

---

---

# INTERNATIONAL LAW STUDIES

PUBLISHED SINCE 1895

U.S. NAVAL WAR COLLEGE



## Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified

*Marco Sassóli*

90 INT'L L. STUD. 308 (2014)

Volume 90

2014

---

---

## Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified

*Marco Sassóli\**

### I. INTRODUCTION

Autonomous weapons are subject to great controversy. They have been defined as weapon systems<sup>1</sup> “that can learn or adapt [their] functioning in response to changing circumstances in the environment in which [they are] deployed.”<sup>2</sup> Such systems, once developed, should, through sensors that give them situational awareness, be able to identify both legitimate targets and hopefully civilians/civilian objects that may potentially suffer incidental

---

\* Professor of International Law at the University of Geneva, Switzerland, Director of the Department of Public International Law and International Organization, Associate Professor at the Université du Québec, Montreal, Canada and Commissioner within the International Commission of Jurists. The author would like to thank Ms. Yvette Issar, doctoral candidate at the University of Geneva, for her valuable research assistance, critical remarks and for having revised this text.

1. Hin-Yan Liu, *Categorization and Legality of Autonomous and Remote Weapons Systems*, 94 INTERNATIONAL REVIEW OF THE RED CROSS 627, 635–36 (2012).

2. International Committee of the Red Cross, *International Humanitarian Law and the Challenges of Contemporary Armed Conflict: Report Prepared for the 31st International Conference of the Red Cross and Red Crescent* 39 (2011), available at <http://www.icrc.org/eng/resources/documents/report/31-international-conference-ihl-challenges-report-2011-10-31.htm>.

effects of attack. Identification would then trigger corresponding action through processors or artificial intelligence that would “decide . . . how to respond . . . and effectors that carry out those ‘decisions.’”<sup>3</sup> Ideally, autonomous weapons would select and engage targets without ongoing human intervention in an open environment under circumstances which are unstructured and dynamic. At present, no weapon system possesses such capabilities. The absence or presence of human intervention is, however, a relative distinction, as is the distinction between humans “in,” “on” or “out of the loop,” which is therefore not very helpful.<sup>4</sup> Despite the system’s autonomy, human beings will inevitably be involved, either in overseeing the operation of the weapon, or at least in producing and programming the weapon systems. There is agreement that, although these systems do not yet exist, they could be developed within twenty years. Many request that they be purely, simply and preventively banned, specifically because their use would not be consistent with international humanitarian law (IHL).<sup>5</sup> The 117 States parties to the UN Conventional Weapons Convention<sup>6</sup> have agreed to hold this year a four-day intergovernmental meeting to explore questions related to lethal autonomous weapon systems with a view to potentially drafting a Protocol VI to the Convention.<sup>7</sup> Even the United States, which is among the most technologically advanced States in this field, currently requires that such weapon systems “be designed to allow commanders and operators to exercise appropriate levels of judgment over the use of

---

3. Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, *Report*, ¶ 39, U.N. Doc. A/HRC/23/47 (Apr. 9, 2013) (by Christof Heyns) [hereinafter Heyns].

4. PETER W. SINGER, *WIRED FOR WAR: THE ROBOTICS REVOLUTION AND CONFLICT IN THE 21ST CENTURY* 124–27 (2009).

5. HUMAN RIGHTS WATCH, *LOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS* (2012), available at <http://www.hrw.org/reports/2012/11/19/losing-humanity>; European Parliament Resolution on the Use of Armed Drones ¶ H.2(d) (2014/2567(RSP)) (Feb. 25, 2014); Peter Asaro, *On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making*, 94 *INTERNATIONAL REVIEW OF THE RED CROSS* 687 (2012).

6. 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, Oct. 10, 1980, 1342 U.N.T.S. 137.

7. 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, Geneva, Switz., Nov. 14–15, 2013, *Final Report*, ¶¶ 18, 32, U.N. Doc. CCW.MSP/2013/10 (Dec. 16, 2013), available at <http://daccess-ods.un.org/TMP/4361535.01272202.html>.

force,”<sup>8</sup> which means that it is not admissible for producers to program machines that make final decisions as to targets against which force shall be used.

It is perhaps because I have been confronted in actual armed conflicts with so many violations committed by human beings, but inevitably never with atrocities by robots<sup>9</sup> (although admittedly, they did not exist in the armed conflicts I witnessed), that my first feeling is not skepticism, but hope for better respect of IHL.<sup>10</sup> Only human beings can be inhuman and only human beings can deliberately choose not to comply with the rules they were instructed to follow. To me, it seems more reasonable to expect (and to ensure) a person who devises and constructs an autonomous weapon in a peaceful workplace to comply with IHL than a soldier on the battlefield or in a hostile environment. A robot cannot hate, cannot fear, cannot be hungry or tired and has no survival instinct.<sup>11</sup> “Robots do not rape.”<sup>12</sup> They can sense more information simultaneously and process it faster than a human being.<sup>13</sup> As the weapons actually delivering kinetic force become increasingly quicker and more complex, it may be that humans become simply too overwhelmed by information and the decisions that must be taken to direct them.<sup>14</sup> Human beings often kill others to avoid being killed themselves. The robot can delay the use of force until the last, most appropriate moment, when it has been established that the target and the attack are legitimate. Certainly, there may be technical failures, but all those who drive cars and every traffic policeman know that

---

8. U.S. Department of Defense, DoD Directive 3000.09, *Autonomy in Weapon Systems* ¶ 3(a) (2012).

9. The terms “atrocious” or “violation,” like many other terms which describe human behavior used in this article, are admittedly not appropriate for a machine.”

10. For a detailed, powerful, but not always entirely convincing plea, see RONALD C. ARKIN, *GOVERNING LETHAL BEHAVIOR IN AUTONOMOUS ROBOTS* (2009). For a positive assessment and call for regulation, see Kenneth Anderson & Matthew Waxman, *Law and Ethics for Autonomous Weapons Systems: Why a Ban Won't Work and How the Laws of War Can*, HOOVER INSTITUTE (Apr. 9, 2013), <http://www.hoover.org/publications/monographs/144241>.

11. Ronald C. Arkin, *Ethical Robots in Warfare*, GEORGIA INSTITUTE OF TECHNOLOGY (Jan. 20, 2009), <http://www.cc.gatech.edu/ai/robot-lab/online-publications/arkin-rev.pdf>; Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, *Interim Report*, ¶ 20 (by Philip Alston), *transmitted by Note of the Secretary-General*, U.N. Doc. A/65/321 (Aug. 23, 2010) [hereinafter Alston].

12. Heyns, *supra* note 3, ¶ 54.

13. SINGER, *supra* note 4, at 127.

14. Arkin, *supra* note 11; SINGER, *supra* note 4, at 128.

most accidents are not due to technical, but human, failures (although drivers, unlike soldiers, are usually not seeking to kill or injure).

As soon as robots have artificial intelligence, one must obviously also make sure that such intelligence is not used in the same way as some human intelligence is used, i.e., to circumvent the rules or to decide—from a purely utilitarian perspective—that non-respect of instructions that conform to IHL may make it easier to achieve the main objective of overcoming the enemy. In addition, States developing and producing autonomous weapon systems must—and will in their own interest—take measures to prevent the enemy from tampering with such systems and directing them against the producing State and its civilians.<sup>15</sup>

It is obviously not up to me as a lawyer to speculate on whether autonomous weapons that (a) do not share the same disadvantages of human beings fighting wars and (b) are able to function in compliance with the principles and rules on distinction, proportionality and precautions as well as human beings can be developed. I will nevertheless first highlight this technical question which is, in my view, decisive, and one to which I have no answer. It is a question to which a definitive answer must be supplied before such weapon systems may be used (Part II). Next I will address some preliminary issues which either go beyond IHL or are raised by autonomous weapon systems just as they are for other means of warfare (Part III). The principal portion of the article will then be devoted to the specific problems autonomous weapon systems may raise under IHL and whether they may be able to comply—or provide even greater compliance—with the core IHL targeting principles of distinction, proportionality and precautions in attack (Part IV).

## II. THE CRUCIAL TECHNICAL ISSUE

As explained above, I consider that weapon systems which do not base the use of force upon an ad hoc human decision offer the advantage of a greater possibility of respecting IHL. However, this presupposes that it is technically possible to make them as accurate as an average soldier in terms of distinction, proportionality and precautions. The main question remains

---

15. Michael N. Schmitt & Jeffrey S. Thurnher, “Out of the Loop”: *Autonomous Weapon Systems and the Law of Armed Conflict*, 4 HARVARD NATIONAL SECURITY JOURNAL 231, 242–43 (2013).

a technical one:<sup>16</sup> is it, or will it one day be, possible to develop a robot able to sense information and thereafter to act such that it distinguishes—in the same manner as a human being—between legitimate targets, i.e., military objectives, combatants, civilians directly participating in hostilities on the one hand, and those protected by IHL from attacks, i.e., civilians and civilian objects, specially protected objects such as cultural property, objects indispensable for the survival of the civilian population, medical units, etc., on the other?

