban on new weapons (e.g. autonomous ones) is superfluous: if they do not comply with IHL principles, their employment is straightaway unlawful, no matter if they have never been formally banned (Egeland 2016).

5 The Martens Clause reads: *"Until a more complete code of the laws of war has been issued, the High Contracting Parties deem it expedient to declare that, in cases not included in the Regulations adopted by them, the inhabitants and the belligerents remain under the protection and the rule of the law of nations, as they result from the usages established among civilized peoples, from the laws of humanity and the dictates of public conscience."*

6 ICJ, Legality of the threat or use of nuclear weapons (1996), para 78.

4.1 Means and methods of war

Article 35 of the First Protocol Additional to the Geneva Conventions introduces the concepts of "means" and "methods" of war. [7] Essentially, the norm states that combatants can use whatever means (i.e. weapons) and methods (i.e. tactics) to fight each other during the war, provided that they respect IHL's dicta. According to art. 36, if new means or methods of war are developed, the State that intends to use them must first ascertain whether their employment would contravene IHL.

Typically, armaments and strategies are considered incompatible with the law of armed conflicts if they cause unnecessary sufferings, or if they end up hurting also protected subjects. Notwithstanding this, the vast majority of armaments is not per se incompatible with IHL, as most weapons do not have physical characteristics that render them absolutely incapable of abiding by the requirements of humanitarian law. Most commonly, their compliance depends, instead, on the way they are employed in practice. In other words, if a given armament possesses certain features, that make it incapable of complying with IHL requirements, it will never be possible to lawfully use that mean, no matter the context. On the contrary, if a weapon's capacity to abide by the law of armed conflicts depends on the practical circumstances of its employment, it will be necessary to meticulously regulate its use (e.g. by restricting the number of situations in which that weapon can be resorted to). Nonetheless, that weapon won't have to be necessarily banned.

To determine if there are circumstances in which autonomous systems could be employed lawfully, first of all, it is necessary to ascertain with which kind of offensive weapons they have been equipped with. For instance, if a blinding laser is mounted on an autonomous weapon, the whole robotic system will not comply with IHL, as it features a weapon that has already been banned.[8] Supposing that this is not the case and that the autonomous weapon is equipped with an armament that is not prohibited, it will then be necessary to further proceed with the analysis of IHL's principles, to try to identify other aspects that might be problematic in relation to the autonomous nature of the weapon.

4.2 Principle of Distinction

One of IHL's most important principles is the one of distinction, stated in article 48 of the first Additional Protocol to the Geneva Conventions. It obliges combatants to always distinguish between those who are taking part in hostilities and those who are not, as only the first are legitimate military targets, which can be lawfully attacked (provided that all other legal requirements are respected). This category comprehends soldiers, military premises, military vehicles etc. On the contrary, non-combatants

7 "*In any armed conflict, the right of the Parties to the conflict to choose methods or means of warfare is not unlimited*" First Protocol Additional to the Geneva Conventions of 12 August 1949, article 35.

8 "*It is prohibited to employ laser weapons specifically designed, as their sole combat function or as one of their combat functions, to cause permanent blindness*"(Protocol on Blinding Laser Weapons 1995).

such as civilians, medical and religious personnel, people *hors de combat* and civilian infrastructures must never be targeted, as they enjoy the protection granted by IHL.

Therefore, pursuant to the principle of distinction, it is forbidden to use weapons that cannot distinguish between military targets and protected subjects, as they would not be able to direct their offensive power only against the first. Similarly, weapons capable of discriminating might still be prohibited, if the effects of their attacks cannot be limited exclusively to the targets, because too widespread in space or time and end up affecting also protected people with collateral damages. For instance, biological and chemical weapons were prohibited precisely because there was no effective way to limit their impact (UN General Assembly 1969). Attacks of this kind are called "indiscriminate attacks" since, despite having been launched against lawful targets, they eventually affect both military and civilian subjects and infrastructures, indiscriminately.[9]

To understand if autonomous weapons abide by the principle of distinction, it is necessary to ascertain whether they have the capacity to target a specific objective and strike it accurately.[10] Then, assuming that they can be aimed precisely against a target, it is essential to assess the range of their attack's effects. Undoubtedly, it is difficult to deny that these armaments are more precise in targeting than any other weapon. As they do not need to rest and do not endanger soldiers' lives, these robotic systems can get really close to their targets and observe them for a longer amount of time, collecting greater amount of data and better intelligence, which, in turns, diminishes the probability of committing mistakes.

