"Out of the loop": autonomous weapon systems and the law of armed conflict

Article

Published Version

ARTICLE

"Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict

Michael N. Schmitt* & Jeffrey S. Thurnher**

I. Introduction

The introduction of autonomous weapon systems into the "battlespace" will profoundly influence the nature of future warfare. This reality has begun to draw the attention of the international legal community. Most notably, in November 2012, Human Rights Watch released *Losing Humanity: The Case against Killer Robots*.[1] The report is a scathing critique of autonomous weapons systems and has sparked a lively and important debate over the lawfulness of taking humans "out of the loop" during lethal targeting.[2]

* Chairman and Professor, International Law Department, United States Naval War College; Honorary Professor, Strategy and Security Institute and Law School, Exeter University (UK); Honorary Professor of International Humanitarian Law, Durham University Law School (UK). This Article derives in part from Michael N. Schmitt, *Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics*, HARV. NAT'L. SEC. J. FEATURES (2013), http://harvardnsj.org/wp-content/uploads/2013/02/Schmitt-Autonomous-Weapon-Systems-and-IHL-Final.pdf. The views expressed are those of the authors alone and should not be understood as necessarily representing those of the U.S. Department of Defense or any other government entity.

** Lieutenant Colonel, United States Army, Military Professor, International Law Department, United States Naval War College.

[1] HUMAN RIGHTS WATCH, LOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS (2012), *available at* http://www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf [hereinafter LOSING HUMANITY].

[2] On the issue of the legality of autonomous weapon systems, see Jeffrey S. Thurnher, *The Law that Applies to Autonomous Weapon Systems*, 17 ASIL INSIGHTS 4 (Jan. 18, 2013), http://www.asil.org/pdfs/insights/insight130118.pdf; Kenneth Anderson & Matthew Waxman, *Law and Ethics for Robot Soldiers*, 176 POL. REV. (Dec. 1, 2012), http://www.hoover.org/publications/policy-review/article/135336; Markus Wagner, *Taking Humans Out of the Loop: Implications for International Humanitarian Law*, 21 J. L. INFO. &

Human Rights Watch's position on "robots," a colloquial rendering for autonomous weapon systems, is forceful and unambiguous: "[F]ully autonomous weapons would not only be unable to meet legal standards but would also undermine essential non-legal safeguards for civilians."[3] The organization accordingly concludes that they "should be banned and . . . governments should urgently pursue that end."[4]

This conclusion has important national security implications. The United States and its allies have a substantial interest in maintaining a technological edge over potential adversaries, in particular by fielding systems that enable them to deliver lethal force while minimizing the risk to their own forces. After all, the objective of warfare is to achieve and exploit advantages to the enemy's detriment. Sensitive to this reality, the law of armed conflict has never been about ensuring a "fair fight"; rather, it comprises prohibitions, restrictions, and obligations designed to balance a State's interest in effectively prosecuting the war (military necessity) with its interest in minimizing harm to those involved in a conflict (humanity).[5] The question for the legal community, therefore, is whether autonomous weapon systems comply with the legal norms that States have put in place to achieve this balance. Any misinterpretation or misapplication of said norms will necessarily skew this delicate balance, thereby placing compliance by States with the law of armed conflict at risk.

Furthermore, as has too often been the case with international calls to ban particular weapons, debates over weapons have had a way of getting off-course, both normatively and operationally. At least from the United States government's perspective, this tendency contributed to its unwillingness to become Party to the Ottawa Antipersonnel Land Mines and the Dublin Cluster Munitions Conventions.[6] A contemporary example

---

SCI. 155, 155 (2011), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1874039.

[3] LOSING HUMANITY, *supra* note 1, at 1–2.

[4] *Id.* at 2.

[5] *See generally* Michael N. Schmitt, *Military Necessity and Humanity in International Humanitarian Law: Preserving the Delicate Balance,* 50 VA. J. INT'L L. 795 (2010).

[6] Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction, Sept. 18, 1997, 2056 U.N.T.S. 211; Convention on Cluster Munitions, Dec. 3, 2008, 48 I.L.M. 357. Although the United States elected not to become Party to the Ottawa Convention, the Obama Administration is presently reviewing its position on the matter. Andrew J. Shapiro, *Briefing on the Release of*

presaging the emerging controversy over autonomous weapon systems is the polemic over drones, which has often been counter-normative and counter-factual.[7] Rather than allowing the discourse over autonomous weapon systems to similarly spiral out of control, what is needed is a vibrant, informed, measured, and mature discussion of the relevant legal issues.

This Article is intended to help infuse granularity and precision into the legal debates surrounding such weapon systems and their future uses. 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 the law of armed conflict's prescriptive norms governing the "conduct of hostilities."[8] This Article concludes that an outright ban of autonomous weapon systems is insupportable as a matter of law, policy, and operational good sense. Indeed, proponents of a ban underestimate the extent to which the law of armed conflict, including its customary law aspect, will control autonomous weapon system operations.[9] Some autonomous weapon systems that might

---

the Tenth Edition of the "To Walk the Earth in Safety" Report, U.S. DEP'T OF ST. (Dec. 19, 2011), http://www.state.gov/t/pm/rls/rm/179143.htm. For the U.S. position on cluster munitions, *see* U.S. DEP'T. OF STATE, CLUSTER MUNITIONS, http://www.state.gov/t/pm/wra/c25930.htm (last visited March 21, 2013).

[7] Michael N. Schmitt, *Unmanned Combat Aircraft Systems (Armed Drones) and International Humanitarian Law: Simplifying the Oft Benighted Debate,* 30 B.U. INT'L L. J. 595, 596 (2012).

[8] The phrase "conduct of hostilities" refers in particular to "attacks," as that term is understood in Article 49 of Additional Protocol I. Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, art. 49, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter Additional Protocol I]. The core conduct of hostility rules are set forth in Part IV, Section I, of the Protocol. Although the United States is not a Party to the Protocol, most of the rules contained in the Section reflect customary international law.

[9] We refer to the following non-binding compilations of customary international humanitarian law rules to draw conclusions as to customary status of the norms referenced in this article: INTERNATIONAL COMMITTEE OF THE RED CROSS, 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), *available at* http://www.iihl.org/iihl/Documents/The%20Manual%20on%20the%20Law%20of%20NIAC.pdf; PROGRAM ON HUMANITARIAN POLICY AND CONFLICT RESEARCH AT HARVARD UNIVERSITY (HPRC), COMMENTARY ON THE MANUAL ON INTERNATIONAL LAW APPLICABLE TO AIR AND MISSILE WARFARE (2010) [hereinafter AMW MANUAL COMMENTARY], TALLINN MANUAL ON THE INTERNATIONAL LAW APPLICABLE TO CYBER WARFARE (Michael N. Schmitt gen. ed., 2013 forthcoming) [hereinafter TALLINN

be developed would already be unlawful *per se* under existing customary law, irrespective of any treaty ban. The use of certain others would be severely limited by that law.

Furthermore, an outright ban is premature since no such weapons have even left the drawing board.[10] Critics typically either fail to take account of likely developments in autonomous weapon systems technology or base their analysis on unfounded assumptions about the nature of the systems. From a national security perspective, passing on the opportunity to develop these systems before they are fully understood would be irresponsible. Perhaps even more troubling is the prospect that banning autonomous weapon systems altogether based on speculation as to their future form could forfeit their potential use in a manner that would minimize harm to civilians and civilian objects when compared to non-autonomous weapon systems.

II. Understanding Autonomous Weapon Systems

Before turning to the law, it is essential to examine the operational and technological context in which that law will be applied. The starting point for any such discussion is definitional. A "weapon system" consists of a weapon and the items associated with its employment.[11] An example is an

---

MANUAL]. Additionally, although they are not compilations of customary international law, scholars and practitioners 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/COMDTPUB P5800.7A (2007), *available at* http://www.usnwc.edu/getattachment/a9b8e92d-2c8d-4779-9925-0defea93325c/1-14M_(Jul_2007)_(NWP) [hereinafter COMMANDER'S HANDBOOK].

[10] The International Committee of the Red Cross sagely suggests taking time to analyze carefully the legal and other issues surrounding autonomous weapon systems before drawing conclusions as to their lawfulness. INT'L COMM. OF THE RED CROSS, INTERNATIONAL HUMANITARIAN LAW AND THE CHALLENGES OF CONTEMPORARY ARMED CONFLICTS 40 (Report of the 31st Conference of the Red Cross and Red Crescent, Geneva, Switzerland 31IC/11/5.1.2 2011) , *available at* http://www.icrc.org/eng/assets/files/red-cross-crescent-movement/31st-international-conference/31-int-conference-ihl-challenges-report-11-5-1-2-en.pdf [hereinafter ICRC REPORT].

[11] In this Article, we use the term "weapon system" as shorthand to refer both to a weapon and a complete weapon system.

F-16 fighter-bomber armed with a Joint Direct Attack Munition (JDAM) bomb.

The Department of Defense (DoD) has 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.[12]

The crux of full autonomy, therefore, is the 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 is capable of operating on its own. Of course, a fully autonomous system is never entirely human-free. Either the system designer or an operator would at least have to program it to function pursuant to specified parameters, and an operator would have to decide to employ it in a particular battlespace.

Fully autonomous weapon systems must be distinguished from other categories of weapons with lesser degrees of autonomy, which Human Rights Watch has not suggested banning. For instance, 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.[13] Another human-supervised

---

[12] U.S. DEP'T OF DEF., Directive 3000.09, AUTONOMY IN WEAPON SYSTEMS 13–14 (Nov. 2, 2012), *available at* http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf [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 1, at 2.

[13] Lockheed Martin, *Aegis Combat System*, http://www.lockheedmartin.com/us/products/aegis.html (last visited Feb. 23, 2013); Federation of American Scientists (FAS), *Patriot TMD*, http://www.fas.org/spp/starwars/program/patriot.htm (last visited Feb. 23, 2013).

autonomous weapon system, Israel's Iron Dome, recently received a great deal of attention as it very effectively destroyed incoming rockets from Gaza.[14] The only substantive distinction between these systems and those that are fully autonomous is that a human with override capability carefully monitors the operation of the human-supervised systems.

Other autonomous systems are sometimes labelled "automatic weapons defense systems."[15] Such systems respond automatically (or near automatically) when they detect incoming threats; their key distinguishing features are that they are designed solely for defense and are fixed. An example is the "close-in weapon system" (CIWS or "Sea Whiz").[16] Used for point-defense of warships against incoming missiles, the Sea Whiz can be programmed to detect and automatically engage inbound missiles based on parameters that include speed and altitude.