Many experts consider that development of robots with that capability is impossible given the current technology. I have seen no official pronouncement indicating that it is possible, nor am I aware of anyone suggesting that autonomous weapon systems may be used even if such development turns out to be impossible. Governments with the possible ability to produce these kinds of weapons indicate that genuinely autonomous deployment is not currently envisaged.<sup>17</sup> William Boothby, the most renowned weapons law expert I know, and who has a strong military background, envisages their autonomy, without human supervision, only in very limited and predetermined environments.<sup>18</sup> It is claimed that no one envisages developing any system “without restrictive engagement parameters, such as limiting the area of operation or nature of the target” (even among lawful targets).<sup>19</sup> It is also true that wars are full of “friction” and unexpected situations.<sup>20</sup> All this makes me skeptical of whether it will be possible to create a machine with the necessary contextual intelligence to adapt to the great variety of situations which may arise in hostilities. Nevertheless, I cannot predict possible future developments.

That autonomous weapon systems will operate in dynamic and unstructured circumstances does not mean that those circumstances are unforeseeable. An important question I am unable to answer concerns whether it is possible to program these weapons to adapt a rule to circumstances not foreseen by the programmer, or whether they can only deal

---

16. *See also* Jakob Kellenberger, President, International Committee of the Red Cross, Keynote Address at the 34th Round Table on Current Issues on International Humanitarian Law: International Humanitarian Law and New Weapons Technologies (Sept. 9, 2011), in 94 INTERNATIONAL REVIEW OF THE RED CROSS 812 (2012).

17. Heyns, *supra* note 3, ¶ 29. SINGER, *supra* note 4, at 128, is skeptical. *See, e.g.*, DOD Directive 3000.09, *supra* note 8, ¶ 4(a).

18. WILLIAM H. BOOTHBY, WEAPONS AND THE LAW OF ARMED CONFLICT 282–86 (2009).

19. Schmitt & Thurnher, *supra* note 15, at 241.

20. Asaro, *supra* note 5, at 691–92.

with circumstances foreseen by the programmer. If the latter is true, they would have to be programmed not to act in other circumstances and/or to notify a human being that he/she must make the necessary decision.<sup>21</sup> However, while it may be impossible to program autonomous weapons for all circumstances which may appear in an armed conflict, might it be possible for them to “learn”?<sup>22</sup>

It will indeed be a challenge to translate some aspects of IHL—for example, the International Committee of the Red Cross’s (ICRC) *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law*<sup>23</sup> or various States’ interpretations of the concept of direct participation in hostilities—into a computer program. This is, however, necessary. It is not simply an issue of distinguishing between combatants and civilians in a Cold War-like situation, but in the real contexts of today’s conflicts. In these conflicts, civilians directly participate in hostilities, rendering application of the principle of distinction difficult for average soldiers, to say nothing of autonomous weapons.

Although I don’t know what is technically feasible, if I take a position in this article it is primarily because I consider that many objections of principle against autonomous weapon systems are either questionable from an IHL point of view, or misunderstand IHL and warfare.

### III. PRELIMINARY ISSUES

#### A. *Questions Going Beyond International Humanitarian Law*

Many objections against autonomous weapons are situated outside IHL. Some are moral. To me, it seems obvious that robots can neither “behave” “morally” nor “immorally.”<sup>24</sup> While I would never claim that robots can be

---

21. DOD Directive 3000.09, *supra* note 8, ¶ 4(a)(1)(b).

22. Alan Backstrom & Ian Henderson, *New Capabilities in Warfare: an Overview of Contemporary Technological Developments and the Associated Legal and Engineering Issues in Article 36 Weapons Reviews*, 94 INTERNATIONAL REVIEW OF THE RED CROSS 483, 493–94 (2012).

23. See INTERNATIONAL COMMITTEE OF THE RED CROSS, INTERPRETIVE GUIDANCE ON THE NOTION OF DIRECT PARTICIPATION UNDER INTERNATIONAL HUMANITARIAN LAW (2009), available at <http://www.icrc.org/eng/resources/documents/publication/p0990.htm>, reprinted in 90 INTERNATIONAL REVIEW OF THE RED CROSS 991 (2008) [hereinafter INTERPRETIVE GUIDANCE].

24. Noel E. Sharkey, *The Evitability of Autonomous Robot Warfare*, 94 INTERNATIONAL REVIEW OF THE RED CROSS 787, 793 (2012).

more humane than humans,<sup>25</sup> I am convinced that only human beings can be inhuman. Moral judgments are reserved for the human beings creating, directing and using weapons.

### 1. May Only a Human Being Decide to Kill Another Human Being?

Many find the very idea that a robot may kill a human being to be horrible. One philosopher even claims there is an implicit requirement in IHL for a human decision to kill a person.<sup>26</sup> Others wonder whether it is “inherently wrong to let autonomous machines decide who and when to kill.”<sup>27</sup> However, if this were true, all mines would be prohibited, as well as all weapons such as missiles that can be directed at a military objective and combatants, but which do not allow the operator to know who exactly will be killed. Moreover, today computers already open the bomb bays of bomber aircraft or decide which targets to engage in the case of the Aegis naval defense system when used in the automatic mode.<sup>28</sup> If the moral requirement of human involvement simply means that a human being must decide what categories of people can be targeted, autonomous weapons, as I understand them, comply because they are programmed by human beings. Sometimes this requirement of human involvement is justified by referring to the requirement under Article I of the Hague Regulations<sup>29</sup> that a combatant must be commanded by a person.<sup>30</sup> I am, however, not aware of anyone suggesting that robots would instruct and supervise humans.

In addition, to be lawful<sup>31</sup> autonomous weapons must still be subject to general instructions given to them by humans. The weapon system would simply apply those instructions autonomously to a given situation.

Beyond that, is it not as horrible to imagine one human being deliberately killing another in the absence of an immediate threat from the latter as it is to imagine a machine doing the same? But this horror is war where killing in the absence of a threat is lawful.

---

25. *Id.* at 793. Here I agree with Sharkey.

26. Asaro, *supra* note 5, at 696–703.

27. Heyns, *supra* note 3, ¶ 92.

28. SINGER, *supra* note 4, at 124.

29. Regulations Concerning the Laws and Customs of War on Land, annexed to Hague Convention (IV) Respecting the Laws and Customs of War on Land, Oct. 18, 1907, 36 Stat. 2227.

30. Heyns, *supra* note 3, ¶ 89.

31. *See infra* Part IV.B.4 for the caveat that weapons that could autonomously choose to “go rogue” would be inherently unlawful.



It would be a misconception of existing IHL to claim that the decision to kill someone in an armed conflict must be taken after a value judgment (which a machine is obviously unable to make and must be made by a human being) is made about that person. Whether a person may be targeted in an armed conflict is dependent on their status (combatant/civilian) and/or the objective impression resulting from their conduct (direct participation in hostilities) and not whether that person is innocent or guilty of a crime. What counts, for example, is, not whether a person wants to surrender, but whether he or she indicates their willingness to surrender and the attacker becomes aware of this indication.

## 2. Do They Make it Easier to Go to War?

Some fear that autonomous weapon systems make it easier for a country to go to war and to use force beyond its borders because it no longer risks the lives of its own soldiers, but only those of civilians on the opposing side.<sup>32</sup> This risk is partly mitigated by the fact that it is highly improbable that one side would fight a war only with robots.<sup>33</sup> Additionally, conflict has moved beyond fighting between the knights or samurai of the Middle Ages; today many weapons and delivery system technologies permit cross-border attacks. This risk must be countered by *ius ad bellum* and disarmament.

Importantly, the very decision to engage in an armed conflict must be reserved for humans.<sup>34</sup> It is true that the motivation for a State to comply with *ius ad bellum* could be lowered if it could expect that only combatants and civilians suffering incidental effects of the enemy will be affected. However, nothing in the law would hinder the opposing side from attacking the human beings—combatants and civilians—who deploy the robots.

Rules of IHL have long been criticized for attempting to “humanize” war, making it more conceivable and more acceptable. War must be horrible—and horrible for both sides—to deter States and human beings from resort to it. If this line of argument worked in reality, war would have disappeared; unfortunately it has not. Beyond that, it may well be that the possibility of secrecy in the use of autonomous weapons and the resulting difficulties of attribution make the implementation of State responsibility and of international criminal responsibility for an act of aggression more

---

32. HUMAN RIGHTS WATCH, *supra* note 5, at 4, 39; Heyns, *supra* note 3, ¶ 58; Alston, *supra* note 11, ¶ 44; Asaro, *supra* note 5, at 692.

