2020

Autonomous Weapons: How Existing Law Can Regulate Future Weapons

Charles P. Trumbull IV

AUTONOMOUS WEAPONS: HOW EXISTING LAW CAN REGULATE FUTURE WEAPONS

Charles P. Trumbull IV[*]

INTRODUCTION

Swarms of miniature drones assassinate U.S. Senators,[1] Terminator-like robots go rogue, and "the Singularity" finally materializes.[2] These could be scenes in a science fiction movie, but they have also been the subject of serious discussion among States, academics, and non-governmental organizations (NGOs). In particular, the potential development and use of autonomous weapons systems—which, broadly defined, are weapons systems that can select and engage targets without further human intervention after activation—have generated significant debate over the past decade.[3]

Inter-governmental discussions of autonomous weapons have occurred primarily at the Convention on Certain Conventional Weapons (CCW). Following the release of a provocative report by Human Rights Watch in 2012,[4] the High Contracting Parties to the CCW convened an informal group of experts to discuss various issues related to "lethal autonomous weapon systems" (LAWS).[5] After three years of meetings, in 2016 the CCW made these

---

[1]   *See, e.g.*, *UC Berkeley Professor Helps Create Viral Video to Warn About Killer Robots,* ABC 7 (Nov. 18, 2017), http://abc7news.com/technology/uc-berkeley-professor-helps-create-viral-video-to-warn-about-killer-robots/2664980/.

[2]   *See, e.g.,* John Markoff, *When is the Singularity? Probably Not in Your Lifetime*, N.Y. TIMES (Apr. 7, 2016), https://www.nytimes.com/2016/04/07/science/artificial-intelligence-when-is-the-singularity.html.

[3]   *See, e.g.*, Philip Alston, *Interim Report of the Special Rapporteur of the Human Rights Council on Extrajudicial, Summary or Arbitrary Executions*, ¶ 20, U.N. Doc. A/65/321 (Aug. 23, 2010) (calling on the international community to "address the legal, political, ethical and moral implications of the development of lethal robotic technologies."); *Autonomous Weapons Are a Game-Changer*, ECONOMIST, (Jan. 25, 2018) ("The fast-approaching revolution in military robotics … poses daunting ethical, legal, policy and practical problems, potentially creating dangers of an entirely new and, some think, existential kind.").

[4]   Bonnie Docherty, *Losing Humanity: The Case Against Killer Robots*, HUMAN RIGHTS WATCH (Nov. 12, 2012), https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots; Philip Alston, *Interim Report of the Special Rapporteur of the Human Rights Council on Extrajudicial, Summary or Arbitrary Executions*, U.N. Doc. A/65/321 (Aug. 23, 2010).

[5]   *See Background on Lethal Autonomous Weapons Systems in the CCW,* UNITED NATIONS: GENEVA, https://www.unog.ch/80256EE600585943/(httpPages)/8FA3C2562A60FF81C1257CE600393DF6?OpenDocument&cntxt=2F594&cookielang=en. (last visited Jan. 29, 2020).

discussions more formal with the establishment of a Group of Governmental Experts (GGE).[6]

The GGE is deeply divided on how to address autonomous weapons.[7] A group of NGOs under the umbrella of the Campaign to Stop Killer Robots, and a growing number of States, have called for a preemptive ban on LAWS or, at a minimum, international regulations on their development and use.[8] This opposition to autonomous weapons is hardly surprising. These weapons raise a perfect storm of concerns regarding civilian casualties, accountability gaps, destabilizing arms races, and the ethics of giving machines the ability to decide over life and death. The U.N. Secretary-General, Antonio Guterres, stated before the General Assembly that the prospect of autonomous weapons "raises multiple alarms" and is "morally repugnant."[9]

A number of leaders in the field of robotics and artificial intelligence (AI) have also voiced concern about autonomous weapons. An open letter to the CCW, signed by dozens of industry experts, claims that these weapons "threaten to become the third revolution in warfare."[10] These weapons, they assert, "will permit armed conflict to be fought at a scale greater than ever, and at timescales faster than humans can comprehend."[11] Corporations are also entering this debate. In 2018, Google announced that it would not "design or deploy AI in …

---

[6]   Fifth Review Conference of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (Convention on Certain Conventional Weapons), *Final Document of the Fifth Review* Conference, sec. III, Decision 1, U.N. Doc. CCW/Conf.V/10 (Dec. 23, 2016) ("To establish an open-ended Group of Governmental Experts (GGE) related to emerging technologies in the area of lethal autonomous weapons systems").

[7]   *See, e.g.*, Rain Liivoja, *Why it's so Hard to Reach an International Agreement on Killer Robots*, THE CONVERSATION (Sept. 11, 2018), https://theconversation.com/why-its-so-hard-to-reach-an-international-agreement-on-killer-robots-102637; Stewart M. Patrick, *Laying Down the LAWS: Strategizing Autonomous Weapons Governance*, COUNCIL ON FOREIGN RELATIONS (Apr. 24, 2019), https://www.cfr.org/blog/laying-down-laws-strategizing-autonomous-weapons-governance.

[8]   *Five years of Campaigning, CCW Continues*, CAMPAIGN TO STOP KILLER ROBOTS (Mar. 18, 2018), https://www.stopkillerrobots.org/2018/03/fiveyears/; *see also* Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions,* at 113, 118, UN Doc. A/HRC/23/47 (Apr. 9, 2013) (Christof Heyns, the U.N. Special Rapporteur on Extrajudicial, Summary, or Arbitrary Executions has similarly called for a moratorium on the development of autonomous weapons); Austria, Brazil and Chile, *Proposal for a Mandate to Negotiate a Legally-Binding Instrument that Addresses the Legal, Humanitarian and Ethical Concerns Posed by Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS)*, CCW/GGE.2/2018/WP.7 (Aug. 30, 2018) (proposal by Brazil, Chile, Austria).

[9]   António Guterres, U.N. Secretary-General, *Address to the General Assembly* (Sept. 25, 2019), https://www.un.org/sg/en/content/sg/speeches/2018-09-25/address-73rd-general-assembly.

[10]   *An Open Letter to the United Nations Convention on Certain Conventional Weapons*, https://www.cse.unsw.edu.au/~tw/ciair/open.pdf (on file with UNSW Sydney, Australia).

[11]   *Id.*

[w]eapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people."[12]

Most major military powers oppose new international regulations on the development or use of autonomous weapons.[13] They argue that that existing international humanitarian law (IHL)—the legal framework applicable during armed conflicts—is sufficient to appropriately regulate new weapons with emerging technologies.[14] States also point to the potential humanitarian benefits of greater autonomy and artificial intelligence in weapons systems. Such weapons may be more discriminate than existing weapons, thus reducing the risk to civilians and friendly forces.[15] Finally, it would be exceedingly difficult, and likely counter-productive, to attempt to regulate military use of a technology that is rapidly evolving. Even if States did agree to prohibit such weapons, much of the underlying technology is dual use and being developed by the private sector. A ban would be difficult to verify or enforce.[16]

The ongoing debate on whether to ban autonomous weapons is unlikely to deter many States from pursuing weapons with increasing degrees of autonomy.[17] The United States, Russia, China, and other military powers are investing heavily in AI[18] because they believe it will provide competitive

---

[12]   Sundar Pichai, *AI at Google: Our Principles*, GOOGLE (June 7, 2018), https://blog.google/topics/ai/ai-principles/ (Google clarified that while it will not develop AI for use in weapons, "we will continue our work with governments and the military in many other areas.").

[13]   Damien Gayle, *UK, US and Russia Among Those Opposing Killer Robot Ban*, GUARDIAN (Mar. 29, 2019), https://www.theguardian.com/science/2019/mar/29/uk-us-russia-opposing-killer-robot-ban-un-ai.

[14]   *Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, CCW/GGE.1/2018/3, ¶ 28 (Oct. 23, 2018) (noting that a number of delegations expressed the view that "[a]s IHL is fully applicable to potential lethal autonomous weapons systems … no further legal measures were needed.") [hereinafter *Report of the 2018 Session*].

[15]   *See* United States Working Paper, *Humanitarian Benefits of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, CCW/GGE.1/2018/WP.4 (Mar. 28, 2018) [hereinafter *Humanitarian Benefits of Emerging Technologies*]. Christof Heyns similarly recognizes the possibility that autonomous weapons may offer some humanitarian benefit. *See* Heyns, *supra* note 8, at ¶ 52 (noting that "future generations of robots may be able to employ less lethal force, and thus cause fewer unnecessary deaths").

