

Autonomous weapon systems and international humanitarian law: a reply to the critics

Article

Published Version

Schmitt, M. (2013) Autonomous weapon systems and international humanitarian law: a reply to the critics. Harvard National Security Journal, 4. pp. 1-37. Available at <http://centaur.reading.ac.uk/89864/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <https://harvardnsj.org/2013/02/autonomous-weapon-systems-and-international-humanitarian-law-a-reply-to-the-critics/>

Publisher: Harvard Law School

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

FEATURES

Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics

Michael N. Schmitt*

“A sword is never a killer, it is a tool in the killer’s hands.”
Seneca¹

Introduction

In November 2012, Human Rights Watch, in collaboration with the International Human Rights Clinic at Harvard Law School, released *Losing Humanity: The Case against Killer Robots*.² Human Rights Watch is among the most sophisticated of human rights organizations working in the field of international humanitarian law. Its reports are deservedly influential and have often helped shape application of the law during armed conflict. Although this author and the organization have occasionally crossed swords,³ we generally find common ground on key issues. This time, we have not.

* Chairman and Professor, International Law Department, United States Naval War College; Honorary Professor of International Humanitarian Law, Durham University Law School (UK). The views expressed in this Article are those of the author alone and should not be understood as necessarily representing those of the U.S. Department of Defense or any other government entity.

¹ Letters to Lucilius, 1st c., cited in WAR AND CONFLICT QUOTES 158 (Michael C. Thomsett & Jean F. Thomsett eds., 1997)

² HUMAN RIGHTS WATCH, LOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS, Nov. 2012, http://www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf [hereinafter LOSING HUMANITY].

³See, e.g., Michael N. Schmitt, *The Conduct of Hostilities During Operational Iraqi Freedom: An International Humanitarian Law Assessment*, 6 Y.B. INT’L HUMANITARIAN L. 73 (2003); Dinah Pokempner, Marc Garlasco & Bonnie Docherty, *Off Target on the Iraq Campaign: A Response to Professor Schmitt*, 6 Y.B. INT’L HUMANITARIAN L. 111 (2003).

“Robots” is a colloquial rendering for autonomous weapon systems. Human Rights Watch’s position on them is forceful and unambiguous: “[F]ully autonomous weapons would not only be unable to meet legal standards but would also undermine essential non-safeguards for civilians.”⁴ Therefore, they “should be banned and . . . governments should urgently pursue that end.”⁵ In fact, if the systems cannot meet the legal standards cited by Human Rights Watch, then they are already unlawful *as such* under customary international law irrespective of any policy or treaty law ban on them.⁶

Unfortunately, *Losing Humanity* obfuscates the on-going legal debate over autonomous weapon systems. A principal flaw in the analysis is a blurring of the distinction between international humanitarian law’s prohibitions on weapons *per se* and those on the unlawful use of otherwise lawful weapons.⁷ Only the former render a weapon illegal as such. To

⁴ LOSING HUMANITY, *supra* note 2, at 1–2. On the issue of the legality of autonomous weapon systems, see Jeffrey S. Thurnher, *No One at the Controls: Legal Implications of Fully Autonomous Targeting* 77, JOINT FORCE Q., Oct. 2012; Kenneth Anderson & Matthew Waxman, *Law and Ethics for Robot Soldiers*, POL’Y REV. 35, Dec. 2012; Markus Wagner, *Taking Humans Out of the Loop: Implications for International Humanitarian Law*, 21 J. L. INFO. & SCI. 155, 155 (2011).

⁵ LOSING HUMANITY, *supra* note 2, at 2.

⁶ The author referred to the following non-binding compilations of customary international humanitarian law rules to draw conclusions as to customary status of norms referenced in this article: INTERNATIONAL COMMITTEE OF THE RED CROSS [ICRC], CUSTOMARY INTERNATIONAL HUMANITARIAN LAW (Jean-Marie Henckaerts & Louise Doswald-Beck eds., 2005) [hereinafter CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY]; MICHAEL N. SCHMITT, CHARLES H.B. GARRAWAY & YORAM DINSTEIN, THE MANUAL ON THE LAW OF NON-INTERNATIONAL ARMED CONFLICT WITH COMMENTARY (2006); HARVARD PROGRAM ON HUMANITARIAN POLICY AND CONFLICT RESEARCH, MANUAL ON INTERNATIONAL LAW APPLICABLE TO AIR AND MISSILE WARFARE, WITH COMMENTARY (2010); MICHAEL N. SCHMITT, TALLINN MANUAL ON THE INTERNATIONAL LAW APPLICABLE TO CYBER WARFARE (forthcoming 2012) [hereinafter TALLINN MANUAL]. Although not compilations of customary international law, scholars and practitioners also often look to the Rome Statute and the U.S. Commander’s Handbook as strong indications of a norm’s customary status, the latter in light of the United States’ Additional Protocol I non-Party status. Statute of the International Criminal Court, July 17, 1998, 2187 U.N.T.S. 90 [hereinafter Rome Statute]; U.S. NAVY/U.S. MARINE CORPS/U.S. COAST GUARD, THE COMMANDER’S HANDBOOK ON THE LAW OF NAVAL OPERATIONS, NWP 1-14M/MCWP 5-12.1/COMDT PUBP5800.7A (2007) [hereinafter COMMANDER’S HANDBOOK].

⁷ As Yoram Dinstein notes, “The fact that certain weapons are used indiscriminately in a particular military engagement does not stain the weapons themselves with an indelible

illustrate, a rifle is lawful, but may be used unlawfully, as in shooting a civilian. By contrast, under customary international law, biological weapons are unlawful *per se*; this is so even if they are used against lawful targets, such as the enemy's armed forces. The practice of inappropriately conflating these two different strands of international humanitarian law has plagued debates over other weapon systems, most notably unmanned combat aerial systems such as the armed Predator. In addition, some of the report's legal analysis fails to take account of likely developments in autonomous weapon systems technology or is based on unfounded assumptions as to the nature of the systems. Simply put, much of *Losing Humanity* is either counter-factual or counter-normative.

This Article is designed to infuse granularity and precision into the legal debates surrounding such weapon systems and their use in the future "battlespace." It suggests that whereas some conceivable autonomous weapon systems might be prohibited as a matter of law, the use of others will be unlawful only when employed in a manner that runs contrary to international humanitarian law's prescriptive norms. This Article concludes that *Losing Humanity's* recommendation to ban the systems is insupportable as a matter of law, policy, and operational good sense. Human Rights Watch's analysis sells international humanitarian law short by failing to appreciate how the law tackles the very issues about which the organization expresses concern. Perhaps the most glaring weakness in the recommendation is the extent to which it is premature. No such weapons have even left the drawing board. To ban autonomous weapon systems altogether based on speculation as to their future form is to forfeit any potential uses of them that might minimize harm to civilians and civilian objects when compared to other systems in military arsenals.

I. Autonomous Weapon Systems

Before turning to the law, it is necessary to frame the issue definitionally. A weapon system consists of a weapon and the items associated with its employment.⁸ An example is the Predator unmanned aerial combat system armed with a weapon such as a Hellfire missile. The

imprint of illegality, since in other operations the same weapons may be employed within the framework of [international humanitarian law]." YORAM DINSTEIN, *THE CONDUCT OF HOSTILITIES UNDER THE LAW OF INTERNATIONAL ARMED CONFLICT* 62 (2d ed., 2010).

⁸ In this Article, the term "weapon system" is used as shorthand to refer to both a weapon and the complete weapon system.

Department of Defense has recently defined an *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.⁹

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.

U.S. forces have operated two human-supervised autonomous systems for many years—the Aegis at sea and the Patriot on land—both designed to defend against short notice missile attacks.¹⁰ Another human-supervised autonomous weapon system, Israel’s Iron Dome, is presently receiving a great deal of attention as it very effectively destroys incoming rockets.¹¹ However, the United States is currently not fielding any fully

⁹ DEP’T OF DEF., DIRECTIVE 3000.09, AUTONOMY IN WEAPON SYSTEMS 13–14 (Nov. 2, 2012) [hereinafter DoD DIRECTIVE 3000.09]. Human Rights Watch distinguishes three categories of systems. A “human in the loop” system requires a human to direct the system to select a target and attack it. The Department of Defense labels these “semi-autonomous systems.” A “human on the loop” weapon is one in which the system selects targets and attacks them, albeit with human operator oversight. The Department of Defense term is “human-supervised autonomous system.” Finally, Human Rights Watch calls a system that can attack without any human interface a “human out of the loop weapon.” The Department of Defense moniker is “fully autonomous weapon system.” LOSING HUMANITY, *supra* note 2, at 2.

¹⁰ Federation of American Scientists (FAS), *Patriot TMD*, <http://www.fas.org/spp/starwars/program/patriot.htm> (last visited Feb. 3, 2012); Lockheed Martin, *Aegis Combat System*, <http://www.lockheedmartin.com/us/products/aegis.html> (last visited Feb. 3, 2012).

¹¹ Iron Dome can operate automatically using programmed parameters, but the system also allows for human operator intervention. Inbal Orpaz, *How does Iron Dome Operate?*, HAARETZ, (Nov. 19, 2012), <http://www.haaretz.com/news/features/how-does-the-iron-dome-work.premium-1.478988>; RAFAEL, IRON DOME, *available at*

autonomous weapon systems.¹² Nor are there any “plans to develop lethal autonomous weapon systems other than human-supervised systems for the purposes of local defense of manned vehicles or installations.”¹³ That said, Human Rights Watch is correct in noting that this fact does “not preclude a change in that policy as the capacity for autonomy evolves.”¹⁴ At some point in the future, such systems will find their way into the battlespace.