On the other hand, some experts believe that autonomous weapons are not technologically advanced enough to effectively distinguish between combatants and protected subjects in every situation. (Melzer 2013, 29). For instance, modern conflicts present complex scenarios, where combatants try to mix with civilians on purpose, to benefit from their protected status. In some regions, guerrillas do not wear uniforms and carrying arms openly is a common behaviour, also for civilians. In such contexts, autonomous weapons would need to rely on a very sophisticated software to be able to recognise civilians and spare them.

9 In particular, indiscriminate attacks have been described as "*Those which employ a method or means of combat the effects of which cannot be limited as required by this Protocol; and consequently, in each such case, are of a nature to strike military objectives and civilians or civilian objects without distinction*" Additional Protocol I, article 51.

10 This criterion has been quoted several times, including in the advisory opinion on the legality of the threat or use of nuclear weapons, where Judge Higgins stated that: "*a weapon will be unlawful per se if it is incapable of being targeted at a military objective only*" (International Court of Justice, "*Advisory opinion on the Legality of the Threat or Use of Nuclear Weapons*", 8th July 1996, Judge Higgins dissenting opinion, 366).

Regarding the range of their attacks' effects, autonomous systems usually employ very precise and accurate ammunitions. Among them it is possible to mention laser-guided bombs, GPS equipped ones or e-bombs, which, thanks to the creation of an intense electromagnetic field, are capable of disabling or destroying electronic circuits, without endangering humans or buildings (Zohuri 2019, 271).

From these considerations, it follows that autonomous weapons possess some features that minimise the risk of collateral damages; however, the difficulties linked with the identification of civilians in certain contexts render them unreliable, especially when the separation between civilians and combatants is blurred. Consequently, according to some scientists, only human beings can understand the intentions of other humans and, hence, properly distinguish between lawful targets and protected subjects in every situation (Human Rights Watch 2012; Szpak 2019). According to others, IHL criteria are objective and can be programmed into a computer (Sassoli M. Interview at Panel Discussion on the Challenges of New Technologies in Warfare 2014). Hence, if technologically advanced enough, autonomous weapons might be able to abide by IHL.

In conclusion, it is not possible to state whether autonomous weapons fully comply with the principle of distinction, as it depends on their degree of technological advancement, on their equipment, on the kind of missions they are supposed to perform and especially on the way they are programmed.

4.3 Principle of proportionality

The second principle that it is necessary to examine is the one of proportionality. Its application entails the execution of a prognostic judgement, aimed at weighing the benefits and the harms that would originate from an attack. In other words, this precept requires commanders to try to foresee the entity of a strike's potential collateral damages and then, to compare it with the relevance of the military advantage they expect to gain. The greater the presumed utility of an attack, the more tolerance towards collateral casualties. On this regard, it is important to specify that the incidental harm should not be "*excessive in relation to the concrete and direct military advantage*" predicted for that specific attack. [11] This means that, while balancing the positive and the negative outcomes of a strike, it should be weighted only the military advantage that would be obtained thanks to that single attack, without considering its impact on the whole war. Additionally, the principle states that, while planning a strike, commanders must evaluate all available options, in terms of modalities, weapons' choices and timing of the offensive and that all precautions should be taken (e.g. releasing evacuation warnings for the civilian population) (Sehrawat, 2017,178).

11 Additional Protocol I, article 51 (5) (b).

In the light of these reflections, do autonomous weapons comply with the principle of proportionality? Considering their current characteristics, it seems unlikely that they could make such complex assessments on their own. In addition, commanders would not be able to take precautions (e.g. by releasing evacuation warnings), as they would not be able to tell exactly where, or when, the autonomous weapon would strike. (Davidson, 2018,7).

Nevertheless, these considerations do not imply that autonomous weapons are per se incompatible with the principle of proportionality. In fact, by accurately defining their operational areas, their inability to perform a proportionality analysis could be bypassed, lifting from them the need to carry out the assessment and placing it onto their human operators. For instance, by employing autonomous weapons exclusively in uninhabited zones, where the probability of causing collateral damages is minimal, the robots would not even need to assess the proportionality of their attack right before performing it. In fact, in such a case, the assessment would have been already done by the human officer who decided to employ an autonomous weapon in that area. Essentially, as it happens for the employment of all other, non-autonomous, weapons.

