HEINONLINE

Citation:
Roni A. Elias, Facing the Brave New World of Killer Robots: Adapting the Development of Autonomous Weapons Systems into the Framework of the International Law of War, 3 Indon. J. Int'l & Comp. L. 101 (2016)

Content downloaded/printed from *HeinOnline*

Mon Jun 11 08:40:04 2018

-- Your use of this HeinOnline PDF indicates your acceptance of HeinOnline's Terms and Conditions of the license agreement available at https://heinonline.org/HOL/License

-- The search text of this PDF is generated from uncorrected OCR text.

-- To obtain permission to use this article beyond the scope of your HeinOnline license, please use:

Copyright Information

Use QR Code reader to send PDF to your smartphone or tablet device

Facing the Brave New World of Killer Robots

Adapting the Development of Autonomous Weapons Systems into the Framework of the International Law of War

Roni A. Elias

Florida A&M University College of Law

E-mail: roni@leroyalusa.com

Many nations have developed weapons system with an advanced capacity for identifying targets and for making their own decisions about when to fire. These systems include both anti-missile defense systems and border-patrolling robotic sentries. As weapons technology becomes more sophisticated, and as computers continue to progress rapidly in their capacity for processing and analyzing information, it seems more and more possible that a country could develop something like a "killer robot," a machine with the power to kill and the independent capacity to identify targets and to decide when to engage them. This prospect creates the possibility of a real-life version of the killer robots portrayed in the series of Terminator *films, starring Arnold Schwarzenegger.*

Weapon systems capable of some level of autonomous decision-making have the potential to transform armed conflict, not only in a technical and strategic sense, but also in a legal one. The existing international law of armed conflict and principles of international human rights impose significant restrictions on how weapons can be used. And a foundational premise of that international

I would like to express my Sincere Appreciation for the entire Florida A&M, College of Law faculty & staff. Especially, I am so blessed and greatly appreciate the love of my life, M.G.S, for making my heart smile everyday! I also would like to thank & appreciate Dr. Elmira Mangum, Distinguished Educator of the Year, President, Florida A&M University & Provost & Vice-President for Academic Affairs and Professor of Law, Marcella David. This article is dedicated to my parents, Dr.'s Aida & Adil Elias, my First Great Teachers In Life. I am truly grateful to the best brother anyone could be blessed to have, my brother, Pierre. To the *Indonesian Journal of International and Comparative Law*, a heartfelt Thank You and Gratitude for your efforts, and laser sharp attention to detail.

The Indonesian Journal of International & Comparative Law
ISSN: 2338-7602; E-ISSN: 2338-770X
http://www.ijil.org
© 2014 The Institute for Migrant Rights Press

legal regime is the presumption that there is a human being with the capacity for moral judgment in control of the weapon.

There is an emerging debate about whether the existing structure of international law can be adapted to the inclusion of autonomous weapon systems. Some participants in this debate contend that autonomous weapon systems cannot be deployed without destroying the human moral agency that is required for compliance with international law. Others contend that autonomous weapon systems differ from other kinds of weapons only in small details, not in essential moral character, and that they can be effectively regulated just like any other weapon.

This paper examines the current state of development of autonomous weapon systems, the existing international law regulating the use of weapons in armed conflict, and the debate about the risks and benefits of integrating autonomous weapon systems into contemporary warfare. On the basis of this examination, it considers what is essential for continued compliance with the guiding principles of international law and about what methods may be adapted to assure that this law evolves with weapons technology. This paper concludes that autonomous weapon systems can only conform to the principles of international law if they are ultimately guided in some respect by a human being who can be morally accountable for how the weapon is used.

Keywords: Law of Wars, Law and Technology, Law and Ethics, International Criminal Responsibility, International Law.

III Indonesian Journal of International & Comparative Law 101-26 (January 2016)

I. INTRODUCTION

Science fiction and fantasy have long provided stories about "killer robots," machines with the power to kill and the independent capacity to identify targets and to decide when to engage them. Perhaps the most popular of these stories was represented in the trilogy of *Terminator* films, starring Arnold Schwarzenegger as a time travelling robot dispatched by an evil government as an assassin. In the end, the Terminator's capacity for independent thinking was sufficiently great that he could reject the instructions programmed into his central processing unit and turn on the other autonomous robots which had sent him on his mission in the first place. In a sense, these films both celebrated and condemned the capacities of autonomous killing machines, showing both the benefits and the dangers that they were capable of creating.

As weapons technology becomes more sophisticated, and as computers continue to progress rapidly in their capacity for processing and analyzing information, the prospect of a real-life "killer robot" moves ever closer. Many nations, especially the United States and several countries in Western Europe, have developed weapons system with an advanced capacity for identifying targets and for making their own decisions about when to fire upon them. These systems include both anti-missile defense systems and border-patrolling robotic sentries. It is not hard to imagine that something like Schwarzenegger's Terminator may not be a pure fantasy after all.

Weapon systems capable of some level of autonomous decision-making have the potential to transform armed conflict, not only in a technical and strategic sense, but also in a legal one. The existing international law of armed conflict and principles of international human rights impose significant restrictions on how weapons can be used. And a foundational premise of that international legal regime is the presumption that there is a human being with the capacity for moral judgment in control of the weapon.

There is an emerging debate about whether the existing structure of international law can be adapted to the inclusion of autonomous weapon systems. Some participants in this debate contend that autonomous weapon systems cannot be deployed without destroying the human moral agency that is required for compliance with international law. Others contend that autonomous weapon systems differ from other kinds of weapons only in small details, not in essential moral character, and that they can be effectively regulated just like any other weapon.

This paper examines the current state of development of autonomous weapon systems, the existing international law regulating the use of weapons in armed conflict, and the debate about the risks and benefits of integrating autonomous weapon systems into contemporary warfare. On the basis of this examination, it considers what is essential for continued compliance with the guiding principles of international law and about what methods may be adapted to assure that this law evolves with weapons technology. This paper concludes that autonomous weapon systems can only conform to the principles of international law if they are ultimately guided in some respect by a human being who can be morally accountable for how the weapon is used.