Weapon systems may also be "semi-autonomous." Their use is commonplace in contemporary warfare.[17] Semi-autonomous weapon systems 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.

The United States is not currently fielding any fully autonomous weapon systems.[18] Nor are there any "plans to develop lethal autonomous

---

[14] 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; *Iron Dome*, RAFAEL, http://www.rafael.co.il/marketing/SIP_STORAGE/FILES/6/946.pdf (last visited Feb. 23, 2013).

[15] LOSING HUMANITY, *supra* note 1, at 9.

[16] *MK-15 Phalanx Close-In Weapons System (CIWS)*, U.S. NAVY (October 19, 2012) http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=487&ct=2.

[17] The weapons are also known as "launch and leave" weapons. Examples include the AGM-130 and AGM-65 missiles used for attacking ground targets. For descriptions of these and other such systems, see *Factsheets (Weapons)*, U.S. AIR FORCE, http://www.af.mil/information/factsheets/index.asp (last visited Feb. 23, 2013).

[18] U.S. DEP'T OF DEF., 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.

weapon systems other than human-supervised systems for the purposes of local defense of manned vehicles or installations."[19] This does "not preclude a change in that policy as the capacity for autonomy evolves."[20] In fact, such a change can be expected. A series of DoD studies, plans, and roadmaps have thoroughly discussed autonomous weapons technology, clearly in anticipation of its possible fielding in the future.[21] A U.S. Joint Forces Command study has likewise acknowledged that, based upon "the explosive developments in technology," it can "envision a world in which humans need not be in the decision loop."[22]

Operational realities will likely drive the United States to discard its practice of keeping a human in the loop for lethal targeting decisions. First, requiring a human in the loop can be personnel intensive. For example, it can take scores of people, from pilots and weapon systems operators to technicians and intelligence analysts, to operate a single tethered unmanned aircraft.[23] As a rule, the more autonomous the system, the fewer personnel needed to operate it.

---

[19] *Id.* at 3. The U.K. Ministry of Defence similarly has no current plans to develop fully autonomous weapon systems. Meredith Hagger & Tim McCormack, *Regulating the Use of Unmanned Combat Vehicles: Are General Principles of International Humanitarian Law Sufficient?*, 21 J.L. INFO. & SCI. 74, 89 (2011).

[20] LOSING HUMANITY, *supra* note 1, at 8.

[21] U.S. DEP'T OF DEF., OFFICE OF THE SEC'Y OF DEF., UNMANNED SYSTEMS ROADMAP 2007–2032 49, 54 (2007), *available at* http://www.fas.org/irp/program/collect/usroadmap2007.pdf; U.S. DEP'T OF DEF.,, OFFICE OF THE SEC'Y OF DEF.,, UNMANNED AIRCRAFT SYSTEMS ROADMAP 2005–2030 52 (2005), *available at* https://www.fas.org/irp/program/collect/uav_roadmap2005.pdf; U.S. AIR FORCE, UNMANNED AIRCRAFT SYSTEMS FLIGHT PLAN 2009–2047 50–51 (2009), *available at* http://www.govexec.com/pdfs/072309kp1.pdf [hereinafter USAF FLIGHT PLAN].

[22] U.S. JOINT FORCES COMMAND, UNMANNED EFFECTS (UFX): TAKING THE HUMAN OUT OF THE LOOP, PROJECT ALPHA STUDY, RAPID ASSESSMENT PROCESS (RAP) REPORT #03-10 4 (2003), *available at* http://edocs.nps.edu/dodpubs/org/JFC/RAPno.03-10.pdf [hereinafter PROJECT ALPHA STUDY]. *See also* USAF FLIGHT PLAN, *supra* note 21, at 41 (2009) (contending that "advances in [artificial intelligence] will enable systems to make combat decisions and act within legal and policy constraints without necessarily requiring human input. Authorizing a machine to make lethal combat decisions is contingent upon political and military leaders resolving legal and ethical questions.").

[23] Some sources estimate that it takes at least fifty-five personnel to operate a Predator. DEFENSE UPDATE, *RQ-1A/MQ-1 Predator UAV*, http://defense-update.com/products/p/predator.htm (last visited Feb. 23, 2013).

Second, unmanned systems tethered to a human operator are vulnerable to satellite communications jamming and cyber-attack; they cannot operate once the communications link with the operator has been severed.[24] Both disruptive capabilities are rapidly improving.[25] Third, a tethered system might prove unacceptably slow. Many nations, including China, are already developing advanced systems with autonomous features.[26] Future combat may therefore occur at such a high tempo that human operators will simply be unable to keep up.[27] Indeed, advanced weapon systems may well "create an environment too complex for humans to direct."[28] A force that does not employ fully autonomous weapon systems will inevitably operate outside its enemy's "OODA loop," thereby ceding initiative on the battlefield.[29] Facing such realities, some DoD studies have

---

[24] *See* JAN VAN TOL ET AL., AIRSEA BATTLE: A POINT-OF-DEPARTURE OPERATIONAL CONCEPT, CENTER FOR STRATEGIC AND BUDGETARY ASSESSMENTS 33–34 (2010), *available at* http://www.csbaonline.org/wp-content/uploads/2010/05/2010.05.18-AirSea-Battle.pdf.

[25] ARMIN KRISHNAN, KILLER ROBOTS: LEGALITY AND ETHICALITY OF AUTONOMOUS WEAPONS 38–39 (2009). Jamming and cyber-attacks may also disrupt or manipulate communications between an unmanned system and GPS navigation satellites. To ensure reliability both as to location and target, therefore, unmanned aerial vehicles (UAVs) should also be equipped with inertial navigation systems.

[26] Darren Stewart, *New Technology and the Law of Armed Conflict: Technological Meteorites and Legal Dinosaurs?*, 87 NAVAL WAR C. INT'L L. STUD. 272, 276–77, 281 (2011), *available at* https://www.usnwc.edu/getattachment/a276a9eb-38b1-47d4-93f9-bd03ca86a5c6/New-Technology-and-the-Law-of-Armed-Conflict.aspx. For example, South Korea has begun development of fully autonomous lethal unmanned ground systems to patrol the demilitarized zone. RONALD C. ARKIN, GOVERNING LETHAL BEHAVIOR IN AUTONOMOUS ROBOTS 10 (2009). The Israelis have similarly developed and deployed an autonomous unmanned ground vehicle to patrol its borders. Brendan Gogarty & Meredith Hagger, *The Laws of Man over Vehicles Unmanned: The Legal Response to Robotic Revolution on Sea, Land and Air*, 19 J. L. INFO. & SCI. 73, 90–91 (2008). Perhaps even more disconcerting is the ease with which terrorist or other dangerous organizations might obtain such technology. For example, Hezbollah reportedly used makeshift armed UAVs against Israel in 2006. ARKIN, *supra* note 26, at 44.

[27] "[A]rmed robots are set to change the pace of battle dramatically in the coming decade. It may not be militarily advantageous to keep a human in control of targeting." Noel Sharkey, *Drones Proliferation and Protection of Civilians*, *in* INTERNATIONAL HUMANITARIAN LAW AND NEW WEAPON TECHNOLOGIES 108, 110 (Wolff Heintschel von Heinegg ed., 2012), *available at* http://www.iihl.org/iihl/documents/IHL%20and%20new%20weapon%20technologies_Sanremo%20(2).pdf. *See also* Stewart, *supra* note 26, at 275.

[28] P. W. SINGER, WIRED FOR WAR 128 (2009) (quoting Thomas Adams, Colonel (Ret.), U.S. Army).

[29] The "OODA loop" refers to the "Observe, Orient, Decide, Act" cycle. The key to the concept is completing the cycle faster than the enemy such that the enemy is always in a

recommended "aggressively" incorporating autonomy into future systems.[30] They have even suggested that autonomous weapons may become the norm on the battlefield in a generation.[31]

A host of autonomous features are presently the subject of research and development efforts. For instance, the U.S. Navy has successfully begun testing its X-47B aircraft, which is designed to take off and land autonomously on an aircraft carrier.[32] The U.S. Army and Marines have developed a version of the so-called K-Max helicopters capable of flying autonomously along a directed route. The helicopters have already flown autonomously in Afghanistan to deliver cargo between forward operating bases.[33] The British military is designing a supersonic attack aircraft, known as the Taranis, which will fly autonomously but not engage targets without human operator approval.[34] While these systems are not yet designed to autonomously attack an enemy, it is not difficult to imagine how such technology could be adjusted to support fully autonomous targeting.

The computing capabilities of future autonomous weapon systems will be exponentially faster and more powerful than those of contemporary military systems and the computers they employ will be physically much smaller. Autonomous systems will also be equipped with advanced general (or strong) artificial intelligence applications. Instead of merely making choices in order to complete specific and defined tasks, general artificial intelligence systems will exhibit human-like cognitive abilities, enabling them to make decisions in response to complex problems and situations.[35]

---

reactive mode and cannot seize the initiative. "[By] 2047 technology will be able to reduce the time to complete the OODA loop to micro or nanoseconds." USAF FLIGHT PLAN, *supra* note 21, at 16.

[30] U.S. DEP'T OF DEF. SCIENCE BD., TASK FORCE REPORT: THE ROLE OF AUTONOMY IN DOD SYSTEMS 1 (July 2012) [hereinafter DSB REPORT], *available at* http://www.fas.org/irp/agency/dod/dsb/autonomy.pdf.

[31] PROJECT ALPHA STUDY, *supra* note 22, at 4. *See also* SINGER, *supra* note 28, at 128.

[32] Jonathan Skilings, *Unmanned X-47B Aircraft Completes Sea Trial*, CNET (Dec. 20, 2012), http://news.cnet.com/8301-11386_3-57560226-76/unmanned-x-47b-aircraft-completes-sea-trial/.

[33] *Autonomous Helicopters: Robocopter Arrives*, THE ECONOMIST, Sept. 15, 2012, http://www.economist.com/node/21562897.

[34] Ryan Gallagher, *Military Moves Closer to Truly Autonomous Drones*, SLATE (Jan. 16, 2013), http://www.slate.com/blogs/future_tense/2013/01/16/taranis_neuron_militaries_moving_closer_to_truly_autonmoous_drones.html.