4.4 Prohibition to cause unnecessary suffering

This principle, already stated in the Preamble of the 1868 St. Petersburg Declaration, in the Hague regulations of 1899 and in article 35(2) of Additional Protocol I, is now considered customary law. It restricts adversaries' freedom of choice, regarding means and methods of warfare, by forbidding combatants from employing weapons, or tactics, designed to inflict superfluous or unnecessary sufferings to the enemies. In other words, although it is almost inevitable to injure or kill adversaries taking part in hostilities, IHL prohibits to make them suffer beyond what is strictly necessary to put them *hors de combat*. This principle, therefore, provides some kind of protection to all those individuals that are identified as legitimate military targets.

The problem with this precept is that it is quite difficult to establish whether a mean or method of warfare causes unnecessary suffering, as the discriminating factor resides in its ability to serve a specific military purpose, rather than simply injuring the enemy. The International Court of Justice simply defined unnecessary suffering as a "*harm greater than that unavoidable to achieve legitimate military objectives*".[12] Consequently, some States deduced that this prohibition requires to carry out a judgement similar to the one prescribed by the proportionality principle. In particular, a weapon would not violate this precept, if its use leads to a legitimate military advantage that outweighs the sufferings that it is expected to cause to enemy fighters. The intensity of the suffering

12 International Court of Justice, "*Advisory opinion on the Legality of the Threat or Use of Nuclear Weapons*", 8th July 1996, Judge Higgins dissenting opinion, para 78.

should be calculated basing on the level of pain, the eventual permanent health problems, and the likelihood of death. (Geneva Academy, 2017). Additionally, the potential availability of alternative means or methods of warfare should also be taken into consideration, during the assessment. (ICRC, 2019).

Due to the prohibition to cause unnecessary suffering, the use of several weapons has been prohibited or restricted. Although there is not unanimous consensus, among forbidden weapons it is possible to find lances or spears with a barbed head, explosive and expanding bullets (also known as “dum-dum” bullets), poison and poisoned weapons, biological and chemical weapons, bullets with fragments that cannot be detected by X-rays, incendiary weapons, blinding lasers and nuclear weapons. (Henckaerts et al., 2005, 243). Nowadays, there are no explicit references to unmanned or autonomous weapons compliance with the prohibition to cause unnecessary suffering. This uncertainty is most likely due to the difficulties linked with the identification of a precise and stable threshold. In fact, according to the legal test described above, to establish if autonomous weapons could lawfully be employed, it would be necessary to calculate the suffering that the victims of the attack would endure and the likelihood of their death. However, it would be extremely difficult to rate the intensity of pain suffered by someone or to predict victims’ survival chances, as the simple advancement of medical techniques constantly increases them, turning banned weapons into acceptable ones.

Furthermore, assessing the compliance of autonomous weapons with the principle of unnecessary suffering is more complicated than assessing the lawfulness of any other weapon. This is due to the fact that a commander has to make all the calculations illustrated above, as soon as he decides to employ autonomous weapons, i.e. a lot before the actual strike. This means that, right before the attack, the situation could radically change, without the possibility for the commander to modify autonomous weapons’ directives; consequently, to be lawfully employed, autonomous weapons should feature an extremely reliable and predictable operativity, even in unexpected situations (Davison, 2018, 10).

4.5 Principle of necessity

The prohibition of unnecessary sufferings examined above is strictly related to the idea of military necessity, as IHL essentially tries to balance the opposite pushes of these two concepts. On this regard, the principle of necessity states that, to accomplish a legitimate military objective, combatants can employ whatever means and methods of warfare, provided that they are not otherwise prohibited by IHL. In armed conflicts, the only legitimate purpose is to weaken the military power of the enemy and not to completely defeat it (ICRC, 2015, 6). Nevertheless, an attack against protected subjects could never be justified claiming that it was useful to weaken the enemy.

In this respect, article 52(2) of Additional Protocol I specifies that:

Attacks shall be limited strictly to military objectives. In so far as objects are concerned, military objectives are limited to those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage”.[13]

Consequently, autonomous weapons should be programmed in a way that enables them not only to distinguish between civilians and combatants, but also to understand the difference between an improvised military camp and a shelter for displaced civilians. Additionally, their software should be capable to respect the rule set forth in article 52 para 3, meaning that, in case of doubts whether an object usually designed for civilian purposes is actually used for military ones, it should always presume that it is not and make the robot refrain from striking.