II. THE NATURE OF AUTONOMOUS WEAPON SYSTEMS

As an initial matter, it is important to define "autonomy" in the context of weapon systems. "Autonomy" must be distinguished from "automatic" or automated."[1] An "automatic" system is one that operates within a structured and predictable environment, such as a household appliance.[2] By contrast, an "autonomous" system is one that is designed to operate in a dynamic and unstructured environment, and its actions, like those of human beings, may be unpredictable, especially in situations that can be chaotic, like armed conflict, and in situations where they may interact with other autonomous systems.[3] But even this relatively large degree of unpredictability does not mean that an autonomous system has anything like "free will" or "moral agency."[4]

This means that "autonomous" weapon systems all currently operate under some significant degree of human control. And this is not likely to change. As one analyst has pointed out, "while the relevant technology is developing at an exponential rate, and full autonomy is bound to mean less human involvement in ten years time compared to today, sentient robots, or strong artificial intelligence are not currently in the picture."[5]

Thus, for the foreseeable future, any discussion of autonomous weapon systems must presuppose some level of human agency in the system's operation. They key to understanding the nature and legal significance of autonomous weapon systems lies in accurately describe the degree of human agency involved in the system and the way in which that agency can be exercise. Most of the current definitions of "autonomous weapon system" are premised on making that kind of description.

1. Christof Heyns, Report of the Special Rapporteur on Extrajudicial, Summary, or Arbitrary Executions, United Nations General Assembly, Human Rights Council 8 (Apr. 9, 2013), *available at* http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf (last visited Jan. 8, 2015).
2. *Id.*
3. *Id.*
4. *Id.*
5. *Id.*

III Indonesian Journal of International & Comparative Law 101-26 (January 2016)

The United States Department of Defense defines "autonomous weapon system" as:

> a weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation. The crux of full autonomy is a capability to identify, target, and attack a person or object without human interface. Although a human operator may retain the ability to take control of the system, it can operate without any control being exercised. Of course, a fully autonomous system is never completely human-free. Either the system designer or an operator would at least have to program the system to function pursuant to specified parameters.[6]

Human Rights Watch, an international organization dedicated to advocacy for human rights, offers an essentially similar definition that involves three different categories of autonomous weapon systems.[7] A "human in the loop" system is one in which a human directs the system to select a target and attack it.[8] A "human on the loop" weapon is one in which the system can select targets and attacks them without human intervention; but a human being does retain oversight over the system's operation and can override the system's decisions about targeting and engagement.[9] The third category includes systems that can attack without any human interface, and it is known as a "human out of the loop weapon."[10]

Nations with technologically advanced military capacities are moving quickly towards developing weapon systems with ever-higher degrees of autonomy. For example, the United States Navy has deployed a "Phalanx" system which defends ships against incoming missiles and

6. United States Dep't of Defense, Directive 3000.09: Autonomy In Weapon Systems 13–14 (Nov. 21, 2012), *available at* http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf (last visited Jan. 8, 2015).
7. Human Rights Watch, Losing Humanity: The Case Against Killer Robots 2 (Nov. 2012), *available at* http://www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf (last visited Jan. 7, 2015).
8. *Id.*
9. *Id.*
10. *Id.*

rockets through automated identification of targets and automated firing commands.[11] Similar air-defense weapon systems include the United States' "Patriot Missile System"[12] and Israel's "Iron Dome."[13] The United Kingdom has developed "fire and forget" Brimstone missiles, which can act on their own to identify tanks, cars, and buses and find targets in a pre-determined region without further human involvement.[14] Another example of a highly automated weapons defense system is the NBS Mantis, which is a "short-range force protection system, [that] will detect, track and shoot the projectiles within a close range of the target base."[15] Within less than five seconds after detecting targets about three kilometers away, it can fire six 35mm automatic guns at 1,000 rounds per minute.[16] The system has a "very high degree of automation, including automatic target detection and engagement processes which the operator only has to monitor."[17]

Autonomy is even coming to aircraft. "KMAX" helicopters, which were developed for the United States Army and Marines, can fly autonomously along pre-programmed routes.[18] The United States Navy has developed a combat aircraft, the X-47B, which can take off and land

11. MK 15—Phalanx Close-In Weapons System (CIWS), *available at* http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=487&ct=2 (last visited Jan. 8, 2015).
12. *See* Patriot Missile Long-Range Air-Defense System, United States of America, *available at* http://www.army-technology.com/projects/patriot/ (last visited Jan. 8, 2015).
13. Alex Gatopoulos, *How Successful Was Israel's Iron Dome?,* Aljazeera. (Sept. 8, 2014) *available at* http://www.aljazeera.com/news/middleeast/2014/08/israel-iron-dome-gaza-rockets-201481712494436388.html (last visited Jan. 8, 2015).
14. John Markoff, *Fearing Bombs that Can Pick Whom to Kill*, N.Y. Times (Nov. 11, 2014), *available at* http://www.nytimes.com/2014/11/12/science/weapons-directed-by-robots-not-humans-raise-ethical-questions.html?hp&action=click&pgtype=Homepage&module=secondcolumn-region.
15. NBS MANTIS Air Defence Protection System, Germany, *available at* http://www.army-technology.com/projects/mantis/ (last visited Jan. 8, 2015).
16. *Id.*
17. *Germany Orders MANTIS C-RAM Base Defense Systems*, Defense Industry Daily (Jan. 17, 2011), *available at* http://www.defenseindustrydaily.com/Germany-Orders-Skyshield-C-RAM-Base-Defense-Systems-05418/ (last visited Jan. 8, 2015).
18. Lockheed Martin, K-MAX, *available at* http://www.lockheedmartin.co.uk/us/products/kmax.html (last visited Jan. 8, 2015).

by itself on an aircraft carrier.[19] The British Royal Air Force is currently developing an attack aircraft called the "Taranis," which is expected to be capable of autonomous supersonic flight.[20]