[35] Human-like cognitive abilities are not the equivalent of human abilities. No consensus exists as to if and when general artificial intelligence shall be available. Noel Sharkey, for

The systems will adapt and learn by observing their environment and their interaction with it.[36] In fact, the President of the International Committee for the Red Cross (ICRC) has posed the prospect of an autonomous system that might "be programmed to behave more ethically and far more cautiously on the battlefield than a human being."[37]

It is clearly a mistake to assume that autonomous weapons systems will resemble contemporary remotely piloted systems. To the contrary, some will represent a radical departure in form and possess almost unimaginable increases in capability. The Defense Advanced Research Projects Agency (DARPA), for instance, is designing an anti-submarine warfare continuous trail unmanned vessel (ACTUV) able to stay at sea autonomously for up to 90 days while it finds, tracks, and attacks enemy submarines.[38] And some autonomous attacks will be mounted by large numbers of small, expendable systems that engage enemy targets collaboratively as part of a swarm. Swarm technology holds tremendous potential for rapidly attacking and overwhelming an enemy.[39]

---

example, doubts that advances in artificial intelligence will achieve human-like abilities in the next fifteen years. Noel Sharkey, *Automating Warfare: Lessons Learned from the Drones,* 21 J.L. INFO. & SCI. 140 (2011).

[36] SINGER, *supra* note 28, at 74; Jacob Kellenberger, *International Humanitarian Law and New Weapon Technologies*, Keynote Address at the 34th Round Table on Current Issues of International Humanitarian Law in San Remo, September 8, 2011, at 5, http://iihl.org/iihl/Documents/JKBSan%20Remo%20Speech.pdf.

[37] Kellenberger, *supra* note 36, at 5.

[38] *DARPA's Anti Submarine Warfare Game Goes Live*, DARPA, Apr. 4, 2011, http://www.darpa.mil/NewsEvents/Releases/2011/2011/04/04_DARPA's_Anti-Submarine_Warfare_game_goes_live.aspx; Spencer Ackerman, *Here's How Darpa's Robot Ship Will Hunt Silent Subs*, WIRED, Dec. 27, 2012, http://www.wired.com/dangerroom/2012/12/actuv/. Note, however, that, at least initially, the ship is designed to require human operator approval before launching an attack. Similar underwater systems may conduct de-mining operations. Examples include the Underwater Influence Sweep System and the Knifefish, currently being developed by the U.S. Navy. *See* Spencer Ackerman, *Navy Preps to Build a Robot Ship That Blows Up Mines*, WIRED, Jan. 3, 2013, http://www.wired.com/dangerroom/2013/01/robot-mine-sweeper/.

[39] "As autonomy and automation merge, [systems] will be able to swarm . . . creating a focused, relentless, and scaled attack." USAF FLIGHT PLAN, *supra* note 21, at 16. *See also* W.J. Henigan, *Boeing Technology Enables Drones To Swarm Like Insects*, L.A. TIMES, Aug. 8, 2012, http://articles.chicagotribune.com/2012-08-09/business/chi-boeing-technology-allows-drones-to-swarm-like-insects-20120809_1_drones-swarm-scaneagle. The U.S. Air Force's Proliferated Autonomous Weapons (PRAWNs) may represent an early prototype of future swarming systems. SINGER, *supra* note 28, at 232–33. *See also* Philip Alston, *Lethal*

Before turning to the legal issues surrounding autonomous weapon systems, it is necessary to debunk a number of myths about autonomous weapon systems that 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."[40] Autonomous weapon systems will be integrated into human warfare, but are highly unlikely to replace it.[41]

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 limiting the area of operation or nature of the target. As the Defense Science Board points out, "all 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*."[42] At least for the foreseeable future, 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

---

Robotic Technologies: The Implications for Human Rights and International Humanitarian Law, 21 J.L. INFO. & SCI. 35, 43 (2011).

[40] DSB REPORT, *supra* note 30.

[41] As one commentator has perceptively noted, "Just as a knife extends the reach and lethality of a hand, sophisticated weapons like [autonomous weapon systems] can be considered extensions of human action, and the primary difference is the increase in time and distance intervening between the action and result." Vik Kanwar, *Post-Human Humanitarian Law: the Law of War in the Age of Robotic Weapons*, 2 HARV. NAT'L SEC. J. 616, 619–620 (2011).

[42] DSB REPORT, *supra* note 30, at 1–2. DOD DIRECTIVE 3000.09, *supra* note 12, ¶ 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."

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 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 [rules of engagement] . . . ; [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.[43]

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. However, the prospect of them "taking on a life of their own" is a fantastical Hollywood invention.[44]

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

---

[43] DOD DIRECTIVE 3000.09, *supra* note 12, encl. 4, ¶ 8.
[44] Interview with Alan Schultz, U.S. Naval Research Laboratory, Laboratory for Autonomous Systems Research, in Washington, D.C. (Sept. 25, 2012).

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."[45] It should be noted that because the risk of tampering is not limited to autonomous weapon systems, it represents an especially significant vulnerability for advanced militaries.[46]

III. The Law Applicable to Autonomous Weapon Systems

There is universal consensus that the law of armed conflict applies to autonomous weapon systems.[47] Indeed, the legal requirement to conduct a review of new weaponry, discussed below, would be absurd were this not the case. What is contentious is *how* particular norms apply to new systems, such as drones, cyber weapons, and autonomous weapon systems.[48]

In *Losing Humanity*, Human Rights Watch 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."[49]

Such conclusions mistakenly conflate two distinct strains of weapons law. One focuses on the legality of the weapon system itself. To illustrate, 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. Legal reviews of new weaponry are intended to identify those systems running afoul of the law on this basis. A separate family of prohibitions (labelled "conduct of hostilities" rules) bears on the use of weapon systems, not their legality as such. To take a simple example, a rifle is lawful, but may be used unlawfully, as in shooting a civilian or a prisoner of war. Since a weapon must successfully navigate both tracks before it may

---

[45] DOD DIRECTIVE 3000.09, *supra* note 12, encl. 4, ¶ 4a(2).
[46] As an example, cyber means may be used to take control of tethered unmanned systems. Cyber means may even be used to manipulate data that conventional weapons use to geolocate targets.
[47] "There can be no doubt that IHL applies to new weaponry and to the employment in warfare of new technological developments." ICRC REPORT, *supra* note 10, at 36.
[48] For an example, see the analysis of the applicability of various law of armed conflict norms to cyber operations in the TALLINN MANUAL, *supra* note 9, part II.
[49] LOSING HUMANITY, *supra* note 1, at 30.

be lawfully used on a battlefield, it is essential to understand the difference before painting a system unlawful with broad strokes.

A. Weapons Unlawful 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.[50] It first appeared in the regulations annexed to the 1899 Hague Convention II and its 1907 counterpart.[51] Article 35(2) of the 1977 Additional Protocol I to the 1949 Geneva Conventions affirms the prohibition: "It is prohibited to employ weapons, projectiles and material and methods of warfare of a nature to cause superfluous injury or unnecessary suffering."[52] Substantively, the article outlaws those means and methods of warfare that unnecessarily aggravate suffering or injury, that is, which cause suffering or injury serving no military purpose. Article 35(2) applies only to harm experienced by combatants; other provisions of the law of armed conflict protect civilians.[53] The article irrefutably reflects customary international law and, therefore, the norm binds even States that are not Party to the

---

[50] *See generally* YORAM DINSTEIN, THE CONDUCT OF HOSTILITIES UNDER THE LAW OF INTERNATIONAL ARMED CONFLICT 63–67 (2d ed., 2010); WILLIAM H. BOOTHBY, WEAPONS AND THE LAW OF ARMED CONFLICT, ch. 5 (2009).

[51] 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.

[52] Additional Protocol I, *supra* note 8, art. 35(2).

[53] For the purposes of this Article, the term "combatants" includes civilians who are directly participating in hostilities since they are subject to attack for such time as they so participate. Additional Protocol I, *supra* note 8, art. 51(3). On the subject of targetability and direct participation, which is of particular importance in the context of autonomous weapon systems, see INTERNATIONAL COMMITTEE OF THE RED CROSS, INTERPRETIVE GUIDANCE ON THE NOTION OF DIRECT PARTICIPATION IN HOSTILITIES UNDER INTERNATIONAL HUMANITARIAN LAW (Nils Melzer ed., 2009), *available at* , http://www.icrc.org/eng/assets/files/other/icrc-002-0990.pdf [hereinafter INTERPRETIVE GUIDANCE]; 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), *available at* http://harvardnsj.org/2010/05/the-interpretive-guidance-on-the-notion-of-direct-participation-in-hostilities-a-critical-analysis/.

Protocol, such as the United States.[54]

Autonomous systems would not automatically violate the prohibition on unnecessary suffering and superfluous injury because Article 35(2) only 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.[55] 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 pre-emptive ban on the systems.

The law of armed conflict also prohibits weapon systems that cannot be aimed.[56] These weapons are unlawful *per se* in that they are of a nature to strike combatants, military objectives, civilians, and civilian objects without distinction. A norm of customary international law,[57] the companion treaty prohibition appears in Article 51(4)(b) of Additional Protocol I: "Indiscriminate attacks are . . . [t]hose which employ a method or means of combat which cannot be directed at a specific military objective."[58]

---

[54] CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, 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 9, art. 8(2)(b)(xx); COMMANDER'S HANDBOOK, *supra* note 9, ¶ 9.1.1. On Article 35(2), see INTERNATIONAL COMMITTEE OF THE RED CROSS, 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 AP 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 195–198 (1982).

[55] 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 9, Rule 79.

[56] *See generally* BOOTHBY, *supra* note 50, ch. 6.

[57] CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, Rules 7, 12. *See also* Rome Statute, *supra* note 9, art. 8(2)(b)(xx); COMMANDER'S HANDBOOK, *supra* note 9, ¶¶ 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.

[58] AP COMMENTARY, *supra* note 54, ¶¶ 1956–1960.

The ban on weapon systems that are indiscriminate because they cannot be aimed at a lawful target is often confused with the prohibition on the use of discriminate weapons in an indiscriminate manner. 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 employment against troops in open areas such as the desert, and they actually struck very large military installations without seriously endangering the civilian population.[59] 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.

