

## A . Report by Professor Christof Heyns, UN special rapporteur on extrajudicial, summary or arbitrary executions for the Office of the High Commissioner for Human Rights

Case prepared by Ms. Sophie Bobillier, Master student at the Faculty of Law of the University of Geneva, under the supervision of Professor Marco Sassòli and Ms. Yvette Issar, research assistant, both at the University of Geneva.

**N.B. As per the disclaimer, neither the ICRC nor the authors can be identified with the opinions expressed in the Cases and Documents.** Some cases even come to solutions that clearly violate IHL. They are nevertheless worthy of discussion, if only to raise a challenge to display more humanity in armed conflicts. **Similarly, in some of the texts used in the case studies, the facts may not always be proven;** nevertheless, they have been selected because they highlight interesting IHL issues and are thus published for didactic purposes.

[Source : HEYNS Christof, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions in A/HRC/23/47, 9 April 2013, available online [http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47\\_en.pdf](http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf)]

### *Summary*

Lethal autonomous robotics (LARs) are weapon systems that, once activated, can select and engage targets without further human intervention.

[...]

### **I. The use of LARs by States outside armed conflict**

82. The experience with UCAVs [unmanned combat aerial vehicles, commonly known as drones] has shown that this type of military technology finds its way with ease into situations outside recognized battlefields.

83. One manifestation of this, whereby ideas of the battlefield are expanded beyond IHL contexts, is the situation in which perceived terrorists are targeted wherever they happen to be found in the world, including in territories where an armed conflict may not exist and IHRL is the applicable legal framework. The danger here is that the world is seen as a single, large and perpetual battlefield and force is used without meeting the threshold requirements. LARs could aggravate these problems.

84. On the domestic front, LARs could be used by States to suppress domestic enemies and to terrorize the population at large, suppress demonstrations and fight “wars” against drugs. It has been said that robots do not question their commanders or stage coups d’état.

85. The possibility of LAR usage in a domestic law enforcement situation creates particular risks of arbitrary deprivation of life, because of the difficulty LARs are bound to have in meeting the stricter requirements posed by IHRL [International Human Rights Law].

## **II. Implications for States without LARs.**

86. Phrases such as “riskless war” and “wars without casualties” are often used in the context of LARs. This seems to purport that only the lives of those with the technology count, which suggests an underlying concern with the deployment of this technology, namely a disregard for those without it. LARs present the ultimate asymmetrical situation, where deadly robots may in some cases be pitted against people on foot. LARs are likely – at least initially – to shift the risk of armed conflict to the belligerents and civilians of the opposing side.

[...]

88. The advantage that States with LARs would have over others is not necessarily permanent. There is likely to be proliferation of such systems, not only to those to which the first user States transfer and sell them. Other States will likely develop their own LAR technology, with inter alia varying degrees of IHL-compliant programming, and potential problems for algorithm compatibility if LARs from opposing forces confront one another. There is also the danger of potential acquisition of LARs by non-State actors, who are less likely to abide by regulatory regimes for control and transparency.

## **III. Taking human decision-making out of the loop**

89. It is an underlying assumption of most legal, moral and other codes that when the decision to take life or to subject people to other grave consequences is at stake, the decision-making power should be exercised by humans. The Hague Convention (IV) requires any combatant “to be commanded by a person”. The Martens Clause, a longstanding and binding rule of IHL, specifically demands the application of “the principle of humanity” in armed conflict. Taking humans out of the loop also risks taking humanity out of the loop.

[...]

## IV. LARs and restrictive regimes on weapons

100. The treaty restrictions placed on certain weapons stem from the IHL norm that the means and methods of warfare are not unlimited, and as such there must be restrictions on the rules that determine what weapons are permissible. The Martens Clause prohibits weapons that run counter to the “dictates of public conscience.” The obligation not to use weapons that have indiscriminate effects and thus cause unnecessary harm to civilians underlies the prohibition of certain weapons and some weapons have been banned because they “cause superfluous injury or unnecessary suffering” to soldiers as well as civilians. The use of still others is restricted for similar reasons.

101. In considering whether restrictions as opposed to an outright ban on LARs would be more appropriate, it should be kept in mind that it may be more difficult to restrict LARs as opposed to other weapons because they are combinations of multiple and often multipurpose technologies. Experts have made strong arguments that a regulatory approach that focuses on technology – namely, the weapons themselves – may be misplaced in the case of LARs and that the focus should rather be on intent or use.

[...]

106. Article 36 of the First Protocol Additional to the Geneva Conventions is especially relevant, providing that, “in the study, development, acquisition or adoption of a new weapon, means or methods of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.”

107. This process is one of internal introspection, not external inspection, and is based on the good faith of the parties. The United States, although not a State party, established formal weapons mechanisms review as early as 1947. While States cannot be obliged to disclose the outcomes of their reviews, one way of ensuring greater control over the emergence of new weapons such as LARs will be to encourage them to be more open about the procedure that they follow in Article 36 reviews generally.