33. Schmitt & Thurnher, *supra* note 15, at 241.

34. Asaro, *supra* note 5, at 692.

difficult. On the other hand, the fact that computer systems register everything—or could be programmed to do so—would make an inquiry leading to accountability easier.<sup>35</sup>

### 3. Unfairness?

A related concern is that the use of autonomous weapons against an enemy not possessing them itself would be unfair, *inter alia*, because it would not involve any risk for the personnel of the attacker, while the defender would have to fight with actual human beings who would be killed. This would include not only combatants, but also civilians as incidental victims of a lawful attack even if IHL is respected. This is, however, a misconception of existing IHL. For a long time war has not been fair in this sense.<sup>36</sup> The idea that two knights fight against each other and all civilians stand by wondering who will win belongs to the past. No one suggests that a party may not use its air force or navy if the enemy has no air force or navy, or has no weapons to combat aircraft. As for the unfairness, contemporary reality shows that the technologically weaker side may prevail over the stronger belligerent and impose its political will over the latter. Some claim that the weaker side tries to compensate for its technological deficit by neglecting some of the rules according to which it should fight.<sup>37</sup> It is difficult, however, to argue that under IHL this risk prevents a party from using technology not available to the enemy.

### 4. The Risk of Proliferation

As with all weapons, there is obviously a risk of proliferation, and those developing autonomous weapons are well advised to take that into account. This risk has an aspect specific to autonomous weapons; a risk which is the downside to one of their main advantages. I mentioned above that autonomous weapon systems will not violate IHL if they were programmed to

---

35. Heyns, *supra* note 3, ¶ 52.

36. Schmitt & Thurnher, *supra* note 15, at 232.

37. Wolff Heintschel von Heinegg, *Asymmetric Warfare*, ¶ 3, in MAX PLANCK ENCYCLOPEDIA OF PUBLIC INTERNATIONAL LAW (Rüdiger Wolfrum ed., 2012), <http://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e1809?rskey=phOWN9&result=1&prd=OPIL>; Michael N. Schmitt, *War, Technology and International Humanitarian Law* 33, 46–47, HPCR OCCASIONAL PAPER SERIES (Summer 2005), <http://www.hpcresearch.org/publications/occasional-paper-series> (then OPS Issue 4 PDF hyperlink).

respect it. They could even be programmed to refuse orders that violate the law. The downside of this absolute “discipline” is that if autonomous weapons fall into the hands of a ruthless leader able to program them, they will never abandon his or her cause as soldiers might because the programmed robots will be unable to come to the conclusion that it is unjust, even when there are repeated and widespread violations of IHL.<sup>38</sup> This may indicate that a legal framework on the production, stockpiling and transfer of autonomous weapons may be necessary. It may be possible to make sure that they are automatically deactivated if they fall into the wrong hands or receive instructions contrary to IHL.

#### 5. The Risk of a “Gameboy Mentality” of the Producer?

Some argue that those who build an autonomous weapon and program it (imagine, for example, a programmer in New Jersey who is the last human being in the loop able to make targeting-relevant decisions, but who has no knowledge of where and in what circumstances the weapon is to be used) are psychologically less committed, feel a greater distance from and less responsibility for use of the weapon system, and will therefore more easily adopt a “computer game mentality” than a soldier who is actually on the battlefield among the human beings he or she will kill.<sup>39</sup> As far as I know, there is no scientific evidence to support this (or the opposing) opinion. In my experience during the conflicts in the former Yugoslavia, I met people who killed other people face-to-face, and who displayed no more inhibition than my son would in a computer game. I am not so sure that the producers and programmers in New Jersey would see the world simply as a computer game if they were adequately trained, supervised and subject to appropriate accountability systems. In a peacetime activity in which human life is at stake, an increasing number of processes are automated and no one claims that those who devise such processes comport themselves as though they were playing a computer game. No one requires that those who produce medical robots work in a hospital and surgery room environment. Similarly, those who produce aircraft parts know that people’s lives are at risk if they make mistakes, even if they see neither the passengers nor the aircraft.

---

38. HUMAN RIGHTS WATCH, *supra* note 5, at 4, 38.

39. Heyns, *supra* note 3, ¶ 57; Alston, *supra* note 11, ¶ 44.

## 6. Lack of Emotions

A related objection is that robots lack “human emotions and the capacity for compassion.”<sup>40</sup> This is true. In response, however, the human beings programming the robots hopefully have compassion and there is no reason why they should have less of it than a soldier in midst of the battle. Second, I have noticed among those who actually fight armed conflicts at least as many negative emotions as the positive emotions this argument envisages.<sup>41</sup> Third, IHL does not seek to promote “love,” “mercy”<sup>42</sup> or “human empathy”<sup>43</sup> (a robot is indeed unable to have such feelings), but respect based upon objective criteria. A soldier may not kill a civilian even if there are good reasons to hate him or her, while he or she may kill a combatant even if that combatant is very nice. If “human empathy” was decisive in war, few combatants would be killed, war would finally disappear and only cases of peacetime self-defense and defense of others would persist.

### B. *International Humanitarian Law Issues which are not Specific to Autonomous Weapons*

When robots are used certain fundamental questions of IHL may become even more important than would be the case if other means of warfare were employed. However, the answer to those questions remains controversial. The first question is always what is an armed conflict? We know that there is no unique definition. The questions are rather: what is an international armed conflict and what is a non-international armed conflict? What is the threshold of violence necessary to make violence between a State and a non-State actor (or between non-State actors) an armed conflict?

Those questions are not specific to robots and even when autonomous weapons are used, the answers must be given—and will perforce be given—by human beings. The answers simply become even more important when autonomous weapons are used. Indeed, outside an armed conflict, lethal robots could only be used if they were able to arrest a person, which is, as opposed to the use of lethal force, always the solution preferred by human rights law. The deployment of autonomous weapons produced for

---

40. HUMAN RIGHTS WATCH, *supra* note 5, at 4, 38.

41. *See also* Schmitt & Thurnher, *supra* note 15, at 249.

42. HUMAN RIGHTS WATCH, *supra* note 5, at 34.

43. *Id.* at 37.

armed conflicts will be unlawful outside armed conflicts. Soldiers, on the other hand, can be more easily sent, for example, to ensure security at a football match. One might obviously imagine programming an autonomous weapon equally capable of distinguishing between hostilities and law enforcement operations and applying human rights law to the latter. However, this will be an extremely difficult task. Under human rights law, for example, alternatives to the use of lethal force must be evaluated and the risk the target presents to human life must be assessed.

Many other questions must find an answer before an autonomous weapon can be programmed. What is the relationship between international human rights law and IHL? What is the geographical scope of application of IHL, that is, the scope of the battlefield? Autonomous weapons raise this latter question more acutely, but, legally, the answer must be the same as for an aerial bombardment. May a belligerent attack a target which would be a legitimate target under IHL at a location far distant from the actual fighting, restrained only by the rules of IHL? In such circumstances, does IHL apply at all, but, if so, does international human rights law prevail as the *lex specialis*? Whether drones, missiles or autonomous weapons are used, the answer must be the same, and however intelligent an autonomous weapon may be, I cannot imagine that it can provide the answer itself. The geographical scope of the field of operation of an autonomous weapon and the circumstances under which it may use force according to its IHL program must be defined by human beings.

#### IV. SPECIFIC IHL ISSUES

##### *A. How to Compare the Performance of Robots with that of Human Beings?*

###### 1. For a Given Attack

There is widespread agreement that the ability to use autonomous weapons in compliance with IHL should not be evaluated against a hypothetical ideal, but instead the comparison should be to human beings.<sup>44</sup> This is particularly true for the question of whether a precautionary measure is feasible, as only feasible precautions must be taken under IHL.<sup>45</sup> What counts, at a

---

44. Heyns, *supra* note 3, ¶ 63; Backstrom & Henderson, *supra* note 22, at 492; Schmitt & Thurnher, *supra* note 15, at 247.

45. See Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts art. 57, June 8, 1977,

minimum, is obviously not whether the precaution is feasible for the robot, but whether it would be feasible for a State employing human soldiers, i.e., whether it would be feasible for a person. If it is technically not feasible to respect certain requirements of IHL with autonomous weapons, this is not a sufficient reason for abandoning those requirements. The use of an autonomous weapon in such cases is simply unlawful. Human Rights Watch is correct in warning that once autonomous weapons are developed their mere availability and military capabilities should never be considered to mean that there is a “military necessity” to use them, even if they are not as able as a human being to comply with IHL in a given attack.<sup>46</sup> This is simply not the meaning of the principle of military necessity as a prohibitive rule in IHL.<sup>47</sup> Conversely, if autonomous systems are better than human beings, such as in taking precautions, and a State and a commander have them in their arsenal and don’t need to reserve their use for other militarily more important tasks or tasks involving higher risks for civilians, they must use them.