Fully autonomous weapon systems must be distinguished from those that are semi-autonomous, which are commonplace in contemporary warfare.¹⁵ The latter engage specific targets or categories of targets that a human operator selects. For instance, a “fire and forget” missile on an aircraft locks onto a target identified by the pilot and then attacks it without human involvement. *Losing Humanity* does not address semi-autonomous systems—appropriately so since the systems differ qualitatively from those that are fully autonomous. Nor does the report examine what it calls “automatic weapons defense systems.”¹⁶ These systems respond automatically (or near automatically) when they detect incoming threats. An example is the “close-in weapon system” (CIWS or “Sea Whiz”).¹⁷ Used for point-defense of warships against incoming missiles, the Sea Whiz can be programmed to detect inbound missiles based on parameters that include speed and altitude, and can automatically engage them.

Before turning to the legal issues surrounding autonomous weapon systems, it is necessary to debunk a number of myths about the systems that

http://www.rafael.co.il/marketing/SIP_STORAGE/FILES/6/946.pdf (last visited Feb. 3, 2012).

¹² DEP’T OF DEF., DOD DIRECTIVE 3000.09: AUTONOMOUS WEAPON SYSTEMS: RESPONSE-TO-QUERY TALKING POINTS 1 (date unknown) (on file with author). The United States fields autonomous weapon systems that use nonlethal and non-kinetic force. An example is the Miniature Air Launched Decoy Jammer (MALD-J), which is launched from an aircraft and flies a preprogrammed mission while jamming enemy radar and serving as a decoy. *Id.* at 2.

¹³ *Id.* at 3.

¹⁴ LOSING HUMANITY, *supra* note 2, at 8.

¹⁵ The weapons are also known as “launch and leave” weapons. Examples include the AGM-130 and AGM-65 missiles used for attacking ground targets. See descriptions of these and other such systems at U.S. AIR FORCE, *Factsheets (Weapons)*, <http://www.af.mil/information/factsheets/index.asp> (last visited Feb. 3, 2012).

¹⁶ LOSING HUMANITY, *supra* note 2, at 9.

¹⁷ U.S. Navy, *MK-15 Phalanx Close-In Weapons System (CIWS)*, http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=487&ct=2 (last visited Feb. 3, 2012).

are clouding public debate. First, the idea of “robot wars” is pure science fiction. As noted by a Department of Defense Task Force, “the true value of these systems is not to provide a direct human replacement, but rather to extend and complement human capability by providing potentially unlimited persistent capabilities, reducing human exposure to life-threatening tasks, and, with proper design, reducing the high cognitive load currently placed on operators/supervisors.”¹⁸ Autonomous weapon systems will be integrated into human warfare, but are highly unlikely to replace it.

Second, neither the United States nor any other country is contemplating the development of any systems that would simply hunt down and kill or destroy enemy personnel and objects without restrictive engagement parameters, such as area of operation or nature of the target. Moreover, the Defense Science Board points out that:

[A]ll autonomous systems are *supervised* by human operators at some level, and autonomous systems’ software embodies the designed limits on the actions and decisions delegated to the computer. Instead of viewing autonomy as an intrinsic property of an unmanned vehicle in isolation, the design and operation of autonomous systems needs to be considered in terms of *human–system collaboration*.¹⁹

As presently envisaged, autonomous weapon systems will only attack targets meeting predetermined criteria and will function within an area of operations set by human operators.

The U.S. Department of Defense is exceptionally sensitive to the human interface issue. It has recently promulgated policy guidance that requires the Secretaries of the military departments, the Commander of U.S. Special Operations Command, and certain other high level officials to:

[d]esign human-machine interfaces for autonomous and semi-autonomous weapon systems to be readily

¹⁸ Undersecretary of Defense for Acquisition, Technology, and Logistics, *Memorandum in DEP’T OF DEF., DEFENSE SCIENCE BOARD, THE ROLE OF AUTONOMY IN DOD SYSTEMS* (July 2012), <http://www.acq.osd.mil/dsb/reports/AutonomyReport.pdf>.

¹⁹ *Id.* at 1–2. DoD DIRECTIVE 3000.09, *supra* note 9, ¶ 4a, similarly provides that “[a]utonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force.”

understandable to trained operators, provide traceable feedback on system status, and provide clear procedures for trained operators to activate and deactivate system functions . . . ; [c]ertify that operators of autonomous and semi-autonomous weapon systems have been trained in system capabilities, doctrine, and [tactics, techniques, and procedures] in order to exercise appropriate levels of human judgment in the use of force and employ systems with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable ROE . . . ; [and e]stablish and periodically review training, and [tactics, techniques, and procedures], and doctrine for autonomous and semi-autonomous weapon systems to ensure operators and commanders understand the functioning, capabilities, and limitations of a system's autonomy in realistic operational conditions, including as a result of possible adversary actions.²⁰

Finally, robots will not “go rogue.” While autonomous and semi-autonomous weapon systems will be susceptible to malfunction, that is also the case with weapon systems ranging from catapults to computer attack systems. Like a missile that “goes ballistic” (loses guidance), future autonomous systems could fall out of parameters. But the prospect of them “taking on a life of their own” is an invention of Hollywood.

The one real risk is tampering by the enemy or non-State actors such as hackers. As an example, the enemy might be able to use cyber means to take control of an autonomous weapon system and direct it against friendly forces or a civilian population. Those developing the systems are acutely aware of such risk. U.S. policy on the matter is that, “[c]onsistent with the potential consequences of an unintended engagement or loss of control of the system to unauthorized parties, physical hardware and software will be designed with appropriate: (a) [s]afeties, anti-tamper mechanisms, and information assurance [and] (b) [h]uman machine interface.”²¹

²⁰ *Id.*, encl. 4, ¶ 8.

²¹ *Id.*, ¶ 4a(2).

II. Unlawful Weapon Systems

Losing Humanity concludes that “[a]n initial evaluation of fully autonomous weapons shows . . . such robots would appear to be incapable of abiding by key principles of international humanitarian law. They would be unable to follow the rules of distinction, proportionality, and military necessity and might contravene the Martens Clause.”²² While it is true that some autonomous weapon systems might violate international humanitarian law norms, it is categorically not the case that all such systems will do so. Instead, and as with most other weapon systems, their lawfulness as such, as well as the lawfulness of their use, must be judged on a case-by-case basis. To assert, as Human Rights Watch does, that “[f]ull autonomy would strip civilians of protections from the effects of war that are guaranteed under the law” is to melodramatically oversimplify international humanitarian law, while failing to accord it the respect it is due.

The international humanitarian law governing weaponry proceeds along two tracks. One focuses on the legality of the weapon system itself. The legal reviews discussed below respond to this issue. A separate and distinct family of prohibitions (labelled “conduct of hostilities” rules) applies to a weapon system’s use irrespective of whether or not the weapon system is lawful *per se*.

Among the earliest prohibitions with respect to the legality of weapons *per se* is the ban on means or methods of warfare that are of a nature to cause superfluous injury or unnecessary suffering.²³ It first appeared in the regulations annexed to the 1899 Hague Convention II and its 1907 counterpart.²⁴ Article 35(2) of Additional Protocol I to the 1949 Geneva Conventions affirms the prohibition and irrefutably reflects

²² LOSING HUMANITY, *supra* note 2, at 30.

²³ See generally DINSTEIN, *supra* note 7, at 63–67; WILLIAM H. BOOTHBY, WEAPONS AND THE LAW OF ARMED CONFLICT, ch. 5 (2009).

²⁴ Convention (II) with Respect to the Laws and Customs of War on Land, pmbl., July 29, 1899, 32 Stat. 1803 [hereinafter 1899 Hague II]; Convention (IV) Respecting the Laws and Customs of War on Land, pmbl., Oct. 18, 1907, 36 Stat. 2277 [hereinafter 1907 Hague IV]. The 1868 St. Petersburg Declaration presaged the prohibition with its condemnation of “the employment of arms which uselessly aggravate the sufferings of disabled men, or render their death inevitable.” Declaration Renouncing the Use, in Time of War, of Explosive Projectiles Under 400 Grammes Weight, pmbl., Nov. 29/Dec. 11, 1868, 18 Martens Nouveau Recueil (ser. 1) 474 [hereinafter St. Petersburg Declaration].

customary international law.²⁵ Therefore, the norm binds even States that are not Party to the Protocol, such as the United States. Substantively, it outlaws those means and methods of warfare that unnecessarily aggravate the suffering of combatants, that is, which cause suffering serving no military purpose. In that it is otherwise unlawful to attack civilians, this norm applies only to suffering or injury experienced by combatants.²⁶

Losing Humanity does not mention the prohibition. Perhaps this is because autonomy is unlikely to present unnecessary suffering and superfluous injury issues since the rule addresses a weapon system's effect on the targeted individual, not the manner of engagement (autonomous). Nevertheless, an autonomous system could be used as a platform for a weapon that would violate the prohibition, such as a bomb containing fragments that are designed to be difficult to locate during the treatment of wounded combatants.²⁷ The combination of the platform and the weapon would render the autonomous weapon system unlawful *per se*. But this *possibility* is not a valid basis for imposing an across-the-board preemptive ban on the systems.

²⁵ Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, art. 35(2), June 8, 1977, 1125 U.N.T.S. 3 [hereinafter Additional Protocol I]; CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rule 70; Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. 226 (July 8), ¶ 78 [hereinafter Nuclear Weapons]. *See also* Rome Statute, *supra* note 6, art. 8(2)(b)(xx); COMMANDER'S HANDBOOK, *supra* note 6, ¶ 9.1.1. On Article 35(2), *see also* ICRC, COMMENTARY ON THE ADDITIONAL PROTOCOLS OF 8 JUNE 1977 TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949 (Yves Sandoz et al. eds., 1987), ¶¶ 1410–1439 [hereinafter ICRC COMMENTARY]; MICHAEL BOTHE, KARL JOSEF PARTSCH & WALDEMAR A. SOLF, NEW RULES FOR VICTIMS OF ARMED CONFLICTS: COMMENTARY ON THE TWO 1977 PROTOCOLS ADDITIONAL TO THE GENEVA CONVENTIONS OF 1949 at 195–198 (1982).