Critics of autonomous weapons claim that "[f]ully autonomous weapons would not have the ability to sense or interpret the difference between soldiers and civilians, especially in contemporary combat environments."[60] But, as with the SCUD, an assertion that a weapon system is unlawful can only pass muster if there are no circumstances in which it can be used without placing civilians at excessive risk. What has been missed in much of the dialogue so far 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 used lawfully in certain environments. Not all battlespaces contain civilians or civilian objects. When they do not, a system devoid of an ability 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 on their lawfulness *per se*.

---

[59] *See* RAND CORPORATION, *Coalition SCUD-Hunting in Iraq, 1991*, *in* SPECIAL OPERATIONS FORCES AND ELUSIVE ENEMY GROUND TARGETS (2001), *available at* http://www.rand.org/content/dam/rand/pubs/monograph_reports/MR1408/MR1408.ch3.pdf; U.S. DEP'T OF DEF., CONDUCT OF THE PERSIAN GULF WAR: FINAL REPORT TO CONGRESS 166–168 (1992), *available at* http://www.ndu.edu/library/epubs/cpgw.pdf.

[60] LOSING HUMANITY, *supra* note 1, at 30.

The claims of the critics are not only counter-normative, they are counter-factual. 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 collect 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 facilitate target engagement when the risk of collateral damage is low. Even software for autonomous weapon systems that enables visual identification of individuals, thereby enhancing accuracy during autonomous "personality strikes" against specified persons, is likely to be developed.[61] These and related technological capabilities auger against characterization of autonomous weapon systems as unlawful *per se* solely based on their autonomous nature.[62]

It must be emphasized that as a matter of law, more may not be asked of autonomous weapon systems than of human-operated systems. For example, some opponents of autonomous weapons contend a ban is necessary because autonomous weapon systems may be deceived, as in the case of "concealing weapons or by exploiting their sensual and behavioural limitations," and thereby have difficulty distinguishing civilians from combatants.[63] Yet, asymmetrically disadvantaged enemies have been feigning civilian or other protected status to avoid being engaged by human-operated weapon systems for centuries.[64] The fact that the techniques sometimes prove successful has never merited classifying those systems as

---

[61] Some opponents of autonomous weapons have a more pessimistic view of the capabilities autonomous weapons possess: "Current sensing apparatus and processing can just about tell us that something resembles a human, but little else." Sharkey, *Automating Warfare, supra* note 35, at 143–44. *See also* Gary E. Marchant et al., *International Governance of Autonomous Military Robots*, 12 COLUM. SCI. & TECH. L. REV. 272, 284–85 (2011).

[62] In particular, these capabilities reveal that Human Rights Watch's concerns that combatants sometimes fail to "wear uniforms or insignia" or are identifiable only through their "direct participation in hostilities" are exaggerated (at least with regard to the issue of distinction by the systems). *See* LOSING HUMANITY, *supra* note 1, at 30.

[63] *Id.* at 31.

[64] *See* Michael N. Schmitt, *Asymmetrical Warfare and International Humanitarian Law,* 62 A.F. L. REV. 1 (2008). The critics might also have cited the use of human shields, a practice that would also complicate autonomous weapon system targeting. *See* Michael N. Schmitt, *Human Shields and International Humanitarian Law*, 47 COLUM. J. TRANSNAT'L L. 292 (2009).

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.

Some critics have also decried the inability of autonomous weapon systems to recognize human intentions.[65] Although improved gesture and intent recognition software for autonomous systems development is underway, the technology remains in its infancy.[66] Yet, while it may be true that human perception of human activity can sometimes enhance identification, human-operated systems already engage targets without the benefit of emotional sensitivity. For example, human-operated "beyond visual range" attacks are commonplace in modern warfare; no serious charge has been levelled that the weapon systems conducting them are unlawful *per se*.[67]

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-15s involved made a close visual pass of the targets before engaging them.[68] Pilot error (and human error aboard the AWACs monitoring the situation) contributed to their misidentification as 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

---

[65] For example, Human Rights Watch contends "fully autonomous weapons would not possess human qualities necessary to assess an individual's intentions." LOSING HUMANITY, *supra* note 1, at 31.

[66] Some autonomous systems, such as those being developed and tested at the U.S. Naval Research Laboratory's new Laboratory for Autonomous System Research, are having initial success in understanding basic human speech, expression, and hand gestures. Jessica L. Tozer, *Robots With Faces*, ARMED WITH SCIENCE (Apr. 5, 2012), http://science.dodlive.mil/2012/04/05/robots-with-faces/.

[67] As an example, the U.S. Navy's AIM-54 air-to-air Phoenix missile has a range in excess of 100 nautical miles. *AIM-54 Phoenix Missile*, U.S. NAVY, http://www.navy.mil/navydata/fact_display.asp?cid=2200&tid=700&ct=2 (last visited Feb. 23, 2013).

[68] U.S. ARMY AIRCRAFT ACCIDENT INVESTIGATION BOARD REPORT, U.S. ARMY UH-60 BLACKHAWK HELICOPTERS 87-26000 AND 88-26060 Vol.1 (Executive Summary) (May 27, 1994), at 3, *available at* http://www.dod.mil/pubs/foi/operation_and_plans/PersianGulfWar/973-1.pdf.

down the aircraft.[69] Such tragedies demonstrate that a human 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.

Critics of autonomous weapon systems sometimes note the unique emotional character of human beings to suggest that "robots would not be restrained by human emotions and the capacity for compassion, which can provide an important check on the killing of civilians."[70] The *Losing Humanity* report even concludes that "[e]motionless robots could, therefore, serve as tools of repressive dictators seeking to crack down on their own people without fear the troops would turn on them."[71] Although emotions can restrain humans, it is equally true that they 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. The International Committee of the Red Cross made precisely this point in a 2011 report on the challenges of contemporary armed conflicts: "After all, emotion, the loss of colleagues and personal self-interest is not an issue for a robot and the record of respect for [the law of armed conflict] by human soldiers is far from perfect, to say the least."[72] An autonomous weapon system's lack of emotions serves as little justification for an outright ban.

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*. By contrast, an autonomous

---

[69] U.S. DEP'T OF DEF., FORMAL INVESTIGATION INTO THE CIRCUMSTANCES SURROUNDING THE DOWNING OF IRAN AIR FLIGHT 655 ON 3 JULY 1988 37, 42–45 (Aug. 19, 1988), *available at* http://www.dod.mil/pubs/foi/International_security_affairs/other/172.pdf. The report concluded that "[s]tress, task fixation, and unconscious distortion of data may have played a major role in this incident." *Id.* at 145. 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.*

[70] LOSING HUMANITY *supra* note 1, at 4

[71] *Id.* at 4.

[72] ICRC REPORT, *supra* note 10, at 40.

weapon system unable to reliably distinguish between civilians and combatants, but planned for use where civilians are not present, would not be unlawful *per se*. However, it would still have to be capable of geographical restriction (either based on system constraints such as maximum range and endurance or on human operator pre-programming) to prevent it from passing into areas where civilians are located. Arguably, autonomous weapon systems must also be capable of temporal limitation since few areas are always completely devoid of civilians or civilian objects.[73]

A second form of prohibition on indiscriminate weapons is codified in Article 51(4)(c) of Additional Protocol I, and reflects customary international law.[74] 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.[75] A biological weapon could not lawfully be mounted on an autonomous platform. Another example is an autonomous weapon 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.[76] In such a case, the autonomous weapon would be unlawful *per se*.

B. Unlawful Use of Weapons

The likelihood of an autonomous weapon system being unlawful *per se* is very low. However, as noted perceptively by the ICRC in a report

---

[73] DOD DIRECTIVE 3000.09, *supra* note 12, ¶ 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."

[74] "Indiscriminate attacks are . . . [t]hose which employ a method or means of combat the effects of which cannot be limited as required by this Protocol." Additional Protocol I, *supra* note 8, art. 51(4)(c). *See also* CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, Rules 12, 71; Rome Statute, *supra* note 9, art. 8(2)(b)(xx); COMMANDER'S HANDBOOK, *supra* note 9, ¶ 5.3.2.

[75] Of course, biological weapons have been rendered unlawful as such, irrespective of Article 51(4)(c), in treaty law. Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on Their Destruction, Apr. 10, 1972, 26 U.S.T. 583, 1015 U.N.T.S. 163; CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, Rule 73.

[76] Injury or physical damage would have to result in the case of the cyber attack. TALLINN MANUAL, *supra* note 9, Rule 43 and accompanying commentary.

addressing, *inter alia*, autonomous weapons, "the crucial question does not seem to be whether new technologies are good or bad in themselves, but instead what are the circumstances of their use."[77] In other words, the law of armed conflict rules that govern the use of weapons are more relevant to the issue of autonomous weapon systems than those rules that limit the nature of weaponry. Any such examination of the use of weapons should naturally begin with the seminal principle of distinction.

1. Distinction

Distinction is one of two principles in the law of armed conflict recognized as "cardinal" by the International Court of Justice, which has also characterized it as "intransgressible."[78] The principle of distinction serves as the fount for the law of armed conflict 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 hostilities. Article 48 of Additional Protocol I codifies this customary law principle: "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."[79] The principle incontrovertibly applies to autonomous weapon systems.[80]

Distinction is operationalized in a number of rules, the two most fundamental being the customary law prohibitions on making civilians[81] and civilian objects[82] the target of attacks. They are codified, respectively, in Articles 51(2) and 52(1) of Additional Protocol I:

---

[77] ICRC REPORT, *supra* note 10, at 40.
[78] Nuclear Weapons, *supra* note 54, ¶¶ 78–79.
[79] Additional Protocol I, *supra* note 8, art. 48, at 25. *See also* CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, Rule 1 (explaining its customary law nature).
[80] A distinguished group of law of armed conflict experts examining the international law applicable to air and missile warfare specifically affirmed that the principle of distinction applied to autonomous weapon systems: "[An autonomous weapons system's] sensors and computer programs must be able to distinguish between military objectives and civilian objects, as well as between civilians and combatants." AMW MANUAL COMMENTARY, *supra* note 9, Rule 17a, ¶ 3.
[81] *See* CUSTOMARY INTERNATIONAL HUMANITARIAN Law, *supra* note 9, Rules 1, 6; Rome Statute, *supra* note 9, art. 8(2)(b)(i); COMMANDER'S HANDBOOK, *supra* note 9, ¶ 8.3.
[82] *See* CUSTOMARY INTERNATIONAL HUMANITARIAN LAW, *supra* note 9, Rules 7, 9, 10; Rome Statute, *supra* note 9, art. 8(2)(b)(ii); COMMANDER'S HANDBOOK, *supra* note 9, ¶ 8.3.