## B. Ban autonomous armed robots

[Source : BOLTON Matthew, NASH Thomas, MOYES Richard: « Ban autonomous armed robots» in article36.org, 5 March 2012, available online <http://www.article36.org/statements/ban-autonomous-armed-robots>]

[...]

[1] Weapons that are triggered automatically by the presence or proximity of their victim can rarely be used in a way that ensures distinction between military and civilian. Despite eventual successes on anti-personnel mines, and more recently cluster munitions, technology develops faster than a humanitarian consensus. A

pressing challenge is the rapid evolution in military systems which are able to select and attack targets autonomously, moving towards the use of fully autonomous armed robots.

[2] Although the relationship between landmines and fully autonomous armed robots may seem stretched, in fact they share essential elements of DNA. Landmines and fully autonomous weapons all provide a capacity to respond with force to an incoming 'signal' (whether the pressure of a foot or a shape on an infra-red sensor). Whether static or mobile, simple or complex, it is the automated violent response to a signal that makes landmines and fully autonomous weapons fundamentally problematic – it is killing by machine.

[...]

## **C. U.S Department of Defense (DoD) Task Force Report on the Role of Autonomy in DoD Systems**

[Source : Department of Defense (Defense Science Board), "Task Force Report : The Role of Autonomous in DoD systems", July 2012, available online [http:-  
//fas.org/irp/agency/dod/dsb/autonomy.pdf](http://fas.org/irp/agency/dod/dsb/autonomy.pdf)]

[...]

### **1.0 Executive Summary**

[1] Unmanned systems are proving to have a significant impact on warfare worldwide. The true value of these systems is not to provide a direct human replacement, but rather to extend and complement human capability in a number of ways. These systems extend human reach by providing potentially unlimited persistent capabilities without degradation due to fatigue or lack of attention. Unmanned systems offer the warfighter more options and flexibility to access hazardous environments, work at small scales, or react at speeds and scales beyond human capability. With proper design of bounded autonomous capabilities, unmanned systems can also reduce the high cognitive load currently placed on operators/supervisors. Moreover, increased autonomy can enable humans to delegate those tasks that are more effectively done by computer, including synchronizing activities between multiple unmanned systems, software agents and warfighters—thus freeing humans to focus on more complex decision making.

[...]

### **2.0 Operational Benefits of Autonomy**

[...]

#### **2.1. Unmanned Aerial Vehicles**

[2] While UAVs [unmanned aerial vehicles] have long held great promise for military operations, the technology has only recently matured enough to exploit that potential. In recent years, the UAV mission scope has expanded from tactical reconnaissance to include most of the capabilities within the ISR [intelligence, surveillance and reconnaissance] and battle space awareness mission areas. Without the constraint of the nominal 12-hour limitation of a human in the cockpit, UAVs can maintain sensors and precision weapons over an area of interest at great distances for longer periods of time, providing situational awareness to all levels of command.

[...]

[3] In addition to expanded persistence, the integration of ISR and strike on the same unmanned platform, coupled with direct connectivity of UAV operators to ground forces, has led to reduced reaction time and is saving lives of U.S. troops on the ground. Moreover, autonomous technology is increasing the safety of unmanned aircraft during auto-takeoff and landing (for those organizations leveraging that technology) and reducing workload via waypoint navigation and orbit management. In addition, due to developments in sense-and-avoid technologies, redundant flight controls, experience and revised procedures, the accident rate for most unmanned systems now mirrors manned aircraft.

[4] Unmanned aircraft clearly have a critical role in the DoD operational future. However, the development of these systems is still in the formative stage, and challenges remain relative to training, integration of command and control and integration of UAVs into the National Air Space.

[...]

## 2.2. Unmanned Ground Systems

[5] Similar to the value UAVs bring to the skies in the form of persistent visibility, Unmanned Ground Systems (UGVs) bring benefits to land in standoff capability. Generally designed as sensory prosthetics, weapons systems or for gaining access to areas inaccessible by humans, UGVs are reducing service member exposure to life threatening tasks by enabling them to identify and neutralize improvised explosive devices (IEDs) from a distance. Today, UGVs are largely used in support of counter-IED and route clearance operations, using robotic arms attached to, and operated by, modified Mine Resistant Ambush Protected (MRAP) vehicles and remotely controlled robotic systems. To a lesser extent, UGVs are being used in dismounted and tactical operations, providing initial and in-depth reconnaissance for soldiers and Marines.

[6] In general, UGVs in combat operations face two primary challenges: negotiating terrain and obstacles on the battlefield and performing kinetic operations within the Rules of Engagement (ROE). Terrain negotiation and obstacle avoidance are driven by mechanical capabilities coupled with pattern recognition and problem solving skills. Operations within the ROE, however, represent a higher order, biomimetic cognitive skill that

must fall within the commander's intent. Going forward, development efforts should aim to advance technologies to better overcome these challenges. Particularly in the latter case, the development of autonomous systems that allow the operator/commander to delegate specific cognitive functions, that may or may not change during the course of a mission or engagement, would appear to be an important milestone in evolution from remotely controlled robotics to autonomous systems.