²⁶ For the purposes of this article, the term combatants includes civilians who are directly participating in the hostilities since they are subject to attack for such time as they so participate. Additional Protocol I, *supra* note 25, art. 51(3). On the subject of targetability and direct participation, *see* ICRC, INTERPRETIVE GUIDANCE ON THE NOTION OF DIRECT PARTICIPATION IN HOSTILITIES UNDER INTERNATIONAL HUMANITARIAN LAW (Nils Melzer ed., 2009); Michael N. Schmitt, *The Interpretive Guidance on the Notion of Direct Participation in Hostilities: A Critical Analysis*, 1 HARV. NAT'L SEC. J. 5 (2010).

²⁷ Protocol (to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects) on Non-detectable Fragments, Oct. 10, 1980, 1342 U.N.T.S. 168; CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rule 79.

International humanitarian law also prohibits weapon systems that cannot be directed at a specific military objective.²⁸ These weapons are unlawful *per se* in that they are of a nature to strike combatants, military objectives, civilians, and civilian objects without distinction. The principle of distinction is a norm of customary international law,²⁹ and the companion treaty prohibition appears in Article 51(4)(b) of Additional Protocol I.³⁰

The prohibition on weapon systems that are indiscriminate because they cannot be aimed at a lawful target is often confused with the ban on use of discriminate weapons in an indiscriminate fashion. The classic case is that of the SCUD missiles launched by Iraq during the 1990-91 Gulf War. While it is true that the missiles were inaccurate, they were not unlawful *per se* because situations existed in which they could be employed discriminately. In particular, the missiles were capable of use against troops in open areas such as the desert, and they actually struck very large military installations without seriously endangering the civilian population.³¹ However, when launched in the direction of cities, as repeatedly occurred during the conflict, their *use* was undeniably unlawful. Even though the cities contained military objectives, the missiles were insufficiently accurate to reliably strike any of them.

The SCUD example has particular resonance with respect to autonomous weapon systems. A misperception exists that without a “man in the loop” there is a high risk of misidentifying civilians or civilian objects as lawful military objectives that either may be directly attacked or that do not have to be considered when assessing whether an attack on a military objective will comply with the rule of proportionality (discussed below). In fact, Human Rights Watch categorically claims that “[f]ully autonomous weapons would not have the ability to sense or interpret the difference

²⁸ See generally Boothby, *supra* note 23, ch. 6.

²⁹ CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rules 12 & 7. See also Rome Statute, *supra* note 6, art. 8(2)(b)(xx); COMMANDER’S HANDBOOK, *supra* note 6, ¶¶ 5.3.2 & 9.1.2. Application of the prohibition evolves over time in relation to the development of increasingly accurate weapon systems. To illustrate, many of the gravity bombs designed for release from high altitudes that were dropped during World War II would today be characterized as indiscriminate.

³⁰ ICRC COMMENTARY, *supra* note 25, ¶¶ 1956–1960.

³¹ DEP’T OF DEF., CONDUCT OF THE PERSIAN GULF WAR: FINAL REPORT TO CONGRESS 166–168, 621–622 (1992); WILLIAM ROSENAU, SPECIAL OPERATIONS FORCES AND ELUSIVE ENEMY GROUND TARGETS, ch. 3 (2001) (entitled “Coalition SCUD- Hunting in Iraq 1991”).

between soldiers and civilians, especially in contemporary combat environments.”³²

What Human Rights Watch appears to have missed is that even an autonomous weapon system that is completely incapable of distinguishing a civilian from a combatant or a military objective from a civilian object can be lawful *per se*. Not every battlespace contains civilians or civilian objects. When they do not, a system devoid of any capacity to distinguish protected persons and objects from lawful military targets can be used without endangering the former. Typical examples would include the employment of such systems for an attack on a tank formation in a remote area of the desert or from warships in areas of the high seas far from maritime navigation routes. The inability of the weapon systems to distinguish bears on the legality of their use in particular circumstances (such as along a roadway on which military and civilian traffic travels), but not their lawfulness *per se*.

Human Rights Watch’s apprehension is also counterfactual. Military technology has advanced well beyond simply being able to spot an individual or object. Modern sensors can, *inter alia*, assess the shape and size of objects, determine their speed, identify the type of propulsion being used, determine the material of which they are made, listen to the object and its environs, and intercept associated communications or other electronic emissions. They can also gather additional data on other objects or individuals in the area and, depending on the platform with which they are affiliated, monitor a potential target for extended periods in order to gather information that will enhance the reliability of identification and permit target engagement when the target is relatively isolated. Even software for autonomous weapon systems that enables visual identification of individuals, thereby enabling precision during autonomous “personality strikes” against specified persons, is likely to be developed. These and related technological capabilities auger against characterization of autonomous weapon systems as unlawful *per se* solely based on their autonomous nature.³³

³² LOSING HUMANITY, *supra* note 2, at 30.

³³ In particular, they reveal Human Rights Watch’s concerns that combatants sometimes fail to “wear uniforms or insignia” or are identifiable only through their “direct participation in hostilities” as exaggerated and simplistic (at least with regard to the issue of distinction by the systems). *Id.*

The organization actually seems to ask more of autonomous weapon systems than of human-operated systems. For example, Human Rights Watch points to the possibility that an autonomous weapon system might be deceived through the concealment of weapons or exploitation of the system's sensor limitations.³⁴ Yet, asymmetrically disadvantaged enemies have been feigning civilian or other protected status to avoid being engaged by human-operated weapon systems for decades (even centuries).³⁵ The fact that the techniques sometimes prove successful has never merited classifying those systems as indiscriminate *per se*. In fact, it would be counter-productive to take such an approach because it would incentivize the enemy's use of the tactic in order to keep weapon systems off the battlefield.

Human Rights Watch also observes that “fully autonomous weapons would not possess human qualities necessary to assess an individual's intentions.”³⁶ While it is true that human perception of human activity can enhance identification in some circumstances, human-operated systems already frequently engage targets without the benefit of the emotional sensitivity cited by the organization. For example, human-operated “beyond visual range” attacks are commonplace in modern warfare; no serious charge has been levelled that the weapon systems that conduct them are unlawful *per se*.³⁷

In fact, human judgment can prove less reliable than technical indicators in the heat of battle. For instance, during the 1994 friendly fire shootdown of two U.S. Army Blackhawks in the no-fly zone over northern Iraq, the U.S. Air Force F-15's involved made a close visual pass of the targets before engaging them.³⁸ Pilot error (and human error aboard the AWACs monitoring the situation) contributed to their misidentification as

³⁴ *Id.* at 31.

³⁵ See generally Michael N. Schmitt, *Asymmetrical Warfare and International Humanitarian Law*, 62 AIR FORCE L. REV. 1, 15–16 (2008). Human Rights Watch might also have cited the use of human shields, a practice that would also complicate autonomous weapon system targeting. See generally Michael N. Schmitt, *Human Shields and International Humanitarian Law*, 47 COLUM. J. TRANSNAT'L L. 292 (2009).

³⁶ LOSING HUMANITY, *supra* note 2, at 31.

³⁷ As an example, the U.S. Navy's AIM-54 air-to-air Phoenix missile has a range in excess of 100 nautical miles. U.S. Navy, *AIM-54 Phoenix Missile*, http://www.navy.mil/navydata/fact_display.asp?cid=2200&tid=700&ct=2 (last visited Feb. 3, 2012).

³⁸ Aircraft Accident Investigation Board Report, *U.S. Army UH-60 Blackhawk Helicopters 87-26000 and 88-26060, vol. 1 (Executive Summary)* 3 (May 27, 1994), available at http://www.dod.mil/pubs/foi/operation_and_plans/PersianGulfWar/973-1.pdf.

Iraqi military helicopters. Similarly, in 1988 the USS Vincennes engaged an Iranian airliner that it mistakenly believed was conducting an attack on the ship. The warship's computers accurately indicated that the aircraft was ascending. Nevertheless, human error led the crew to believe it was descending in an attack profile and, in order to defend the ship, they shot down the aircraft.³⁹ Such tragedies demonstrate that a man in the loop is not a panacea during situations in which it may be difficult to distinguish civilians and civilian objects from combatants and military objectives. Those who believe otherwise have not experienced the fog of war.

Losing Humanity also highlights the unique emotional character of human beings. It suggests “robots would not be restrained by human emotions and the capacity for compassion, which can provide an important check on the killing of civilians.” The report concludes that “[e]motionless robots could, therefore, serve as tools of repressive dictators seeking to crack down on their own people without fear their troops would turn on them.”⁴⁰ While it is certainly correct that emotions can restrain humans, it is equally true that emotions can unleash the basest of instincts. From Rwanda and the Balkans to Darfur and Afghanistan, history is replete with tragic examples of unchecked emotions leading to horrendous suffering. Reliance on this factor by Human Rights Watch is empirically suspect.

An autonomous weapon system only violates the prohibition against weapons incapable of being directed at a lawful target if there are no circumstances, given its intended use, in which it can be used discriminately. Consider an autonomous anti-personnel weapon system designed for employment in urban areas. Because it is contemplated for use where civilians and combatants are regularly co-located, the system must have sufficient sensor and artificial intelligence capability to distinguish them; otherwise, it qualifies as indiscriminate *per se*. An autonomous weapon system unable to reliably distinguish between civilians and combatants but planned for use where civilians are not present would still have to be capable of geographical restriction (either based on system constraints such

³⁹ Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988, Aug. 19, 1988, at 37, 42–45 *available at* <http://homepage.ntlworld.com/jksonc/docs/ir655-dod-report.html>. The report concluded that “[s]tress, task fixation, and unconscious distortion of data may have played a major role in this incident.” *Id.* at 45. It also noted that “scenario fulfillment,” that is, the distortion of “dataflow in an unconscious attempt to make available evidence fit a preconceived scenario.” *Id.*

⁴⁰ LOSING HUMANITY, *supra* note 2, at 4.

as maximum range or on human operator pre-programming). This would be needed to prevent it from passing into areas where civilians are located. Systems can be limited temporally to achieve the same end.⁴¹

A second form of prohibition on indiscriminate weapons is codified in Article 51(4)(c) of Additional Protocol I, and reflects customary international law.⁴² It disallows weapon systems that, despite being able to strike their targets accurately, have uncontrollable effects. The paradigmatic example is a biological contagion used to infect combatants, the subsequent spread of which is uncontrollable. A weapon with such effects could be mounted on an autonomous platform. Another example is an autonomous weapon system that searches for and conducts cyber attacks against dual-use infrastructure (cyber infrastructure used by both the military and civilians). The malware used to conduct the attacks could be indiscriminate if designed in a way that makes it likely to spread into the civilian network.⁴³

III. Unlawful Use of Lawful Weapon Systems

As should be apparent, the likelihood of an autonomous weapon system being unlawful *per se* is very low. This being so, the question becomes whether international humanitarian law provides sufficient safeguards with respect to the use of these weapon systems. Sadly, the unlawful use of lawful weapons is far from rare during armed conflicts.