> Article 51(2): The civilian population as such, as well as individual civilians, shall not be the object of attack. Acts or threats of violence the primary purpose of which is to spread terror among the civilian population are prohibited.[83]
>
> Article 52(1): Civilian objects shall not be the object of attack or of reprisals.[84]

Obviously, it would be unlawful to use an autonomous weapon system to directly attack civilians or civilian objects or to intentionally terrorize the civilian population. 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.[85] 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,[86] as does the controversy over whether war-sustaining objects qualify as military objectives.[87] As these issues are common to all weapon systems, they will not be addressed here.[88]

---

[83] Additional Protocol I, *supra* note 8, art. 51(2), at 26. *See also* AP COMMENTARY, *supra* note 54, ¶¶ 1938–41; BOTHE, *supra* note 54, at 300–01.

[84] Additional Protocol I, *supra* note 8, art. 52(1), at 27. *See also* AP COMMENTARY, *supra* note 54, ¶¶ 2011–13; BOTHE *supra* note 54, at 322–23.

[85] *See* Additional Protocol I, *supra* note 8, art. 51(3), at 26; CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, Rule 6; INTERPRETIVE GUIDANCE, *supra* note 53.

[86] "In so far as objects are concerned, military objectives are limited to those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military of advantage." Additional Protocol I, *supra* note 8, art. 52(2), at 27.

[87] Although the United States accepts the definition of military objectives set forth in 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. *See* COMMANDER'S HANDBOOK, *supra* note 9, ¶¶ 8.2, 8.2.5. *See also* TALLINN MANUAL, *supra* note 9, commentary accompanying Rule 38 (discussing the controversy); Michael N. Schmitt, *Targeting in Operational Law*, *in* THE HANDBOOK OF THE INTERNATIONAL LAW OF MILITARY OPERATIONS 245, 254 (Terry D. Gill & Dieter Fleck eds., 2010) (same).

[88] For application to another contemporary category of weapons, *see* TALLINN MANUAL, *supra* note 9, chs. 4–5 and accompanying commentary.

Article 51(4)(a), which reflects customary international law, sets forth a further prohibition that is particularly relevant to autonomous systems. By the article, attacks that are not directed at a specific target, and, as a result, are of a nature to strike lawful targets and civilians or civilian objects without distinction, are banned.[89] Unlike the ban on indiscriminate weapons, this prohibition involves a weapon that is capable of being aimed at a lawful target, but the attacker does not do so.[90] As an example, it would clearly be unlawful to employ a programmable autonomous weapon system capable of distinguishing military from civilian aircraft without activating the relevant sensors in an environment where civilian aircraft might be present. Reduced to basics, the prohibition on indiscriminate use of a weapon capable of discrimination requires the use of any sensors associated with an autonomous weapon system capable of enhancing the ability to distinguish in a situation where civilians may be present. Of course, use of the sensors would not alone render the attack lawful. Even a discriminate attack will be unlawful if it is in violation of the rule of proportionality or the requirement to take precautions in attack.

2. Proportionality

An important element of the principle of distinction is the rule of proportionality. This customary law rule, codified in Articles 51(5)(b) and 57(2)(a)(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."[91] It is among the most complex and misunderstood norms in the law of armed conflict with respect to both interpretation and application.[92] While the

---

[89] CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, Rules 11–12. *See also* COMMANDER'S HANDBOOK, *supra* note 9, ¶ 5.3.2. On Article 51(4)(a), *see* AP COMMENTARY, *supra* note 54, ¶¶ 1951–1955.

[90] As noted by Yoram Dinstein, "[t]he key to finding that a certain attack has been indiscriminate is the nonchalant state of mind of the attacker." DINSTEIN, *supra* note 50, at 127.

[91] Additional Protocol I, *supra* note 8, arts. 51(5)(b) & 57(2)(a)(iii), at 26, 29. As to its customary nature, see CUSTOMARY INTERNATIONAL LAW STUDY, *supra* note 9, Rule 14; COMMANDER'S HANDBOOK, *supra* note 9, ¶ 5.3.3.

[92] To take one common example, the collateral damage caused during an attack, or the failure to achieve an attack's military aim, are often considered 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

discussion that follows is limited to those aspects of proportionality having particular relevance to the issue of autonomy, it must be cautioned that other issues – such as whether the survival of the weapon system conducting the attack qualifies as "military advantage" – that have plagued application of the rule in other contexts are no less applicable with respect to autonomous weapon systems.[93]

At the core of the rule of proportionality lies the notion of "excessiveness." The law of armed conflict contains no accepted definition of the term excessive. However, excessiveness "is not a matter of counting civilian casualties and comparing them to the number of enemy combatants that have been put out of action."[94] Rather, it is the product of a case-by-case assessment that is evaluated in terms of its reasonableness given the attendant circumstances. All that can be said of excessiveness in terms of quantification is that the greater the reasonably anticipated military advantage likely to accrue from an attack, the more the law of armed conflict will tolerate the expected collateral damage.

This begs the question of whether autonomous weapon systems are capable of performing proportionality calculations. Such calculations require consideration of both expected collateral damage and anticipated military advantage. An effective 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

---

reasonably expects to cause five incidental deaths, but the strike unpredictably causes 15, 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).

[93] Although the authors disagree, some commentators suggest that the security of one's own forces and "avoiding the placement of its military forces unnecessarily in harm's way" may be part of the military advantage component of proportionality, at least with respect to the proportionality provisions of Additional Protocol I, art. 57(2)(a)(iii). Jonathan David Herbach, *Into the Caves of Steel: Precaution, Cognition and Robotic Weapon Systems Under the International Law of Armed Conflict*, 4(3) AMSTERDAM L. F. 3, 7–8 (2012), *available at* http://ssrn.com/abstract=2154466. The use of autonomous systems may well be "the most effective way of ensuring security of the attacking forces." *Id.* Including that factor in the proportionality calculation could possibly lead to an increased use of autonomous weapons. The authors take the more generally accepted position that force protection is an issue to be considered in the context of precautions in attack, discussed below.

[94] AMW MANUAL COMMENTARY, *supra* note 9, Rule 14, ¶ 7.

near the target, and the composition of structures to estimate the number of civilian casualties likely to be caused during an attack.[95] 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 since the analysis is performed using objective data and scientific algorithms.

Yet, the analysis would not resolve whether a particular attack complies with the rule of proportionality because it is necessary to consider expected collateral damage in light of the military advantage anticipated to result from the attack. Instead, CDEM is merely a policy-related instrument used to determine the level of command at which an attack harming civilians must be authorized; the greater the likelihood of such harm according to the CDEM calculations, the higher the required approval authority. A commander with authority to authorize the attack must still make the proportionality determination as part of the attack's approval process. It is this individual who evaluates, and factors in, the anticipated military advantage.[96] In an attack that is not pre-planned, it is often the individual conducting or controlling the attack, such as a pilot or a supported ground force commander, who performs the proportionality calculation by assessing risk to civilians and civilian objects and likely military advantage.

The question is whether an autonomous weapon system would be capable of gauging military advantage. Doing so will be challenging because military advantage determinations are always 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

---

[95] For a discussion of the methodology, see Jeffrey Thurnher & Timothy Kelly, *Panel Discussion: Collateral Damage Estimation*, YOUTUBE (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 Estimation Methodology (CDM)* (Nov. 10, 2009), *available at* http://www.aclu.org/files/dronefoia/dod/drone_dod_ACLU_DRONES_JOINT_STAFF_SLIDES_1-47.pdf.

[96] 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 9, art. 33.

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.

Although they have not yet been developed, "military advantage" algorithms could theoretically 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. An autonomous weapon system could be programmed with a base maximum collateral damage level of X for a tank; a human would have to have already made the determination that X generally comports with the proportionality rule. Although determining the appropriate threshold would be a very subjective endeavour, as noted in 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."[97] Therefore, the "quantification" of the military value of a target in terms of collateral damage does not require surgical precision. Rather, the test is one of reasonableness.

Because military advantage is such a context specific value, compliance with the rule of proportionality would require that the base maximum collateral damage threshold either be very conservative or be adjustable based on the engagement context. For instance, with regard to the former, valuing an enemy tank at one civilian would generally be reasonable during high order conflict. To the extent destruction of a tank did not justify any civilian casualties in light of the intensity of the conflict at a particular point in time, the commander would simply decide not to employ the "non-adjustable" autonomous weapon system.

Being able to adjust values would provide much greater flexibility since autonomous weapon systems could be programmed prior to launch based on the current situation or even reprogrammed remotely while it is hunting for targets should the situation change. As the technology advances, algorithms that would permit the autonomous weapon system to itself adjust the base level threshold to account for specified variables it encountered on a mission will likely be developed. As an example, it would be reasonable to

[97] AP COMMENTARY, *supra* note 54, ¶ 2208; LOSING HUMANITY, *supra* note 1, at 33.

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 battlefront.

Given the complexity and fluidity of the modern battlespace, it is unlikely that, despite impressive advances in artificial intelligence, "machines" will soon be programmable to perform robust assessments of a strike's likely military advantage on their own. 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."[98] But, again, it is inappropriate to ask more of machines than the humans whom the law of proportionality was originally designed to address. While autonomous weapon systems would likely not be able to account for all imaginable scenarios and variables that might present themselves during hostilities, the same is true of a human confronted with unexpected or confusing events who must nonetheless make a time sensitive decision in combat. Neither the human nor the machine is held to a standard of perfection; in the law of armed conflict the standard is always one of reasonableness.

For the immediate future, though, the actual proportionality decision will continue to be made by humans—by deciding to launch the system into a particular environment, by deciding on how to preprogram the system, or by revising the engagement criteria remotely. These humans remain fully responsible for compliance with the rule of proportionality. They must determine whether the ensuing attacks are likely to cause excessive collateral damage by considering both the system's capabilities and the environment in which it will operate. For example, the operator will have violated the rule if he or she approved use of an autonomous system with weapons that were insufficiently precise to be used in a particular setting, such as a city, and, as a result, harm to civilians and civilian objects was reasonably likely to be excessive to the anticipated military gains.

---

[98] LOSING HUMANITY, *supra* note 1, at 33.