[...]

### **2.3. Unmanned Maritime Vehicles**

[7] Mission areas for unmanned maritime vehicles (UMVs) can generally be categorized into surface and underwater domains (unmanned surface vehicles (USVs) and unmanned underwater vehicles (UUVs), respectively). Unmanned surface vehicles "operate with near-continuous contact with the surface of the water, including conventional hull crafts, hydrofoils and semi-submersibles. Unmanned underwater vehicles are made to operate without necessary contact with the surface (but may need to be near surface for communications purposes) and some can operate covertly."

[8] USV missions may include antisubmarine warfare (ASW), maritime security, surface warfare, special operations forces support, electronic warfare and maritime interdiction operations support. The Navy has identified a similarly diverse, and often overlapping, range of missions for UUVs, which include ISR, mine countermeasures, ASW, inspection/identification, oceanography, communication/navigation network node, payload delivery, information operations and time-critical strike.

[...]

### **2.4. Unmanned Space Systems**

[...]

[9] Two promising space system application areas for autonomy are the increased use of autonomy to enable an independent acting system and automation as an augmentation of human operation. In such cases, autonomy's fundamental benefits are to increase a system's operational capability and provide cost savings via increased human labor efficiencies, reducing staffing requirements and increasing mission assurance or robustness to uncertain environments. The automation of human operations, that is, transformation from control with automatic response to autonomy for satellite operations, remains a major challenge. Increased use of autonomy—not only in the number of systems and processes to which autonomous control and reasoning can be applied, but especially in the degree of autonomy that is reflected in these systems and processes—can provide the Air Force with potentially enormous increases in its capabilities. If implemented correctly, this increase has the potential to enable manpower efficiencies and cost reductions.

[10] A potential, yet largely unexplored benefit from adding/increasing autonomous functions could be to increase the ability of space systems to do on-board maintenance via auto-detect, auto-diagnose and auto-tune. Increasing presence of such functionality in space and launch systems can be imagined to reduce the cost of mission assurance by making the systems more adaptive to operational and environmental variations and anomalies.

## **2.5. Conclusion**

[11] Unmanned vehicle [UxV] technologies, even with limited autonomous capabilities, have proven their value to DoD operations. The development and fielding of air and ground systems, in particular, have helped save lives and extend human capabilities.

[...]

[12] The Task Force observes that autonomy has a role in advancing both collection and processing capabilities toward more efficient, integrated ends, such as: operating platforms (from two to many) in concert to improve look angles at priority targets, merging sensor data from multiple vehicles and alternative sources and using both mixed (human/computer) teams and heterogeneous, autonomous agents.

[...]

[13] The Task Force also notes that key external vulnerability drivers for unmanned systems include communication links, cyber threats and lack of self- defense. Internally generated limitations are dominated by software errors, brittleness of physical systems and concerns with collateral damage.

[...]

## **Appendix A – Details of Operational Benefits by Domain**

### **A. 1 Aerial Systems Strategy**

[...]

[14] Findings: Unmanned aircraft clearly have a critical role in the future. Admittedly, the development of unmanned systems is still in the formative stage with more focus being given to sensors, weapons, and manned/unmanned operations than in the past [...]. [A]s other nations continue to develop and proliferate unmanned systems, there is a growing need for counter adversary unmanned systems weapon tactics. Key Task Force findings are:

Autonomy can accelerate safe operations in the national air space Mission expansion is growing for all

unmanned system groups Precision weapons are being added to almost all UAV medium and large unmanned aircraft systems [...] Big data has evolved as a major problem at the National Geospatial Intelligence Agency (NGA). Over 25 million minutes of full motion video are stored at NGA Unmanned systems are being used more and more in natural and manmade disasters [...] Homeland Security and other government agencies are increasing their investments in unmanned systems

[15] Benefits: Unmanned systems will need to make use of their strengths and opportunities. As DoD continues to become more experienced in the employment of unmanned systems, operational concepts and tactics, and cultural and Service obstacles will become more manageable. The Department should be able to capitalize on system synergies and economies of scale. A better understanding of how best to employ the systems leads to a better understanding of the optimum mix of manned and unmanned systems as well as a better understanding of how best to employ them against a complex and changing threat environment. Key benefits include:

Extend and complement human capabilities: The greatest operational attribute is endurance. The greatest programmatic attribute is affordability. Resilience: Unmanned systems offer incomparable resilience in terms of cross-decking sensors, replacement costs, and timely deployment. Reduced manpower: Creation of substantive autonomous systems/platforms will create resourcing and leadership benefits. The automation of the actual operation/fighting of platforms will decrease the need for people to crew them, while the personnel needed to simply maintain the vehicles is likely to increase. Reduce loss of life: The original concept for a fleet of unmanned systems was to have a mix of highly capable and moderately survivable systems as well as highly survivable and moderately capable systems. In high-threat environments, the need for manned aircraft will become diminished as sensor and weapons capabilities on unmanned systems increase. Hedge against vulnerabilities: Unmanned systems have an unprecedented advantage in persistence. Low-technology adversary missions such as cruise missile defense and countering of IEDs represent ideal growth missions for unmanned systems. Greater degree of freedom: The ability to function as either an ISR platform or strike platform in anti-access and denied areas represents a major breakthrough in mission flexibility and adaptability.