The seminal principle of international humanitarian law is distinction. It is one of two principles in this body of law recognized as “cardinal” by the International Court of Justice, which has also characterized it as “intransgressible.”⁴⁴ The principle of distinction serves as the fount for the international humanitarian law rules, including those regarding the use of weapon systems, that seek to safeguard civilians, civilian objects, and other protected persons and places during the conduct of

⁴¹ DoD DIRECTIVE 3000.09, *supra* note 9, ¶ 4a(1)(b), requires measures to be taken to ensure that autonomous weapon systems “[c]omplete engagements in a timeframe consistent with commander and operator intentions and, if unable to do so, terminate engagements or seek additional human operator input before continuing the engagement.”

⁴² CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rules 12 & 71. *See also* Rome Statute, *supra* note 6, art. 8(2)(b)(xx); COMMANDER’S HANDBOOK, *supra* note 6, ¶ 5.3.2.

⁴³ Injury or physical damage would have to result in the case of the cyber attack. TALLINN MANUAL, *supra* note 6, rule 43 and accompanying commentary.

⁴⁴ Nuclear Weapons, *supra* note 25, ¶¶ 78–79,

hostilities. Article 48 of Additional Protocol I codifies this customary law principle⁴⁵: “In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives.”

Distinction is operationalized in a number of rules, the two most fundamental being the customary law prohibitions on making civilians and civilian objects the object of attack.⁴⁶ They are codified in Articles 51(2) and 52(1) of Additional Protocol I respectively.⁴⁷ Self-evidently, it would be unlawful to use an autonomous weapon system to directly attack civilians or civilian objects. In this regard, note that the same issues that present themselves with regard to other weapons systems also appear in the case of autonomous weapon systems. For instance, the exception to the prohibition on attacking civilians that exists for those who directly participate in hostilities also applies to the use of autonomous weapon systems against them.⁴⁸ Similarly, the universally accepted definition of military objectives found in Article 52(2) of Additional Protocol I pertains equally to attacks by autonomous weapon systems on objects, as does the controversy over whether war-sustaining objects qualify as military objectives.⁴⁹

⁴⁵ CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rule 1.

⁴⁶ As to civilians, see *id.*, rules 1 & 6; Rome Statute, *supra* note 6, art. 8(2)(b)(i); COMMANDER’S HANDBOOK, *supra* note 6, ¶ 8.3. As to civilian objects, see CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rules 7, 9 & 10; Rome Statute, *supra* note 6, art. 8(2)(b)(ii); COMMANDER’S HANDBOOK, *supra* note 6, ¶ 8.3.

⁴⁷ On Article 51(2), see ICRC COMMENTARY, *supra* note 25, ¶¶ 1938–1941; Bothe, *supra* note 25, at 300. On Article 52(1), see ICRC COMMENTARY, *supra* note 25, ¶¶ 2011–2013; Bothe, *supra* note 25, at 322. Note that Article 51(2), in a provision that reflects customary law, prohibits “acts or threats of violence the primary purpose of which is to spread terror among the civilian population.” CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rule 2. This prohibition would apply equally to attacks with autonomous weapon systems designed to spread terror.

⁴⁸ Additional Protocol I, *supra* note 25, art. 51(3); CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rule 6.

⁴⁹ Although the United States accepts the definition of military objectives set forth in Additional Protocol I, Article 52(2), as an accurate articulation of the customary law norm, its explanation of the concept in the Commander’s Handbook extends the definition to objects that sustain the war effort, such as oil exports. COMMANDER’S HANDBOOK, *supra* note 6, ¶ 8.2. For a discussion of the controversy, see TALLINN MANUAL, *supra* note 6, commentary accompanying rule 38; Michael N. Schmitt, *Targeting in Operational Law*, in THE HANDBOOK OF THE INTERNATIONAL LAW OF MILITARY OPERATIONS 245, 254 (Terry Gill & Dieter Fleck eds., 2010).

Article 51(4)(a), which reflects customary international law, sets forth a further prohibition on indiscriminate attacks by banning those that are not directed at an unlawful target, and, as a result, are of a nature to strike lawful targets and civilians or civilian objects without distinction.⁵⁰ This prohibition differs from that on indiscriminate weapons in that in this case the weapon is capable of being aimed at a lawful target, but the attacker does not do so.⁵¹

Doubt—the lack of certainty that a person is a lawful target—is of particular importance in respect to this prohibition. During an attack, doubt as to status must be resolved in favor of treating the individual in question as immune from attack. Article 50(1) of Additional Protocol I codifies this presumption, which is generally characterized as customary in nature.⁵²

The mere existence of some doubt does not bring the presumption into operation.⁵³ Rather, the degree of doubt that bars attack is that which would cause a reasonable attacker in the same or similar circumstances to hesitate before attacking. Restated, attackers must act responsibly as a matter of law when conducting military operations. They must consider “the information from all sources which is reasonably available to them at the relevant time,”⁵⁴ as well as factors like force protection, the military value of the target, and the likelihood that subsequent opportunities to conduct an attack will present themselves.

The fact that the doubt threshold is framed in terms of human reasonableness complicates translation into the autonomy context. Obviously, development of an algorithm that can both precisely meter

⁵⁰ CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rules 11–12; see also COMMANDER’S HANDBOOK, *supra* note 6, ¶ 5.3.2. On Article 51(4)(a), see ICRC COMMENTARY, *supra* note 25, ¶¶ 1951–1955.

⁵¹ As noted by Yoram Dinstein, “The key to finding that a certain attack has been indiscriminate is the nonchalant state of mind of the attacker.” Dinstein, *supra* note 7, at 127.

⁵² CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, commentary accompanying rule 6. On Article 50(1), see ICRC COMMENTARY, *supra* note 25, ¶¶ 1911–1921; Bothe, *supra* note 25, at 295–296.

⁵³ TALLINN MANUAL, *supra* note 6, commentary accompanying rule 33.

⁵⁴ U.K. Statement made upon Ratification of Additional Protocols I and II, ¶ (h), *reprinted in* DOCUMENTS ON THE LAWS OF WAR 511 (Adam Roberts & Richard Guelff eds., 3d ed. 2000) [hereinafter U.K. Ratification Statement]; U.K. MINISTRY OF DEFENCE, THE JOINT SERVICE MANUAL OF THE LAW OF ARMED CONFLICT, JSP 383 (2004), ¶ 5.4.3.

doubt and reliably factor in the unique situation in which the autonomous weapon system is being operated will prove highly challenging. After all, artificial intelligence is artificial.

Detailed discussion of the technical mechanisms for determining doubt is beyond the scope of this article. Nevertheless, algorithms attributing values to sensor data, thereby enabling the autonomous weapon system to compute doubt (or, since it is a machine, the likelihood of being a lawful target), are theoretically achievable. For instance, autonomous weapon systems could be equipped with sensors that enable them to determine when a potential target is a child. Such a determination would substantially decrease the probability that the target is a combatant. On the other hand, if the sensors ascertain that a potential target is carrying a weapon or engaging in hostilities (for instance, by launching a missile or firing a weapon), the likelihood of the target being a combatant increases. These are overly simplistic examples offered for the sake of illustration; the actual sensor capabilities of autonomous weapon systems will be much more advanced.

Since such determinations are highly contextual, it will prove more problematic to decide upon the doubt threshold at which an autonomous weapon system will be programmed to refrain from attack. For instance, more doubt can be countenanced on a “hot” battlefield than in a relatively benign environment. The key is human interaction with the system. In theory, human operators could program these and other factors into an autonomous weapon system. Should they set unreasonably high thresholds of doubt (that is, the point where the systems will not attack), the system would violate the prohibition on indiscriminate attacks.

No equivalent presumption exists in the *lex scripta* for objects. Still, clearly an attack based on an unreasonable conclusion that an object is a military objective violates international humanitarian law. The mode of analysis described above for persons would apply equally to objects. In addition, Article 52(3) of Additional Protocol I sets forth a separate rule with regard to objects normally dedicated to civilian purposes: “In case of doubt whether [such] an object . . . is being used to make an effective contribution to military action, it shall be presumed not to be so used.”⁵⁵ Some difference

⁵⁵ Additional Protocol I, *supra* note 25, art. 52(3). On Article 52(3), see ICRC COMMENTARY, *supra* note 25, ¶¶ 2029–2037; Bothe, *supra* note 25, at 326–327.

of opinion exists over whether this presumption reflects customary law.⁵⁶ In light of that disagreement, the international group of experts that drafted the *Tallinn Manual* took the admittedly tautological position that in case of doubt such an object may only be attacked “following a careful assessment.”⁵⁷

Ultimately, there is little difference between application of the rules governing attacks on individuals and objects. An autonomous weapon system that lacks any capability to distinguish lawful from unlawful targets may not be used where the two are co-located. In addition, human operators must be able to program acceptable levels of doubt, based on the circumstances in which they will be used, into the systems. Failure to comply with these requirements would constitute an indiscriminate attack.