Autonomy is not reserved for aircraft and missile defense applications. Some new systems perform functions previously reserved for individual foot soldiers on the battlefield. In 2010, South Korea launched a system that performed the functions of a sentry for use along the boundaries of its demilitarized zone ("DMZ") with North Korea—the SGR-1. This system was comprised of a series of stationary robots with the capacity to sense the presence of humans within the DMZ through heat and motion sensors.[21] When a person is detected in the DMZ, these robot sentries transmit a warning to a command center.[22] In the command center, human soldiers can communicate with the person identified in the DMZ and decide whether to fire the robot's 5.5mm machine gun or its 40mm grenade launcher.[23] The SGR-1's sensors can detect people two miles away during the day and one mile away at night, and its guns have a lethal range of two miles.[24] Currently, only the surveillance functions of the system are entirely automated; a human command is still required to fire on a target.[25] But the journal of the Institute of Electrical and Electronics Engineers has reported, however, that "the robot does have an automatic mode, in which it can make the decision."[26]

19. Northrop Grumman, X-47B UCAS, *available at* http://www.northropgrumman.com/Capabilities/X47BUCAS/Pages/default.aspx (last visited Jan. 8, 2015).
20. Jeffrey S. Thurnher, *Examining Autonomous Weapon Systems from a Law of Armed Conflict Perspective*, *in* New Technologies and the Law of Armed Conflict 213-18 (H. Nasu & R. McLaughlin eds. 2014).
21. Jon Rabiroff, *Machine Gun-Toting Robots Deployed on DMZ*, Stars and Stripes (July 12, 2010), *available at* http://www.stripes.com/machine-gun-toting-robots-deployed-on-dmz-1.110809 (last visited Jan. 8, 2015).
22. *Id.*
23. *South Korea's Military Technologies: Defensive Robots and Urine Powered Batteries*, Korea IT Times (July 14, 2010), *available at* http://www.koreaittimes.com/story/9750/south-koreas-military-technologies-defensive-robots-and-urine-powered-batteries (last visited Jan. 8, 2015).
24. Rabiroff, *supra* note 21.
25. *Id.*
26. Jean Kumagai, *A Robotic Sentry for Korea's Demilitarized Zone*, IEEE Spectrum (Mar. 1, 2007), *available at* http://spectrum.ieee.org/robotics/mili-

Israel has developed a similar sentry system. It has deployed the Guardium, "a semi-autonomous unmanned ground system," which is reportedly used for patrolling Israel's border with Gaza.[27] According to a brochure published by its manufacturer, G-NIUS, the Guardium "was designed to perform routine missions, such as programmed patrols along border routes, but also to autonomously react to unscheduled events, in line with a set of guidelines specifically programmed for the site characteristics and security doctrine."[28] While the brochure implies there is some level of human oversight because it refers to stationary, mobile, and portable control terminals, it also notes that the Guardium can have "autonomous mission execution."[29] This seems to mean that the Guardium could identify targets and fire on them without human direction.

None of these weapons are currently capable of selecting and attacking targets entirely without direct human action. But even in these systems, human participation can be significantly limited. In discussing Germany's C-RAM, one commentator has noted that "[t]he human is certainly part of the decision making but mainly in the initial programming of the robot. During the actual operation of the machine, the operator really only exercises veto power, and a decision to override a robot's decision must be made in only half a second, with few willing to challenge what they view as the better judgment of the machine."[30]

But progress in autonomous technologies indicates that the time is not far off when weapons will be capable of targeting and attacking autonomously, without immediate human control—that is, without humans "in the loop," and with, at the most, human supervision.[31] The

tary-robots/a-robotic-sentry-for-koreas-demilitarized-zone (last visited Jan. 8, 2015).

27. Human Rights Watch, Losing Humanity: The Case Against Killer Robots 2 (Nov. 2012), *available at* http://www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf (last visited Jan. 7, 2015).

28. G-NIUS Unmanned Ground Systems, Guardium UGV, *available at* http://g-nius.co.il/pdf/brochures/GuardiumUGV.pdf (last visited Jan. 8, 2015).

29. *Id.*

30. P.W. Singer, *War of the Machines: A Dramatic Growth in the Military Use of Robots Brings Evolution in Their Conception*, Scientific American 63 (July 2010).

31. Thurnher, *supra* note 20, at 226-27.

Ministry of Defense in the United Kingdom has stated that it "currently has no intention to develop systems that operate without human intervention in the weapon command and control chain, but it is looking to increase levels of automation where this will make systems more effective."[32] Similarly, the United States Defense Department anticipates making further substantial advances in weapons automation, as long as any new systems would "allow commanders and operators to exercise appropriate levels of human judgment over the use of force."[33]

III. EXISTING LAW GOVERNING THE USE OF WEAPONS

The law governing the use of weapons in armed conflict comes from a variety of international agreements and treaties, which can be classified into two principal categories: the law of armed conflict, which is principally shaped by the Geneva Conventions; and the law of international human rights, which is shaped by principles and rules adopted in numerous international conventions, especially those organized in connection with the United Nations.[34] These laws apply to nations that are signatories to or participants in the relevant conventions, and they are often reinforced by legislation adopted within the signatory or participant nations.[35] The rules relevant to the use of any weapon system, including autonomous weapon systems, relate to rules governing the kinds of weapons that can be used as means and methods of warfare

32. United Kingdom Ministry of Defense, Development, Concepts and Doctrine Centre, The UK Approach to Unmanned Aircraft Systems, Joint Doctrine Note 2/11 para. 508 (Mar. 30, 2011), *available at* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/33711/20110505JDN_211_UAS_v2U.pdf (last visited Jan. 8, 2015).
33. United States Dep't of Defense, *supra* note 6; see also Thurnher, *supra* note 20, at 213–18.
34. International Committee of the Red Cross, Handbook on International Rules Governing Military Operations 21-22 (2013), *available at* https://www.icrc.org/eng/assets/files/publications/icrc-002-0431.pdf (last visited Jan. 8, 2015).
35. *Id.* at 22.

and the way in which those weapons can be targeted.[36]