[16]   *Autonomous Weapons Are a Game-Changer*, *supra* note 3 (noting that "the most dramatic advances in AI and autonomous machines are being made by private firms with commercial motives" and that "the technology enabling autonomous weapons will be both pervasive and easily transferable."); *see also* Andrew Ilachinski, *AI, Robots, and Swarms: Issues, Questions, and Recommended Studies*, 7 CNA (Jan. 2017), https://www.cna.org/cna_files/pdf/DRM-2017-U-014796-Final.pdf ("most key innovations in AI, robotics, and autonomy are now being driven by the *commercial sector*"); Kenneth Anderson et al., *Adapting the Law of Armed Conflict to Autonomous Weapons Systems*, 90 INT'L STUD. 386, 397–98 (2014).

[17]   Alan Schuller, *At the Crossroads of Control,* 8 HARV. NAT'L SEC. J. 381, 390 (2017) ("It is a safe assumption that autonomy will continue to increase in modern weapons."); *see also* Anderson, et al., *supra* note 16, at 390–91 (describing the supply and demand for weapons with greater autonomy or automation).

[18]   *See, e.g.*, Drew Harwell, *Defense Department Pledges Billions Toward Artificial Intelligence Research*,

advantages.[19] Vladimir Putin predicted that whichever country leads the field of AI "will be the ruler of the world."[20] In June 2018, the then-Deputy Secretary of Defense wrote, "to preserve and expand our military advantage … we must pursue AI applications with boldness and alacrity while ensuring strong commitment to military ethics and AI safety."[21] In future wars, victory may depend on "the quality of each side's algorithm" rather than on the skill or bravery of a State's armed forces.[22] Just as technology has profoundly affected the role of humans in other professions—such as medicine, finance, and transportation,[23] there is little doubt that advances in autonomy and AI will transform the nature of, and humans' role in, warfare.[24]

The rules and principles of IHL have thus far been able to adapt to the use of increasingly sophisticated weapons in warfare, such as unmanned aerial

---

WASH. POST (Sep. 7, 2018), https://www.washingtonpost.com/technology/2018/09/07/defense-department-pledges-billions-toward-artificial-intelligence-research/?noredirect=on&utm_term=.f9bcbf0833d5; *see also Russia's New State Armament Program May Focus on Autonomous Weapons*, TASS (Mar. 28, 2017), https://tass.com/defense/937984.

[19] *See, e.g.,* PAUL SCHARRE, AUTONOMOUS WEAPONS AND THE FUTURE OF WAR, ARMY OF NONE 4 (2018) ("Militaries around the globe are racing to deploy robots at sea, on the ground, and in the air …. These robots are increasingly autonomous and many are armed."); *Summary of the 2018 National Defense Strategy*, DEP'T OF DEF., https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf (last visited Oct. 12, 2019) ("The Department will invest broadly in military application of autonomy, artificial intelligence, and machine learning, including rapid application of commercial breakthroughs, to gain competitive military advantages."). The UK Ministry of Defense stated that its pursuit of AI for military purposes is "non-negotiable." It emphasized the importance of investing in education and research in AI, which it stated, "may be as valuable an advantage as the ability to fabricate high grade steel during the Victorian age." UK Ministry of Defense, *Joint Concept Note 1/18: Human Machine Teaming,* DEVELOPMENT, CONCEPTS & DOCTRINE CENTRE, at 8 (May 2018) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/709359/20180517-concepts_uk_human_machine_teaming_jcn_1_18.pdf [hereinafter *UK Joint Concept Note*].

[20] *Putin: Leader in Artificial Intelligence Will Rule World*, CNBC (Sept. 24, 2107), https://www.cnbc.com/2017/09/04/putin-leader-in-artificial-intelligence-will-rule-world.html.

[21] Patrick M. Shanahan, *Establishment of the Joint Artificial Intelligence Center*, DEP'T OF DEF. (June 27, 2018), https://admin.govexec.com/media/establishment_of_the_joint_artificial_intelligence_center_osd008412-18_r….pdf (Deputy Secretary Shanahan announced the establishment of a new Joint Artificial Intelligence Center "with the overarching goal of accelerating the delivery of AI-enabled capabilities, scaling the Department-wide impact of AI, and synchronizing DoD AI activities to expand Joint Force advantages").

[22] *War at Hyperspeed: Getting to Grips with Military Robots,* ECONOMIST (Jan. 27, 2018), https://www.economist.com/special-report/2018/01/25/getting-to-grips-with-military-robotics.

[23] *See generally* DEF. SCI. BOARD, SUMMER STUDY ON AUTONOMY (June 2016) (documenting the increasing use of autonomy in the commercial sector, including to identify fraud in financial transaction, recommend treatment for cancer patients, and self-driving cars).

[24] *UK Joint Concept Note*, *supra* note 19, at 53 ("Robotic and artificial intelligence systems are likely to revolutionise the battlespace."); *see* Heyns, *supra* note 8, ¶ 28, at 5 (stating that "the robotics revolution has been described as the next major revolution in military affairs, on par with the introduction of gunpowder and nuclear bombs").

vehicles (UAVs).[25] The pace of technological advancement and its effect on the conduct of hostilities, however, is rapidly outpacing the more glacial evolution of IHL.[26] Advances in AI and machine learning will pose new and more difficult challenges to current interpretations and applications of IHL.[27] In particular, the ability of machines to make certain decisions that have traditionally been made exclusively by humans will force States to reconcile IHL's focus on *human* decision-making with this new technology.

This Article provides the first comprehensive analysis of how fundamental principles of IHL can and should be interpreted and applied in light of rapidly changing advances in warfare and the correlated humanitarian risks. Although humans' role in warfare may be increasingly removed from the physical battlefield, human judgment and decision-making over the use of force remains the focus of IHL. IHL imposes obligations on humans, and these obligations cannot be delegated to machines.[28] This Article will accordingly focus on the challenge of using weapons with significant degrees of autonomy consistent with IHL, as well as the challenges for IHL in regulating the use of this new technology. This Article does not address the more technical and speculative question of whether future weapons could independently make the same judgments that are required of humans in order to comply with IHL.[29]

This Article proceeds as follows. Part II describes the different levels of autonomy in weapons systems and seeks to clarify some of the ongoing confusion regarding the term "lethal autonomous weapons systems." Part III discusses the military benefits of autonomy, the foreseeable applications of this technology in warfare, and the humanitarian concerns that it poses. Part IV explains why autonomous weapons are not categorically prohibited by IHL. Part

---

[25]   *See* Harold Koh, *The Obama Administration and International Law, Keynote Speech at the Annual Meeting of the American Society of International Law*, (Mar 25, 2019), https://2009-2017.state.gov/s/l/releases/remarks/139119.htm.

[26]   P. W. Singer, *The Five Deadly Flaws of Talking About Emerging Military Technologies and the Need for New Approaches to Law, Ethics, and War*, *in* DRONE WARS: TRANSFORMING CONFLICT, LAW, AND POLICY 215, 228 (Peter L. Bergen & Daniel Rothenberg eds., 2014) ("What unites the complex moral and policy issues associated with transformative military technologies and killer applications is how the rapid rate of innovation makes it difficult for our all too-human institutions, including those involved in managing ethics and law, to keep pace.").

[27]   *See* Schuller *supra* note 17, at 383 ("There remains a death of practical guidance on how states should regulate [autonomous weapons systems] development").

[28]   Neil Davison, *A Legal Perspective: Autonomous Weapon Systems Under International Humanitarian Law*, INT'L COMM. OF RED CROSS (Jan. 31, 2018), https://www.icrc.org/en/document/autonomous-weapon-systems-under-international-humanitarian-law.

[29]   A number of articles have explored this issue. *See, e.g.,* Christopher M. Ford, *Autonomous Weapons and International Law*, 69 S.C. L. REV. 413 (2017).

V analyzes why autonomy will challenge core understandings of how IHL regulates human decision-making over uses of force. Part VI then turns to how core provisions of IHL—including those relating to weapons development, targeting, and accountability—can be applied to the use of autonomous weapons.