An attack that is directed against a lawful target must nevertheless comply with the rule of proportionality. The customary international humanitarian law rule of proportionality, codified in Articles 51(5)(b) and 57(2)(iii) of Additional Protocol I, prohibits “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 in relation to the concrete and direct military advantage anticipated.”⁵⁸ This rule is among the most complex and misunderstood in international humanitarian law with respect to both interpretation and application.⁵⁹

Selection of targets by autonomous weapon systems would not absolve humans of responsibility for attacks conducted in violation of the rule of proportionality. Most importantly, a human operator who launches

⁵⁶ See CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, commentary accompanying rule 10.

⁵⁷ TALLINN MANUAL, *supra* note 6, rule 40 and accompanying commentary.

⁵⁸ CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rule 14; see also COMMANDER’S HANDBOOK, *supra* note 6, ¶ 5.3.3.

⁵⁹ To take one common example, the collateral damage caused during an attack or the failure to achieve an attack’s military aim are often relied upon when characterizing a particular attack as violating the rule. Such an approach is counter-normative because the rule of proportionality is evaluated *ex ante*, not *post factum*. For instance, if an attacker reasonably expects to cause five incidental deaths, but the strike causes fifteen, the proportionality rule was not violated so long as five is not excessive in light of the anticipated military advantage. On proportionality generally, see William J. Fenrick, *The Rule of Proportionality and Protocol I in Conventional Warfare*, 98 MIL. L. REV. 91 (1982), available at http://www.loc.gov/rr/frd/Military_Law/Military_Law_Review/pdf-files/277C87~1.pdf.

an autonomous weapon system into a situation in which the ensuing attacks are likely to cause excessive collateral damage has violated the rule. When making judgments as to possible violations, both the system's capabilities and the environment in which it will operate have to be considered. For instance, a violation would occur if the autonomous system served as a platform for weapons that are insufficiently precise to be used in the particular setting, such as a city, and, as a result, harm to civilians and civilian objects was probably going to be excessive to the military gains.

A matter that has drawn a great deal of attention in discussions about autonomous weapon systems is their capability to perform proportionality calculations. Recall that autonomous weapon systems employed in an area where civilians and civilian objects are present must possess some means of distinguishing them from lawful targets. If an autonomous weapon system locates a person or object, but cannot satisfactorily identify it as a lawful target, the system must treat that person or object as civilian. Any harm expected to befall the person or object during an attack on a valid military target would have to be factored into the proportionality calculation as collateral damage.

Proportionality calculations require consideration of both expected collateral damage and anticipated military advantage. A system already exists for determining the likelihood of collateral damage to objects or persons near a target. The "collateral damage estimate methodology" (CDEM) is a procedure whereby an attacking force considers such factors as the precision of a weapon, its blast effect, attack tactics, the probability of civilian presence in structures near the target, and the composition of structures to estimate the number of civilian casualties likely to be caused during an attack.⁶⁰ CDEM does not resolve whether a particular attack complies with the rule of proportionality. Instead, it is a policy-related instrument used to determine the level of command at which an attack causing collateral damage must be authorized. Oversimplified, the policy of

⁶⁰ For a discussion of the methodology, see *Collateral Damage Estimation Brief: Panel Discussion: Major Jeffrey Thurnher and Major Timothy Kelly* (U.S. Naval War College Oct. 23, 2012), http://www.youtube.com/watch?v=AvdXJV-N56A&list=PLam-yp5uUR1YEwLbqC0IPrP4EhWOeTf8v&index=1&feature=plpp_video; see also Defense Intelligence Agency General Counsel, *Briefing: Joint Targeting Cycle and Collateral Damage Estimate Methodology (CDM)*, (Nov. 10, 2009), http://www.aclu.org/files/dronefoia/dod/drone_dod_ACLU_DRONES_JOINT_STAFF_SLIDES_1-47.pdf.

those armed forces utilizing the methodology is that the higher the likely collateral damage, the higher the required approval authority.

The commander with authority to authorize the attack makes the proportionality determination as part of the attack's approval process. It is this individual who factors in the other essential element of a proportionality calculation, that is, the anticipated military advantage of the attack.⁶¹ Such determinations are contextual. For instance, an attack on a command-and-control facility expected to cause five civilian deaths at an early stage of the conflict yields greater military advantage than an attack on the same facility that occurs after enemy forces are in disarray and nearing defeat. Similarly, the destruction of a tank that is distant from the frontlines does not yield as much military advantage as destruction of one effectively firing on friendly forces. Because it is contextual, the military advantage element of the proportionality rule generally necessitates case-by-case determinations.

There is no question that autonomous weapon systems could be programmed to perform CDEM-like analyses to determine the likelihood of harm to civilians in the target area. Moreover, these weapon systems would usually be no less likely to generate a reliable result than CDEM since the latter is heavily reliant on scientific algorithms. The more difficult task for the autonomous weapon system would be assessing military advantage. Given the complexity and fluidity of the modern battlespace, it is unlikely in the near future that, despite impressive advances in artificial intelligence, "machines" will be programmable to perform robust assessments of a strike's likely military advantage. In part, this leads Human Rights Watch to conclude that the proportionality test "requires more than a balancing of quantitative data, and a robot could not be programmed to duplicate the psychological processes in human judgment that are necessary to assess proportionality."⁶²

Yet, military advantage algorithms could in theory be programmed into autonomous weapon systems. For example, the systems could be *pre*-programmed with unacceptable collateral damage thresholds for particular target sets or situations. As an example, an autonomous weapon system

⁶¹ A word of caution is necessary: a commander's decision does not relieve others involved in an attack of their own responsibility for compliance with international humanitarian law. Even a commander's order must be disobeyed if it is manifestly unlawful. *See, e.g.*, Rome Statute, *supra* note 6, art. 33.

⁶² LOSING HUMANITY, *supra* note 2, at 33.

could be pre-programmed with a base maximum collateral damage level of X for a tank; a human would have already made the determination that X generally comports with the proportionality rule. Such thresholds would have to be adjustable by human operators based on the military situation at a particular phase in the conflict, in a particular area of operations, and so forth. Of course, determining the appropriate threshold would be a very subjective endeavour. However, as noted by the ICRC commentary to Additional Protocol I, and as acknowledged in *Losing Humanity*, proportionality determinations necessarily involve a “fairly broad margin of judgment” and “must above all be a question of common sense and good faith for military commanders.”⁶³

Because military advantage is such a context specific value, compliance with the rule of proportionality would require that the base maximum collateral damage threshold be very conservative. Algorithms could be then be developed that would permit the autonomous weapon system to refine the base level threshold to account for specified variables it encountered on a mission. As an example, it would be reasonable to allow the system to increase the level of acceptable collateral damage if it identifies a concentration of enemy tanks, as distinct from a single tank. The concentration poses a greater threat and therefore the military advantage of destroying individual tanks making up the concentration is greater than that of destroying the same tanks when they are operating alone. Similarly, it would be reasonable for the system to adjust the level of acceptable collateral damage based on whether a targeted tank is headed towards or away from the battlefield.

These highly simplistic examples are offered solely for the sake of illustrating the point that the requirement to assess military advantage need not be an insurmountable obstacle. Obviously, the actual algorithms used would need to be much more sophisticated. While they would likely not be able to account for all imaginable scenarios and variables that might occur during hostilities, the same is true of a human confronted with unexpected or confusing events when making a time sensitive decision in combat. Neither the human nor the machine is held to a standard of perfection; on the contrary, in international humanitarian law the standard is always one of reasonableness.

⁶³ LOSING HUMANITY, *supra* note 2, at 33; ICRC COMMENTARY, *supra* note 25, ¶ 2208;

Human Rights Watch also cites the principle of “military necessity” as a basis for finding autonomous weapon systems unlawful.⁶⁴ In the author’s view, the organization mischaracterizes military necessity as a distinct rule of international humanitarian law, rather than a foundational principle that undergirds the entire body of law.⁶⁵ Despite the different approaches to military necessity adopted by the author and Human Rights Watch, it is clear that even military necessity as understood by that organization would not render autonomous weapon systems unlawful. They would not be unlawful *per se* because it is clear that autonomous weapon systems may be used in situations in which they are valuable militarily—that is, militarily necessary. As to prohibitions based on use, the condition that military objectives yield some military advantage would make any separate requirement for military necessity superfluous.⁶⁶ With regard to situations raising proportionality issues, any strike lacking military advantage but causing harm to civilians or civilian objects would violate the rule.⁶⁷ Taking these observations together, the result is that military necessity has no independent valence when assessing the legality of autonomous weapon systems or their use.

Losing Humanity somewhat inexplicably fails to examine a central element in every assessment of a weapon system’s use: the international humanitarian law requirement that the attacker take precautions in attack.⁶⁸ Set forth in Article 57 of Additional Protocol I, the rule, which reflects customary international law, requires an attacker to exercise “constant care . . . to spare the civilian population, civilians and civilian objects.”⁶⁹ The

⁶⁴ LOSING HUMANITY, *supra* note 2, at 34–35. Military necessity was originally described in the “Lieber Code”: “the necessity of those measures which are indispensable for securing the ends of [the] war, and which are lawful according to the modern law[s] and usages of war.” U.S. WAR DEP’T, INSTRUCTIONS FOR THE GOVERNMENT OF ARMIES OF THE UNITED STATES IN THE FIELD, General Orders No. 100 (Apr. 24, 1863), art. 14, *available at* <http://www.icrc.org/ihl.nsf/FULL/110?OpenDocument>.

⁶⁵ The author’s views on the subject are set forth in *Military Necessity and Humanity in International Humanitarian Law: Preserving the Delicate Balance*, 50 VA. J. INT’L L. 795, 795–839 (2010).

⁶⁶ Additional Protocol I, *supra* note 25, art. 52(2).

⁶⁷ *Id.*, arts. 51(5)(b) & 57(2)(iii).

⁶⁸ *See generally* A.P.V. ROGERS, LAW ON THE BATTLEFIELD, ch. 5 (3d. ed., 2012).