5. Concluding remarks

Although the debate on the possibility to ban autonomous weapons has become more and more relevant lately, the issue of their compliance with IHL has not yet received a definitive answer. Contrary to “simple” unmanned weapons, which feature the presence of a human operator who, although from afar, can (and must) make all the evaluations prescribed by the law, to ensure that the attack is not disproportionate and does not lead to civilian casualties, autonomous weapons present several issues, that question their conformity with the law of armed conflicts.

As first thing, since it wouldn’t be a human to fire the weapon, the timing and location of the attack would not be predictable, making it impossible to issue evacuation warnings or to abort the mission if, at the very last moment, it is discovered that the attack would be indiscriminate, disproportionate or would cause excessive suffering to its victims. Then, leaving aside the ethical concerns related to the possibility that machines take life and death decisions over humans, most of the other arguments leverage on the allegedly insufficient technological advancement of these weapon systems (Human Rights Watch 2012).

Undoubtedly, to distinguish protected subjects from combatants, especially in situations where the latter try on purpose to disguise as civilians, autonomous weapons should possess an extremely sophisticated software. In addition, even supposing that an autonomous system could actually understand if a soldier is only pretending to be a non-combatant, there would still be the danger that the robot could slowly start to identify real civilians as fighters, due to the way machine learning works. In fact, to be able to react unassisted also to unexpected situations, autonomous weapons are provided with a sort of database,

13 Additional Protocol I, article 52(2).

containing some examples on how to behave in different situations. Over time, the robots add new examples, based on their experiences, developing new behavioural patterns. (Feldman et al. 2019). What would happen if an autonomous weapon understood that, sometimes, even individuals which seem to be civilians, are actually lawful targets? In doubtful contexts, it might start attacking civilians, contravening to the prescriptions of article 50 para 1 of the first Additional Protocol.[14]

Furthermore, even without picturing such a complex scenario, machines might not be able to tell the difference between an active combatant and a soldier who wants to surrender. On this regard, humanitarian law prescribes that anyone who is defenceless, because of unconsciousness, wounds or sickness, or who has clearly expressed the intention to yield, must be granted protection, provided that he/she does not try to escape or commit hostile acts.[15] It is not difficult to imagine that, thanks to biometric sensors, autonomous weapons might be able to recognise an unconscious person and, perhaps, even a seriously wounded or sick one. Consequently, they should have no problems in realising that these subjects, although belonging to enemy troops, should be spared. Probably, robots could even be programmed to register and communicate the location of injured soldiers to the Red Cross or to similar organisations, facilitating rescuing operations. However, it does not appear as easy to imagine autonomous weapons recognising the intention to surrender of a healthy soldier, who begs for mercy in an unknown language, wearing an enemy uniform and maybe with his gun still at hand.

This inability to perceive the will to yield, would almost inevitably lead to violations of the prohibition of denial of quarter, which requires combatants to remain receptive to a declaration of surrender (Gill et al. 2015). Of course, this prescription does not entail that only means and methods of war capable of capturing prisoners alive comply with IHL, nor that enemies must always be provided with a chance to surrender, before a lethal attack. However, employing robots that are not capable to understand that a person wants to surrender is equal to ordering a platoon that there shall be no survivors.

On the other hand, it is also true that autonomous weapons would not feel hate nor resentment against the enemies. Hence, robots could somehow act more humanely than real soldiers, considering also that they would not be pushed by the self-preservation instinct, or by fear of reporting damages. Indeed, such emotions sometimes led to the commission of mistakes that autonomous systems would not have made. An example is the case of the USS Vincennes, a military ship that was notified by its defence system that an approaching flying object was a civilian aircraft. The system, after having correctly identified the airplane

14 The article states that "*In case of doubt whether a person is a civilian, that person shall be considered to be a civilian*".

15 This prescription is contained in common article 3 of the Geneva Conventions and in article 41 para 2 of Additional Protocol I.

as civilian, disregarded it. However, the officers on board of the ship were fearing an attack and, pushed by the stressful situation, forced the system to shoot down the plane, killing all the passengers and the crew (Rochlin 1991). This case shows that, sometimes, autonomous systems provide better protection to non-combatants than humans.