I.   WHAT ARE AUTONOMOUS WEAPONS SYSTEMS?

A significant impediment to substantive discussions of autonomous weapons among States and in the academic literature is the lack of universally agreed terms of reference. The term "lethal autonomous weapons systems" was coined as the title of the GGE and has since been adopted more broadly.[30] The term, however, is an artificial and somewhat misleading construct. Weapon systems may incorporate varying degrees of autonomy distributed among a number of different functions. For this reason, it is "helpful to think of autonomy as a spectrum or series of spectrums."[31] Paul Scharre, a leading expert on autonomy in warfare, refers to three "dimensions" of autonomy: the human machine relationship, the type of task to be performed by a machine, and the sophistication of the machine's decision-making capabilities.[32]

This Part discusses the spectrum of autonomy[33] in weapons systems, including the defining characteristics of semi-autonomous and autonomous weapons. This Article generally disfavors the term "LAWS" that is used in the CCW discussions, given the confusion regarding what it encompasses and the political baggage that it carries. The term, and the ongoing discussions in Geneva, also tend to oversimplify the complex issues presented by autonomy in weapons by suggesting that it is possible to draw a bright line along this

---

[30]   *See, e.g.*, Adam Cook, *Taming Killer Robots: Giving Meaning to the "Meaningful Human Control" Standard for Lethal Autonomous Weapon Systems*, JAG SCHOOL PAPERS (June 2019), https://media.defense.gov/ 2019/Jun/18/2002146749/-1/-1/0/JP_001_COOK_TAMING_KILLER_ROBOTS.PDF.; *2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)*, UNITED NATIONS: GENEVA https://www.unog.ch/80256EE600585943/(httpPages)/7C335E71DFCB29D1C1258243003E8724?OpenDocu ment (last visited Jan. 29, 2020). The GGE on LAWS followed an informal group of experts on LAWS convened by the CCW in 2013.

[31]   Schuller, *supra* note 17, at 392; *see also* Ford, *supra* note 29, at 417, 421.

[32]   SCHARRE, *supra* note 19, at 27.

[33]   The term "autonomy" is itself somewhat ambiguous, and the line between automation and autonomy is not always clear. In general, autonomous systems are capable of goal directed behavior, and can take into account different variables in determining how best to achieve that programmed objective. *See, e.g., id.* at 32; *see also* Brian K. Hall, *Autonomous Weapons Systems Safety*, 86 JOINT FORCE QTR., 4 (June 21, 2017) ("Autonomy is better understood as a technologically advanced capability or capabilities that enables the larger human-machine system to accomplish a given mission by the performance a key actions—with or without human intervention.").

spectrum of autonomy, both in terms of degree and function. After Part II, this Article will generally use the term "autonomous weapons," which for the purpose of this Article may include weapons with varying degrees of autonomy across different functions.[34]

A. Semi-Autonomous

Semi-autonomous weapons have autonomous functions to identify, track, and maneuver to targets.[35] These weapons, however, can "only engage individual targets or specific target groups that have been selected by a human operator."[36] Semi-autonomous weapons are often described as having a human "in the loop"—a reference to the Orient, Observe, Decide and Act (OODA) loop.[37] A semi-autonomous weapon cannot complete this loop absent a human decision.[38] Semi-autonomous weapons can be used to attack known targets or to identify unknown targets, but they cannot attack unknown targets.[39] For example, homing munitions may be used to search and engage a target within a confined geographic area.[40] The homing munition uses autonomous capabilities to identify and track the intended target, but it cannot select unintended targets.[41] A semi-autonomous weapon may also be used to identify and track (but not engage) unknown targets.[42] For example, the South Korean SGR A-1 Sentry is a border defense system that uses sensors to identify and track potential targets along the De-Militarized Zone with North Korea.[43] Although this weapon system can independently locate potential targets, it is considered semi-autonomous because a human must complete the OODA loop by initiating the use of lethal force.[44]

---

[34] Importantly, this term should not be equated with "fully autonomous weapons systems," which as discussed in Part II(D), are unlikely to exist in the foreseeable future, if ever.

[35] *See* Paul Scharre & Michael C. Horowitz, *An Introduction to Autonomy in Weapon Systems*, CENTER FOR A NEW AM. SECURITY (Feb. 2015), https://www.files.ethz.ch/isn/188865/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf.

[36] Dep't of Def. Directive 3000.09, Autonomy in Weapon Systems, at 14 (DoD 2017) [hereinafter DoDD 3000.09].

[37] Scharre & Horowitz, *supra* note 35.

[38] *Id.*

[39] *Id.* at 13.

[40] *Id.*

[41] *Id.*

[42] *Id.*

[43] Rebecca Crootof, *War Torts: Accountability for Autonomous Weapons*, 164 U. PENN. L. REV. 1347, 1367 (2016).

[44] *Id.*

B. Human Supervised

Human-supervised autonomous weapon systems can select and engage targets, but are designed to give human operators the opportunity to intervene and terminate engagements in the event of malfunction or unforeseen circumstances.[45] These systems are often described as having a human "on the loop."[46] Human supervised weapons are generally employed for defensive purposes, and against objects rather than persons, such as incoming precision guided missiles (PGMs) and artillery.[47] In these defensive situations, "the reaction time required for engagement is so short that it would be physically impossible for humans to remain 'in the loop[.]'"[48] Nevertheless, a human operator can supervise the weapon system and intervene as appropriate.[49] Examples of such systems include the ship-based U.S. Aegis,[50] the land-based Counter Rocket, Artillery, and Mortar System (C-RAM),[51] and the Israeli Iron Dome.[52] Although these weapons have autonomous functions, they operate within human defined parameters.[53] Humans determine what threats the weapon system will, or will not respond to, even though a human does not identify each specific object to be attacked in a given engagement.[54]

C. Autonomous Weapons

The Department of Defense (DoD) Directive on Autonomy in Weapon Systems defines an autonomous weapon as any weapon system that, "once activated, can select and engage targets without further intervention by a human operator."[55] With autonomous weapons the human is "out of the loop" because

---

[45]   DoDD 3000.09, *supra* note 36 ("An autonomous weapon system that is designed to provide human operators with the ability to intervene and terminate engagements, including in the event of a weapon system failure, before unacceptable levels of damage occur").

[46]   Ford, *supra* note 29, at 424.

[47]   Scharre & Horowitz, *supra* note 35, at 8, 12, 18.

[48]   *Id.* at 12.

[49]   *Id.* at 6.

[50]   Gary Marchant et al., *International Governance of Autonomous Military Robots*, 12 COLUM. SCI. & TECH. L. REV. 272, 276 (2011).

[51]   SCHARRE, *supra* note 19, at 323–25.

[52]   Michael Schmitt, *Autonomous Weapons and International Humanitarian Law: A Reply to Critics*, HARV. NAT'L SEC. J. 4 (2013). Scharre and Horowitz estimate that at least thirty countries employ or are developing human supervised autonomous weapons. Scharre & Horowitz, *supra* note 35, at 3.

[53]   *See* Nicholas W. Mill, *The Roboticization of Warfare with Lethal Autonomous Weapon Systems (LAWS); Mandate of Humanity or Threat to It?* 40 HOUSTON J. INT'L L. 461, 485 (2018) (noting that in these systems engagement only occurs with the approval of a human operator).

[54]   Scharre & Horowitz, *supra* note 35, at 12.

[55]   DoDD 3000.09, *supra* note 36. The ICRC similarly defines autonomous weapons as: "Any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect,

the weapon independently selects and engages targets without human intervention after its deployment.

The ambiguity in the terms "select" and "engage" poses some definitional challenges.[56] For example, a weapon deployed against a group of incoming missiles may have the ability to select which missile to attack first. In this sense, the weapon system "selects" one like target over another, but only within the parameters set by the human programmer or operator.[57] Most commentators would not consider such a weapon to be autonomous because its selection capability is constrained to a specific and pre-determined group of military objectives.

One commonly cited example of an extant autonomous weapon is the Israeli Harpy.[58] The Harpy is similar to many semi-autonomous weapons in the sense that it is programmed to search for and destroy specific military objectives (i.e., radar installations).[59] Its prolonged loitering capabilities, however, arguably render it an autonomous weapon.[60] Unlike semi-autonomous weapons, it need not be launched against a specific target in a particular location. Whereas semi-autonomous anti-radar weapons have a flight time of approximately five minutes, the Harpy can search for enemy radars for up to two and half hours and cover up to 500 kilometers.[61] The Harpy's loiter and search capabilities mean that it can successfully eliminate targets even when the human operator has little knowledge of whether those targets exist or where they might be located.