The law of armed conflict proceeds from the several foundational premises. One premise is that combatants do not have an unlimited right to choose means and methods of warfare.[37] "This is sometimes referred to as the principle of limitation, which is reflected in a series of treaty-based rules restricting specific means and methods of warfare."[38] Another cardinal principle is one of "distinction." The principle of distinction requires that the parties to a conflict distinguish between the civilian population and combatants and between civilian objects and military objectives, so that they focus their attacks only on combatants and legitimate military objectives.[39]

These principles lead to certain specific rules about the use of weapons. Combatants may not use any weapon that, by its very nature causes superfluous injury or unnecessary suffering to combatants, or to use any weapon in a manner that causes superfluous injury or unnecessary suffering.[40] International humanitarian law also prohibits the use of any weapon that cannot be directed at a specific military objective.[41] Thus, a weapon or weapon system is unlawful *per se* if it will strike combatants, military objectives, civilians, and civilian objects without distinction. As the Red Cross puts it, "it is prohibited to use a weapon which is by its very nature 'indiscriminate.'"[42] It is also unlawful to use a weapon in a disproportionate manner.[43] According to the Red Cross, "It is prohibited to engage in an attack which 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

36. *Id.* at 143, 177.
37. *Id.* at 177.
38. *Id.*
39. Michael N. Schmitt, *Autonomous Weapons Systems and International Humanitarian Law: A Reply to the Critics*, Harv. Int'l Security J. Features 14-15 (2012), *available at* http://harvardnsj.org/wp-content/uploads/2013/02/Schmitt-Autonomous-Weapon-Systems-and-IHL-Final.pdf (last visited Jan. 7, 2015).
40. Yoram Dinstein, The Conduct of Hostilities Under the Law of International Armed Conflict 63-67 (2d ed., 2010).
41. International Committee of the Red Cross, *supra* note 34, at 180.
42. *Id.*
43. *Id.* at 181.

in relation to the concrete and direct military advantage anticipated."[44]

International law also establishes rules about how targets are chose for weapons. In general, the law of targeting focuses on three principles: distinction, proportionality and precautions in attack.[45] These principles apply in both international armed conflict, as well as in armed conflict within a single nation.[46]

With respect to distinction, combatants must distinguish between military and civilian[47] persons and objects, and combatants may not target any civilian person or object unless that person or object is taking a direct part in hostilities.[48] In an international armed conflict, lawful military targets include military objectives and individual combatants who are actively involved in hostilities.[49] On the other hand, it is unlawful to target combatants that are "hors de combat," medical and religious personnel, and, of course, civilians—except for those directly participating in the conflict.[50] Targeting rules also include principles that are the inverse of those prohibiting indiscriminate attacks. Targeting is consistent with international law when it: is directed at a specific lawful military target; employs means and methods of combat that are capable of being directed at a specific lawful military target; employs means and methods of combat that can be limited as required by the law of armed conflict.[51] In the event that an attack is aimed at clearly distinct military objectives within an area containing a concentration of civilians or civilian objects, the attack must be subdivided into separate attacks on each of the multiple military objectives.[52]

With respect to proportionality, an attack is prohibited by international law "when it may be expected to cause incidental harm to civilians and civilian objects which would be excessive in relation

44. *Id.*
45. *Id.* at 143-47.
46. *Id.*
47. "Civilians" are defined as all persons who are not members of the armed forces of a party to a conflict. *Id.* at 144.
48. *Id.* at 143-47.
49. *Id.* at 146.
50. *Id.*
51. *Id.* at 147.
52. *Id.*

to them concrete and direct military advantage anticipated."[53] While the theory behind the principle of proportionality is clear enough, it can be difficult to translate that theory into practice. While there is no uncertainty about the idea that civilian harm should not be excessive in relation to military advantages gained, it can be difficult to evaluate the cost of any civilian harm in terms of the benefits of military advantage because those costs and benefits are measured on entirely different scales.[54] As a practical matter, there is no objective formula that can be used to determine when any particular use of force is "excessive," and the analysis of when this principle has been violated usually depends upon drawing analogies with truly outlandish cases, a method that hardly lends itself to objectivity.[55] Thus, the method for assessing proportionality typically involves examining what was done in prior situations and determining a course of future action based on similarities and differences between the prior situation and the current one.[56]

Finally, with respect to precautions in attack, international law requires "that constant care must be taken by all those involved in the conduct of military operations to spare the civilian population, civilians and civilian objects.[57] Principles relating to precautions in attack require a combatant to take feasible precautions in the circumstances to spare the civilian population.[58] The concepts of "precautions" and "feasibility" are both terms of art in the law of armed conflict, providing military commanders with reasonable discretion in undertaking attacks. The

53. *Id.* at 147.

54. *See* WILLIAM H. BOOTHBY, THE LAW OF TARGETING 96–97 (2012).

55. *See id.*

56. In the United States military, officers charged with making broad-scale targeting decisions sometimes employ sophisticated software programming to estimate likely collateral damage from the employment of a particular weapon in particular circumstances. See Schmitt, *supra* note 39 at 19. But such software is a fairly blunt instrument because it depends upon specific assumptions that do not apply in every case and only address the likely collateral damage—not the comparison to military advantage. *Id*; *see also* GEOFFREY S. CORN ET AL., THE LAW OF ARMED CONFLICT: AN OPERATIONAL APPROACH 194 (2012).

57. INTERNATIONAL COMMITTEE OF THE RED CROSS, *supra* note 34, at 149.

58. DINSTEIN, *supra* note 40, at 138–40; *see also* IAN HENDERSON, THE CONTEMPORARY LAW OF TARGETING: MILITARY OBJECTIVES, PROPORTIONALITY, AND PRECAUTIONS UNDER ADDITIONAL PROTOCOL I 157-96 (2009).