Nevertheless, the very same episode imposes a reflection on the issue of accountability. In fact, while in this case it is fairly easy to identify the individual responsible for the accident, who could have been held accountable, if it had been a fully autonomous system the one that committed a mistake, causing several victims? Clearly, the fact that autonomous weapons operate without human inputs, does not imply the ability to bear responsibility for their own actions.

A first solution to this accountability gap could consist in holding accountable the weapons' producing company or the State which employed them. However, no action for individual criminal responsibility could be brought against these subjects, as they are not individuals. Then, it could be argued that the engineers who programmed the weapons must be personally responsible in case of IHL violations committed due to malfunctions. While, prima facie, this might appear as an acceptable solution, it is necessary to consider that autonomous systems operate thanks to complex lines of coding, written by teams of people in which no single programmer has a clear overview of how the whole software will operate in the end. This circumstance represents an insurmountable obstacle for the attribution of criminal responsibility, as the latter only plays a role when there is something an individual can be blamed for (Heyns 2016). The last option is based on the concept of command responsibility, which implies that a commander has criminal responsibility for the crimes committed by his subordinates if he knew, or should have known, that such crimes were being committed and he did not take all feasible measures to prevent or repress such conducts.[16] Consequently, if it was established that autonomous weapons must be considered as human soldiers, commanders would bear full responsibility for all the breaches committed by these machines (Press 2017). Nevertheless, if the violations were unpredictable, because committed due to sudden malfunctioning or enemy cyber-attacks, it would be unreasonable to punish the commanders.

As a result, depending on the circumstances, it is sometimes impossible to identify an accountable person for the behaviour of autonomous weapons, creating an accountability gap that contravenes to one of IHL's fundamental principles (Cavallaro 2012).

All things considered, it is fair to say that all the issues that undermine autonomous weapons' capacity to comply with IHL can eventually be linked to one core problem: their inability to accurately distinguish combatants from protected subjects, in every situation. However, in the future, thanks to technological progress, robots' discrimination skills will

16 See Rome Statute of the International Criminal Court, art. 28; Additional Protocol I, art 86 (2).

probably improve significantly, perhaps even allowing them to determine the status of an individual better than humans.

Should, then, autonomous weapons be banned until that moment? Many academics support this idea, which seems to reflect also the position of the UN. In fact, while the 2013 and 2014 reports of the UN Special Rapporteur on extrajudicial, summary or arbitrary executions only called for a moratorium on autonomous weapons, the 2016 joint UN report on the proper management of assemblies, in para 67 (f), explicitly calls upon States to refrain from using systems that operate "without meaningful human control" to police assemblies (Kiai et al 2016). Notwithstanding this trend, at least in the immediate future, it seems unlikely that a binding prohibition of autonomous weapons will be adopted, considering the large amount of money invested in the development of these weapons.(Leys 2018).

Therefore, instead of arguing on the advisability of a ban on autonomous weapons, the priority should be the adoption of specific norms, aimed at meticulously regulating their use. For instance, if autonomous weapons were to be employed solely in areas where there are minimal possibilities of encountering civilians (e.g. the high sea or uninhabited zones), the probability of incidents would radically decrease. Alternatively, equipping sentinel robots exclusively with non-lethal arms would make eventual incidents non-fatal.[17] Thanks to measures like these ones, the risk of collateral damages would be significantly lowered and complying with the requirements set by the proportionality principle would become much easier. Furthermore, the absence of incidents would be a (provisory) patch also to the accountability gap problem.

To conclude, a total ban on autonomous weapons would certainly be the safest solution, but it is also a utopic one. IHL itself was created because it was recognised that simply prohibiting war was unrealistic and that laying down some ground rules, which guarantee minimal protection standards, was better than nothing. Similarly, the introduction of precise norms, aimed at regulating the employment of autonomous weapons, is probably the best result that it is possible to achieve at the moment.