The fact that the Harpy is often characterized as an autonomous weapon reveals the grey area between semi-autonomous and autonomous weapons. In many cases, the difference may be one of degree rather than of kind.

---

identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention." *Views of the International Committee of the Red Cross (ICRC) on Autonomous Weapons System,* INT'L COMM. OF THE RED CROSS (Apr. 11, 2016), https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system; *see also* SCHARRE, *supra* note 19, at 46 ("These weapons can search for, decide to engage, and engage targets on their own and no human can intervene.").

[56]   *See* Ford, *supra* note 29, at 419 (noting the different interpretations of "select and engage").

[57]   United States Working Paper, *Human-Machine Interaction in the Development, Deployment, and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems: Ensuring that Machines Effectuate Human Intent in Using Force,* CCW/GGE.2/2018/WP.4, ¶ 53 (Aug. 28, 2018) ("The fact that the projectile might also 'select' a target that has been identified by a human operator or that has been programmed into it and autonomously maneuver itself toward a target does not amount to a delegation of decision-making from humans to machines.").

[58]   *See e.g.,* Shelby Smith, *Technology Explainer: Automated Defense Technology,* 3 GEO. L. TECH REV. 492, 499 (2019).

[59]   *Id.*

[60]   *See id.* (citing the Harpy as an example of an autonomous weapon).

[61]   SCHARRE, *supra* note 19, at 48.

Autonomous weapons may perform the same kind of functions as semi-autonomous weapons but exercise a greater degree of autonomy in carrying out their assignment. With semi-autonomous weapons, the human operator must have fairly specific knowledge of where the objective is located, even if he does not fire the weapon toward a specific geographic coordinate. Autonomous weapons, by contrast, can be deployed with far less certainty regarding the existence or whereabouts of the intended targets because they have greater "freedom in time and space."[62] Accordingly, it is the knowledge of the *human* operator that is required to achieve an intended military effect that often determines the characterization of the weapon system.

D. Fully Autonomous Weapons

The definition of "autonomous weapon" used in the DoD Directive and in much of the academic literature is not synonymous with the more politically charged term "lethal autonomous weapon system" that is the focus of the CCW. There is little agreement among parties to the CCW on what should be characterized as a LAWS, which complicates efforts to meaningfully discuss the potential costs and benefits of these weapons, as well as possible legal and policy approaches to regulating them.[63]

Some States propose a definition of LAWS similar to that set forth above. The Netherlands defines LAWS as, "A weapon that, without human intervention, selects and engages targets matching certain predefined criteria, following a human decision to deploy the weapon on the understanding that an attack, once launched, cannot be stopped by human intervention."[64] Other States take the position that LAWS refer solely to "fully" autonomous weapons systems, meaning that the weapons system has complete (or at least near complete) independence from human decision-making. Such weapons may be understood as having "general" rather than "narrow" AI.[65] France stated, for

---

[62]   *Id.* at 52.

[63]   The range of definitions of LAWS is likely driven, in part, by political motivations. States tailor their definitions of LAWS in light of the growing movement to ban such weapons. Accordingly, some have elected to define LAWS to encompass only those weapons that they have no desire, or ability, to develop. Others seek a broad definition of LAWS so as to expand the category of weapons that would be prohibited by such a treaty.

[64]   Netherlands, *Examination of Various Dimensions of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, in the Context of the Objectives and Purposes of the Convention*, CCW/GGE.1/2017/WP.2, ¶ 5 (Oct. 9, 2017).

[65]   *War at Hyperspeed*, *supra* note 22, at 14 (narrow AI allows a machine to permit a specific task at a level that matches or exceeds human capabilities; machines with general AI would have the ability "reason, plan, solve problems, think abstractly, and learn quickly from experience"). General AI does not currently exist and is likely decades away. *See* SCHARRE, *supra* note 19, at 232 ("A majority of AI experts predict AGI could be

example, "LAWS should be understood as implying a total absence of human supervision, meaning there is absolutely no link (communication or control) with the military chain of command."[66] Similarly, the UK defines LAWS as a system "which is capable of understanding, interpreting, and applying higher level intent and direction based on a precise understanding and appreciation of what a commander intends to do and perhaps more importantly why."[67] The Israeli Harpy would not be considered a LAWS under these latter two definitions. The Harpy autonomously tracks, selects, and engages its targets without human intervention, but it can only attack a specific class of military objective.[68] It is not capable of understanding a commander's intent or deciding to attack different types of military objectives.

The difference between "autonomous" and "fully autonomous" weapons is not always clear, but it is significant.[69] Weapon systems that can select and engage targets once activated may nevertheless incorporate significant degrees of human judgment or control in the targeting process. Similar to a driver that programs an autonomous vehicle to drive to a "grocery store" without having a particular destination in mind, a human operator could theoretically instruct an autonomous weapon only to attack "tanks." In both cases, the machine can exercise autonomy in determining how to achieve the objective set by a human operator. The definitions of fully autonomous weapon systems, by contrast, do not contemplate a role for the human operator in defining the parameters of the targets to be attacked. Rather, such a weapon might be understood to search for and attack whatever the weapon itself determines to be a military objective, perhaps based on some understanding of the commander's general intent or strategy.

Critics of autonomous weapons seem most concerned, and reasonably so, with "fully autonomous" weapons.[70] A preemptive ban on such weapons, however, would likely have little practical effect. There is no legitimate military

---

possible by 2040 and likely by the end of the century, but no one really knows").

[66]   France, *Characterization of a LAWS* (Apr. 2016) Working Paper to Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) (Apr. 2016).

[67]   UK Foreign & Commonwealth Office, *Statement to the Informal Meeting of Experts on Lethal Autonomous Weapons Systems* (Apr. 2016).

[68]   Crootof, *supra* note 43, at 1368.

[69]   SCHARRE, *supra* note 19, at 190 ("The key difference between semi-, supervised, and fully autonomous weapons is amount of damage the system can cause until the next opportunity for a human to intervene.").

[70]   *See* Docherty, *supra* note 4, at 3–4. Human Rights Watch's report, *Losing Humanity*, generally refers to the concerns presented by "fully autonomous weapons," which it claims would "eliminat[e] human involvement in the decision to use lethal force in armed conflict." *Id.* Docherty acknowledges that such fully autonomous weapons do not yet exist. *Id.*

rationale for developing weapons that cannot be controlled.[71] No State has expressed an interest in pursuing fully autonomous weapons,[72] and a number of major military powers have stated publicly that they will not do so.[73] States have reaffirmed the need for human control or responsibility over the use of force, although there is debate regarding the necessary degree and nature of human involvement.[74]

A precise definition of autonomous weapons is not necessary to consider the significant questions raised by the development and use of weapons with high degrees of autonomy, but which may not be considered fully autonomous. Weapons will increasingly have various degrees of autonomy across a number of different functions. Some of these autonomous functions (such as navigation) may not pose any concern regarding compliance with IHL, while autonomy in functions relating to selecting and engaging targets may require greater attention.[75] The important question, discussed in Part VI, is whether and how existing IHL can appropriately regulate the use of weapons incorporating this emerging technology.

II.   AUTONOMY IN WARFARE: ADVANTAGES, APPLICATIONS, AND CONCERNS

This Part briefly describes the military advantages presented by autonomy and how militaries are likely to use this technology in the foreseeable future. Much of the debate regarding autonomous weapons, and in particular the concerns raised by NGOs and States that seek to ban autonomous weapons, focuses on hypothetical technologies and military applications. A reality-based

---

[71]   Crootof, *supra* note 43, at 1369 ("Because of the military's interest in foreseeable results, one may credibly suggest that states will have little incentive to develop, let alone deploy, potentially unpredictable autonomous weapons."); France, *Mapping of Technological Developments*, Working Paper to Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) (Apr. 2016) (stating that "the use of a totally autonomous weapons system … would be militarily useless.").

[72]   Heyns, *supra* note 8, at ¶ 29 ("Official statements from Governments with the ability to produce [lethal autonomous robots] indicate that their use during armed conflict or elsewhere is not currently envisioned."); SCHARRE, *supra* note 19, at 98 (former Deputy Secretary of Defense, Robert Work, noting that he does not envision the U.S. military even deploying weapons with general artificial intelligence).

[73]   UK Working Paper, *Human Machine Touchpoints: The United Kingdom's Perspective on Human Control Over Weapon Development and Targeting Cycles*, CCW/GGE.2/2018/WP.1, ¶ 6 (Aug. 8, 2018) [hereinafter UK Working Paper, *Human Machine Touchpoints*] ("[T]he UK does not possess fully autonomous weapons and is not developing them.").