The bold statement that autonomous weapons may only be used if and where they are as good as human beings in complying with IHL is, however, subject to two nuances. First, artillery and missiles are not—in the same manner as a sniper—able to cancel or suspend an attack at the last moment based on changing circumstances.<sup>48</sup> Nevertheless, no one claims that such weapons are inherently unlawful. What counts is that either the system itself through technical means, or the human beings using it, are able to acquire information indicating that the attack must be interrupted and either the machine or its human operators are able to react to such information.

---

1125 U.N.T.S. 3 [hereinafter Protocol I]. For its customary character in both international and non-international armed conflicts, see 1 CUSTOMARY INTERNATIONAL HUMANITARIAN LAW r. 15–21 (Jean-Marie Henckaerts & Louise Doswald-Beck eds., 2005).

46. HUMAN RIGHTS WATCH, *supra* note 5, at 35.

47. GABRIELLA VENTURINI, NECESSITÀ E PROPORZIONALITÀ NELL’ USO DELLA FORZA MILITARE IN DIRITTO INTERNAZIONALE 127 (1988); William V. O’Brien, *The Meaning of Military Necessity in International Law*, in 1 WORLD POLITY: A YEARBOOK OF STUDIES IN INTERNATIONAL LAW AND ORGANIZATION 109, 138–63 (1957); 2 GEORG SCHWARZENBERGER, INTERNATIONAL LAW AS APPLIED BY INTERNATIONAL COURTS AND TRIBUNALS 135 (1968); Robin Geiss, *Military Necessity: A Fundamental “Principle” Fallen into Oblivion*, in 2 SELECT PROCEEDINGS OF THE EUROPEAN SOCIETY OF INTERNATIONAL LAW 554 (Hélène Ruiz Fabri, Rüdiger Wolfrum & Jana Gogolin eds., 2008); Robert Kolb, *La Nécessité Militaire dans le Droit des Conflits Armés—Essai de Clarification Conceptuelle*, in SOCIÉTÉ FRANÇAISE DE DROIT INTERNATIONAL, LA NÉCESSITÉ EN DROIT INTERNATIONAL: COLLOQUE DE GRENOBLE 151(2007).

48. BOOTHBY, *supra* note 18, at 285.

Second, it may well be that the average soldier is better able to respect certain aspects of IHL than an autonomous weapon, while such a weapon is better in other respects. The question then arises whether a consolidated assessment of advantages and disadvantages is admissible or whether an autonomous weapon must be as good as an average soldier with respect to every rule of IHL. In my view, the choice of means is always based upon a consolidated assessment, but such assessment must be made for every attack. It is not sufficient that when compared with alternatives in most, normal or ideal circumstances a weapon minimizes incidental civilian losses. An autonomous weapon would have to make such a determination in relation to the specific circumstances of each attack, and indicate, if necessary, that it cannot execute that attack, but that the attack must, if it is to be carried out, be executed by a human being.

## 2. When the Autonomous Weapon is Reviewed

In addition, an *in abstracto* assessment of all possible uses is equally necessary to determine whether the autonomous weapon concerned may be produced in the first place. Article 36 of Protocol Additional I prescribes:

In the study, development, acquisition or adoption of a new weapon, means or method 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.

Several States, including the United States which is not a party to Protocol I, have implemented this requirement through specific procedures.<sup>49</sup> Because it is implicit in its substantive obligations—“[t]he faithful and responsible application [of which] would require a State to ensure that the new weapons, means and methods of warfare it develops or acquires will not violate these obligations,”<sup>50</sup>—the ICRC considers that this obligation binds all States. The establishment of an international body monitoring

---

49. For commentaries, see INTERNATIONAL COMMITTEE OF THE RED CROSS, A GUIDE TO THE LEGAL REVIEW OF NEW WEAPONS, MEANS AND METHODS OF WARFARE: MEASURES TO IMPLEMENT ARTICLE 36 OF ADDITIONAL PROTOCOL I OF 1977 (2006) [hereafter ICRC GUIDE]; BOOTHBY, *supra* note 18, at 340–52.

50. ICRC GUIDE, *supra* note 49, at 4.

weapons development would obviously be highly desirable,<sup>51</sup> but secrecy concerns are a barrier.<sup>52</sup>

It is obvious that before autonomous weapons may be deployed, such an assessment must be made.<sup>53</sup> It may be that “reviews should take place at the stage of the conception/design of the weapon, and thereafter at the stages of its technological development (development of prototypes and testing), and in any case before entering into the production contract.”<sup>54</sup> However, an evaluation is only possible once the technical capabilities of a weapon are known, which is not the case at present.<sup>55</sup> Admittedly, there is a risk that once the technology has been developed at great expense vested interests will make it nearly impossible politically to conclude that the result is unlawful. The solution may be to accompany the development process with constant reviews. In addition, one has to make sure that as much effort is invested in developing the weapon’s capacity to respect IHL as in its lethal capacity, including the development of safeguards against technical and communication errors.<sup>56</sup> Fortunately, to be accurate and to sense and process as much information as possible is both a military and a humanitarian imperative. As for the actual testing, it is obviously only useful if it can be excluded that autonomous weapons will act unpredictably in unforeseen circumstances,<sup>57</sup> but here again my technical assumption is that machines act according to algorithms and, therefore, according to a plan established by humans, even if that plan instructs them to adapt in a certain way to certain circumstances.

---

51. YORAM DINSTEIN, *THE CONDUCT OF HOSTILITIES UNDER THE LAW OF INTERNATIONAL ARMED CONFLICT* 88 (2d ed., 2010).

52. BOOTHBY, *supra* note 18, at 343.

53. See Isabelle Daoust, Robin Coupland & Rikke Ishoey, *New Wars, New Weapons? The Obligation of States to Assess the Legality of Means and Methods of Warfare*, 84 *INTERNATIONAL REVIEW OF THE RED CROSS* 345 (2002); Justin McClelland, *The Review of Weapons in Accordance with Article 36 of Additional Protocol I*, 85 *INTERNATIONAL REVIEW OF THE RED CROSS* 397 (2003). Specifically for autonomous weapons, see Backstrom & Henderson, *supra* note 22, 483–514; Schmitt & Thurnher, *supra* note 15, at 271–76.

54. ICRC GUIDE, *supra* note 49, at 23.

55. As submitted by HUMAN RIGHTS WATCH, *supra* note 5, at 22.

56. Alston, *supra* note 11, ¶ 40.

57. Asaro, *supra* note 5, at 692–93.



B. *Machines are not Bound by International Humanitarian Law: Problems of Attribution and of Accountability*

1. Only Humans are Responsible

Only human beings are subject to legal rules. In the case of autonomous weapons, IHL is addressed to those human beings who devise, produce and program them, as well as those who decide upon their use.<sup>58</sup> I reject the idea that IHL is inadequate to regulate autonomous weapons because they would be situated somewhere between weapon systems and combatants, and further reject the suggestion that a new category with new rules should be created to regulate them.<sup>59</sup> The difference between a weapon system and a human being is not quantitative but qualitative; the two are not situated on a sliding scale, but on different levels—subjects and objects. A combatant is a human being, only he or she is an addressee of legal obligations. However far we go into the future and no matter how artificial intelligence will work, there will always be a human being at the starting point.<sup>60</sup> In my understanding, an autonomous weapon system will always operate within the limits of its software; software designed by humans.<sup>61</sup> It is the human being who will decide whether a machine will be created and who will create it. Even if one day robots construct other robots, there will still be the need for a human being to develop the first robot and instruct it as to how to construct new robots. This human being is bound by the law; the machine is not bound by the law.

Human Rights Watch writes that it would be unclear who would be held accountable for unlawful actions a robot commits: “Options include the military commander that deployed it, the programmer, the manufacturer, and the robot itself, but all are unsatisfactory. It would be difficult and arguably unfair to hold the first three actors liable and the actor that actually committed the crime—the robot—would not be punishable.”<sup>62</sup> I agree with the last part of this statement and I find some suggestions that robots

---

58. BOOTHBY, *supra* note 18, at 283.

59. Liu, *supra* note 1, at 629.

60. Schmitt & Thurnher, *supra* note 15, at 235, 277.

61. DEFENSE SCIENCE BOARD, U.S. DEPARTMENT OF DEFENSE, TASK FORCE REPORT: THE ROLE OF AUTONOMY IN DOD SYSTEMS 1, 21 (July 2012).