⁶⁹ Additional Protocol I, art. 57(1). *See also* CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rule 15; COMMANDER’S HANDBOOK, *supra* note 6, ¶ 8.1. Other treaty instruments include the requirement. *See* Second Protocol to the Hague Convention of 1954 for the Protection of Cultural Property in the Event of Armed Conflict, art. 7(b), Mar. 26, 1999, 2253 U.N.T.S. 212; Protocol (to the Convention on Prohibitions or

article goes on to articulate the means by which this obligation is to be carried out. In particular, an attacker is required to “do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects and are not subject to special protection but are military objectives”; cancel an attack if it becomes apparent that the rule of proportionality will be breached; provide “effective advance warning” of an attack if it may affect the civilian population, “unless circumstances do not permit”; “[w]hen a choice is possible between several military objectives for obtaining a similar military advantage, [select] that the attack on which may be expected to cause the least danger to civilian lives and to civilian objects”; and “take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects.”⁷⁰

Each of these obligations applies fully to the use of autonomous weapon systems. The requirement to do everything feasible to verify that the target is a military objective would, for example, require full use of on-board or external sensors that could boost the reliability of target identification. In fact, an autonomous weapon system could not be used in isolation if additional external means of identifying the target would measurably improve identification and their use was militarily feasible in the circumstances. As an illustration, such a situation might present itself if an unmanned aerial system could be used to narrow down the location of enemy forces before the autonomous weapon system is launched into that area. This would reduce the likelihood of the system’s misidentification of civilians as combatants.

The fulcrum of the verification requirement is the term “feasible.” Feasible has been interpreted as that which is “practicable or practically possible, taking into account all circumstances ruling at the time, including

Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects) on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices as amended on May 3, 1996, art. 3(10), 2048 U.N.T.S. 133; Protocol (to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects) on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices, art. 3(4), Oct. 10, 1980, 1342 U.N.T.S. 168.

⁷⁰ Additional Protocol I, *supra* note 25, art. 57 (2)–(3); CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, rules 16–21. On Article 57, see ICRC COMMENTARY, *supra* note 25, ¶¶ 2184–2238; BOTHE, *supra* note 25, at 359–369.

humanitarian and military considerations.”⁷¹ Military considerations include both technical and operational factors. Thus, in the previous example, sending an unmanned aerial system into an area of operations that could place it at a degree of risk not justified by the extent of enhanced identification capability would not be feasible. This might be because the aerial system is needed for operations elsewhere of greater military value or because its use in other operations may have a greater prospect for the avoidance of civilian casualties.

The requirement to select among military objectives to minimize civilian casualties and damage to civilian objects would likewise apply to autonomous weapon systems. For instance, an autonomous weapon system could not be employed to attack electrical substations if attacking transmission lines was militarily feasible, would achieve the same military objective (such as temporarily disrupting enemy command and control during friendly operations), and placed civilians and civilian objects at less risk.

However, it is the requirement to select the means of warfare likely to cause the least harm to civilians and civilian objects without sacrificing military advantage that is the key to the controversy over autonomous weapon systems. Indeed, it is the oft-ignored linchpin to various other weapon controversies, such as that surrounding the use of unmanned aerial combat systems. Consider the practical implications of this prescriptive norm: if the use of an autonomous weapon system can be expected to cause greater collateral damage than the use of a weapon system under human control, and the use of the latter is neither likely to diminish the probability that the desired military objective will be achieved nor poses a significant risk to the human operator, use of the autonomous weapon system would be forbidden as a matter of law. Restated, the only situation in which an autonomous weapon system can lawfully be employed is when its use will realize military objectives that cannot be attained by other available systems that would cause less collateral damage. Of course, there is a fair degree of elasticity in application of the norm given that it is based on the feasibility of

⁷¹ Protocol (to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects) on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices, *amended* May 3, 1996, art. 3(10), 2048 U.N.T.S. 133; U.K. Ratification Statement, *supra* note 54, ¶ (b). *See also* Commander’s Handbook, *supra* note 6, ¶ 8.3.1; CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 6, commentary accompanying rule 15.

the competing systems' use. Nevertheless, the requirement to select among means of warfare should significantly temper the concerns of those who would prophylactically prohibit use of autonomous weapon systems.

Indeed, contemplate the consequences of prohibiting autonomous weapon systems completely. What critics miss is that an autonomous weapon system may be able to achieve a military objective with *less* threat of collateral damage than a human controlled system. For example, an autonomous weapon system could be armed with non-lethal weapons unavailable on manned systems, its sensor suite could be more precise or discriminatory than one available on manned systems, or its decision-making capability could be better than that of a human in a particular environment (such as a very dangerous one). If the use of the human controlled system in question comports with the rule of proportionality, it would be lawful for an attacker to use it in the absence of the autonomous weapon system. Therefore, the prohibition of autonomous weapon systems would actually place civilians and civilian property at greater risk of incidental harm than if the autonomous weapon system had been available to the attacker.

Some States are beginning to set forth autonomous weapon systems guidelines that are meant to foster compliance with international humanitarian law in addition to achieving other objectives, such as avoiding mistaken engagements. For instance, the current U.S. policy, which distinguishes between semi-autonomous weapon systems, human-supervised autonomous weapon systems, and autonomous weapon systems, provides:

- (1) Semi-autonomous weapon systems (including manned or unmanned platforms, munitions, or sub-munitions that function as semi-autonomous weapon systems or as subcomponents of semi-autonomous weapon systems) may be used to apply lethal or non-lethal, kinetic or non-kinetic force. Semi-autonomous weapon systems that are onboard or integrated with unmanned platforms must be designed such that, in the event of degraded or lost communications, the system does not autonomously select and engage individual targets or specific target groups that have not been previously selected by an authorized human operator.
- (2) Human-supervised autonomous weapon systems may be used to select and engage targets, with the exception of

selecting humans as targets, for local defense to intercept attempted time-critical or saturation attacks for:

- (a) Static defense of manned installations.
- (b) Onboard defense of manned platforms.
- (3) Autonomous weapon systems may be used to apply non-lethal, non-kinetic force, such as some forms of electronic attack, against materiel targets in accordance with [applicable directives].

The policy acknowledges that autonomous or semi-autonomous weapon systems might be intended for use in a manner falling outside these policies. In such cases, it mandates high-level approval before formal development and then again before fielding the system.⁷² This requirement is in addition to the legal review requirements set forth below.

⁷² DoD DIRECTIVE 3000.09, *supra* note 9, ¶ 4d. Approval by the Under Secretary of Defense for Policy, the Under Secretary of Defense for Acquisition, Technology, and Logistics, and the Chairman of the Joint Chiefs of Staff is required. Prior to formal development the following are required:

- (1) The system design incorporates the necessary capabilities to allow commanders and operators to exercise appropriate levels of human judgment in the use of force.
- (2) The system is designed to complete engagements in a timeframe consistent with commander and operator intentions and, if unable to do so, to terminate engagements or seek additional human operator input before continuing the engagement.
- (3) The system design, including safeties, anti-tamper mechanisms, and information assurance . . . addresses and minimizes the probability or consequences of failures that could lead to unintended engagements or to loss of control of the system.
- (4) Plans are in place for [verification and validation] and [test and evaluation] to establish system reliability, effectiveness, and suitability under realistic conditions, including possible adversary actions, to a sufficient standard consistent with the potential consequences of an unintended engagement or loss of control of the system.
- (5) A preliminary legal review of the weapon system has been completed, in coordination with the General Counsel of the Department of Defense . . . and in accordance with [the relevant policy guidance].

Id., encl 3, ¶ 1a. Before fielding, the review must assess:

- (1) System capabilities, human-machine interfaces, doctrine, TTPs, and training have demonstrated the capability to allow commanders and

IV. Legal Review of Weapon Systems

Since the prospect of autonomous weapon systems is so new, the requirement to conduct a review of their legality looms large.⁷³ Codified in Article 36 of Additional Protocol I, the rule provides that “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.”⁷⁴ Means of warfare are weapons and weapon systems, whereas method of warfare refers to the tactics, techniques and procedures (TTP) by which hostilities are conducted. An autonomous weapon system is a means of warfare. Employing multiple autonomous

operators to exercise appropriate levels of human judgment in the use of force and to employ systems with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable ROE.

(2) Sufficient safeties, anti-tamper mechanisms, and information assurance in accordance with Reference (a) have been implemented to minimize the probability or consequences of failures that could lead to unintended engagements or to loss of control of the system.

(3) V&V and T&E assess system performance, capability, reliability, effectiveness, and suitability under realistic conditions, including possible adversary actions, consistent with the potential consequences of an unintended engagement or loss of control of the system.

(4) Adequate training, TTPs, and doctrine are available, periodically reviewed, and used by system operators and commanders to understand the functioning, capabilities, and limitations of the system’s autonomy in realistic operational conditions.

(5) System design and human-machine interfaces are readily understandable to trained operators, provide traceable feedback on system status, and provide clear procedures for trained operators to activate and deactivate system functions.

(6) A legal review of the weapon system has been completed, in coordination with the [the DoD General Counsel and relevant policy guidance].

Id., encl. 3, ¶ 1b.

⁷³ On weapons review generally, see BOOTHBY, *supra* note 23, at 340–52; W. Hays Parks, *Conventional Weapons and Weapons Reviews*, 8 Y.B. INT’L HUMANITARIAN L. 55 (2005).

⁷⁴ On Article 36, see ICRC Commentary, *supra* note 25, ¶¶ 1463–1482; BOTHE, *supra* note 25, at 199–201.

weapon systems to conduct, for example, a siege by targeting all vehicular traffic into or out of a populated area illustrates their use as a method of warfare.