[74]   *See, e.g., Report of the 2018 Session*, *supra* note 14, ¶ 21(b) ("Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines.").

[75]   *Views of the International Committee of the Red Cross (ICRC) on Autonomous Weapons System*, *supra* note 55 (noting the need to further examine the implications of autonomy in the critical functions of weapons systems).

discussion focused on current or likely uses of autonomy can better inform our understanding of the potential challenges and benefits of this technology.

A. Military Advantages of Autonomy

Autonomous weapons will offer a number of advantages in warfare, including faster data analysis and reaction times, greater endurance, the ability to launch scaled attacks and to operate in communications degraded environments, and potentially greater discrimination in the use of force.[76]

Artificial Intelligence[77] (and its subfields including machine learning[78]) allows weapons to process data and make decisions at speeds that far exceed human capabilities. Human reactions, for example, are simply "too slow to deal with multiple, inbound, high-speed missiles."[79] Artificial intelligence can help solve this problem. "Autonomy reduces the human workload required to operate systems … and allows human decision making to focus on points where it is most needed."[80] Autonomous weapons will thus permit armed forces to "integrate more information from more sources far faster before responding with lethal force than a human possibly could in real-time."[81] The next generation of algorithms will "transform computers from tools to problem-solving partners."[82]

Advances in robotics and AI will lead to weapons with far greater endurance than humans. Autonomous weapons can "stay out on the battlefield far beyond

---

[76] Hall, *supra* note 33, at 5.

[77] *See generally* M.L. Cummings, *Artificial Intelligence and the Future of Warfare*, CHATHAM HOUSE, at 2 (Jan. 2017) (AI can generally be understood as "the capability of a computer system to perform tasks that normally require human intelligence"); U.S. ARMY, ROBOTIC AND AUTONOMOUS SYSTEMS STRATEGY 3 (2017) ("Artificial Intelligence (AI) is the capability of computer systems to perform tasks that normally require human intelligence such as perception, conversation, and decision-making.").

[78] *See generally* PETER FLACH, MACHINE LEARNING: THE ART AND SCIENCE OF ALGORITHMS THAT MAKE SENSE OF DATA 3 (2012) (Machine learning refers to "algorithms and systems that improve their knowledge or performance with experience.").

[79] Kenneth Anderson & Matthew Waxman, *Law and Ethics for Robot Soldiers*, HOOVER INST. (Dec. 1, 2012), https://www.hoover.org/research/law-and-ethics-robot-soldiers; United States Working Paper, *Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons* Systems, CCW/GGE.2/2018/WP.4, ¶ 10, (Aug. 28, 2018) (noting that certain weapons with autonomous functions can "strike incoming projectiles with much greater speed and accuracy than a human gunner could achieve manually.") [hereinafter *Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies*].

[80] SCHARRE, *supra* note 19, at 16 (citing DoD 2011 roadmap).

[81] Ronald Arkin, *Lethal Autonomous Systems and the Plight of the Non-Combatant*, 137 AISB QTR. 3 (2013).

[82] Peter Highnam, Deputy Dir., Defense Advanced Research Projects Agency, Transcript of Testimony Before the Emerging Threats and Capabilities SubCommittee (Mar. 12, 2019) (transcript *available at* https://www.armed-services.senate.gov/imo/media/doc/Highnam_03-12-19.pdf).

the limits of human endurance, for weeks, months, or even years at a time without rest."[83] Machines "do not get tired, frightened, bored, or angry."[84] They do not suffer the effects of post-traumatic stress disorder or seek revenge after witnessing their fellow soldiers killed in action.[85] Accordingly, autonomous weapons are not susceptible to the human frailties that often lead to war crimes.[86]

Autonomous weapons can be deployed in hostile environments that would pose an unacceptable risk to humans. Autonomous weapons "do not need to have self-preservation as a foremost drive, if at all. They can be used in a self-sacrificing manner if needed and appropriate without reservation by a commanding officer."[87] These weapons, some of which may be relatively cheap, can also be deployed in greater numbers than human combatants. As the UK Ministry of Defense observed, "The confluence of AI and robotics will allow us to scale physical mass and battlefield points of presence increasingly independently of numbers and locations of human combatants."[88] This permits saturation attacks that would be incredibly costly and dangerous if conducted by soldiers.[89]

Autonomous weapons may also allow commanders to achieve mission objectives in communications degraded environments. "One technological response [to a degraded or jammed communications link might be] to reduce the vulnerability of the communications link by severing it, thus making the robot dependent on executing its own programming, or even [rendering it] genuinely autonomous."[90] This capability will allow militaries to operate in areas that are too dangerous for humans and to complete missions that might previously have

---

[83]   SCHARRE, *supra* note 19, at 13.

[84]   *Getting to Grips with Military Robotics*, ECONOMIST (Jan. 25, 2018), https://www.economist.com/special-report/2018/01/25/getting-to-grips-with-military-robotics.

[85]   Sergeant Robert Bales killed sixteen civilians in Afghanistan, arguably the most heinous war crime committed by a U.S. soldier since the Vietnam War. His rampage was reportedly fueled by a combination of alcohol, steroids, sleep deprivation, the stress of multiple deployments, and a desire to avenge a recent Taliban IED attack that mutilated a fellow soldier. Brendan Vaughan, *Robert Bales Speaks: Confessions of America's Most Notorious War Criminal*, GQ (Oct. 21, 2015), https://www.gq.com/story/robert-bales-interview-afghanistan-massacre.

[86]   *See* Crootof, *supra* note 43, at 1372 (noting that autonomous weapons would not be affected by human emotions, fears, or prejudices, and thus may be more humane than humans).

[87]   Arkin, *supra* note 81, at 3.

[88]   *UK Joint Concept Note*, *supra* note 19, at 21.

[89]   Robert O. Work & Shawn Brimley, *20YY Preparing for War in the Robotic Age*, CENTER FOR A NEW AMERICAN SECURITY (Jan. 2014), at 1, 8 ("as more and more adversaries begin to employ guided munitions and as large numbers of effective and low-cost unmanned systems proliferate, mass will likely once become more prominent in U.S. military force-on-force calculations.").

[90]   Anderson & Waxman, *supra* note 79, at 7.

been aborted. According to one DoD official, "The ability to operate in fast-paced, contested, nonpermissive, force-on-force engagements, particularly under conditions of degraded communications, will drive the need for increased autonomy."[91]

Lastly, advances in autonomy have the potential to reduce civilian harm by minimizing opportunities for human error.[92] The humanitarian benefits of autonomy are well documented in the civilian sector. Most commercial airliners are flown almost exclusively in auto-pilot mode, which has led to a decrease in crashes. Self-driving vehicles promise to reduce the number of accidents, which in 2016 killed approximately 40,000 people in the United States alone.[93] Advances in autonomy may provide similar humanitarian benefits in warfare.[94] As the United States noted in a written submission to the CCW GGE on LAWS, "Emerging technologies in the area of lethal autonomous weapon systems could be used to create entirely new capabilities that would increase the ability of States to reduce the risk of civilian casualties in applying force."[95] For one, "automated target identification, tracking, selection, and engagement functions can allow weapons to strike military objectives more accurately and with less risk of collateral damage."[96] This technology will also allow "personnel to set the parameters for when, where, and how force is deployed without manually controlling the weapons system at all times."[97] The increased discrimination capabilities offered by smart weapons allow a commander to deploy fewer weapons and smaller warheads to achieve the intended military effect. This reduces the potential for unintended collateral damage. In other words,

---

[91]   Hall, *supra* note 33, at 7.

[92]   *See* Marchant et al., *supra* note 50, at 279–80 (describing reasons that autonomous weapons may be more discriminate than humans); Nicholas W. Mull, *The Roboticization of Warfare with Lethal Autonomous Weapon Systems (LAWS): Mandate of Humanity or Threat to It*, 40 HOUS. J. INT'L L. 461, 498 (noting that an autonomous weapon that has proven capability to reduce collateral damage would provide a military advantage).

[93]   Neal E. Boudette, *U.S. Traffic Deaths Rise for a Second Straight Year,* N.Y. TIMES (Feb. 15, 2017), https://www.nytimes.com/2017/02/15/business/highway-traffic-safety.html.