62. HUMAN RIGHTS WATCH, *supra* note 5, at 4. *See also id.* at 40; Asaro, *supra* note 5, at 693; Sharkey, *supra* note 24, at 790–91; Liu, *supra* note 1, at 632.

could be scrapped or disabled as a kind of punishment absurd.<sup>63</sup> As for the first options, it is as fair to hold a commander of a robot accountable as it would be to hold accountable a commander who instructs a pilot to bomb a target he describes as a military headquarters, but which turns out to be a kindergarten. It is obvious that a commander deploying autonomous weapons must understand how they function, just as for any other means and method of warfare. In my view, the responsibility of such a commander is not a case of—nor is it analogous to—command responsibility,<sup>64</sup> but a case of direct responsibility, just as that of a soldier firing a mortar believing that it can land only on the targeted tank, but which will kill civilians he knows are following the tank. This is a question of the *mens rea*, intent and recklessness with which criminal lawyers are familiar, just as it is for a surgeon using a medical robot or, for that matter, prescribing a medicine. Based on their Protocol I, Article 36 assessment, States deploying robots must give military commanders and operators clear instructions as to when and under what circumstances the robots may be actually be used. The operator need not understand the complex programming of the robot,<sup>65</sup> but must understand the result, that is, what the robot is able and unable to do.<sup>66</sup> As for the manufacturer and the programmer, domestic criminal laws often hold criminally responsible those who deliberately, recklessly or negligently construct defective buildings or machines that lead to the loss of human life.<sup>67</sup> I do not think that the possession of autonomous decision-making capacity breaks the causal chain allowing attribution and responsibility,<sup>68</sup> because I assume that it is always humans who define how this au-

---

63. See HUMAN RIGHTS WATCH, *supra* note 5, at 45 (referring to some who seriously suggest this).

64. As argued by HUMAN RIGHTS WATCH, *id.* at 42–43; Heyns, *supra* note 3, ¶ 78.

65. As claimed by Heyns, *supra* note 3, ¶ 78.

66. DoD Directive 3000.09, *supra* note 8, ¶ 4(a)(3)(a); Schmitt & Thurnher, *supra* note 15, at 267.

67. See, e.g., Article 229 of the Swiss Criminal Code criminalizing conduct, either deliberate or negligent, by “[a]ny person engaged in the management or execution of construction or demolition work who wilfully disregards the accepted rules of construction.” SCHWEIZERISCHES STRAFGESETZBUCH [STGB] [Criminal Code] Dec. 21, 1937, SR 757 (1938) (Switz.), available at <http://www.admin.ch/opc/en/classified-compilation/19370083/index.html>. Schmitt and Thurnher consider the suggestion that the producer could be criminally responsible “spurious.” They do not explain who would be criminally responsible if the commander and operator trust that the system was correctly programmed, when, in fact, it had not been so programmed. Schmitt & Thurnher, *supra* note 15, at 278–79.

68. Liu, *supra* note 1, at 650.

tonomy will function. This implies the need for drafting specific standards of due diligence for both manufacturers and commanders.

The further question of whether a robot could distinguish lawful from unlawful orders<sup>69</sup> is equivalent to that of whether they are able to apply rules to a complex situation without human intervention. If they cannot, they may not be used. If they can, it will be easy to program them not to follow unlawful orders.<sup>70</sup> None of the reasons for which soldiers often obey unlawful orders apply to them.

## 2. The Problem of the Temporal Field of Application

The discussion above may raise problems concerning the temporal field of application of IHL, as many of those last human interventions—including development of an IHL compliant program instructing the autonomous weapon when to use lethal force—may occur before an armed conflict exists. In terms of criminal accountability, war crimes can only be committed in armed conflicts, thus raising the question of whether pre-conflict conduct is governed by IHL. I would suggest that IHL applies to all conduct of a State aimed at having effects during an armed conflict. But, in any event, a State using a weapon which was programmed in peacetime not to comply with IHL has not taken all feasible precautionary measures in wartime to avoid incidental civilian losses.

For criminal responsibility, the issue is trickier. It might be possible to treat the person who intentionally programs a system, which is used in good faith by an operator during an armed conflict, not to comply with IHL to be the indirect perpetrator of the war crime committed during the conflict. Another option would be to consider the programmer as a guarantor who is obliged to intervene in wartime to avoid the commission of violations of IHL, thus committing a war crime by omission if he or she does not intervene. Obviously, if the operator is conscious of the defect of the weapon, the programmer would be an accessory to the war crime.<sup>71</sup>

---

69. Heyns, *supra* note 3, ¶ 55.

70. Alston, *supra* note 11, ¶ 34.

71. I am grateful to my colleague, Ursula Cassani, Professor of Criminal Law at the University of Geneva, for these suggested solutions.

### 3. Technology Increases Transparency

On the other hand, technology possessed by autonomous weapons enables more precise reconstructions of events for criminal inquiries, although under IHL, in contrast to human rights law, there is no “mandatory ex-post facto review of all footage in cases of lethal use, regardless of the status of the individual killed.”<sup>72</sup> If an inquiry is necessary under the obligation to search for persons alleged to have committed war crimes,<sup>73</sup> an electronic record will facilitate accountability.<sup>74</sup> This presupposes that producers must program robots to record such information.

### 4. May Robots “Go Rogue”?

A further technical question arising in this context is whether there is a risk of an autonomous weapons “going rogue.” Might a weapon one day have the technical ability, either of its own “will” or as a result of system imperfections, to override its programs and instructions and start “deciding” in a truly autonomous manner to violate IHL? A philosopher quoted by Human Rights Watch writes, “[T]he possibility that an autonomous system will make choices other than those predicted and encouraged by its programmers is inherent in the claim that it is autonomous.”<sup>75</sup> I understand “autonomy,” however, as including equally autonomous decisions within a framework the robot is unable to override. If my understanding is correct, it would not be unfair to hold the developer of the weapon responsible for such a design flaw. The risk that the robot “goes rogue” must be avoided in the way the robot is devised. If this is not possible, such weapons must be outlawed. The entire implementation and accountability system of international law—as with all law—is exclusively addressed to humans. Ac-

---

72. Heyns, *supra* note 3, ¶ 81.

73. See Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field art. 49, Aug. 12, 1949, 6 U.S.T. 3114, 75 U.N.T.S. 31; Convention for the Amelioration of the Condition of the Wounded, Sick and Shipwrecked Members of Armed Forces at Sea art. 50, Aug. 12, 1949, 6 U.S.T. 3217, 75 U.N.T.S. 85; Convention Relative to the Treatment of Prisoners of War art. 129, Aug. 12, 1949, 6 U.S.T. 3217, 75 U.N.T.S. 135; Convention Relative to the Protection of Civilian Persons in Time of War art. 146, Aug. 12, 1949, 6 U.S.T. 3516, 75 U.N.T.S. 287.

74. Alston, *supra* note 11, ¶ 30 (while in paragraph 36 he writes mysteriously that unmanned systems do not archive information).

75. Robert Sparrow, *Killer Robots*, 24 JOURNAL OF APPLIED PHILOSOPHY 62, 70 (2007), quoted in HUMAN RIGHTS WATCH, *supra* note 5, at 43. *Contra* Schmitt & Thurnher, *supra* note 15, at 242.

countability mechanisms and sanctions directly addressed to machines are absurd, because machines are not capable of moral choice.

C. *Are Distinction and Proportionality Evaluations Based upon Rational Decisions or Subjective Judgments?*

1. Difficulties in Discrimination and Distinction

Concerning discrimination and distinction, advanced technology still remains to be developed before a robot can distinguish as well as a human being between legitimate and unlawful targets. The problem is not only the technological inadequacy of sensors, but also of translating IHL into a computer language.<sup>76</sup> Human beings may make a lot of mistakes; even more mistakes than a machine would have technical failures. Nevertheless, there are many elements that make a human being understand what is/is not a legitimate target, and those factors must be reproduced in a computer program. A robot must be able to sense all the necessary information in order to distinguish between targets in the same manner as a person. A “civilian with a large piece of metal in his hands” must obviously be distinguished from “a combatant with a rifle in plain clothes.”<sup>77</sup> It may be particularly difficult to automate the indicators that convince a human being a certain person belongs to a category (combatants; possibly members of an organized armed group who assume a continuous combat function) or is engaged in conduct (direct participation in hostilities) which makes them a legitimate target.<sup>78</sup> Pending revolutionary technological innovations, this problem may be solved by allowing a weapon system to target autonomously only those categories of objects that are, without question, targetable.<sup>79</sup> It is suggested, therefore, that autonomous weapon systems incapable of distinguishing meaningfully may be used in an environment where no civilians could be endangered.<sup>80</sup> However, this solution does not fully work when persons are targets, as anyone may surrender making it unlawful (and militarily unnecessary) to target them.

---

76. Heyns, *supra* note 3, ¶ 67.

77. *Id.*, ¶ 74.