The need to tackle this problem rapidly is emphasised by the fact that autonomous systems are registering rapid technological advancements lately. Although fully autonomous weapons (other than defence systems) are not yet widespread, the risk of ending up with an entire non-regulated warfare sector is an imminent one. Hence, the need for the legislation to keep up with scientific progresses and to adopt a legal framework to regulate the employment of autonomous weapons. Considering also that, ultimately, for the people affected by armed conflicts, the way an arm is used in practice has much more relevance than its nature's theoretical compliance with IHL, as *"[a] sword is never a killer*" but simply "*a tool in the killer's hand*".[18]

17 Some companies already produced robots capable of issuing warnings and, if ignored, of releasing tear gas or non-lethal electric shock or of spraying paint on protesters to mark them. (Heyns 2016, 360).

Bibliography

Birkeland J. (2018), “The Concept of Autonomy and the Changing Character of War”, *Oslo Law Review*, 5, 2.

Boulanin V., Verbruggen M. (2017), *Mapping the Development of Autonomy in Weapon Systems*', Stockholm International Peace Research Institute (SIPRI), November 2017.

Cavallaro J., Sonnenberg S., Knuckey S. (2012), *Living Under Drones: Death, Injury and Trauma to Civilians from US Drone Practices in Pakistan*, International Human Rights and Conflict Resolution Clinic, Stanford Law School, Stanford; NYU School of Law, Global Justice Clinic, New York.

Davison, N. (2018), "A legal perspective: Autonomous weapon systems under international humanitarian law", in UNODA Occasional Papers No. 30, *Perspectives on Lethal Autonomous Weapon Systems*, UN, New York.

Egeland K. (2016), “Lethal Autonomous Weapon Systems under International Humanitarian Law*”, Nordic Journal of International Law*, 85, 2 pp. 89 - 118.

Feickert, A. (2016), *Army and Marine Corps Active Protection System (APS) Efforts’,* Congressional Research Service Report for Congress R44598, Congress of the United States.

Feldman P., Dant A., Massey A. (2019), *Integrating Artificial Intelligence into Weapon Systems*, arXiv preprint arXiv:1905.03899.

First Additional Protocol to the Geneva Conventions of 12 August 1949, relating to the Protection of Victims of International Armed Conflicts (1977).

Florian R. (2003), *Autonomous artificial intelligent agents*, Technical Report, Center for Cognitive and Neural Studies (Coneural) Cluj-Napoca, Romania.

General Dynamics – Ordnance and Tactical Systems, *Phalanx CIWS*, General Dynamics Website. Available at: https://www.gd-ots.com/armaments/naval-platforms-system/phalanx/

Geneva Academy of International Humanitarian Law and Human Rights (2017), *Weaponslaw Encyclopedia. Available at:* http://www.weaponslaw.org/glossary/superfluous-injury-or-unnecessary-suffering.

Geneva Conventions of 1949.

18 Seneca, ‘*Ad Lucilium Epistulae Morales*’, Letter 87, (62-65 AC).

Gill T., Fleck D. (2015), *The Handbook of the International Law of Military Operations'*, Oxford University Press, Oxford.

Guzman A. (2005), "Saving Customary International Law", *Michigan Journal of International Law*, 27,1, pp. 115-176.

Hague Conventions of 1899 and 1907.

Henckaerts J., Doswald-Beck L. (2005), *Customary International Humanitarian Law. Volume I and II: Rules*, ICRC, Cambridge.

Heyns C. (2016), "Human Rights and the Use of Autonomous Weapons Systems (AWS) during Domestic Law Enforcement", *Human Rights Quarterly*, 38, pp. 350-378.

Human Rights Watch, 'Losing Humanity: The Case Against Killer Robots" (2012).

ICRC (2015), '*International Humanitarian Law: Answers to Your Questions*', ICRC, Geneva.

ICRC (2016), *Views of the ICRC on autonomous weapon systems*, paper submitted to the Convention on Certain Conventional Weapons Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), 11th April 2016.

ICRC, *Rule 70. Weapons of a Nature to Cause Superfluous Injury or Unnecessary Suffering*, IHL database. Available at: https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule70)

International Court of Justice (1996), *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion of 8th July 1996.

Ioanes E. (2019), "The US military is buying Israel's battle-proven Iron Dome that destroys rockets. Here's how it works", *Business Insider*, 14th August. Available at: https://www.businessinsider.com/how-the-uss-new-iron-dome-missile-defense-system-works-2019-8?IR=T

Israel Aerospace Industries (IAI), *Harpy*, IAI Website. Available at: https://www.iai.co.il/p/harpy

Jewish Virtual Library (2019), *U.S.-Israel Strategic Cooperation: Iron Dome Missile Defense System*. Available at: https://www.jewishvirtuallibrary.org/the-iron-dome

Johansen S. (2018), "So Man Created Robot in His Own Image: The Anthropomorphism of Autonomous Weapon Systems and the Law of Armed Conflicts", *Oslo Law Review*, 5,2.