[94]   This does not mean that we should assume that autonomy will necessarily make war more humane. As Peter Singer notes, "New technologies in war are often described as a way to reduce the costs of war, mitigate its passions, and limit the possibility of excessive acts or crimes." Peter W. Singer, *The Five Deadly Flaws of Talking About Emerging Military Technologies and the Need for New Approaches to Law, Ethics, and War in* DRONE WARS TRANSFORMING CONFLICT, LAW AND POLICY 215, 219 (Bergen & Rothenberg eds., 2015).

[95]   *Humanitarian Benefits of Emerging Technologies*, *supra* note 15, ¶ 39; *see also* Schmitt, *supra* note 52, at 25; Anderson et al., *supra* note 16, at 393 (describing how autonomous weapons can reduce the risks to civilians).

[96]   *Humanitarian Benefits of Emerging Technologies*, *supra* note 15, ¶ 26.

[97]   United States Working Paper, *Human-Machine Interaction in the Development, Deployment, and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, CCW/GGE.2/2018/WP.4, ¶ 51 (Aug. 28, 2018).

autonomy can better effectuate the commander's intent, which is to achieve a specific military effect with as little collateral damage as possible.

B. *Military Applications*

The military advantages outlined in Section A make it inevitable that weapons with significant degrees of autonomy will be deployed on the battlefield. The employment of autonomy in warfare is likely to mirror how autonomy has been utilized in civilian sectors. Humans will assign to machines those tasks that machines can perform more effectively than humans.[98] Routine tasks that entail if/or decision-making are most likely to be automated. Similarly, tasks that require decision-making speeds that exceed human capabilities will increasingly be given to machines.[99] By contrast, tasks that require contextual awareness, common sense, creativity, or abstract decision-making will be performed by humans for the foreseeable future.[100] The introduction of weapons with autonomous features will almost certainly be "deliberate and incremental" as militaries seek to ensure the safety and effectiveness of this technology before deploying it on the battlefield.[101] Militaries will also seek to maximize the strengths of machines and humans by designing weapons systems that collaborate with soldiers.[102]

Environmental factors will play a significant role in determining where and how autonomous weapons are used. Militaries will seek to deploy these weapons in environments that optimize the advantages of autonomous decision-making and minimize the risk of unintended engagements. Militaries will initially deploy autonomous weapons in uncluttered environments where the risk of civilian contact is low, such as underwater or in "non-complex areas of battlespace."[103] While our ability to predict how emerging technology will be

---

[98]   *See generally* Thomas H. Davenport & Rajeev Ronanki, *Artificial Intelligence for the Real World*, HARV. BUS. REV., Jan.–Feb. 2018 (describing how AI can help business automate business process, gain insight on customers through data analysis and engage with customers and employees).

[99]   *See* Work & Brimley, *supra* note 89, at 24 (noting advantages of machine and human decision-making).

[100]   *UK Joint Concept Note*, *supra* note 19, at 42 (noting that "the last roles likely to be automated will be where personnel conduct activities that demand contextual assessment …"); Alex Williams, *Will Robots Take Our Children's Jobs?,* N.Y. TIMES (Dec. 11, 2017), https://www.nytimes.com/2017/12/11/style/robots-jobs-children.html.

[101]   Hall, *supra* note 33, at 9. *See also* Anderson et al., *supra* note 16, at 389 (predicting that "new autonomous systems will develop incrementally as more functions … are automated").

[102]   *See generally* U.S. ARMY, *supra* note 77.

[103]   *UK Joint Concept Note*, *supra* note 19, at 33. *See also* James Farrant & Christopher M. Ford, *Autonomous Weapons and Weapon Reviews: The UK Second International Weapon Review Forum*, 93 INT'L L. STUD. 389, 398 (2017) (noting that the "U.S. Department of the Navy sees autonomy as a core technology to their future success").

used is constrained, we can expect that autonomy will be used in the following ways in the near term.

Autonomous weapons will primarily be used against other machines in the foreseeable future. Machines may generate distinct signatures that a weapon system can identify through sophisticated sensors and programming, reducing the risk of unintended engagements. Weapons systems similar to the C-RAM and AEGIS will defend military platforms and bases from high-speed attacks. Weapons, such as the sensor fuzed BONUS Munition will be deployed to track and engage enemy combat vehicles.[104] Untethered weapons that can search for and destroy enemy objects, such as sea mines or radar installations, will also become more common.[105]

Autonomous systems will work in collaboration with soldiers to augment and extend human capabilities.[106] The Department of Defense's AI Strategy states that DoD will "prioritize the fielding of AI systems that augment the capabilities of our personnel by offloading tedious cognitive or physical tasks and introducing new ways of working."[107] Autonomous systems may improve soldiers' situational awareness, facilitate better decision-making, conduct reconnaissance, or provide supply services.[108] On the ground, autonomous target

---

[104]   *See Bofors 155mm BONUS Munition,* BAE SYSTEMS, https://www.baesystems.com/en-us/product/155-bonus (last visited Jan. 30, 2020). Similarly, the Long Range Anti-Ship Missile (LRASM) uses autonomous functions to engage enemy ships. The LRASM "uses onboard sensors and a semiautonomous guidance system to reduce its dependence on intelligence, surveillance, and reconnaissance (ISR) platforms, network links and GPS navigation." *Lockheeds Long-Range Anti-Ship Missile Looks Like a Killer And China Should be Afraid*, NAT'L INT. (Mar. 23, 2018), http://nationalinterest.org/blog/the-buzz/lockheeds-long-range-anti-ship-missile-continues-impress-25057; *Lockheed Martin's Long Range Anti-Ship Missile Marks Sixth Successful Flight Mission*, LOCKHEED MARTIN (Mar. 19, 2018), https://news.lockheedmartin.com/2018-03-19-Lockheed-Martins-Long-Range-Anti-Ship-Missile-Marks-Sixth-Successful-Flight-Mission#assets_20295_128436-117.

[105]   The U.S. Navy, for example, is developing autonomous systems to detect, identify, and neutralize sea mines. *See* Aaron Gregg, *Defense Contractors Race to Build Self-Driving Submarines That Can Clear Sea Mines*, WASH. POST (Oct. 5, 2018, 5:34 PM), https://www.washingtonpost.com/business/2018/10/05/defense-contractors-race-build-self-driving-submarines-that-can-clear-sea-mines/?utm_term=.1e972f3630ef.

[106]   *UK Joint Concept Note*, *supra* note 19, at 12; JARNA M. PETMAN, AUTONOMOUS WEAPONS SYSTEMS AND INTERNATIONAL HUMANITARIAN LAW: 'OUT OF THE LOOP'? 9 (2018) ("At present, the prevailing view seems to be that robots will be used only to augment and extend [a] soldier's involvement in war."); Heyns, *supra* note 8, at ¶ 47.

[107]   U.S. DEP'T OF DEF., SUMMARY OF THE 2018 DEPARTMENT OF DEFENSE ARTIFICIAL INTELLIGENCE STRATEGY, HARNESSING AI TO ADVANCE OUR SECURITY AND PROSPERITY 7 (2018).

[108]   U.S. ARMY, *supra* note 77. This strategy paper states that development and employment of autonomous systems will be guided by the objectives of increasing situational awareness, lightening the soldiers' physical and cognitive workloads, sustaining the force, facilitating movement, and protecting the force. As a near term objective, the Army will procure more man portable robotics and autonomous systems that "enable tactical forces to make contact with threats on their own terms." *Id.* at 4–5.

acquisition technology may identify targets for combat vehicles.[109] In the air, swarms of UAVs may accompany manned aircraft, providing a cost effective force multiplier.[110] The pilot may define targets for the UAV, which will then determine how to meet that goal "by selecting from a predefined set of actions."[111]

Militaries will harness AI to process the massive amount of real time data that UAV technology provides.[112] UAVs' persistent surveillance capabilities generate far more data than humans can analyze. AI systems can process this "deluge of data" and help commanders better understand the battle space, including the location of potential targets and protected civilians or objects.[113] A UK Ministry of Defense concept paper notes that the confluence of AI and robotics will "extend[] the reach of and persistence of our intelligence, surveillance and reconnaissance (ISR)."[114] AI driven software "will be able to pre-filter, fuse and classify all data flows, eliminate paralyzing information overload, and accelerate the observe, orient, decide and act (OODA) loop of decision-makers."[115] The U.S. DoD is similarly designing AI programs to assist in analyzing UAV imagery.[116] Project Maven, for example, will use neural networks to help analysts sift through massive amounts of video surveillance to identify significant military objects.[117]

---

[109] Patrick Tucker, *US Military Changing 'Killing Machine' Robo-tank Program After Controversy*, DEFENSE ONE (Mar. 1, 2019), https://www.defenseone.com/technology/2019/03/us-military-changing-killing-machine-robo-tank-program-after-controversy/155256/; *see also Industry Day for the Advanced Targeting and Lethality Automated System ATLAS Program*, FED. BUS. OPPORTUNITIES (Feb. 11, 2019), https://www.fbo.gov/index.php?s=opportunity&mode=form&id=29a4aed941e7e87b7af89c46b165a091&tab=core&_cview=0.