## **A.2. Maritime systems**

[...]

[16] Summary: Unmanned maritime systems are poised to make a big impact across naval operations. Though in its infancy, there is significant opportunity for this impact to grow. Autonomy's main benefits are to extend/complement human performance providing platforms to do the "dull, dirty, and dangerous" and the capacity to deal with growing volumes of ISR data and potentially reducing/aligning workforce. The requirements-driven development and transition of UUVs and USVs into the fleet can be expected to result in a more cost-efficient mix of manned and unmanned systems.

[...]

### **A.3 Ground Systems**

[17] Autonomous systems, defined broadly as Unmanned Ground Vehicle (UGV), which may include remotely controlled vehicles, have been used on the battlefield as early as 4000 B.C. by the Egyptians and the Romans, in the form of military working dogs. Today, military working dogs are still employed on the battlefield [...] as sensory prosthetics. Additional autonomous ground systems within the U.S. inventory include missiles, such as the Tube-launched, Optically-tracked, Wire command, (TOW) guided missile, introduced in the later stages of the Vietnam Conflict and still in the current U.S. inventory. In all UGV, the system is designed as either a sensory-prosthetic weapon system or for gaining accessibility to areas inaccessible by humans.

[18] Currently, the use of UGVs on the battlefield is not as commonly known as the use of UAVs. Further, UGVs in service have less autonomous capability than the range of UAVs primarily due to challenges in mobility, where the terrain of the battlefield is variable and more difficult to navigate than the air. Nonetheless, UGVs are desired by both the Army and Marine Corp to achieve:

Risk mitigation; Accessibility to areas on the battlefield that are inaccessible by humans; Enhanced sensing capabilities coupled with unmanned mobility; A capability for the application of violence that is not humanly possible; Biotic/abiotic battle formations, where combat units are composed of both human war fighters and automation components.

[...]

### **A.4. Space Systems**

[...]

## **D. Loosing humanity: The Case against Killer Robots**

[Source : Human Rights Watch (HRW), International Human Rights Clinic (IHRC), "« Losing Humanity : The Case against Killer Robots" » ; p. 30-39 available online [http://www.hrw.org/sites/default/files/reports/arms1112\\_ForUpload.pdf](http://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf)]

### **I. Challenges to compliance with International Humanitarian Law**

[1] An initial evaluation of fully autonomous weapons shows that even with the proposed compliance mechanisms, such robots would appear to be incapable of abiding by the key principles of international humanitarian law. They would be unable to follow the rules of distinction, proportionality, and military necessity and might contravene the Martens Clause. Even strong proponents of fully autonomous weapons

have acknowledged that finding ways to meet those rules of international humanitarian law are “outstanding issues” and that the challenge of distinguishing a soldier from a civilian is one of several “daunting problems.” Full autonomy would strip civilians of protections from the effects of war that are guaranteed under the law.

## II. Distinction

[2] The rule of distinction, which requires armed forces to distinguish between combatants and noncombatants, poses one of the greatest obstacles to fully autonomous weapons complying with international humanitarian law. Fully autonomous weapons would not have the ability to sense or interpret the difference between soldiers and civilians, especially in contemporary combat environments.

[3] Changes in the character of armed conflict over the past several decades, from state-to-state warfare to asymmetric conflicts characterized by urban battles fought among civilian populations, have made distinguishing between legitimate targets and noncombatants increasingly difficult. States likely to field autonomous weapons first—the United States, Israel, and European countries—have been fighting predominately counterinsurgency and unconventional wars in recent years. In these conflicts, combatants often do not wear uniforms or insignia. Instead they seek to blend in with the civilian population and are frequently identified by their conduct, or their “direct participation in hostilities.” Although there is no consensus on the definition of direct participation in hostilities, it can be summarized as engaging in or directly supporting military operations. Armed forces may attack individuals directly participating in hostilities, but they must spare noncombatants.

[4] It would seem that a question with a binary answer, such as “is an individual a combatant?” would be easy for a robot to answer, but in fact, fully autonomous weapons would not be able to make such a determination when combatants are not identifiable by physical markings. First, this kind of robot might not have adequate sensors. Krishnan writes, “Distinguishing between a harmless civilian and an armed insurgent could be beyond anything machine perception could possibly do. In any case, it would be easy for terrorists or insurgents to trick these robots by concealing weapons or by exploiting their sensual and behavioral limitations.”