78. Sharkey, *supra* note 24, at 788–89, and Liu, *supra* note 1, at 641, argue that this is impossible.

79. Backstrom & Henderson, *supra* note 22, at 492.

80. Schmitt & Thurnher, *supra* note 15, at 246.

As for objects, the definition of a military objective depends on its “effective contribution to military action” and the “definite military advantage” the attack offers “in the circumstances ruling at the time.”<sup>81</sup> These requirements imply a need to be aware of plans and the overall development of a military operation. An autonomous weapon system could therefore not be allowed to remain fully autonomous over time without receiving a constant update on those elements, which determine whether an object may be targeted.

It will be equally difficult to formalize factors which convince a human being that he or she must cancel or suspend an attack because the target is not lawful. For a machine to autonomously make such decisions may be even more difficult because the enemy could feign those indicators which cause the robot determine that it is confronted with a legitimate target. As for the reverse situation, that is, the enemy artificially fulfills the indicators which make a robot calculate that it may not attack under IHL, the fascinating question arises as to whether a machine can be “led to believe” something, or whether it is possible to “invite the confidence” of a machine—two elements of the prohibited act of perfidy.<sup>82</sup>

Another issue of distinction is that an autonomous weapon must be able to recognize when a legitimate human target surrenders or is wounded and abstaining from any act of hostility.<sup>83</sup> Beyond that, distinction has, in any case, become a very difficult task in counterinsurgency warfare for both machines and human beings. I simply do not see any reason of principle why a machine could never become better at fulfilling this task than a human being. Admittedly, however, as long as the necessary technology does not exist, human beings must be involved in the process.<sup>84</sup>

## 2. Difficulties in Determining who Directly Participates in Hostilities

It is unfortunately still controversial as to exactly what constitutes direct participation in hostilities. The ICRC’s *Interpretive Guidance* has suggested an understanding which, as mentioned previously, will be difficult to translate into instructions for a machine. The rule suggested by some critics of the *Interpretive Guidance*, that is, to make any member of an armed group tar-

---

81. See Protocol I, *supra* note 45, art. 52(2).

82. *Id.*, art. 37; 1 CUSTOMARY INTERNATIONAL HUMANITARIAN LAW, *supra* note 45, at 223.

83. Heyns, *supra* note 3, ¶ 67; BOOTHBY, *supra* note 18, at 286.

84. BOOTHBY, *supra* note 18, at 233.

getable,<sup>85</sup> will be equally difficult to translate because neither a machine nor a human being can determine whether an unknown person is a member of an organization. It is, however, certainly not necessary “to interpret intentions and emotions” in making that determination,<sup>86</sup> which would be an unrealistic criterion for both robots and soldiers when confronted with an enemy who is not in their control. In any case, if a robot has to be programmed, it cannot be instructed that “it all depends” or to “use reasonable judgment,” or, in other words, to be programmed that the appropriate action depends on the given situation without clarifying what action should be undertaken in which situation. In writing a computer program on the factors on which the targeting decision depends and how those factors can be determined, clarity will be required.

The definition of direct participation in hostilities raises another problem in connection with autonomous weapons. According to the *Interpretive Guidance*, the direct causation of harm, which is one of the requirements for an act to constitute direct participation in hostilities, “should be understood as meaning that the harm in question must be brought about in one causal step.”<sup>87</sup> Obviously only human steps can count (because the concept of direct participation defines who and not what may be targeted). Therefore, if autonomous weapons are used, the last causal human step leading to the harm caused may be geographically and temporally removed from the effect. The human being performing this step must be considered as directly participating in hostilities. In addition, the *Interpretive Guidance* correctly recalls:

The required standard of direct causation of harm must take into account the collective nature and complexity of contemporary military operations. For example, attacks carried out by unmanned aerial vehicles may simultaneously involve a number of persons . . . . While all of these persons are integral to that operation and directly participate in hostilities, only a few of them carry out activities that, in isolation, could be said to directly cause the required threshold of harm. The standard of direct causation must therefore be interpreted to include conduct that causes harm only in conjunction with other acts. More precisely, where a specific act does not

---

85. Michael N. Schmitt, *The Interpretive Guidance on the Notion of Direct Participation in Hostilities: A Critical Analysis*, 1 HARVARD NATIONAL SECURITY JOURNAL 5, 24 (2010). For a discussion of the criticism of the *Interpretive Guidance* in this particular regard, see *id.* at 14–24.

86. As Heyns, *supra* note 3, ¶ 68, suggests.

87. INTERPRETIVE GUIDANCE, *supra* note 23, at 54.

on its own directly cause the required threshold of harm, the requirement of direct causation would still be fulfilled where the act constitutes an integral part of a concrete and coordinated tactical operation that directly causes such harm.<sup>88</sup>

This interpretation must probably be even wider in cases involving the use of autonomous weapons than in cases, such as the one provided by the ICRC in this example, where human-controlled unmanned aerial vehicles are used.<sup>89</sup> In addition, while direct participation in hostilities does not, in my view, constitute a violation of IHL, IHL inherently requires that States only use combatants to directly participate in hostilities.<sup>90</sup> This implies that States must entrust to members of their armed forces the last human intervention in the determination of respect or non-respect for the principles of distinction, proportionality and precaution. By requiring this determination to be made by soldiers, it should facilitate training, supervision and accountability.

### 3. Difficulties in Applying the Principle of Military Necessity

As for the evaluation of military necessity, both soldiers and those who program autonomous weapons have to deal with the controversial issue of whether or not this principle involves an obligation to capture—rather than kill—legitimate targets.<sup>91</sup> For the same reasons discussed below with regard to the proportionality evaluation, I am not convinced that a human being is always better at making this decision, and that a robot is inherently incapable of “determin[ing] 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.”<sup>92</sup>

---

88. *Id.* at 55.

89. See also BOOTHBY, *supra* note 18, at 287–88.

90. For a detailed discussion, see LINDSEY CAMERON & VINCENT CHETAIL, *PRIVATIZING WAR, PRIVATE MILITARY AND SECURITY COMPANIES UNDER PUBLIC INTERNATIONAL LAW* 92–107 (2013)

91. On this controversy, see Ryan Goodman, *The Power to Kill or Capture Enemy Combatants*, 24 *EUROPEAN JOURNAL OF INTERNATIONAL LAW* 819 (2013); Michael N. Schmitt, *Wound, Capture, or Kill: A Reply to Ryan Goodman's 'The Power to Kill or Capture Enemy Combatants'*, 24 *EUROPEAN JOURNAL OF INTERNATIONAL LAW* 855; Ryan Goodman, *The Power to Kill or Capture Enemy Combatants: A Rejoinder to Michael N. Schmitt*, 24 *EUROPEAN JOURNAL OF INTERNATIONAL LAW* 863.

92. HUMAN RIGHTS WATCH, *supra* note 5, at 35.



#### 4. Difficulties in Translating the Proportionality Principle into a Computer Program

The need to translate the proportionality principle into a computer program for autonomous weapons may present an opportunity to improve objectivity. This principle, which is codified in Article 51(5)(b) of Protocol I, prohibits attacks, even if directed at a military objective, if they “may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated.” Despite the qualities the rule ascribes to military advantage, it is very difficult to conceptualize. Comparing military advantage anticipated against expected civilian losses is a process riddled with inevitably subjective value judgments, especially if there is not an absolute certainty that the advantage gained outweighs the effects on the civilian population, but, instead, the judgment is less certain. It might, however, be possible to identify, with the help of both military and humanitarian experts, indicators and criteria to evaluate proportionality, and to make the implied judgment slightly more objective. So far, such suggestions have been rejected by military lawyers who insisted that evaluations on the matter are dependent on the circumstances of particular situations and the good faith of military commanders.<sup>93</sup> A machine, however, needs clear criteria and a formula to calculate proportionality. The need to program autonomous weapons to respect proportionality (assuming in the first place that they are technically capable of acquiring, through sensors, the information necessary to apply the principle) may have the advantage of obliging States to agree on how exactly proportionality must be calculated and also on which parameters influence this calculation.

In my view, the greatest difficulty an autonomous weapon system will have in applying the proportionality principle is not linked to the evaluation of the risks for civilians and civilian objects, but to the evaluation of the military advantage anticipated.<sup>94</sup> I could imagine a robot gathering the necessary information to evaluate risks to civilians and even to proceed to the necessary evaluation if objective formulas are adopted. However, the “con-

---

93. Schmitt and Thurnher so state, although they then somewhat mysteriously claim that autonomous systems are not in principle unable to make such an evaluation. Schmitt & Thurnher, *supra* note 15, at 256, 257.