III Indonesian Journal of International & Comparative Law 101-26 (January 2016)

commander must exercise this discretion with reasonableness and good faith, and in "planning, deciding upon or executing attacks, the decision taken by the person responsible has to be judged on the basis of all information available to him at the relevant time, and not on the basis of hindsight." [59]

In addition to these rules, states have an obligation to review new weapon systems to assure that they will be capable of complying with these rules. The International Committee of the Red Cross has contended that "the faithful and responsible application of its international law obligations would require a State to ensure that the new weapons, means and methods of warfare it develops or acquires will not violate these obligations."[60] Although not all states have accepted this proposition, many weapons-producing states have agreed to abide by it, including the United States.[61]

59. *See* W. Hays Parks, *The Protection of Civilians from Air Warfare*, 27 ISR. Y.B. ON HUM. RTS 65, 110 (1997) (quoting German declaration on Article 57, Protocol I).

60. 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 5-6 (2006), *available at* http://www.icrc.org/eng/assets/files/other/icrc_002_0902.pdf (last visited Jan. 8, 2015).

61. The governments that have adopted formal review mechanisms include Australia, Belgium, the Netherlands, Norway, Sweden, the United Kingdom, and the United States. *Id.* at 5-6, n. 8. For the United State's description of its own policy concerning weapons review, see UNITED STATES DEP'T OF DEFENSE, DIRECTIVE 5000.1: THE DEFENSE ACQUISITION SYSTEM, DEFENSE ACQUISITION GUIDEBOOK, E1.1.15 (May 12, 2003), *available at* https://acc.dau.mil/CommunityBrowser.aspx?id=314790&lang=en-US (last visited Jan. 8, 2015).

IV. THE CURRENT DEBATE OVER WHETHER AUTONOMOUS WEAPONS SYSTEMS CAN COMPLY WITH EXISTING RULES OF INTERNATIONAL LAW FOR THE USE OF WEAPONS

There is currently an intensive debate about whether autonomous weapon systems can be effectively regulated by the existing regime of international law with regard to the use of weapons in armed conflict. This debate involves both government agencies and commentators and advocates about military policy. In particular, the United Nations has sponsored numerous explorations of different perspectives on this subject through its Convention on Certain Conventional Weapons,[62] under the auspices of its Human Rights Council,[63] as well as its Institute for Disarmament Research.[64] Other organizations have sponsored discussions of how best to regulate the use of autonomous weapon systems, including the International Committee of the Red Cross ("ICRC"),[65]

62. Report of the 2014 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) (June 11, 2014), *available at* http://daccess-dds.ny.un.org/doc/UNDOC/GEN/G14/048/96/PDF/G1404896.pdf?OpenElement (last visited Jan. 8, 2015).

63. *See* Christof Heyns, Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions (2013), *available at* https://papersmart.unmeetings.org/media2/703363/statement-by-christof-heyns-item-69b.pdf (last visited Jan. 9, 2015).

64. The Weaponization of Increasingly Autonomous Technologies: Implications for Security and Arms Control, *available at* http://www.unidir.org/programmes/security-and-society/the-weaponization-of-increasingly-autonomous-technologies-implications-for-security-and-arms-control (last visited Jan. 8, 2015).

65. Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects 26–28 (Nov. 2014), *available at* ttps://www.icrc.org/en/document/report-icrc-meeting-autonomous-weapon-systems-26-28-march-2014#.VK9Q64rF8jE (last visited Jan. 8, 2015).

III Indonesian Journal of International & Comparative Law 101-26 (January 2016)

academic institutions,[66] and think tanks.[67]

Numerous common points were have been made in the many different forums for this debate. At one end of the spectrum, some have argued that autonomous weapons systems have the capacity to make compliance with international law easier. At the other end of the spectrum, advocates have expressed serious concerns that the development of autonomous weapons systems will make such compliance much more difficult, chiefly by removing the moral dimension of decision-making about the use of deadly force.

Those who are optimistic about the prospects for integrating autonomous weapon systems into the existing international legal regime have focused on several points. First, they note that such systems will be faster at sensing and processing information.[68] Second, the increased capacity for acquiring and processing information will increase thereby increasing the likelihood of making accurate decisions about targeting, as well as the flexibility and speed with which those decisions can be made.[69] Third, commentators have suggested that there are significant advantages to removing the dynamic of human emotion from decisions about how to target and apply deadly force. According to this view, autonomous weapon systems will be more effective than human beings at undertaking dull, dirty, and dangerous tasks.[70] Moreover, the absence of particular emotions such as fear, the desire for revenge, or the desire to promote individual self-interest may lead to outcomes that are, on the whole, less harmful than those that

66. Geneva Academy of International Humanitarian Law and Human Rights, Summary of Discussions: Experts Meeting on Armed Drones and Robots under International Law 3–5 (Dec. 2013), *available at* http://www.geneva-academy.ch/docs/expert-meetings/Weapons%20Law%20Review_Summary%20of%20Expert%20Meeting%20Discussions_FINAL-1.pdf (last visited Jan. 8, 2015.

67. Autonomous Military Technologies: Policy & Governance for Next Generation Weapon Systems, (Feb. 2014), *available at* http://www.chathamhouse.org/Autonomous (last visited Jan. 8, 2015).

68. *See, e.g.,* W. H. Boothby, Conflict Law: The Influence of New Weapons Technology, Human Rights and Emerging Actors 104-07 (2014); and M. Sassòli, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified*, 90 Int'l L. Stud., Naval War College 310 (2014).