[5] An even more serious problem is that fully autonomous weapons would not possess human qualities necessary to assess an individual's intentions, an assessment that is key to distinguishing targets. According to philosopher Marcello Guarini and computer scientist Paul Bello, “[i]n a context where we cannot assume that everyone present is a combatant, then we have to figure out who is a combatant and who is not. This frequently requires the attribution of intention.” One way to determine intention is to understand an individual's emotional state, something that can only be done if the soldier has emotions. Guarini and Bello continue, “A system without emotion ... could not predict the emotions or action of others based on its own states because it has no emotional states.” Roboticist Noel Sharkey echoes this argument: “Humans understand one another in a way that machines cannot. Cues can be very subtle, and there are an infinite number of circumstances where lethal force is inappropriate.” For example, a frightened mother may run after

her two children and yell at them to stop playing with toy guns near a soldier. A human soldier could identify with the mother's fear and the children's game and thus recognize their intentions as harmless, while a fully autonomous weapon might see only a person running toward it and two armed individuals. The former would hold fire, and the latter might launch an attack. Technological fixes could not give fully autonomous weapons the ability to relate to and understand humans that is needed to pick up on such cues.

### III. Proportionality

[6] The requirement that an attack be proportionate, one of the most complex rules of international humanitarian law, requires human judgment that a fully autonomous weapon would not have. The proportionality test prohibits attacks if the expected civilian harm of an attack outweighs its anticipated military advantage. Michael Schmitt, professor at the US Naval War College, writes, "While the rule is easily stated, there is no question that proportionality is among the most difficult of LOIAC [law of international armed conflict] norms to apply." Peter Asaro, who has written extensively on military robotics, describes it as "abstract, not easily quantified, and highly relative to specific contexts and subjective estimates of value."

[7] Determining the proportionality of a military operation depends heavily on context. The legally compliant response in one situation could change considerably by slightly altering the facts. According to the US Air Force, "[p]roportionality in attack is an inherently subjective determination that will be resolved on a case-by-case basis." It is highly unlikely that a robot could be pre-programmed to handle the infinite number of scenarios it might face so it would have to interpret a situation in real time. Sharkey contends that "the number of such circumstances that could occur simultaneously in military encounters is vast and could cause chaotic robot behavior with deadly consequences." Others argue that the "frame problem," or the autonomous robot's incomplete understanding of its external environment resulting from software limitations, would inevitably lead to "faulty behavior." According to such experts, the robot's problems with analyzing so many situations would interfere with its ability to comply with the proportionality test.

[8] Those who interpret international humanitarian law in complicated and shifting scenarios consistently invoke human judgment, rather than the automatic decision making characteristic of a computer. The authoritative ICRC commentary states that the proportionality test is subjective, allows for a "fairly broad margin of judgment," and "must above all be a question of common sense and good faith for military commanders." International courts, armed forces, and others have adopted a "reasonable military commander" standard. The International Criminal Tribunal for the Former Yugoslavia, for example, wrote, "In determining whether an attack was proportionate it is necessary to examine whether a reasonably well-informed person in the circumstances of the actual perpetrator, making reasonable use of the information available to him or her, could have expected excessive civilian casualties to result from the attack." The test requires more than a balancing of quantitative data, and a robot could not be programmed to duplicate the psychological processes in human judgment that are necessary to assess proportionality.

[9] A scenario in which a fully autonomous aircraft identifies an emerging leadership target exemplifies the

challenges such robots would face in applying the proportionality test. The aircraft might correctly locate an enemy leader in a populated area, but then it would have to assess whether it was lawful to fire. This assessment could pose two problems. First, if the target were in a city, the situation would be constantly changing and thus potentially overwhelming; civilian cars would drive to and from and a school bus might even enter the scene. As discussed above, experts have questioned whether a fully autonomous aircraft could be designed to take into account every movement and adapt to an ever-evolving proportionality calculus. Second, the aircraft would also need to weigh the anticipated advantages of attacking the leader against the number of civilians expected to be killed. [...] Humans are better suited to make such value judgments, which cannot be boiled down to a simple algorithm.

[10] Proponents might argue that fully autonomous weapons with strong AI [artificial intelligence] would have the capacity to apply reason to questions of proportionality. Such claims assume the technology is possible, but that is in dispute as discussed above. There is also the threat that the development of robotic technology would almost certainly outpace that of artificial intelligence. As a result, there is a strong likelihood that advanced militaries would introduce fully autonomous weapons to the battlefield before the robotics industry knew whether it could produce strong AI capabilities. Finally, even if a robot could reach the required level of reason, it would fail to have other characteristics—such as the ability to understand humans and the ability to show mercy—that are necessary to make wise legal and ethical choices beyond the proportionality test.

#### **IV. Military necessity**

[11] Like proportionality, military necessity requires a subjective analysis of a situation. It allows “military forces in planning military actions ... to take into account the practical requirements of a military situation at any given moment and the imperatives of winning,” but those factors are limited by the requirement of “humanity.” One scholar described military necessity as “a context-dependent, value-based judgment of a commander (within certain reasonableness restraints).” Identifying whether an enemy soldier has become hors de combat, for example, demands human judgment. A fully autonomous robot sentry would find it difficult to determine whether an intruder it shot once was merely knocked to the ground by the blast, faking an injury, slightly wounded but able to be detained with quick action, or wounded seriously enough to no longer pose a threat. It might therefore unnecessarily shoot the individual a second time. Fully autonomous weapons are unlikely to be any better at establishing military necessity than they are proportionality.