94. *See id.* at 255–57.

crete and direct military advantage anticipated”<sup>95</sup> from an attack on a legitimate target constantly changes according to the plans of the commander and the development of military operations on both sides. Except where no, or clearly negligible, effects upon civilians can be anticipated, a machine, even if perfectly programmed, could, therefore, not be left to apply the proportionality principle unless constantly updated about military operations and plans. This is in my view the most serious IHL argument against the even theoretical possibility of deploying weapons that remain fully autonomous over considerable periods of time.

##### 5. Must Targeting Decisions be Subjective?

Everything discussed thus far implies that there are considerable challenges for developers of autonomous weapons, and I am not sure they will be ever able to overcome them. Critics object for reasons of principle, however, that should those challenges be overcome, autonomous weapons would still violate IHL because “[e]ven if the development of fully autonomous weapons with humanlike cognition became feasible, they would lack certain human qualities, such as emotion, compassion, and the ability to understand humans.”<sup>96</sup> Their main objection to the theoretical possibility that a robot could distinguish between targets and make proportionality evaluations in a more objective and reliable way than a human being is that those rules involve subjective judgments,<sup>97</sup> which can indeed only be made by human beings. I would agree that a machine could never evaluate whether a witness in a trial is reliable. The question is simply whether targeting decisions are subjective. Many individuals in the military apparently agree that such decisions are subjective when they reject—the discussion on autonomous weapons aside—very detailed rules on proportionality, precautions and what constitutes direct participation in hostilities. I agree that “justice cannot be autonomous.”<sup>98</sup> To target a person is, however, definitely not to render justice or more precisely, it is not a determination that the person deserves the death penalty, but involves exclusively a categorization of the

---

95. Protocol I, *supra* note 45, art. 51(5)(b).

96. HUMAN RIGHTS WATCH, *supra* note 5, at 29; Heyns, *supra* note 3, ¶ 55.

97. HUMAN RIGHTS WATCH, *supra* note 5, at 4; Asaro, *supra* note 5, at 696–700; Sharkey, *supra* note 24, at 789–90; Heyns, *supra* note 3, ¶¶ 70–74 (more nuanced). Astonishingly, even Schmitt and Thurnher, who fail to explain how a machine could make subjective judgments, merely argue that the subjective decision remains that of the human being involved. Schmitt & Thurnher, *supra* note 15, at 256, 265–68.

98. Asaro, *supra* note 5, at 700.

person (as a combatant) or their conduct (direct participation in hostilities) without any determination of fault or culpability.

Human Rights Watch gives the following example: “Distinguishing between a fearful civilian and a threatening enemy combatant requires a soldier to understand the intentions behind a human’s actions, something a robot could not do.”<sup>99</sup> I agree only with the last part of the sentence, while the remainder misconstrues IHL.<sup>100</sup> Even a human being engaged in hostilities will never know, and is not required to inquire into, the intent of another human being, but instead will be receptive only to objective indications of the danger a person represents.<sup>101</sup> In the example, no one would criticize a soldier who fired upon a civilian who points, out of fear, a weapon at him or her.

A second example given by Human Rights Watch raises more serious questions. “[A] frightened mother may run after her two children and yell at them to stop playing with toy guns near a soldier.” Human Rights Watch is certain that “[a] human soldier could identify with the mother’s fear and the children’s game and thus recognize their intentions.”<sup>102</sup> First, I am not so sure that in some of today’s fighting environments, where there are child soldiers, fanaticized mothers, and differences in language and culture between soldiers and local populations, a human soldier would always be able to easily identify what was happening in such a situation. Second, what the soldier must determine is not the intent of the children (becoming a kind of child psychologist), but the objective risk of harm to him, his comrades and his mission, based upon objective indicators. Even if the mother was inciting the children to hate and the children were crying out in hate and subjectively willing to kill the soldier, the latter could not fire if it was apparent that the pistols were toy guns. Conversely, if the children simply wanted to play, but fired live ammunition as part of their game, the soldier could use force. Third, the example nevertheless shows how difficult devising an autonomous weapon capable of replacing a soldier in all circumstances will be.

---

99. HUMAN RIGHTS WATCH, *supra* note 5, at 4; Heyns, *supra* note 3, ¶ 55.

100. It is not surprising that Human Rights Watch references writings by a philosopher and a computer scientist, not those of a lawyer, to support this proposition. HUMAN RIGHTS WATCH, *supra* note 5, at 31.

101. International Committee of the Red Cross, *Fifth Expert Meeting on the Notion of Direct Participation in Hostilities: Summary Report* 37, 42, Geneva, Switz., Feb. 5–6, 2008 (May 9, 2008), available at <http://www.icrc.org/eng/assets/files/other/2008-05-report-dph-2008-icrc.pdf>.

102. HUMAN RIGHTS WATCH, *supra* note 5, at 31.

It is true that “common sense,” meaning the ability to make a sound practical evaluation of social situations, is based upon human experience. It would be particularly difficult to create a machine to reproduce this,<sup>103</sup> but herein could lay an opportunity to create “objective common sense” based, not upon the life history of one individual, but of several persons—ideally of mankind. I do not consider that autonomous weapons would only be lawful if they could demonstrate omniscience, but only if they can, on average, do a better job than an average soldier.

As for the proportionality rule, several authors and military manuals indicate that its application involves a subjective determination.<sup>104</sup> The question is, however, whether this is simply a description of the unfortunate reality, while recognizing that the determination should ideally be as objective as possible, or whether this is a normative proposition that the determination should be subjective. I admit that States have not yet been able to quantify how the risk of losing one civilian life compares with the potential of gaining a certain military advantage, nor what relation between the risk and the advantage would be excessive. I think, however, that both for human operators and for autonomous weapons it would be desirable if a formula for such a calculation, together with indicators of the elements that should/should not be taken into account, could be agreed upon.

Obviously, the determination must be made on a case-by-case basis (and modeling and determining indicators for the infinite variety of possible situations will be a perhaps insurmountable difficulty for producers of genuinely autonomous weapons<sup>105</sup>), but I do not see why it should be “subjective.”<sup>106</sup> According to one definition, subjective means

relating to the way a person experiences things in his or her own mind; based on feelings or opinions rather than facts; . . . a characteristic of or belonging to reality as perceived rather than as independent of mind; relating to or being experience or knowledge as conditioned by personal mental characteristics or states; peculiar to a particular individual; modified or affected by personal views, experience or background . . . ; arising from conditions within the brain or sense organs and not directly caused by external stimuli . . . .<sup>107</sup>

---

103. SINGER, *supra* note 4, at 131; Sharkey, *supra* note 24, at 789.

104. HUMAN RIGHTS WATCH, *supra* note 5, at 32; Heyns, *supra* note 3, ¶ 70.

105. Alston, *supra* note 11, ¶ 39; Sharkey, *supra* note 24, at 789–90.

106. HUMAN RIGHTS WATCH, *supra* note 5, at 32; Asaro, *supra* note 5, at 701.

107. *Subjective*, MERRIAM-WEBSTER DICTIONARY, <http://www.merriam-webster.com/dictionary/subjective> (last visited Apr. 16, 2014).

But why should a certain civilian be better protected under the law from incidental effects arising from an attack by one soldier than by another soldier? Why should the soldier's youth, education, values, religion or ethics matter at all? Should not the only consideration be the military advantage to be gained and the incidental effect upon civilians?

When the ICRC *Commentary* states that a commander must use "common sense and good faith,"<sup>108</sup> this does not, in my view, mean that the decision must be subjective. Taking the wide latitude of possible common sense evaluations into account, I would clearly prefer more precise criteria upon which to base decisions. I would not exclude robots from making such evaluations simply because they would do so in a necessarily objective way. Similarly, in my view, the International Criminal Court for the former Yugoslavia (ICTY) did not require that the determination of proportionality be subjective when it held: "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."<sup>109</sup> This is due to the fact that it tries individual human beings and not, as Human Rights Watch argues, because a proportionality evaluation requires "psychological processes in human judgment."<sup>110</sup> To the contrary, the ICTY formula attempts to liberate the evaluation from the purely subjective judgment of the attacker (which must nevertheless be taken into account when assessing his or her *mens rea*).

D. *Autonomous Weapons make Additional Precautions Possible*

Precautions in attack, in particular those listed in Article 57 of Protocol I, must be taken only if they are feasible. Obviously, whether a certain precautionary measure is feasible has to be measured against the alternatives available to those who plan and decide upon an attack or who execute it,

---

108. COMMENTARY ON THE ADDITIONAL PROTOCOLS OF 8 JUNE 1977 TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949, at 182 (Yves Sandoz, Christophe Swinarski & Bruno Zimmermann eds., 1987). *See also* Heyns, *supra* note 3, ¶ 72 (who uses the same terms).

109. Prosecutor v. Galic, Case No. IT-98-29-T, Judgment, ¶ 58 (Int'l Crim. Trib. for the former Yugoslavia Dec. 5, 2003).