69. Boothby, *supra* note 68; Sassòli, *supra* note 68.

70. Boothby, *supra* note 68.

could be accomplished through human agency.[71] As one scholar put it, "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 [international human rights law] than a soldier on the battlefield or in a hostile environment."[72] Finally, the optimists about autonomous weapon systems expect that, by replacing human fighters, those systems have the capacity to save lives.[73]

Those who are more skeptical about autonomous weapon systems contend that their increasing use will diminish the force of international human rights law and the law of armed conflict by greatly diminishing, if not completely removing, the moral dimension of weapons use. According to this perspective, reliance on autonomous weapon systems remove positive human emotions, such as compassion or mercy, from the decision-making process about targeting or engagement.[74] Along similar lines, the value of human life may be diminished if machines are in a position to make essentially independent decisions about who should be killed in armed conflict.[75] These skeptics also point out that the use of a technology that creates little risk for the party using it, but real risk for the fighters and civilians of the user's enemy is asymmetrical and unfair.[76] Finally, they are concerned about the reliability of these systems and that the speed and complexity with which they process information will make it difficult for any human being supervising the system to correct mistakes that might occur.[77]

Along with these contrary viewpoints, some commentators have

71. International Humanitarian Law and the Challenges of Contemporary Armed Conflicts (2011), *available at* https://www.icrc.org/eng/assets/files/red-cross-crescent-movement/31st-international-conference/31-int-conference-ihl-challenges-report-11-5-1-2-en.pdf (last visited Jan. 8, 2015); K. Anderson, et al., *Adapting the Law of Armed Conflict to Autonomous Weapon Systems*, 90 Int'l L. Stud., Naval War College, 393 (2014); Boothby, *supra* note 68.
72. Sassòli, *supra* note 68.
73. Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects, *supra* note 65, at 9.
74. *Id.*
75. Sassòli, *supra* note 68.
76. *Id.*
77. Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects, *supra* note 65, at 8.

III Indonesian Journal of International & Comparative Law 101-26 (January 2016)

pointed out that there are open questions about how autonomous weapon systems will work and that these questions make it difficult to assess in advance whether and to what extent those systems will comply with international law. For example, one of the requirements for the lawful use of autonomous weapon systems is that they have the ability to distinguish lawful from unlawful targets.[78] The capacity for making these distinctions might vary enormously from one weapon system's technology to another and with the battlefield environment in which a weapon system is used.[79] Some algorithms, sensors or analytic capabilities might perform well, others poorly.[80] Moreover, these capacities are meaningful only in the context of particular battlefield environments; the "context and environment in which the weapon system operates play a significant role" in any analysis of whether a weapon system is in compliance with international law.[81]

V. ARE AUTONOMOUS WEAPON SYSTEMS INHERENTLY UNLAWFUL UNDER THE EXISTING RULES OF INTERNATIONAL LAW FOR THE USE OF WEAPONS?

The debate about the advantages and disadvantages of autonomous weapon systems raises a vital question: are such systems inherently incompatible with the fundamental principles of international law? To put the question another way, is it possible to develop an autonomous weapon system that conforms to the principles requiring a distinction between lawful and unlawful targets, proportionality in the use of deadly force, and taking precautions in attack, as well as to principles

78. *See* Jeffrey S. Thurnher, *The Law that Applies to Autonomous Weapon Systems*, 17 Am. Soc. of Int'l L. Insights (Jan. 18, 2013), *available at* www.asil.org/insights/volume/17/issue/4/law-applies-autonomous-weapon-systems (last visited Jan. 8, 2015).

79. *Id.*

80. *Id.*

81. *Id.*

regarding the indiscriminate use of deadly force?

Some critics of autonomous weapons systems contend that it is not possible to use such systems in a manner consistent with international law. This contention focuses on the proposition that it the lawful use of any weapon requires a high degree of moral agency.[82] According to this viewpoint, the entire edifice of the international law of war rests on the presumption that the agents doing the fighting are *moral agents*—that is, agents who can assume moral responsibility for specific actions.[83] The international law of war does not merely require that soldiers have the capacity to distinguish a combatant from a noncombatant or calculate whether the damages resulting from a particular use of weapons are proportional to the military advantage to be gained from that use.[84] Rather, the law demands moral decision-making by actors in every instance.[85] This kind of moral accountability for decision-making is essential for the law to be able to make its own distinctions between conduct that is morally blameworthy and conduct that is, at the very least, not unlawful.

If moral agency is identified as the core principle of the international law of war, these critics of autonomous weapon systems believe that a weapon system is necessarily unlawful if it involves any significant degree of autonomy in a machine's decision-making about how to select a target or about whether to attack that target. These critics suggest that delegating morally significant decisions to machines is a form of "ethical nihilism," which undermines the essential foundation of the international law of war.[86] Thus, even if an autonomous weapon system is capable of making more accurate decisions about targeting and engagement and other crucial matters, its decisions still cannot be morally evaluated, making the international law of war beside the point.[87]

82. *See* Anthony Beavers, *Moral Machines and the Threat of Ethical Nihilism* in P. Lin, *et al.*, eds.Robot Ethics: The Ethical and Social Implications of Robotics 333-44 (2012).
83. *See id.*
84. *See id.*
85. *See id.*
86. *See id.*
87. R. Sparrow, *Robotic Weapons and the Future of War* in Paolo Tripodi & Jessica Wolfendale (eds.) New Wars and New Soldiers: Military Ethics in the

This problem of moral agency is especially difficult when there is a disparity of technological capacity between combatants in an armed conflict. In other words, the problem of moral agency makes it even harder to apply the principles of the international law of war to a conflict if one side has sophisticated autonomous weapon systems that are making largely independent decisions, while the other side does not. This problem becomes apparent when it is recognized that the validity of the international law of war depends upon the principle of the moral equivalence of soldiers.[88] This principle presumes that soldiers on both sides of a conflict moral agents and that they have enough individual freedom to be held accountable for their actions. After all, if a soldier lacks any freedom at all to make his decisions about how to act, he cannot be held morally responsible for his actions and cannot, therefore, be subject to the law of war. At least in theory, and probably in practice in the very near future, autonomous weapon systems are capable of making many decisions and taking many actions without any meaningful human participation. Therefore, in a conflict between one party that used autonomous weapon systems and another party that did not, the individual combatants would not have moral equivalence. To put it bluntly, one side would be using robots that were not morally accountable, while the other was using human beings who would be. In such a conflict, it would be difficult to justify a universal application of the international law of war when, as a practical matter, that rule of law could only be effectively enforced against one side.