[12] Military necessity is also relevant to this discussion because proponents could argue that, if fully autonomous weapons were developed, their use itself could become a military necessity in certain circumstances. Krishnan warns that the development of “[t]echnology can largely affect the calculation of military necessity.” He writes: “Once [autonomous weapons] are widely introduced, it becomes a matter of military necessity to use them, as they could prove far superior to any other type of weapon.” He argues such a situation could lead to armed conflict dominated by machines, which he believes could have “disastrous consequences.” [...] v. Martens Clause

[13] Fully autonomous weapons also raise serious concerns under the Martens Clause. The clause, which encompasses rules beyond those found in treaties, requires that means of warfare be evaluated according to the “principles of humanity” and the “dictates of public conscience.” Both experts and laypeople have expressed a range of strong opinions about whether or not fully autonomous machines should be given the power to deliver lethal force without human supervision. While there is no consensus, there is certainly a large number for whom the idea is shocking and unacceptable. States should take their perspective into account when determining the dictates of public conscience.

[14] Ronald Arkin, who supports the development of fully autonomous weapons, helped conduct a survey that offers a glimpse into people’s thoughts about the technology. [...] Arkin concluded, “People are clearly concerned about the potential use of lethal autonomous robots. Despite the perceived ability to save soldiers’ lives, there is clear concern for collateral damage, in particular civilian loss of life.” Even if such anecdotal evidence does not create binding law, any review of fully autonomous weapons should recognize that for many people these weapons are unacceptable under the principles laid out in the Martens Clause.

[...]

## **VI. The lack of human emotion**

[15] Proponents of fully autonomous weapons suggest that the absence of human emotions is a key advantage, yet they fail adequately to consider the downsides. Proponents emphasize, for example, that robots are immune from emotional factors, such as fear and rage, that can cloud judgment, distract humans from their military missions, or lead to attacks on civilians. They also note that robots can be programmed to act without concern for their own survival and thus can sacrifice themselves for a mission without reservations. Such observations have some merit, and these characteristics accrue to both a robot’s military utility and its humanitarian benefits.

[16] Human emotions, however, also provide one of the best safeguards against killing civilians, and a lack of emotion can make killing easier. In training their troops to kill enemy forces, armed forces often attempt “to produce something close to a ‘robot psychology,’ in which what would otherwise seem horrifying acts can be carried out coldly.” This desensitizing process may be necessary to help soldiers carry out combat operations and cope with the horrors of war, yet it illustrates that robots are held up as the ultimate killing machines.

[17] Whatever their military training, human soldiers retain the possibility of emotionally identifying with civilians, “an important part of the empathy that is central to compassion.” Robots cannot identify with humans, which means that they are unable to show compassion, a powerful check on the willingness to kill. For example, a robot in a combat zone might shoot a child pointing a gun at it, which might be a lawful response but not necessarily the most ethical one. By contrast, even if not required under the law to do so, a human soldier might remember his or her children, hold fire, and seek a more merciful solution to the situation, such as trying to capture the child or advance in a different direction. Thus militaries that generally

seek to minimize civilian casualties would find it more difficult to achieve that goal if they relied on emotionless robotic warriors.

[18] Fully autonomous weapons would conversely be perfect tools of repression for autocrats seeking to strengthen or retain power. Even the most hardened troops can eventually turn on their leader if ordered to fire on their own people. A leader who resorted to fully autonomous weapons would be free of the fear that armed forces would rebel. Robots would not identify with their victims and would have to follow orders no matter how inhumane they were.

[19] Several commentators have expressed concern about fully autonomous weapons' lack of emotion. [...] US colonel [...] Krishnan writes: One of the greatest restraints for the cruelty in war has always been the natural inhibition of humans not to kill or hurt fellow human beings. The natural inhibition is, in fact, so strong that most people would rather die than kill somebody.... Taking away the inhibition to kill by using robots for the job could weaken the most powerful psychological and ethical restraint in war. War would be inhumanely efficient and would no longer be constrained by the natural urge of soldiers not to kill.

[20] Rather than being understood as irrational influences and obstacles to reason, emotions should instead be viewed as central to restraint in war. vii. Making war easier and shifting the burden to civilians

[21] Advances in technology have enabled militaries to reduce significantly direct human involvement in fighting wars. The invention of the drone in particular has allowed the United States to conduct military operations in Afghanistan, Pakistan, Yemen, Libya, and elsewhere without fear of casualties to its own personnel. [...] The gradual replacement of humans with fully autonomous weapons could make decisions to go to war easier and shift the burden of armed conflict from soldiers to civilians in battle zones.

[22] While technological advances promising to reduce military casualties are laudable, removing humans from combat entirely could be a step too far. Warfare will inevitably result in human casualties, whether combatant or civilian. Evaluating the human cost of warfare should therefore be a calculation political leaders always make before resorting to the use of military force. Leaders might be less reluctant to go to war, however, if the threat to their own troops were decreased or eliminated. In that case, "states with roboticized forces might behave more aggressively.... [R]obotic weapons alter the political calculation for war." The potential threat to the lives of enemy civilians might be devalued or even ignored in decisions about the use of force.