Although Human Rights Watch suggests that disagreement exists as to whether Article 36 restates customary international law,⁷⁵ the obligation to conduct legal reviews of new *means* of warfare before their use is generally considered—and correctly so—reflective of customary international law.⁷⁶ Consensus is lacking as to whether an analogous requirement exists to perform legal reviews of *methods* of warfare.⁷⁷

States that are Party to Additional Protocol I are clearly required to conduct a legal review of both the autonomous weapon system and any TTP that its user develops. Non-Party States are arguably only bound by the obligation to review the system itself in light of its envisaged usage. This author is of the opinion that both reviews—whether or not legally mandated—are well-advised whenever feasible.

Human Rights Watch also advises that reviews of autonomous weapon systems “should take place at the earliest stage possible and continue through any development that proceeds.”⁷⁸ For States Party to Additional Protocol I, this is, as is clear from the plain text of Article 36, a legal requirement. Since there is no corresponding customary international humanitarian law requirement, non-Party States, such as the United States, are only required to ensure weapons are lawful before use. Nevertheless, the organization is right to stress that early legal reviews can shape the development stage of a weapon system and resultantly avoid unnecessary effort and cost associated with components and capabilities that may not pass legal muster. It is U.S. policy to conduct two legal reviews, once prior to taking the decision to enter into formal development, and again before an autonomous weapon system is fielded.⁷⁹

⁷⁵ LOSING HUMANITY, *supra* note 2, at 21.

⁷⁶ See, e.g., Air and Missile Warfare Manual, *supra* note 6, rule 9; TALLINN MANUAL, *supra* note 6, rule 48.

⁷⁷ See TALLINN MANUAL, *supra* note 6, commentary accompanying rule 48.

⁷⁸ LOSING HUMANITY, *supra* note 2, at 22.

⁷⁹ DOD DIRECTIVE 3000.09, *supra* note 9, encl. 3, ¶¶ 1a(5) and 1b(6). Additional U.S. policy guidance on legal reviews is contained in DEP’T OF DEF., DIRECTIVE 5000.01, THE DEFENSE ACQUISITION SYSTEM (May 12, 2003) [hereinafter DOD DIRECTIVE 5000.01]; DEP’T OF DEF., INSTRUCTION 5000.02, OPERATION OF THE DEFENSE ACQUISITION SYSTEM (Dec. 8, 2008); DEP’T OF DEF., DIRECTIVE 3000.03, POLICY FOR NON-LETHAL

Losing Humanity appears to infer that the United States has taken the position that only a weapon—and not the weapon system—need be examined. Putting individual cases where the United States may or may not have complied with the requirement for legal reviews aside, it is presently U.S. policy to review both weapons and weapon systems. This obligation is unambiguously confirmed in a Department of Defense directive which provides that “[t]he acquisition and procurement of DoD weapons and weapon systems shall be consistent with all applicable domestic law and treaties and international agreements . . . , customary international law, and the law of armed conflict An attorney authorized to conduct such legal reviews in the Department shall conduct the legal review of the *intended* acquisition of *weapons or weapons systems*.”⁸⁰

Human Rights Watch expressed particular concern regarding the possibility of autonomous weapon systems modification. It noted “robots . . . are complex systems that often combine a multitude of components that work differently in different combinations.”⁸¹ The organization also highlighted “the fact that some robotic technology, while not inherently harmful, has the potential one day to be weaponized.”⁸² A fair reading of the international humanitarian law norm is that any significant modification to a weapon system requires legal review. The United States agrees. For example, the U.S. Air Force policy on weapon reviews specifically mandates “a timely legal review of all weapons and cyber capabilities, whether a new weapon or cyber capability at an early stage of the acquisition process, or a *contemplated modification* of an existing weapon or cyber capability, to ensure legality under LOAC, domestic law and international law prior to their acquisition for use in a conflict or other military operation.”⁸³ To summarize, U.S. policy is to review all weapons, their associated delivery systems, and any significant modification of them. This policy unquestionably applies to autonomous weapon systems.

WEAPONS (July 9, 1996); and DEP’T OF DEF., DIRECTIVE 2311.01E, DOD LAW OF WAR PROGRAM (May 9, 2006).

⁸⁰ DOD DIRECTIVE 5000.01, *supra* note 79, encl. 1, ¶ E1.1.15 (emphasis added).

⁸¹ LOSING HUMANITY, *supra* note 2, at 23.

⁸² *Id.*

⁸³ U.S. AIR FORCE, INSTRUCTION 51-402, LEGAL REVIEW OF WEAPONS AND CYBER CAPABILITIES para. 1.3.1 (July 27, 2011) [hereinafter AIR FORCE INSTRUCTION 51-402] (emphasis added).

Losing Humanity cites an array of legal prohibitions for consideration during the legal review. However, the report is overinclusive in that a number of the prohibitions cited bear primarily on the unlawful use of lawful weapons. Legal reviews do not generally consider use issues since they are contextual by nature, whereas the sole context in a determination of whether a weapon is lawful *per se* is its intended use in the abstract.⁸⁴ For instance, the rule of proportionality does not factor into a weapons review because compliance depends on the situational risk to civilians and civilian objects and the anticipated military advantage in the attendant circumstances. Because the assessment is contextual, it is generally inappropriate to make *ex ante* judgments as to a weapon's compliance with the rule. Only if the weapon system were necessarily employed in situations where injury to civilians or harm to civilian objects is inevitable and predictable in scope—as in a cyber malware weapon developed for a particular attack—would such an assessment have to be made prior to fielding of the weapon. The requirement that an attacker take feasible precautions in attack to minimize harm to civilians and civilian objects is likewise context specific and, therefore, any assessment of compliance with the norm can only occur with respect to its use in particular circumstances, not as part of the legal review.

The Air Force guidance delineates the legal issues that must be examined when determining the legality of a weapon system being considered for acquisition:

3.1.1. Whether there is a specific rule of law, whether by treaty obligation of the United States or accepted by the United States as customary international law, prohibiting or restricting the use of the weapon or cyber capability in question.

⁸⁴ The Committee Report on the Article presented to the Diplomatic Conference emphasizes that the rule:

[I]s not meant to imply an obligation “to foresee or analyze all possible misuse of a weapon, for any weapon can be misused in a way that would be prohibited.” The meaning of the phrase is to require a determination whether the employment *for its normal or expected use* would be prohibited under some or all circumstances.

BOTHE, *supra* note 25, at 200–201.

3.1.2. If there is no express prohibition, the following questions are considered:

3.1.2.1. Whether the weapon or cyber capability is calculated to cause superfluous injury, in violation of Article 23(e) of the Annex to Hague Convention IV; and

3.1.2.2. Whether the weapon or cyber capability is capable of being directed against a specific military objective and, if not, is of a nature to cause an effect on military objectives and civilians or civilian objects without distinction.⁸⁵

The requirements set forth in paragraph 3.1.2 mirror analogous provisions appearing in Additional Protocol I, all of which are customary in nature.⁸⁶

The Air Force guidance's extension of its substantive requirements to cyber capabilities, a relatively recent revision to the basic document, is noteworthy. It illustrates the principle that the rules of international humanitarian law regarding the legality of weapon systems apply fully to weapons that did not exist at the time a particular treaty norm was crafted or customary law crystallized.⁸⁷ It is incontrovertible that all of the norms discussed apply equally to autonomous weapon systems.

Finally, Human Rights Watch asserts that legal reviews "should assess a weapon under the Martens Clause," a proposition echoed by the International Committee of the Red Cross.⁸⁸ The clause originally appeared in the 1899 Hague Convention II and was subsequently included in both the 1907 version of that treaty and Additional Protocol I.⁸⁹ The International Court of Justice recognizes it as customary in nature and has observed that the Martens Clause "proved to be an effective means of

⁸⁵ AIR FORCE INSTRUCTION 51-402, *supra* note 83, ¶¶ 3.1.1.–3.1.2.2. Note that the reference to effects in the Air Force guidance covers both immediate effects and those which spread, thereby encompassing both the situations envisaged in Additional Protocol I, art. 51(4)(b) & art. 51(4)(c).

⁸⁶ Additional Protocol I, *supra* note 25, arts. 35(2), 51(4)(b)–(c).

⁸⁷ Nuclear Weapons, *supra* note 25, ¶ 86.

⁸⁸ LOSING HUMANITY, *supra* note 2, at 25; ICRC, *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*, ¶ 1.2.2.3 (2006).

⁸⁹ 1899 Hague II, *supra* note 24, pmbli; 1907 Hague IV, *supra* note 24, pmbli; Additional Protocol I, *supra* note 25, art. 1(2).

addressing rapid evolution of military technology.”⁹⁰

By its own terms, though, the clause applies only in the absence of treaty law.⁹¹ In other words, it is a failsafe mechanism meant to address lacunae in the law; it does not act as an overarching principle that must be considered in every case. Today, a rich fabric of treaty law governs the legality of weapon systems. Certain of these treaties bear directly on the development of autonomous weapon systems. The restrictions on incendiary weapons, air delivered antipersonnel mines, and cluster munitions, for example, limit their employment on autonomous weapon systems by States Party to the respective treaties.⁹² As discussed above, general principles and rules of international humanitarian treaty law, particularly those contained in Additional Protocol I, further restrict weaponry. Emergence of many customary international humanitarian law norms since 1899 also measurably diminishes the significance of the clause. By the turn of the 21st century, the likelihood that future weapon systems, including those that might be autonomous, would not violate applicable treaty and customary law, but be unlawful based on the Martens Clause, had become exceptionally low.

Losing Humanity is correct to accentuate the importance of weapons reviews in the process of developing and fielding new weaponry. However, it must be cautioned that such reviews examine only the legality of a weapon system as such, not its use in any particular circumstance.

⁹⁰ Nuclear Weapons, *supra* note 25, ¶¶ 78, 84.

⁹¹ The text of the clause refers to “cases not covered by this Protocol or by other international agreements.” Additional Protocol I, *supra* note 25, art. 1(2).