In this respect, the critics of autonomous weapon systems make a powerful point. They do not argue that it is impossible to adapt the particular requirements of the international law of war to machines. Rather, they contend that it is impossible to apply any form of law to a machine in any way. They emphasize the point that law, by its very nature, is a moral instrument, and it is impossible to apply a moral instrument to a machine. If a fatal auto accident occurs because the brakes malfunctioned on someone's car, the car is not prosecuted under the criminal law, nor is it even subject to civil liability. The critics of any use of autonomous weapon systems contend that trying to apply the international law of war to an autonomous weapon system would be

Contemporary World 123 (2012).

88. David R. Mapel, *Coerced Moral Agents? Individual Responsibility for Military Service*, 6 J. Polit. Phil. 171 (1998).

the functional equivalent of trying to apply the criminal law or the law of negligence to an automobile.

This is a powerful critique, but there are two very important problems with it. One of those problems is conceptual. This critique presumes that moral agency is entirely removed from the use of an automated weapon system just because a moral agent is not making every part of the decision about targeting and firing. This presumption is not well founded. Many ordinary weapons include features that remove human beings from each and every decision about the targeting and firing of a weapon. For example, even an automatic rifle or a machine gun makes it impossible for a human being to make a discrete moral decision about each and every round that is fired; the discharge of individual rounds happens too fast for the person controlling the weapon to make a distinct choice about where each round will go. But no-one suggests that automatic rifles are not capable of being regulated through the law of war. To be sure, contemporary (not to mention future) autonomous weapon systems are exponentially more complicated than automatic rifles in the number and kind of decisions that are removed, deliberately or effectively, from the direct control of the human beings supervising the machine. But the principle is the same. A weapon is not inherently ungovernable just because a person cannot direct every single action that the weapon takes. Weapons are within the control of human beings as long as those human beings have the capacity to direct the process by which the weapon executes its actions.

The second problem with the critique is a practical one. Even if one agrees as a matter of theory that autonomous weapon systems cannot be governed by the international law of war, it will be impossible to create and enforce any rule of law prohibiting their use. For one thing, it will be difficult to distinguish between types of weapons systems that are merely "automated" and "autonomous." Moreover, even if that distinction can be made in a meaningful way, there is no way to prevent combatants from deploying these weapons—especially not when it can be difficult to detect when a weapon system might be operating in an autonomous mode as opposed to an automatic one. It seems more effective to accept autonomous weapons as an inevitable development and to try to regulate them rather than outlaw them altogether.

The next question to arise is how to effect that regulation of automated weapon systems. In particular, and especially in light of the

important points made about moral agency in the use of weapons, the question arises about how best to regulate autonomous weapons so that their use necessarily involves moral agency. In other words, the crucial question in regulating autonomous weapons is how to assure that there is always a human being involved in making a morally responsible decision about how, when, and where the weapon is deployed.

Some have suggested that it would be possible to implement a technological solution to this problem, one in which moral decision-making would be automated along with the decisions about targeting and firing. This would involve the creation of an artificially intelligent machine. Creating a weapon system that was capable of making something like a moral judgment would require "strong" artificial intelligence, a level of intelligence that would match or even exceed human intelligence.[89] This kind of "strong" artificial intelligence might also have to include intentions, the capacity for reflection, and something like "consciousness."[90] Developing this kind of artificial intelligence would likely require systems to "learn" in a manner very similar to the way that human beings learn—through different patterns and experiences.[91]

One scholar has described what this kind of artificial intelligence might look like, calling this control system an "ethical governor."[92] An ethical governor would involve the imposition of a two-step process that the weapon system would have to complete before firing.[93] The first step would require the system to evaluate the information in terms of international humanitarian law and the rules of engagement. If the attack would violate one of those rules, the process stops and the machine cannot fire. If the system determined that there was no violation, it would proceed to the second step, in which it would determine whether attacking the target was *required* by operational

89. *See* Armin Krishnan, KILLER ROBOTS 46-53 (2009); see also Andreas Matthias, *The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata*, 6 ETHICS & INFORMATION TECHNOLOGY 175 (2004)
90. Kirshnan, *supra* note 89, at 46-53.
91. Matthias, *supra* note 89 at 175-83.
92. RONALD C. ARKIN, GOVERNING LETHAL BEHAVIOR IN AUTONOMOUS ROBOTS 69 (2009).
93. *Id.*

orders.[94]

The most obvious problem with this kind of artificial intelligence solution to the problem of moral agency is a technical one. Although scholars can imagine what such a solution might look like, there is no indication that anyone is close to realizing it. Indeed, the prospects for developing artificial moral judgment are much more remote than the prospects for developing sophisticated autonomous targeting and firing decisions.

More fundamentally, the problem of moral agency is not solved by giving autonomous weapon systems artificial moral judgment, even if such a capacity were technologically possible. The point of moral agency is that there must be a human being behind the decision to use force who can be accountable if the decision ends up being contrary to international law. Giving a machine the capacity to engage in moral reasoning does not accomplish that objective. Indeed, it only moves the responsible human being farther back in the process.

In the end, the ultimate requirement of international law is that a person be morally accountable for decisions about how to use weapons. The only way to assure this kind of accountability is to involve a human being in the decision. This does not mean that a human being must control every assessment and action made by an autonomous weapon system; but it does mean that a human being must be "in the loop" or "on the loop" in some way. Anything less would be inconsistent with the requirements of the international law of war.

VI. PROPOSALS FOR GOING FORWARD

Assuring the compliance of autonomous weapon systems with international law demands that there be a system of rules to assure active human participation in the operation of those systems. Those rules will vary with the technology employed in the system, especially the degree of autonomy of which the system is capable in its targeting and firing decisions. Thus, it may not be possible to establish a set of rules in advance that will assure that autonomous weapon systems will continue

94. *Id.* at 183-84.

III Indonesian Journal of International & Comparative Law 101-26 (January 2016)

to comply with the fundamental principles of international law. But the prospect of rapid technological advances means that international law cannot wait until new weapons are developed before making rules to control them. It is incumbent upon the international law community to develop some method for guiding the development of autonomous weapon systems and for establishing a framework for creating and modifying new rules expeditiously.