[23] The effect of drone warfare offers a hint of what weapons with even greater autonomy could lead to. [...] The proliferation of unmanned systems, which according to Singer has a "profound effect on 'the impersonalization of battle,'" may remove some of the instinctual objections to killing. Unmanned systems create both physical and emotional distance from the battlefield, which a number of scholars argue makes killing easier. [...] As D. Keith Shurtleff, Army chaplain and ethics instructor for the Soldier Support Institute at Fort Jackson, pointed out, "[A]s war becomes safer and easier, as soldiers are removed from the horrors of

war and see the enemy not as humans but as blips on a screen, there is a very real danger of losing the deterrent that such horrors provide.” Fully autonomous weapons raise the same concerns.

[24] The prospect of fighting wars without military fatalities would remove one of the greatest deterrents to combat. It would also shift the burden of armed conflict onto civilians in conflict zones because their lives could become more at risk than those of soldiers. Such a shift would be counter to the international community’s growing concern for the protection of civilians. While some advances in military technology can be credited with preventing war or saving lives, the development of fully autonomous weapons could make war more likely and lead to disproportionate civilian suffering. [...]

## E. Accountability for the use of AWS

[Source : ICRC, "« Report of the ICRC Expert Meeting on ‘Autonomous weapon systems: technical legal and humanitarian aspects’, 26-28 March 2014, Geneva" », 9 May 2014, available online <http://www.icrc.org/eng/assets/files/2014/expert-meeting-autonomous-weapons-icrc-report-2014-05-09.pdf>]

[...]

The discussion on accountability for serious IHL violations committed by autonomous weapon systems raised a number of issues, including concern about a possible ‘accountability gap’ or ‘accountability confusion’. Some suggested that such an accountability gap would render the machines unlawful. Others were of the view that a gap will never exist as there will always be a human involved in the decision to deploy an autonomous weapon system to whom responsibility could be attributed. However, it is unclear how responsibility could be attributed in relation to ‘acts’ of autonomous machines that are unpredictable. How can a human be held responsible for a weapon system over which they have no control? In addition, error and malfunction, as well as deliberate programming of an autonomous weapon system to violate IHL, would require that responsibility is apportioned to persons involved in various stages ranging from programming and manufacturing through to the decision to deploy the weapon system.

[...]

## Discussion

### I. Compliance with IHL

1. Which functions and advantages mentioned in Document C are genuinely autonomous? Which simply refer to unmanned, but human-guided vehicles or functions? Which constitute a mix of the two?
2. (Document A, paras 106-7; D, para.[1]) Is it, in your opinion, presently possible to say whether the use of autonomous weapon systems could comply with IHL? Are autonomous weapon systems unlawful as such under existing IHL? (CIHL, Rules 1, 3, 5-24, and 70-71; P I, Arts 35, 36, 48, 50-52 and 57)
3. (Document D) How can you evaluate whether an autonomous weapon system complies with IHL? By

comparing its performance with that of a human being? With that of an average soldier? May one take into account that human beings can deliberately violate the rules they were instructed to follow? Can machines deliberately violate the rules they were programmed to follow?

4. (Document D) What information must an autonomous weapon system be able to sense, collect and process to comply with IHL?
5. What are some of the advantages of using autonomous weapon systems in warfare? (See, in particular, Document C). What are the disadvantages?
6. (Document D, paras [1]-[14]) Which IHL principles may be challenged by the deployment of autonomous weapon systems? In what ways? (CIHL, Rules 1, 3, 5-24, and 70-71; P I, Arts 1(2), 48, 50, 51, 57 and 58)
7. (Document A, paras 106-7; D, paras [1]-[14]) How can the conformity of these autonomous weapon systems with IHL be evaluated before they are deployed? Is there an obligation for States to review new technologies before deploying them in battle? Is the obligation the same for States not parties to Protocol I? What should be taken into account in the review? When should such review be undertaken? (P I Art. 36)
8. (Document D, paras [4], [5], [7], [8] and [10]) Is it currently possible to program an autonomous weapon system to comply with IHL rules governing attacks?
9. (Document A, paras 86 and 88) Does IHL recognize the right of the parties to equality in terms of sophistication of weaponry? Is your answer the same for IACs and NIACs?
10. (Document A, para. 88) Since a large part of IHL is concerned with the protection of human beings involved in armed conflict, do you think this protection would apply to a hypothetical battle involving solely autonomous weapon systems? Would the concepts of inhumane treatment, unnecessary suffering, and superfluous injury be applicable in such a situation?
11. (Document C, para. [17]) Do you agree with the comparison the Task Force's Report makes between autonomous ground vehicles and military working dogs? Dogs, as well as horses and other animals have long formed part of militaries, with relatively little controversy attached to their use in battle. Why, then, is the use of automated robots cause for more concern? If autonomous machines were limited to activities that did not involve the use of lethal force (such as mine detection, for example), would they be more likely to be IHL-compliant? Turning the tables, could other animate life forms (dogs, elephants, bees) be used with lethal effect in battle? Would their use in this way be compliant with IHL? Are the concerns in the latter case the same as those raised by the use of machines?
  - a. Exercising judgment; the capacity for emotion.
12. (Document D, paras [2]-[5] and [15]-[24]) Is an autonomous system capable of exercising judgment, or is that an inherently human capability? Does targeting involve subjective human judgment or only the objective application of parameters? (CIHL, Rules 1, 6, 7, 10-22; P I, Arts 48, 51, 52, 57)
13. How could an autonomous weapon system determine whether a person is a combatant, a civilian taking direct part in hostilities and/or a member of an armed group with a continuous fighting function?
14. How could an autonomous weapon system determine whether a person has surrendered? Whether a person is wounded? (CIHL Rules 46 and 47; GCI-IV, Art. 3; P I, Arts 40 and 41)
15. Does IHL have anything to say about human emotion? Is empathy implied in the principle of humanity? What do you make of the argument that combatants must be human for IHL to be respected? (P I, Art. 1(2))
  - a. Military objectives and proportionality
16. How could an autonomous weapon system determine the concrete and direct military advantage