3. Military Necessity

Critics sometimes rely on the principle of "military necessity" as a basis for finding autonomous weapon systems unlawful. Human Rights Watch cites the example of a system continuing to attack an individual who is no longer fighting.[99] In the authors' view, these assertions mischaracterize military necessity as a distinct rule of the law of armed conflict, rather than a foundational principle that undergirds the entire body of law.[100] Moreover, the very example cited by the organization has already sparked a virulent debate in the context of the ICRC's *Interpretive Guidance on the Notion of Direct Participation in Hostilities*.[101] In light of that controversy, it is hardly an effective example to cite.

Even military necessity as understood by the critics would not render autonomous weapon systems unlawful. They would not be unlawful *per se* because there are clearly situations in which they are valuable militarily. In other words, autonomous weapon systems are not superfluous, if for no other reason than the fact that unlike manned systems they can attack the enemy without placing an operator at risk. As to prohibitions based on use, the requirement that military objectives yield some military advantage would make any separate condition for military necessity redundant.[102] With regard to situations raising proportionality issues, any strike lacking military advantage but causing harm to civilians or civilian objects would violate the rule.[103] And as to situations such as those cited by Human Rights Watch, the law of armed conflict already prohibits attacks on those who

---

[99] *Id.* at 34–35. Military necessity was originally described in the "Lieber Code": "the necessity of those measures which are indispensible for securing the ends of the war, and which are lawful according to the modern laws 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.

[100] *See* Schmitt, *supra* note 5.

[101] The example was a situation where an enemy soldier could be captured instead of killed. INTERPRETIVE GUIDANCE, *supra* note 53, at 78–82. For a contrary view, *see* W. Hays Parks, *Part IX of the ICRC "Direct Participation in Hostilities" Study: No Mandate, No Expertise, and Legally Incorrect,* 42 N.Y.U. J. INT'L L. & POL. 769 (2010). Those adopting the contrary view pointed out that the individual could avoid being killed by the simple act of surrendering, and that the law of armed conflict prohibition on killing a surrendering soldier was the appropriate norm to apply in such cases.

[102] Additional Protocol I, *supra* note 8, art. 52(2).

[103] *Id.*, arts. 51(5)(b), 57(2)(iii).

have surrendered or are otherwise *hors de combat.*[104] Taking these observations together, the result is that military necessity has little or no independent valence when assessing the legality of autonomous weapon systems or their use.

4. Feasible Precautions in Attack

Another central component of the principle of distinction is the law of armed conflict requirement that an attacker take precautions in attack to minimize harm to civilians and civilian objects.[105] 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."[106] The 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

---

[104] *Id.,* art. 41; CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, Rule 47.

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

[106] Additional Protocol I, *supra* note 9, art. 57(1). *See also* CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, Rule 15; COMMANDER'S HANDBOOK, *supra* note 9, ¶ 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 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.

view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects."[107]

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 capable of boosting the reliability of target identification. In fact, some scholars have suggested that sensors and targeting programs for autonomous weapon systems must at least be "comparable" to that of manned or remotely piloted aircraft.[108] Furthermore, 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 humanitarian and military considerations."[109] Military considerations

---

[107] Additional Protocol I, *supra* note 9, art. 57 (2)–(3); *see also* CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, Rules 16–21. On Article 57, *see* AP COMMENTARY, *supra* note 54, ¶¶ 2184–2238; BOTHE, *supra* note 54, at 359–369. One commentator has curiously opined that "the principle of precaution would dictate that belligerents not field autonomous combat robots as a means of warfare because the prospects for avoidance or minimisation of incidental injury to or loss of civilian life or damage to civilian objects could simply not be gauged in the feasibility calculation." Herbach, *supra* note 93, at 19. The authors agree with neither his analysis of the facts nor the law in this regard.

[108] AMW MANUAL COMMENTARY, *supra* note 9, Rule 39, ¶ 4.

[109] 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, art. 3(10), 2048 U.N.T.S. 133; U.K. Statement made upon Ratification of Additional Protocols I and II, at ¶ (b), *reprinted in* DOCUMENTS ON THE LAW OF WAR 510 (Adam Roberts & Richard Guelff eds., 3rd ed. 2000) [hereinafter Ratification Statement]. *See also* COMMANDER'S HANDBOOK, *supra* note 9, ¶ 8.3.1; CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, commentary accompanying Rule 15; AMW MANUAL COMMENTARY, *supra* note 9, Rule 1(q).

include both technical and operational factors, in particular the survivability of the weapon system and any associated systems or personnel involved in its employment. Feasibility is an issue of reasonableness. The law of armed conflict would only require an attacker to assume greater risk to avoid collateral damage if a reasonable attacker in the same or similar circumstances would do so. Thus, in the previous example, if sending an unmanned aerial system into the area of operations could place it at a degree of risk not justified by the extent of enhanced identification capability, then its use would not be feasible. This might be because the aerial system is needed for operations elsewhere against targets of greater military value or because its use in other operations may have a greater prospect for the avoidance of civilian casualties.

A requirement to select among military objectives to minimize civilian casualties and damage to civilian objects likewise applies to autonomous weapon systems. As an example, 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 weaponry 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 a human controlled system, and using the human controlled system will not lower the probability of mission success or pose 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 readily 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 the competing systems' use. Nevertheless, the norm 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* risk of collateral damage than a human controlled system. For example, an autonomous weapon system could be armed with non-lethal weapons unavailable on available manned systems, its sensor suite could be more precise or discriminate than available manned systems, or its decision-making capability could be better than that of a human in a particular environment (such as a very dangerous or rapidly changing 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 because such systems had been banned. Therefore, the prohibition 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.

5. Doubt

As with human identification, autonomous weapon systems may be unable to identify a particular person or object as a combatant or civilian, or military objective or civilian object, with absolute certainty. For instance, an autonomous weapon system may lack the cognitive or sensor capability to positively identify a target in the particular circumstances, as might be the case with its employment in a cluttered battlespace or an electronic warfare intense environment. Determining how the systems must handle such uncertainty is essential in light of the prohibition on attacking civilians and civilian objects, the protection afforded civilians and civilian objects by the rule of proportionality, and the requirement to take precautions in attack to minimize harm to civilians and civilian objects.

Law of armed conflict norms as to doubt apply fully in the autonomous weapons context.[110] During an attack, doubt as to status of a person must be resolved in favor of treating that individual as a civilian. Article 50(1) of Additional Protocol I codifies this presumption, which is

---

[110] "The standards . . . regarding doubt apply equally to . . . autonomous or manned [attacks]." AMW MANUAL COMMENTARY, *supra* note 9, Rule 39, ¶ 5.

generally characterized as customary in nature.[111] The presumption applies both to doubt regarding the status of a targeted individual and as to whether a person is to be considered a civilian in making proportionality calculations and taking feasible precautions in attack.

Mere existence of some doubt does not bring the presumption into operation.[112] 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,"[113] 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. However, as a rule of thumb, an autonomous weapon system should not be permitted to engage a target in situations in which a reasonable attacker would elect not to attack because of the degree of uncertainty. In other words, if all information available to the autonomous weapon system from its sensors and other sources were available to a reasonable human attacker, would he or she engage the target or not?

Obviously, development of algorithms that can both precisely meter doubt and reliably factor in the particular situation in which the autonomous weapon system is being operated will prove highly challenging. Nevertheless, algorithms could theoretically attribute 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). For instance, autonomous weapon systems could be equipped with sensors that enable them to determine when a target is possibly a child (for example, based on size). Such a determination would substantially decrease the probability that

---

[111] "In case of doubt whether a person is a civilian, that person shall be considered to be a civilian." Additional Protocol I, *supra* note 9, art. 50(1). *See also* CUSTOMARY INTERNATIONAL HUMANITARIAN LAW STUDY, *supra* note 9, commentary accompanying Rule 6; AP COMMENTARY, *supra* note 54, ¶¶ 1911–1921; BOTHE, *supra* note 54, at 295–296.

[112] TALLINN MANUAL, *supra* note 9, commentary accompanying Rule 33.

[113] U.K. MINISTRY OF DEFENCE, THE JOINT SERVICE MANUAL OF THE LAW OF ARMED CONFLICT, JSP 383 (2004), ¶ 5.3.4. *See also* Ratification Statement, *supra* note 109, ¶ (h).

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 would likely be much more advanced.

Even if values can be attributable to such variables, it will still be necessary to set the doubt threshold at which an autonomous weapon system will refrain from attack. In a sense, doing so will resemble programming autonomous weapon systems to refrain from attack because of the risk of violating the proportionality rule. Although it is challenging, it is possible to envision how this might be done. For instance, the system could be programmed using the doubt values just discussed; it would not attack if a particular level of doubt were reached. The system could not be used in situations in which the pre-programmed threshold would be "unreasonable." For instance, more doubt can be countenanced on a "hot" battlefield than in a relatively benign environment. In light of the pre-set doubt values, it might only be reasonable to employ the system in the former situation.

An autonomous weapon system may also have adjustable doubt thresholds that can be set before launch to account for the circumstances in which it will be employed (for example, for use in an area where enemy forces have been highly active as distinct from one where they have not). Advanced autonomous weapon systems of the future will possibly be capable of being adjusted remotely while operations are underway to account for unexpected developments, such as intelligence reports of displaced civilians in the area. Systems may even be developed that can adapt *sua sponte* to pre-programmed variables, such as the movement of vehicles towards or away from friendly forces. Greater doubt would be acceptable in the former case.

Autonomous weapon systems arguably possess advantages over humans with respect to doubt. As with other unmanned systems, they are not constrained by the notion of self-preservation. Therefore, the systems could, in some conceivable circumstances, be programmed to either hold their fire until being fired upon or essentially sacrifice themselves to "reveal the presence of a combatant."[114] Such tactics could greatly reduce the

[114] ARKIN, *supra* note 26, at 46.

degree of doubt before striking a target. Similarly, consider automatic "hack-back" measures. When an intrusion into the autonomous weapon system's computers is noted, a clear warning could be issued. If the intrusion continues after being afforded an opportunity to desist, there is a higher likelihood that the intrusion has been intentionally conducted by a combatant or a civilian who is directly participating in hostilities, thereby making it more reasonable for the autonomous weapon system to mount an automatic hack-back attack on the originator.[115]

No equivalent presumption regarding doubt exists in the *lex scripta* for objects. Still, it is clear that an attack based on an unreasonable conclusion that an object is a military objective violates the law of armed conflict. Additionally, 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."[116] Some difference of opinion exists over whether this presumption reflects customary law.[117] In light of that disagreement, one key commentary on the law of armed conflict takes the admittedly tautological position that in case of doubt such an object may only be attacked "following a careful assessment."[118] This is a distinction without a difference. Reduced to basics, as a general rule the mode of analysis described above for persons would apply equally to objects.