Participants in the debate about autonomous weapons have made different suggestions about what this kind of framework should look like. Some have suggested that it is impossible to develop any autonomous weapon system that will comply with the spirit of international law, and that, therefore, autonomous weapon systems should be banned. Others contend that these systems are like every other weapon system before them in the sense that they are ultimately controlled by human beings. On the basis of this contention, these advocates take the position that existing rules will only have to be modestly adapted to meet new the new challenges presented by new weapons. The following section of this paper sets forth some fundamental principles that should be incorporated into any regulatory framework within which autonomous weapon systems will be developed and used into the future.

As an initial matter, any regulatory framework will have to be developed on multiple jurisdictional levels. It would, of course, be helpful and necessary to establish standards, principles, and rules at an international level. But a single international treaty or convention would not be enough to assure compliance because the development of weapon systems is a highly decentralized process, in both a geographic and technological sense. In other words, new technologies for autonomous weapon systems can come from any part of the world or from any part of the scientific community, even those parts that are not directly concerned with weapons development. A single international agreement would not have influence on every geographic and scientific community where developments might occur. Consequently, it would be necessary to have a multilayered framework for controlling the development and use of autonomous weapon systems.

Even if a single international agreement would not be enough, in itself, the effective control of autonomous weapon systems must begin with the implementation of such an agreement. At the core of such an agreement would be the principle that autonomous weapon

systems are governed by the existing rules of warfare, both the laws of armed conflict and the principles of international human rights.[95] According to one scholar, this kind of international agreement would "include some form of interpretative application of the law of armed conflict to commanders deploying autonomous systems, explaining what information such commanders must have and what questions such commanders must ask before deciding to field the weapons in a given situation."[96] In addition, this agreement would "include rules and guidelines for the development of autonomous systems. Such rules and guidelines could be based not only on legal requirements, but also on policy considerations."[97] This latter category of rules would have to address questions such as whether and in what form autonomous weapon systems would permit human intervention. Should that intervention involve active control or direct supervision of the machine's decision-making? Or should it involve a kind of "kill switch" that would allow a human being to terminate the system's operation?[98]

This agreement need not be a finished product "A better approach is to reach consensus on some core minimum standards, but at the same time to retain some flexibility for international standards and requirements to evolve as technology evolves."[99] If an international governance standard is capable of evolution and adaptation to specific conditions, it is much more likely to invite compliance by a larger number of national governments.

An international agreement would not be enough, however. As an international consensus developed around an international agreement, and as that international agreement evolved, individual nations would have to develop their own legal frameworks. Like the international agreement, these national rules would have to include an interpretive application of the law of armed conflict and rules for the development of weapons systems.[100] These national rules would also be bound up with each nation's secret military technologies. "Undoubtedly many of

95. *See* Kenneth Anderson, et al., *Adapting the Law of Armed Conflict to Autonomous Weapons Systems*, 90 INT'L L. STUD. 386, 406 (2014).
96. *Id.* at 407.
97. *Id.*
98. *Id.* at 408.
99. *Id.* at 409.
100. Anderson, *supra* note 95, at 408-09.

the details of national rules and policies will need to remain secret, as they will involve sensitive matters of military capabilities and practices. States should be urged, however, to publish openly their general policies and to promote sharing of best practices."[101] The United States Department of Defense has already implemented a policy that would look very much like this kind of national rule, identifying limits and procedural requirements with regard to research, development and deployment of autonomous weapon systems.[102]

The final level of a legal framework would include regulatory processes that would apply to both formal military and private sector developers of relevant technologies, including those both expressly applicable to weapon systems and those that could be applicable.[103] Military developers "will need clear guidance as to what types of systems they should and should not be developing."[104] Private developers "need to start thinking about methodologies, operating procedures, rules of engagement and other operational and doctrinal level rules for the use of autonomous systems."[105] Although these "local" rules will be influenced from the top down by the existing international and national standards, they will also feed back into the evolution of those higher-order rules.[106] In particular, if there are going to be testing and evaluation systems for autonomous weapon systems that are effective on an international basis, there must be collaboration between private-sector weapons developers and military and government officials, "each drawing upon the other in the complicated dialogue involving international and national interests that generally characterizes the development of international law."[107]

In the final analysis:

> The fundamental principle underlying the gradual development of these standards and rules alongside the evolution of automation technologies, however, should be that what matters is ever greater compliance with the core obligations of the law

101. *Id.*
102. *See* UNITED STATES DEP'T OF DEFENSE, DIRECTIVE 3000.09, *supra* note 6.
103. Anderson, *supra* note 95, at 409-10.
104. *Id.* at 409.
105. *Id.*
106. *Id.*
107. Id.

> of armed conflict: necessity, distinction, proportionality and humanity. Whether the actor on the battlefield is a "who" or a "what" is not truly the issue, but rather how well that actor performs according to the law of armed conflict. Debate over standards or rules for automated or autonomous systems should remain scrupulously neutral as between human or machine, and should affirmatively reject any a priori preference for human over machine. . . .The principle of humanity is fundamental, but it refers, not to some idea that humans must operate weapons, but instead to the promotion of means or methods of warfare that best protect humanity within the lawful bounds of war, irrespective of whether the means to that end is human or machine or some combination of the two.[108]

VII. CONCLUSION

Law regulates people, not machines. If the principles of international law are to be preserved as autonomous weapon systems gain increasing use and develop ever-more sophisticated operations, then human beings will have to maintain the ultimate control over the decisions about how, when, and where those weapon systems are deployed. Because it is impractical to attempt to outlaw autonomous weapon systems, international law must develop a method by which it can be assured that human control is always present at a significant level in autonomous weapon systems. That method must include international and national standards that are constantly evolving along with technology and the changing nature of warfare. In the final analysis, human beings cannot create a set of killer robots and hope that those creations will, like the *Terminator*, eventually conform themselves to established moral principles. The only way to assure that autonomous weapon systems conform to law and morality is for human beings to direct them.

III Indonesian Journal of International & Comparative Law 101-26 (January 2016)

108. Anderson, *supra* note 95 at 411.