anticipated from an attack it is about to launch, under the circumstances ruling at the time, which is necessary to determine whether the target is a military objective? (CIHL, Rule 8; P I, Art. 52)

17. Could an autonomous weapon system evaluate the proportionality of an attack it launched? Could it evaluate the military advantage anticipated? The risks of civilian death, injury or damage? (CIHL, Rule 14; P I, Arts 51(5)(b) and 57(2)(a)(iii))
  - a. Distinction (Document A, para. 86; C, paras [3], [5], [11] and [15]; D para. [24])
18. Do both attacking and defending belligerents share the same obligations when it comes to “protecting lives”? Is the argument that new technologies would save lives of combatants relevant under IHL? Do you agree with the analysis presented in Document D para. [24] about the prospect of so-called “riskless wars”?
  - a. Precautions
19. Would it be lawful to launch an attack with an autonomous weapon system if the system allowed for additional precautionary measures to be taken that would not be feasible for a human attacker if, at the same time, an autonomous weapon system did not perform as accurately as a human attacker in other respects? (CIHL, Rule 15; P I, Art. 57)
20. If an autonomous weapon system decides when to attack, who is the addressee of the precautionary rules that bind those who plan or decide upon an attack? To satisfy those rules, would it be sufficient that human beings fix the parameters according to which the system bases its decisions? (P I, Art. 57(2)(a))
21. (Document D, paras [6]-[12]) An attack must be cancelled or suspended if it becomes apparent that it is unlawful. To whom must any facts that would render an attack unlawful need to become apparent in the case where an autonomous weapon system is used? The machine itself? The human being deploying the machine? Must the machine be programmed so that it can react as well as a human being could to such circumstances? (P I, Art. 57(2)(b))
  - a. Accountability (Document E)
22. Can a machine violate IHL? Who would be responsible for a violation committed by an autonomous weapon system? The machine? The commander deploying it? Even if he or she trusts that the machine will respect IHL? The producer? The programmer? Under what circumstances? What practical difficulties arise in holding any of the aforementioned categories accountable?
23. Would a programmer who, in peacetime, deliberately misprogrammed an autonomous weapon system so that it attacked civilians in the event of an armed conflict violate IHL? Could his actions constitute a war crime? Would a state be responsible for violations of IHL if such a programmer were attributable to it? (CIHL, Rules 149 and 151; GC I-IV, Art. 2; GC IV Arts 146-147; P I Arts 1, 85 and 86)

## **II. Developing the legal framework**

1. (Document A, paras [101]-[107]; Is the current legal framework sufficient to respond to the development and use of these new technologies? Is it sufficient to simply consider that autonomous weapon systems may not be used if they do not comply with the existing rules, or are new rules needed? (P I, Arts 1(2), 35, 36 and 48-58; CIHL, Rules 1, 3, 5-24 and 70-71)
  - a. Is an additional protocol to the Convention on Certain Conventional Weapons of 10 October 1980 necessary to regulate potential new autonomous weapon systems?
  - b. What are the advantages of a total prohibition of autonomous weapon systems? The disadvantages? In your opinion, should a total prohibition include all new technologies, or only

certain kinds? If so, what criteria do you think are important in determining what kinds of weapons should be subject to a total prohibition?

- c. Would it be sufficient to prescribe that there must be meaningful human oversight over any lethal attack executed by an autonomous weapon system?
- d. What are the advantages of additional control or regulation of autonomous weapons? The disadvantages? Should regulation include all new technologies, or only certain kinds? What is the basis for your response to the former question?
- e. What are the advantages of taking no further legislative action on autonomous weapons? What are the risks of waiting until the technology is sufficiently developed to see whether the weapons systems would be able to comply with the principles of distinction, proportionality and precautions?

© International Committee of the Red Cross