Johansson L. (2018), "Ethical Aspects of Military Maritime and Aerial Autonomous Systems", *Journal of Military Ethics*, 17, pp. 140-155.

Kiai, M. Heyns C. (2016), "Joint report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association and the

Special Rapporteur on extrajudicial, summary or arbitrary executions", A/HRC/31/66.

Lark M. (2017), "The Future of Killing: Ethical and Legal Implications of Fully Autonomous Weapon Systems", *Salus Journal*, 5, 1, pp. 297-309.

Leys N. (2018), "Autonomous Weapon Systems and International Crises", *Strategic Studies Quarterly*, 12, 1, pp. 48-73.

Maslen S., Weizmann N., Homayounnejad M., Stauffer H. (2018), *Drones and Other Unmanned Weapons Systems under International Law*, Koninlijke, Brill.

Melzer, N. (2013), *Human rights implications of the usage of drones and unmanned robots in warfare,* DG for External Policies, Policy Department, European Parliament, Brussels.

Miller M. (1989), "One dead, one injured in Navy accident", *UPI*, 12th October. Available at: https://www.upi.com/Archives/1989/10/12/One-dead-one-injured-in-Navy-accident/3176624168000/

Panel discussion on the challenges of new technologies in warfare (2014) available at: https://www.icrc.org/en/doc/resources/documents/event/2014/03-06-research-and-debate-inaugural-event.htm

Press M. (2017), "Of Robots and Rules: Autonomous Weapon Systems in the Law of Armed Conflict", *Georgetown Journal of International Law*, 48, pp. 1337-1366.

Prigg M. (2014), "Who goes there? Samsung unveils robot sentry that can kill from two miles away", *DailyMail Online*, 16th September. Available at: https://www.dailymail.co.uk/sciencetech/article-2756847/Who-goes-Samsung-reveals-robot-sentry-set-eye-North-Korea.html

Protocol on Blinding Laser Weapons (IV Additional Protocol to the Convention on Prohibitions or restrictions on the Use of Certain Conventional Weapons which may be deemed to be Excessively Injurious or to have Indiscriminate Effects 1995)

Raytheon Company, *Phalanx Close-In Weapon System*, Phalanx Producing Company's website. Available at: https://www.raytheon.com/capabilities/products/phalanx

Rochlin G. (1991), "Iran Air Flight 655 and the USS Vincennes", in: La Porte T., *Social Responses to Large Technical Systems*, NATO ASI Series (Series D: Behavioural and Social Sciences), 58, Springer, Dordrecht.

Roff H. (2014), "The Strategic Robot Problem: Lethal Autonomous Weapons in War", *Journal of Military Ethics*, 13, 3, 211-227.

Rome Statute of the International Criminal Court (1998).

Second Additional Protocol to the Geneva Conventions of 12 August 1949, relating to the Protection of Victims of Non-International Armed Conflicts (1977).

Sehrawat V. (2017), "Legal Status of Drones under Law of Armed Conflicts and International Law", *Penn State Journal of Law & International Affairs*, 5, 1, pp. 164-206.

Singh N., McWhinney, E. (1989), "Nuclear Weapons and Contemporary International Law", 2nd ed., Martinus Nijhoff, Dordrecht.

Statute of the International Court of Justice (1945).

Stoner R. (2017), *R2D2 with Attitude: The Story of the Phalanx Close-In Weapons*, Naval Technology and Naval Reunions Website. Available at: http://www.navweaps.com/index_tech/tech-103.php

Szpak A. (2019), "Legality of Use and Challenges of New Technologies in Warfare", *European Review*, 28, 1-14.

UN Convention on Conventional Weapons (2013), *Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, CCW/MSP/2013/CRP.1.

UN General Assembly (1969), *Resolution on Question of Chemical and Bacteriological (Biological) Weapons*, 16th December 1969, UN Doc A/RES/2603(XXIV).

Zohuri B. (2019), *Directed-Energy Beam Weapons*, Springer, Cham.