[110] Lara Seligman, *How Swarming Drones Could Change the Face of Air Warfare*, DEFENSE NEWS (May 17, 2016), https://www.defensenews.com/2016/05/17/how-swarming-drones-could-change-the-face-of-air-warfare/; SCHARRE, *supra* note 19, at 72–73 ("With cooperative behavior, one person can tell a group of drones to achieve a goal, and the drones can divvy up the tasks on their own.").

[111] *War at Hyperspeed*, *supra* note 22, at 14.

[112] *Id.* (noting that AI will play a significant role in data analytics, given that military analysts are overwhelmed with the amount of data provided by surveillance drones).

[113] Tobin Harshaw, *Relax, Google, the Robot Army Isn't Here Yet: A Q&A on the Morality and Practicality of Artificial Intelligence Use by the Military*, BLOOMBERG OPINION (June 16, 2018, 8:00 AM), https://www.bloomberg.com/opinion/articles/2018-06-16/google-s-wrong-military-ai-isn-t-necessarily-evil.

[114] *UK Joint Concept Note*, *supra* note 19, at 2.

[115] *Id.* at 16.

[116] Daniel Cebul, *Google is Helping the Pentagon Learn How to Analyze Drone Footage,* DEFENSENEWS (Mar. 7, 2018), https://www.defensenews.com/home/2018/03/07/google-is-helping-the-pentagon-learn-how-to-analyze-drone-footage/.

[117] Cheryl Pellerin, *Project Maven to Deploy Computer Algorithms to War Zone by Year's End*, DEFENSE ONE (July 21, 2017), https://www.defense.gov/News/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/.

C. Humanitarian and Ethical Concerns

As with many advances in warfare, autonomy raises legitimate humanitarian concerns. These concerns are heightened by modern conflicts' disproportionate toll on civilians.[118] As wars are increasingly fought in urban areas rather than in trenches, civilian deaths have greatly outpaced those of combatants.[119] This Section focuses primarily on the jus in bello (i.e., the rules governing the conduct of hostilities) concerns presented by autonomous weapons, including the potential for indiscriminate or excessive attacks resulting from unintended engagements, the ethical considerations, and the potential accountability gap.[120]

A primary concern with autonomous weapons is that they may malfunction in a manner that causes mass harm to civilians.[121] Even if autonomy can enhance discrimination in the vast majority of cases, the consequences of machine error could be tragic.[122] Malfunctions are also inevitable.[123] Artificial intelligence can exceed human capabilities when performing a defined task in a controlled environment,[124] but it is brittle.[125] AI systems are not good at adapting to unforeseen events.[126] Enemy forces may also cause AI systems to fail by

---

[118]   Geoffrey S. Corn, *Humanitarian Regulation of Hostilities: The Decisive Element of Context*, 51 VAND. J. TRANSNAT'L L. 763, 763–64 (2018).

[119]   *See, e.g.*, *id.* at 764 ("[T]here is no question that beginning with World War II, the ration of civilian to military casualties in war has steadily increased."); Valerie Epps, *Civilian Casualties in Modern Warfare: The Death of the Collateral Damage Rule*, 41 GA. J. INT'L & COMP. L. 307, 309 (2013) ("[M]odern wars invariably result in far more civilian deaths than military deaths."); *Highest Recorded Civilian Deaths from Conflict at Mid-Year Point*, UNAMA (July 15, 2018), https://unama.unmissions.org/highest-recorded-civilian-deaths-conflict-mid-year-point-latest-unama-update.

[120]   Critics of autonomous weapons also warn that this technology will destabilize international peace and security by creating arms races and lowering the thresholds for the use of force. Heyns, *supra* note 8, ¶ 58 (stating that autonomous weapons "may thus lower the threshold for States for going to war or otherwise using lethal force, resulting in armed conflict no longer being a measure of last resort").

[121]   Paul Scharre, *A Million Mistakes a Second*, FOREIGN POL'Y (Sept. 12, 2018, 8:00 AM), https://foreignpolicy.com/2018/09/12/a-million-mistakes-a-second-future-of-war/.

[122]   DEF. SCI. BOARD, *supra* note 23, at 19 (noting the potential for high-regret outcomes in complex scenarios).

[123]   *See* Crootof, *supra* note 43, at 1373–74 (noting the likelihood of malfunction in complex systems).

[124]   AI systems have defeated human champions in contests such as Jeopardy, chess, and Go. The AlphaGo AI system that defeated the Go world champion, Lee Sedol, in 2016 trained based on compiling expert human moves. Subsequently, the AlphaGo Zero "use[d] a method call reinforcement learning, free of human guidance." AlphaGo Zero defeated the human-trained AlphaGo system 100-0 in Go matches, representing a major advance in AI learning. Satinder Singh, *Learning to Play Go from Scratch*, 550 NATURE 336, 336 (2017); *see also* Sarah Knapton, *AlphaGo Zero: Google DeepMind Supercomputer Learns 3,000 Years of Human Knowledge in 40 Days*, TELEGRAPH (Oct. 18, 2017, 6:00 PM), https://www.telegraph.co.uk/science/2017/10/18/alphago-zero-google-deepmind-supercomputer-learns-3000-years/.

[125]   Scharre, *supra* note 121.

[126]   *Id.*

corrupting training data sets or by hacking their software.[127] As Scharre cautions, "When [AI systems] do fail, they often fail badly."[128] This brittleness poses a significant concern in warfare.[129] It is impossible to program the myriad situations that a weapon system may confront in battle, and the consequences of an autonomous weapon malfunction could be catastrophic, "causing large-scale fratricide, civilian casualties, or even unintended attacks on adversaries."[130] Autonomous weapons that cannot be aborted or recalled in the event of malfunction pose particular concerns given that they may continue to attack until depleted of energy or ammunition.[131]

Opponents of autonomous weapons also warn that machine warfare may eliminate the last vestiges of compassion or honor in warfare.[132] Combatants may elect not to kill an enemy, even if legally permitted, if they believe it would violate a moral norm.[133] A soldier may find it repugnant, for example, to kill an enemy in a completely defenseless position, such as when he is sleeping or bathing.[134] Machines likely cannot make such moral judgments. A former U.N. Special Rapporteur, Christof Heyns, argues that machines "lack morality and mortality, and should as a result not have life and death powers over humans."[135] Distancing or even removing the human from these lethal decisions may have broader social consequences. As one commentator notes, "[t]he psychological distance eliminates the application of the shared sense of humanity all humans have, whether conscious or not."[136] Professor Wagner similarly notes that autonomous weapons may contribute to the "dehumanization of killing."[137]

A related ethical argument is that giving machines the ability to make life-and-death decisions disregards the human dignity of combatants.[138] Delegating

---

[127]   Highnam, *supra* note 82, at 3.
[128]   SCHARRE, *supra* note 19, at 146; *see also* Hall, *supra* note 33, at 86, 89 (noting that autonomous systems are "vulnerable to an array of potential failures" and that as "the complexity of a system increases, so does that operational risk").
[129]   Scharre, *supra* note 121.
[130]   *Id.*
[131]   *See* Amitai Etzioni & Oren Etzioni, *Pros and Cons of Autonomous Weapons Systems*, 2017 MIL. REV., at 72, 79.
[132]   *See, e.g.*, Mull, *supra* note 92, at 521–27 (arguing that autonomous weapons may undermine the principle of honor and ethical considerations in warfare).
[133]   *Id.* at 524–25.
[134]   MICHAEL WALZER, JUST AND UNJUST WARS 138–44 (2003 3rd ed.).
[135]   Heyns, *supra* note 8, para. 94.
[136]   Mull, *supra* note 92, at 525.
[137]   Markus Wagner, *The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems*, 47 VAND. J. TRANSNAT'L L. 1371, 1410 (2014).
[138]   Int'l Comm. of the Red Cross, Ethics and Autonomous Weapon Systems: An Ethical Basis for Human

such decisions to machines eliminates human agency and reduces combatants to objects.[139] As the ICRC states, "the central argument here is that it matters not just if a person is killed and injured but *how* they are killed and injured."[140] Human intent needs to be linked to the outcome of an attack in order to preserve moral accountability and the ability to determine whether an individual was "justly" killed.[141]

Finally, there is a concern that autonomous weapons will "result in a potential accountability gap or vacuum."[142] This accountability gap will be driven by several factors. First, determining how or why an autonomous weapon malfunctioned may require examining its coding as well as the information available to the commander.[143] Militaries may legitimately refuse to release this information if it is classified.[144] Second, as the causal chain in uses of force becomes longer or more complex, it becomes increasingly difficult to identify individual liability in the event that tragedies occur.[145] In many cases, it may even be impossible to understand why the machine acted as it did. As Scharre notes, "It's almost certain that as AI becomes more complicated, we'll understand it less and less."[146] Third, to state the obvious, machines do not have legal personality and cannot be held accountable.[147] Machine malfunction may lead to human tragedy, but no human may be accountable.[148] This Article will return to the issue of accountability in Part VI.