110. HUMAN RIGHTS WATCH, *supra* note 5, at 33.

and not against the possibility for a machine to take a certain measure. Conversely, however, an autonomous weapon could be a means to render certain precautions, which would not be available to a soldier, feasible.<sup>111</sup> Because the human life of the pilot or weapons operator is not at risk, using autonomous weapons may result in the ability to take additional precautions. “Robots can thus act ‘conservatively’ and ‘can shoot second.’ Moreover, powerful sensors and processing powers . . . can potentially lift the fog of war for human soldiers . . . thus save lives.”<sup>112</sup> Obviously, this advantage is reduced if the enemy is prepared to defend against robot attacks. Those who produce robots will then program them to shoot first to avoid being destroyed by anti-robot weapons. We are then back to the classical situation in warfare.

Intelligent weapons have an additional advantage. The feasibility of precautions evolves through experience. When precautions taken in the past proved unsuccessful, that may imply the need to learn lessons (and belligerents have, in my view, an obligation to foresee pertinent procedures) to avoid such incidents in the future. If artificial intelligence can be created, it is essential to make sure that weapons with such intelligence can be recalled and reprogrammed, and that human beings monitor the development of that intelligence in order to quickly take advantage of lessons learned.

A legal issue arising is that important precautions, such as the obligations to verify the nature of the target and the lawfulness of the attack, to choose means and methods avoiding or minimizing incidental effects on civilians and to respect the proportionality principle, are addressed by Article 57(2)(a) of Protocol I only to “those who plan or decide upon an attack.” Some wonder whether this means that a human being must plan and decide to conduct the attack.<sup>113</sup> As discussed above, in my view, all rules of IHL are addressed only to human beings. This does not, however, preclude human planners and decision makers from being temporally and geographically removed from the attack, as long as they define the parameters according to which the robot attacks, make sure that it complies with them and has the necessary information to apply such parameters.

A major problem in terms of precautions is the obligation to cancel or suspend an attack

---

111. Kellenberger, *supra* note 16, at 813.

112. Heyns, *supra* note 3, ¶ 69. *See also* Alston, *supra* note 11, ¶ 30.

113. BOOTHBY, *supra* note 18, at 233.

if it becomes apparent that the objective is not a military one or is subject to special protection or that the attack may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated.<sup>114</sup>

Does this mean because an autonomous weapon is used and no human perceives the change in circumstances that the rule cannot, therefore, be violated?<sup>115</sup> In my view, the State's obligation to ensure that commanders do everything feasible to verify the lawfulness of targets implies that weapon systems must be designed to allow such verification—either through the system itself or by a human. Even those who favor lethal autonomous weapons do not seem confident that such weapons could be intelligent enough for it to become apparent to them that an attack is unlawful. They therefore suggest that there should be “some version of a human override.”<sup>116</sup> This may, however, raise the problem that machines process information so quickly based upon such a vast store of information that a human will be unable to really comprehend, evaluate and, if necessary, interrupt the machine's operation.<sup>117</sup> In addition, humans will have a tendency to trust the computer and hesitate to override it, as the example of *USS Vincennes* shooting down Iran Air Flight 655 demonstrated. Finally, communications between the human who is supposed to conduct the override and the machine may be disrupted or manipulated by the enemy.<sup>118</sup> It is my understanding, logically, if autonomous weapons are able to distinguish in the first place, they should be equally able to sense changes in their situational context and to cancel an attack if the given information indicates it is unlawful. If they are not able to distinguish in the first place, it would be inconsistent with IHL to deploy them autonomously.

For attacks involving autonomous weapons, as for any other attack, greater transparency about the precautionary measures taken and those dis-

---

114. Protocol I, *supra* note 45, Art. 57(2)(b).

115. As rhetorically asked by BOOTHBY, *supra* note 18, at 285.

116. Jeffrey S. Thurnher, *No One at the Controls: Legal Implications of Fully Autonomous Targeting*, 67 JOINT FORCES QUARTERLY, 4th Qtr. 2012, at 83.

117. Heyns, *supra* note 3, ¶ 41; Alston, *supra* note 11, ¶ 41; BOOTHBY, *supra* note 18, at 286; DEVELOPMENT, CONCEPTS AND DOCTRINE CENTRE, MINISTRY OF DEFENCE, JDN 2/11, THE UK APPROACH TO UNMANNED AIRCRAFT SYSTEMS 5–10 (2011).

118. Singer, *supra* note 4, at 125, 127.

carded as not feasible, would be highly desirable.<sup>119</sup> In accordance with international human rights law, States could also initiate inquiries, the outcomes of which would be made public. Planning and decision making are by definition secret, and it is often impossible to determine what the commander knew and what alternatives, if any, were available at the time of the attack. However, a higher degree of transparency after the fact would be useful for both the defendant and the prosecutor before an international or domestic court if criminal charges were brought. It would help prove the unlawfulness of certain behavior or, conversely, prove that IHL was respected. Greater transparency would be particularly useful in setting forth the facts for events that unnecessarily make the headlines or are used for propaganda purposes. The credibility of IHL would in turn be reinforced. Hence, it is regrettable that States and military lawyers often refuse to even start discussing such proposals, perhaps in fear of potential criticism and criminal prosecution.<sup>120</sup> The fact that operations by autonomous weapons always leave an electronic trail would facilitate such transparency and inquiries.

## V. CONCLUSION

I assume even autonomous weapon systems with artificial intelligence, though capable of learning, cannot do what the human beings who created them do not want them to do—or that it is at least possible to limit their autonomy in this regard. Such must be the case because they are not addressees of the law. If this is true, I cannot exclude that it may one day be possible to construct autonomous weapon systems which are capable of perceiving the information necessary to comply with IHL (this appears to me as the main challenge moving forward) and then to apply IHL to that information. For the time being, and pending evidence of revolutionary technical developments, it may be wise to limit the use of autonomous weapons to situations in which no proportionality assessment is needed and where the enemy consists of forces declared hostile in high-intensity

---

119. Marco Sassòli & Lindsey Cameron, *The Protection of Civilians Objects—Current Status of the Law and Issues de lege ferenda*, in *THE LAW OF AIR WARFARE: CONTEMPORARY ISSUES* 35, 71 (Natalino Ronzitti & Gabriella Venturini eds., 2006).

120. See Marco Sassòli & Julia Grignon, *Les Limites du Droit International Pénal et de la Justice Pénale Internationale dans la mise en Ouvre du Droit International Humanitaire*, in *LE DROIT INTERNATIONAL HUMANITAIRE FACE AUX DÉFIS DU XXI<sup>e</sup> SIÈCLE* 133, 142 (Abdelwahab Biad & Paul Tavernier eds., 2012).



conflicts. I believe it will still take time before they can be used in counter-insurgency operations. However, the IHL on targeting does not require subjective value judgments that machines are unable to make, but depends on an objective assessment of facts.

If my two technical assumptions and my understanding of IHL are correct, an attack executed by autonomous weapons would have many advantages in terms of distinction, proportionality and precautions over an attack directly executed by human beings. The development of autonomous weapons may even lead, because of programming needs, to a clarification of many rules that have so far remained vague and whose protective utility currently depends upon subjective value judgments. Most arguments of principle against autonomous weapons either do not withstand comparison with other means and methods of warfare—although the risk of proliferation has to be taken very seriously in this case—or they are based upon an erroneous understanding of IHL.

There are, nevertheless, challenges when applying existing IHL to autonomous weapons which necessitate agreement on the proper interpretation of IHL by every State using them and between States. As with all weapons, users must be appropriately trained and subject to accountability mechanisms, which requires that autonomous weapon systems keep records. Accountability must, however, equally apply to the producers. Such a requirement of accountability implies the need for agreed professional standards, but also that lawyers agree on the temporal field of application of IHL and the application of international criminal law to conduct occurring in peacetime which produces results during armed conflict.

In addition, in an environment in which the proportionality principle must be applied—and theoretically even to determine what is a military objective—autonomous weapon systems must be constantly updated on the plans and progress of operations to enable them to evaluate whether a definite military advantage can be anticipated to result from the attack and whether the risks for the civilian population are excessive compared with the concrete and direct military advantage anticipated.

Finally, parameters must be established for comparing the performance of autonomous weapon systems with that of human beings in carrying out attacks. The feasibility of precautions must be understood to refer to what would be feasible for human beings employing the machine, not to the possibilities available to the machine. A particularly tricky issue is to determine, in case autonomous weapons are used, the meaning of the obligation to cancel an attack when it becomes apparent that the attack is unlawful.

This implies that autonomous weapons possess sensing capability and ability to change behavior.

It is to be hoped that these issues will be discussed in the forthcoming experts' and States' meetings addressing the possible need for new rules on autonomous weapons.