6. Subjectivity

A few scholars have questioned whether an autonomous weapon system must be capable of making subjective determinations in order to be lawful.[119] Subjectivity permeates application of the prescriptive norms

---

[115] TALLINN MANUAL, *supra* note 9, Rule 66 and accompanying commentary.
[116] On Article 52(3), *see* AP COMMENTARY, *supra* note 54, ¶¶ 2029–2037; BOTHE, *supra* note 54, at 326–327.
[117] *See* CUSTOMARY INTERNATIONAL LAW STUDY, *supra* note 9, commentary accompanying Rule 10.
[118] TALLINN MANUAL, *supra* note 9, Rule 40 and accompanying commentary.
[119] Tony Gillespie & Robin West, *Requirements for Autonomous Unmanned Air Systems Set by Legal Issues*, 4(2) INT'L C2 J. 1, 13 (2010); Herbach, *supra* note 93, at 17; Markus Wagner, *The Dehumanization of International Humanitarian Law: Legal, Ethical and Political Implications of Autonomous Weapon Systems* (April 22, 2012), http://robots.law.miami.edu/wp-content/uploads/2012/01/Wagner_Dehumanization_of_international_humanitarian_law.pdf.

discussed above. For instance, proportionality, calculations require subjective determinations as to the "value" of the anticipated military advantage to be gained from an attack and the civilian harm expected to be caused, as well as the point at which the latter is excessive in light of the former. Implementing the requirement to take precautions in attack requires subjective determinations as to the feasibility of verification measures and the availability of alternative targets, weapons, and tactics. Ascertaining the degree of doubt regarding the identity of a target or of objects or persons that might become collateral damage is subjective in nature, as is setting the point where sufficient doubt exists to make an engagement unreasonable.

Despite potential advances in artificial intelligence, autonomous weapons systems will be unlikely to be capable of performing such subjective evaluations for the foreseeable future. This reality has led some to conclude that autonomous systems as a class cannot therefore comply with the law of armed conflict.[120] This conclusion is unfounded. In an attack involving autonomous weapon systems, these and other targeting decisions remain subjective and continue to be made by humans. What may differ is the phase in the targeting process when the subjective determinations occur. An autonomous engagement must be conceptualized somewhat differently than one controlled by a human when considering the issue of subjectivity, but it can still meet the requirements of the law.

The first point where human subjectivity enters the process is when the autonomous weapon system is programmed. For non-adjustable systems, this occurs during development. For adjustable systems, it may occur either before launch or remotely as the system is underway. In all these cases, a human has considered relevant variables and made a subjective determination as to a threshold beyond which the system will not attack its target. The human then subjectively assigns numerical or other values to variables that the system's software can assess; qualitative decisions have been translated into quantitative engagement criteria.

---

[120] Markus Wagner states that because of proportionality and distinction concerns, autonomous weapon systems can only be "employed in the rarest of circumstances" and that their "potential deployment is limited to such an extent as to render them useless." Wagner, *supra* note 119, at 5. Jonathan Herbach states that "humans in the loop could be the only way to ensure the basic ability to adhere" to the rules with respect to subjectivity. Herbach, *supra* note 93, at 17.

The objective decisions the autonomous weapon systems make do not replace the subjective ones required under the law. Rather, the objective decisions reached by the autonomous system merely set "shut-down criteria" for when the system is either to disengage or seek additional guidance from a human operator. If, for example, an autonomous weapon system objectively determines that a possible attack will exceed its established parameters for collateral damage, it would not mount the attack. These parameters are more analogous to Rules of Engagement (ROE) with which the autonomous weapon system must comply than to legal thresholds. Like ROE for human operations, policy and operational concerns may result in much more stringent thresholds for engagement than would be allowed by the law of armed conflict.[121] Indeed, thresholds can be expected to be set high if for no other reason than the absence of a human in the loop.

The decisive juncture with respect to subjectivity is when the commander (or authorized officer) orders the autonomous weapon system into battle. He or she must make a subjective decision to do so based upon the attendant circumstances. This requires an understanding not only of the physical capabilities and limitations of the system (the maximum range, the effectiveness of the weaponry, the blast radius of its weapons, etc.), but also the subjective values embedded in it. Ideally, the commander would be able to dial those values up or down as deemed appropriate to the given circumstances. If that is not an option, the commander must subjectively determine whether the autonomous weapon will perform lawfully in the particular battlespace given its preprogramed values.

Ultimately, the reasonableness of the commander's decision to employ the autonomous weapon system is measured against the legal requirements. Many factors will determine whether the decision was reasonably made. How many variables present in the battlespace could the autonomous weapon system evaluate? With what accuracy? How static or dynamic was the environment? Were the programmed thresholds of proportionality and precautions appropriate for the environment? How much doubt was present in the environment? Did the programmed parameters appropriately account for that doubt?

---

[121] Some commentators have called for ethical brakes to be designed into autonomous weapon systems. Michael J. Barnes & A. William Evans III, *Soldier-Robot Teams in Future Battlefields: An Overview*, *in* HUMAN-ROBOT INTERACTIONS IN FUTURE MILITARY OPERATIONS 9, 23 (Michael Barnes & Florian Jentsch eds., 2010).

A particularly critical issue to be considered is time. During the targeting cycle for human operated systems, subjective decisions about the environment and the lawfulness of a lethal strike can sometimes be made up to the moment of impact.[122] For autonomous weapon systems, sensors may continuously feed data about the environment to the system that affects the system's decision to attack based on the programmed subjective parameters. But in most cases the engagement parameters are fixed once the system is launched or, for those systems that can be adjusted remotely during a mission, when the system is no longer capable of having its inputted "subjective values" adjusted.[123] The longer the time since the last human programming, the greater the risk of attack results that the commander did not anticipate and therefore the less reasonable the decision to employ the autonomous weapon system generally becomes. For instance, a system may be programmed in part based on "pattern of life analysis" of the target area.[124] However, patterns of life change over time. This could mean that the programmed parameters are based on out-dated pattern of life analysis at the time the autonomous weapon system conducts an attack.

The summary point regarding subjectivity is that humans, not the autonomous weapon systems, will continue to make the subjective determinations resident in the law of armed conflict, at least absent revolutionary change in artificial intelligence capabilities. Thus, it is somewhat misleading to label the systems "human out of the loop." And it is certainly wrong to assert that the fact that the autonomous weapon system itself cannot make subjective determinations necessarily renders autonomous weapon systems unlawful.

---

[122] For instance, an air-launched AGM-129A has a range of up to 2,000 miles and therefore its employment often involves a significant flight time. U.S. Air Force, AGM-129A Advanced Cruise Missile (Jan. 17, 2011), http://www.af.mil/information/factsheets/factsheet.asp?id=3548.

[123] It is less likely that systems that operate underwater or in areas where communications jamming is prevalent will be able to have their subjective values adjusted during a mission. Commanders who know in advance that the systems will be operating in such environments should recognize that the last opportunity to provide subjective decision making will most likely be the moment of launch.

[124] U.S. ARMY, THE TARGETING PROCESS, FM 3-60, Nov. 26, 2010, at B-3 & B-4.

7. Policy Restrictions

Some states are beginning to set forth autonomous weapon systems guidelines that are meant to foster compliance with the law of armed conflict and achieve 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].[125]

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.[126] This requirement is in addition

---

[125] DOD DIRECTIVE 3000.09, *supra* note 18, ¶ 4c.

[126] DOD DIRECTIVE 3000.09, *supra* note 18, ¶ 5d. 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 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.

to the legal review requirements set forth below. As autonomous weapon system capabilities improve, policy restrictions will evolve in lock step.

C. Legal Review of Autonomous Weapon Systems

Since the prospect of autonomous weapon systems is so new, the requirement to conduct a review of their legality looms large.[127] 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."[128] 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 a swarm of autonomous 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.

While some commentators suggest that a disagreement exists as to whether Article 36 restates customary international law,[129] 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.[130] Consensus is lacking as to whether an analogous requirement exists to perform legal reviews of new *methods* of warfare.[131]

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. The authors are of the opinion that both reviews, whether or not legally mandated, are well advised whenever feasible.

---

[127] On weapons review generally, *see* BOOTHBY, *supra* note 50, at 340–352; W. Hays Parks, *Conventional Weapons and Weapons Reviews,* 8 Y.B. INT'L HUMANITARIAN L. 55 (2005).
[128] On Article 36, *see* AP COMMENTARY, *supra* note 54, ¶¶ 1463–1482; BOTHE, *supra* note 54, at 199–201.
[129] LOSING HUMANITY, *supra* note 1, at 21.
[130] *See*, for example, its treatment as such in AMW MANUAL COMMENTARY, *supra* note 9, Rule 9, and the TALLINN MANUAL, *supra* note 9, Rule 48.
[131] *See* discussion in TALLINN MANUAL, *supra* note 9, commentary accompanying Rule 48.

For States Party to Additional Protocol I, it is clear from the plain text of Article 36 that legal reviews "should take place at the earliest stage possible and continue through any development that proceeds."[132] Because 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, early legal reviews can shape the development stage of a weapon system and resultantly avoid the 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.[133]

Legal reviews are required for both individual weapons and weapons systems. Autonomous weapon systems critics have questioned the United States' commitment to conduct reviews on both. Putting aside individual cases where the United States may or may not have complied with the requirement for legal reviews, 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:

> The 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*.[134]

Modifications of weapon systems can trigger additional requirements to conduct legal reviews. Given that "some robotic technology, while not

---

[132] LOSING HUMANITY, *supra* note 1, at 22.
[133] DOD DIRECTIVE 3000.09, *supra* note 18, 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 WEAPONS (July 9, 1996); and DEP'T OF DEF., DIRECTIVE 2311.01E, DOD LAW OF WAR PROGRAM (May 9, 2006).
[134] DOD DIRECTIVE 5000.01, *supra* note 133, encl. 1, ¶ E1.1 .15. (emphasis added).

inherently harmful, has the potential one day to be weaponized,"[135] this issue is of particular concern. A fair reading of the law of armed conflict 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 [the law of armed conflict], domestic law and international law prior to their acquisition for use in a conflict or other military operation."[136] To summarize, U.S. policy is to review all weapons, their associated delivery system, and any significant modification of them. This policy unquestionably applies to autonomous weapon systems.