⁹² See 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; Protocol (to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects) on Prohibitions and Restrictions on the Use of Incendiary Weapons, Oct. 10, 1980, 1342 U.N.T.S. 171; Protocol (to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects) on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices, Oct. 10, 1980, 1342 U.N.T.S. 168; Protocol (to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects) on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices as amended on May 3, 1996, 2048 U.N.T.S. 133; Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Antipersonnel Mines and on Their Destruction, Sept. 18, 1997, 36 I.L.M. 1507; Convention on Cluster Munitions, Dec. 3, 2008, 48 I.L.M. 357.

Therefore, it is doubtful whether the requirement for the reviews will serve as an impediment to the development of autonomous weapon systems as a class of weapons.

V. Accountability

Human Rights Watch expresses anxiety regarding accountability for the activities of fully autonomous weapons. It asks the very reasonable question: “If the killing were done by a fully autonomous weapon, . . . the question would become: whom to hold responsible?” The organization concludes that, “[s]ince there is no fair and effective way to assign legal responsibility for unlawful acts committed by fully autonomous weapons, granting them complete control over targeting decisions would undermine yet another tool for promoting civilian protection.”⁹³

The problem with this conclusion is that it is based on a false premise.⁹⁴ The mere fact that a human might not be in control of a particular engagement does not mean that no human is responsible for the actions of the autonomous weapon system.⁹⁵ A human must decide how to program the system. Self-evidently, that individual would be accountable for programming it to engage in actions that amounted to war crimes. Moreover, the commander or civilian supervisor of that individual would be accountable for those war crimes if he or she knew or should have known that the autonomous weapon system had been so programmed and did nothing to stop its use, or later became aware that the system had been employed in a manner constituting a war crime and did nothing to hold the individuals concerned accountable.⁹⁶

⁹³ LOSING HUMANITY, *supra* note 2, at 42.

⁹⁴ On accountability, see generally Rogers, *supra* note 68, at 360 & ch. 11.

⁹⁵ William J. Fenrick, *The Prosecution of International Crimes in Relation to the Conduct of Military Operations*, in THE HANDBOOK OF THE INTERNATIONAL LAW OF MILITARY OPERATIONS 501–505 (Terry Gill & Dieter Fleck eds., 2010).

⁹⁶ *See, e.g.*, Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, art. 49, Aug. 12, 1949, 75 U.N.T.S. 31; Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea, art. 50, Aug. 12, 1949, 75 U.N.T.S. 85; Convention (III) Relative to the Treatment of Prisoners of War, art. 129, Aug. 12, 1949, 75 U.N.T.S. 135; Convention (IV) Relative to the Protection of Civilian Persons in Time of War, art. 146, Aug. 12, 1949, 75 U.N.T.S. 287 [hereinafter Geneva Convention IV]; Additional Protocol I, *supra* note 25, arts. 86–87; Hague Convention for the Protection of Cultural Property in the Event of Armed Conflict with Regulations for the Execution of the Convention, art. 28, May 14, 1954, 249 U.N.T.S. 240; Second Protocol to the Hague Convention of 1954 for the

It is hopefully improbable that an autonomous weapon system would be programmed to commit war crimes. Much more likely would be a case in which a system that has not been so programmed is nevertheless used in a manner that constitutes such crimes. For example, the operator of an autonomous weapon system that cannot distinguish civilians from combatants who employs the system in an area where the two are intermixed has committed the war crime of indiscriminate attack. Any commander or supervisor who ordered the attack would likewise be criminally responsible for committing a war crime. So too would a commander or supervisor who knew the operation was about to be mounted and failed to suppress it or who later learned of the operation and failed to take action to hold those responsible accountable.

The United States accepts the premise that those involved in autonomous weapon system operations may be held accountable for their decisions. In its most recent guidance on the use of the systems, the Department of Defense has emphasized that “[p]ersons who authorize the use of, direct the use of, or operate autonomous and semi-autonomous weapon systems must do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement (ROE).”⁹⁷ The policy imposes identical requirements on Commanders of the U.S. Combatant Commands.⁹⁸

Protection of Cultural Property in the Event of Armed Conflict, art. 15(2), Mar. 26, 1999, 2253 U.N.T.S. 212; Rome Statute, *supra* note 6, arts. 25(3)(b), 28; Statute of the International Criminal Tribunal for the Former Yugoslavia, (as amended on May 17, 2002), art. 7(1), S.C. Res. 827 annex, U.N. Doc. S/RES/827 (May 25, 1993); Statute of the International Criminal Tribunal for Rwanda, art. 6(1), S.C. Res. 955 annex, U.N. Doc. S/RES/955, (Nov. 8, 1994); Prosecutor v. Blaškić, Case. No. IT-95-14-T, Trial Chamber Judgment, ¶¶ 281–282 (Int’l Crim. Trib. for the Former Yugoslavia Mar. 3, 2000); Prosecutor v. Krstić, Case No. IT-98-33-T, Judgment, ¶ 605 (Int’l Crim. Trib. for the Former Yugoslavia Aug. 2, 2001); Prosecutor v. Kayishema & Ruzindana, ¶223, Case No. ICTR 95-1-T, Judgment (May 21, 1999); COMMANDER’S HANDBOOK, *supra* note 6, ¶ 6.1.3.

⁹⁷ DoD DIRECTIVE 3000.09, *supra* note 9, ¶. 4b.

⁹⁸ *Id.*, encl. 4, ¶ 10b. The United States has nine Combatant Commands: U.S. Africa Command (USAFRICOM); U.S. Central Command (USCENTCOM); U.S. European Command (USEUCOM); U.S. Northern Command (USNORTHCOM); U.S. Pacific Command (USPACOM); U.S. Special Operations Command (USSOCOM); U.S. Southern Command (USSOUTHCOM); U.S. Strategic Command (USSTRATCOM); U.S. Transportation Command (USTRANSCOM).

Conclusion

This Article has demonstrated that autonomous weapon systems are not unlawful *per se*. Their autonomy has no direct bearing on the probability they would cause unnecessary suffering or superfluous injury, does not preclude them from being directed at combatants and military objectives, and need not result in their having effects that an attacker cannot control. Individual systems could be developed that would violate these norms, but autonomous weapon systems are not prohibited on this basis *as a category*.

International humanitarian law's restrictions on the use of weapons would nevertheless limit their employment in certain circumstances. This is true of every weapon, from a rock to a rocket. Of course, the fact that autonomous weapon systems locate and attack persons and objects without human interaction raises unique issues. These challenges are not grounds for banning the systems entirely. On the contrary, international humanitarian law's restrictions on the use of weapons (particularly the requirements that they be directed only against combatants and military objectives, they not be employed indiscriminately, their use not result in excessive harm to civilians or civilian objects, and they not be used when other available weapons could achieve a similar military advantage while placing civilians and civilian objects at less risk) are sufficiently robust to safeguard humanitarian values during the use of autonomous weapon systems.

But might States ban the weapons in the foreseeable future? As noted in the 1868 St. Petersburg Declaration, international humanitarian law fixes "the technical limits at which the necessities of war ought to yield to the requirements of humanity."⁹⁹ Virtually every rule therein reflects a balance States have made between these two seminal factors—humanitarian concerns and military necessity. The humanitarian concerns that are factored into the equation reflect the interest States have in maximizing international humanitarian law's protection of their combatants and civilian population during armed conflict. In that States are self-interested entities, these concerns are tempered by their desire to retain the ability to fight effectively in order to achieve national interests. The result of this dialectic interplay is international humanitarian law, either in the form of treaty law that has been negotiated by States on the basis of

⁹⁹ St. Petersburg Declaration, *supra* note 24.

their assessment of the balance or customary law resulting from State practice and *opinio juris* that reflects the balancing.

Given this process of norm formulation, the Human Rights Watch position is unlikely to find traction. Clearly, and as illustrated in the new Department of Defense directive on autonomous weapon systems, States are sensitive to the humanitarian implications of these systems. Yet, autonomy in combat is in its infancy. Until both their potential for unintended human consequences and their combat potential are better understood, it is unlikely that any State would seriously consider banning autonomous weapon systems. Indeed, there is little historical precedent for banning weaponry before it has been fielded.¹⁰⁰

Counterintuitive though it may seem, it would arguably be irresponsible to prohibit autonomous weapons at this stage in their development. As noted, such weapons may offer the possibility of attacking the enemy with little risk to the attacker. Although this “value” has sometimes been criticized with respect to unmanned combat aerial systems like the Predator, there is no basis in international humanitarian law for suggesting that attacking forces must assume risk. On the contrary, what is often forgotten is that international humanitarian law affirmatively protects combatants. The paradigmatic example is the “cardinal” prohibition on weapons that cause unnecessary suffering.¹⁰¹ In fact, international humanitarian law was almost exclusively concerned with the protection of combatants until adoption of the fourth Geneva Convention in 1949.¹⁰² It runs counter to the object and purpose of this body of law to suggest that a weapon system that reduces harm to combatants in situations in which its use does not aggravate civilian risk should be unlawful.

An even more compelling argument is that banning autonomous weapon systems before their potential is understood may have the effect of denying commanders a tool for minimizing the risk to civilians and civilian objects in certain attack scenarios. Admittedly, autonomous weapon system

¹⁰⁰ The only contemporary exception is the ban on permanently blinding lasers, one that deprived States of very little militarily because temporarily blinding lasers can generally serve the same military purpose as the former. Additional Protocol on Blinding Laser Weapons (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), Oct. 13, 1995, 1380 U.N.T.S. 370.

¹⁰¹ Nuclear Weapons, *supra* note 25, ¶ 78.

¹⁰² Geneva Convention IV, *supra* note 96.

development is not at the point where one can authoritatively conclude the systems will offer less harmful options than human-operated systems. However, it is equally not at the point where such a possibility can be ruled out.

Human Rights Watch is to be commended for drawing attention to the issue of fully autonomous weapon systems. However, in the absence of even a single such system being fielded, it is premature to draw conclusions either as to their legality or to the broader issue of whether they should be banned as a matter of policy. Understanding of the systems' potential for both positive and negative ends is simply too primitive at this time to comfortably draw conclusions as to their legal, moral, and operational costs and benefits.