The applicable scope of required legal reviews is worth examining closely. The primary intent of the legal review is to determine whether a weapon itself is unlawful under international law. Given the technological advances likely to be embedded in autonomous weapons, this straightforward task may be challenging. Lawyers conducting the reviews will need to work closely with computer scientists and engineers to obtain a better appreciation for the measures of reliability and the testing and validation methods used on the weapons.[137] While significant, these challenges are substantively similar to those facing reviewers of other complicated, modern weapon systems, which are routinely being conducted without fanfare or criticism.

The sole context in a determination of whether the weapon is lawful *per se* is its intended, not possible, use. Legal reviews do not generally consider use issues since they are always contextual by nature.[138] For

---

[135] LOSING HUMANITY, *supra* note 1, at 23.

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

[137] For a discussion of a "multi-disciplinary approach" to weapons reviews of autonomous weapons, see Alan Backstrom & Ian Henderson, *New Capabilities in Warfare: An Overview of Technological Developments and the Associated Legal and Engineering Issues* 28–35, Oct. 22, 2012, http://ssrn.com/abstract=2198826.

[138] The Committee Report on the Article presented to the Diplomatic Conference emphasizes that the rule "is 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

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.[139] 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 was 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.

By way of illustration, the Air Force guidance delineates those legal issues that must be examined when determining whether a weapon system being considered for acquisition is lawful:

> 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.
> 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.[140]

---

employment *for its normal or expected use* would be prohibited under some or all circumstances." BOTHE, *supra* note 54, at 200–201.

[139] BOOTHBY, *supra* note 50, at 79.

[140] AIR FORCE INSTRUCTION 51-402, *supra* note 136, ¶¶ 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)–(c).

The requirements set forth in paragraph 3.1.2 mirror analogous provisions appearing in Additional Protocol I, all of which are customary in nature.[141]

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 law of armed conflict's rules 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.[142] It is incontrovertible that all of the norms discussed apply equally to autonomous weapon systems.

Finally, the interplay between the Martens Clause and legal reviews bears mention. 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.[143] 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.[144] As set forth in the Protocol, it provides: "In cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience." The International Court of Justice recognizes the Martens Clause as customary in nature and has observed that it "proved to be an effective means of addressing rapid evolution of military technology."[145]

By its own terms, though, the clause applies only in the absence of treaty law.[146] 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

---

[141] Additional Protocol I, *supra* note 8, arts. 35(2), 51(4)(b)–(c).
[142] *See* Nuclear Weapons, *supra* note 54, ¶ 86.
[143] LOSING HUMANITY, *supra* note 1, at 25; INTERNATIONAL COMMITTEE OF THE RED CROSS, A GUIDE TO THE LEGAL REVIEW OF NEW WEAPONS, MEANS AND METHODS OF WARFARE ¶ 1.2.2.3 (2006),
[144] 1899 Hague II, *supra* note 51, pmbl; 1907 Hague IV, *supra* note 51, pmbl; Additional Protocol I, *supra* note 8, art. 1(2).
[145] Nuclear Weapons, *supra* note 54, ¶¶ 78, 84.
[146] The text of the clause refers to "cases not covered by this Protocol or by other international agreements." Additional Protocol I, *supra* note 8, art. 1(2).

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.[147] As discussed above, general principles and rules of the law of armed conflict 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 autonomous ones, would not violate applicable treaty and customary law, but nevertheless be unlawful based on the Martens Clause, had become exceptionally low.

Autonomous weapon systems critics are 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. 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.

D. Accountability

Given that fully autonomous weapons represent a somewhat greater distancing of human operators from the battlefield, issues related to determining responsibility and assessing accountability for the activities of these systems are particularly acute.[148] With respect to accountability,

---

[147] Protocol on Prohibitions and Restrictions on the Use of Incendiary Weapons, 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 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 (1997); Convention on Cluster Munitions, Dec. 3, 2008, 48 I.L.M. 357 (2009).

[148] Alston, *supra* note 39, at 51 ("One of the most important issues flowing from increased automation is the question of responsibility for civilian casualties or harm or violations of the laws of war."). *See also* Timothy Coughlin, *The Future of Robotic Weaponry and the Law of*

Human Rights Watch has made the very reasonable point that "[i]f 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."[149]

The problem with this conclusion is that it is based on a false premise.[150] 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.[151] A human must decide how to program the system and when to launch it. 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 the person 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.[152]

Armed Conflict: Irreconcilable Differences?, 17 U. C. LONDON JURISPRUDENCE REV. 67, 86–90 (2011).

[149] LOSING HUMANITY, *supra* note 1, at 42.

[150] On accountability, see generally ROGERS, *supra* note 105, at 360 & ch. 11.

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

[152] *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; Additional Protocol I, *supra* note 8, 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 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 9, arts. 25(3)(b) & 28; Statute of the International Criminal Tribunal for the Former Yugoslavia, 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. Blaški□, Case. No. IT-95-14-T, Trial Chamber Judgment, ¶¶ 281–282 (Int'l Crim. Trib. for the Former Yugoslavia Mar. 3,

Hopefully, autonomous weapon systems will never 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 intermingled 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.[153]

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."[154] The policy imposes identical requirements on Commanders of the U.S. Combatant Commands.[155]

Finally, it is sometimes suggested that individuals who design autonomous weapon systems might be at risk of being held accountable for war crimes committed with those systems. This assertion is spurious, at least to the extent that the system is not specifically designed to commit a war crime. Instead, only individuals who knowingly use the system, and any

---

2000); Prosecutor v. Krstic, Case No. IT-98-33-T, Trial Chamber Judgment, ¶ 605 (Int'l Crim. Trib. for the Former Yugoslavia Aug. 2, 2001); Prosecutor v. Kayishema & Ruzindana, Case No. ICTR 95-1-T, Trial Chamber Judgment, ¶ 223 (Int'l Crim. Trib. for Rwanda May 21, 1999); COMMANDER'S HANDBOOK, *supra* note 9, ¶6.1.3.

[153] The law of armed conflict imposes a duty to investigate possible war crimes. Michael N. Schmitt, *Investigating Violations of International Law in Armed Conflict,* 2 HARV. NAT. SEC. J. 31, 43 (2011).

[154] DOD DIRECTIVE 3000.09, *supra* note 12, ¶ 4b.

[155] *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).

responsible superiors, may be held accountable under the law of armed conflict.

Conclusion

Autonomous technology is poised to revolutionize warfare. Nations will undoubtedly clamour for the weapon systems that this technology will make possible. However, as with prior technological leaps, those intent on using autonomous technology in combat must be mindful of the tenets and teachings of the law of armed conflict—a robust set of laws that exists in a synergistic relationship to warfare.

Skeptical of the willingness of states to comply with this law, critics of autonomous weapon systems have embarked on a campaign to pre-emptively ban lethal autonomous weapons and targeting. Unfortunately, the effort has failed to capture accurately or adequately the significant legal issues at stake. This Article has attempted to refocus the debate. In so doing, four core conclusions emerged.

First, 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*.

Second, the use of autonomous weapon systems in certain circumstances would be lawful under the law of armed conflict, whereas in others it would not. This is true of every weapon which is not unlawful *per se*, from rocks to rockets. Of course, the fact that autonomous weapon systems will 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, the law of armed conflict's restrictions on the use of weapons (particularly the requirements that they be directed only against combatants and military objectives, that they not be employed indiscriminately, that their use not result in excessive harm to civilians or civilian objects, and that 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. After all, as the ICRC suggests, "new technologies do not change existing law, but rather must abide by it."[156]

Third, humans are never really "out of the loop." While autonomous weapon systems will increasingly be capable of solving complex problems, absent dramatic improvements in artificial intelligence, humans will decide when and where to deploy the system and what parameters to embed within it. Human operators, not machines or software, will therefore be making the subjective determinations required under the law of armed conflict, such as those involved in proportionality or precautions in attack calculations. Although the subjective decisions may sometimes have to be made earlier in the targeting cycle than has traditionally been the case, this neither precludes the lawfulness of the decisions, nor represents an impediment to the lawful deployment of the systems. Similarly, the lawfulness of an autonomous weapon system is not undermined simply because the system itself is incapable of making truly subjective decisions.

Finally, humans will always be accountable for the employment of autonomous weapon systems. Although they will gradually delegate more tasks to autonomous systems, the responsibility for the appropriate use of the systems will nevertheless remain with the human operators and commanders. Orders to deploy the system and judgments about how to program it will come from a human. Any recklessness or criminal misuse will result in accountability through the same war crimes mechanisms that already exist under the law of armed conflict.

Assuming these conclusions are correct, calls for a ban on autonomous weapon systems are unlikely to gain much traction with states. Virtually every rule of the law of armed conflict reflects a balancing by States of two seminal factors—military necessity and humanitarian concerns.[157] The result of this dialectic interplay is the law of armed conflict, either in the form of treaty law that has been negotiated by states based on

---

[156] ICRC REPORT, *supra* note 10, at 40.

[157] As noted in the 1868 St. Petersburg Declaration, the law of armed conflict fixes "the technical limits at which the necessities of war ought to yield to the requirements of humanity." St. Petersburg Declaration, *supra* note 51. The humanitarian concerns that are factored into the equation reflect the interest States have in maximizing the law of armed conflict'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.

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

Taking this process of norm formulation into account, achieving an international consensus to prohibit autonomous weapons is highly implausible. While, as illustrated in the new Department of Defense directive on autonomous weapon systems, states are sensitive to the humanitarian implications of the systems, until both their potential for unintended human consequences and their combat capabilities are better understood, it is improbable that any state would seriously consider banning them.[158]

It would 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 the law of armed conflict for suggesting that attacking forces must assume risk. Moreover, banning autonomous weapon systems may, as discussed, have the effect of denying commanders a valuable tool for minimizing the risk to civilians and civilian objects in certain attack scenarios. Until they are better understood, it would be naively premature to draw definitive legal, moral, and operational conclusions as to the use of autonomous weapon systems.

---

[158] There is little historical precedent for banning weaponry before it has been fielded. The only contemporary example 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. *See* Additional Protocol to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be deemed to be Excessively Injurious or to have Indiscriminate Effects, Oct. 13, 1995, 1380 U.N.T.S. 370.