Control?, CCW/GGE.1/2018/WP.5, ¶ 27 (Mar. 29, 2018) [hereinafter Ethics and Autonomous Weapon Systems].

[139] Mull, *supra* note 92, at 524–25.

[140] Ethics and Autonomous Weapon Systems, supra note 138, ¶ 27.

[141] *Id.*

[142] Heyns, *supra* note 8, ¶ 77; *see also* Wagner, *supra* note 137, at 1402 (referring to the potential for "organized irresponsibility" in which no individual is responsible for actions of an autonomous weapon system).

[143] *See* Int'l Comm. of the Red Cross, Expert Meeting on Autonomous Weapon Systems: Technical, Military, Legal, and Humanitarian Aspects (Mar. 26-28, 2014); Bonnie Docherty, *The Lack of Accountability for Killer Robots*, HUMAN RIGHTS WATCH (Apr. 9, 2015), https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots.

[144] *Cf.* Wagner, *supra* note 137, at 1406–07 (noting that states would likely not release weapons reviews results due to the "highly sensitive nature" of autonomous weapons systems).

[145] Andrew Leveringhaus, *Autonomous Weapons Mini-Series: Distance, Weapons Technology and Humanity in Armed Conflict*, HUMANITARIAN LAW & POL'Y (Oct. 6, 2017), https://blogs.icrc.org/law-and-policy/2017/10/06/distance-weapons-technology-and-humanity-in-armed-conflict/ (noting that "causal distance, in particular, poses a longstanding challenge to IHL").

[146] SCHARRE, *supra* note 19, at 186.

[147] *See generally* Bartosz Brożek & Marek Jakubiec, *On the Legal Responsibility of Autonomous Machines*, 25 ARTIFICIAL INTELLIGENCE & L. 293, 293 (2017).

[148] Docherty, *supra* note 143.

The potential costs and benefits of autonomy in warfare explain why the ongoing discussions among States at the CCW, and in the academic literature, are so divided.[149] Autonomous weapons have the potential to provide immense military capabilities.[150] Technologically advanced States will seek to develop and militarize this technology to gain competitive advantages over their adversaries.[151] The concerns surrounding autonomous weapons, however, are undeniably valid. As with many advances in weapons technology, AI has the potential to undermine international peace and security and inflict harm on countless individuals. The debate regarding whether autonomous weapons should be banned or regulated will likely continue for the foreseeable future. This Article does not seek to enter that debate, as it assumes that weapons with significant degrees of autonomy will be deployed in armed conflict.[152] The more pressing question is how to address this looming reality. How can IHL regulate this new technology that will fundamentally change how war is waged?

III. AUTONOMOUS WEAPONS ARE NOT CATEGORICALLY UNLAWFUL

The starting point for any inquiry regarding how IHL can regulate autonomous weapons is to determine whether these weapons are categorically prohibited by any rule of customary IHL or treaty. This Part explains why the answer to that question is almost certainly no.

Customary IHL regulates both the types of weapons employed in armed conflict and the use of those weapons.[153] The latter is subject to much greater regulation than the former. Only two categories of weapons are per se banned as

---

[149] The significant disparity in views regarding the risks and benefits of autonomous weapons can be attributed, at least in part, to the vastly different conceptualizations of how these weapons would function. NGOs such as Human Rights Watch are focused primarily on the concerns presented by fully autonomous weapons. States seeking to defend autonomous weapons, by contrast, are more focused on the types of weapons that they would seek to develop.

[150] *See, e.g.*, Scharre, *supra* note 121.

[151] *See* Kenneth Anderson & Matthew Waxman, *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, HOOVER INST. (Apr. 9, 2013), https://www.hoover.org/research/law-and-ethics-autonomous-weapon-systems-why-ban-wont-work-and-how-laws-war-can.

[152] There is a possibility that some States may conclude a treaty to ban autonomous weapons amongst the parties to that treaty. Such a treaty would have limited practical effect, as the major military powers are exceedingly unlikely to agree to any additional limits on their ability to develop and deploy weapons with technologies that offer a military advantage. *See, e.g.*, *id.* at 21 ("For starters, limitations on autonomous military technologies, although quite likely to find wide superficial acceptance among some states and some non-governmental groups and actors, will have little traction among those most likely to develop and use them.").

[153] *See, e.g.*, Int'l Comm. of the Red Cross, *Rule 70. Weapons of a Nature to Cause Superfluous Injury or Unnecessary Suffering*, IHL DATABASE, https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_cha_chapter20_rule70 (last visited Jan. 30, 2020) [hereinafter *ICRC Rule 70*].

a matter of customary IHL in both international armed conflicts (IACs) and non-international armed conflicts (NIACs).[154]

First, customary IHL categorically prohibits weapons that are of a nature to cause "superfluous injury" or "unnecessary suffering."[155] This rule does not prohibit weapons merely because they cause tremendous suffering, provided such suffering is necessary to achieve a military purpose.[156] The objective of warfare is to kill or injure enemy combatants; even weapons that may cause horrific injuries in pursuit of this objective (such as incendiary weapons) are generally considered lawful when directed against combatants.[157] Nor does this rule prohibit weapons that may occasionally cause unnecessary suffering.[158] The key question is "whether the weapon inevitably breaches the principle in all designed or intended applications, not whether it is capable of use in a way that would breach the principle."[159] In other words, the rule focuses on the weapon's design rather than its particular effects in any given case.[160] An example of a weapon prohibited by this rule is one whose primary effect is to "injure by fragments which in the human body escape detection by x-rays."[161] Undetectable fragments aggravate injuries of soldiers already out of combat and thus cause suffering that does not serve a military purpose.[162]

---

[154] WILLIAM H. BOOTHBY, WEAPONS AND THE LAW OF ARMED CONFLICT 46–47 (2nd ed. 2016); *ICRC Rule 70*, *supra* note 153; Int'l Comm. of the Red Cross, *Rule 71. Weapons That Are By Nature Indiscriminate*, IHL DATABASE https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule71 (last visited Jan. 30, 2020) [hereinafter *ICRC Rule 71*].

[155] Convention Respecting the Laws and Customs of War on Land art. 23(e), Oct. 18, 1907, 36 Stat. 2277, T.S. No. 539 [hereinafter Hague IV]; Protocol Additional to the Geneva Conventions of 12 August 1949 and Relating to the Protection of Victims of International Armed Conflicts art. 35(2), *opened for signature* July 8, 1977, 1125 U.N.T.S. 3. [hereinafter Additional Protocol I]; *ICRC Rule 70, supra* note 153 ("The use of means and methods of warfare which are of a nature to cause superfluous injury or unnecessary suffering is prohibited.").

[156] BOOTHBY, *supra* note 154, at 50.

[157] Michael W. Meier, *Lethal Autonomous Weapons Systems (LAWS): Conducting a Comprehensive Weapons Review*, 30 TEMP. INT'L & COMP. L.J. 119, 128 (2016). ("The prohibition of unnecessary suffering constitutes acknowledgment that necessary suffering to combatants is lawful …"); Protocol III to the Convention on Certain Conventional Weapons, however, imposes restrictions on the use of incendiary weapons to protect against harm to civilians. Protocol on Prohibitions or Restrictions on the Use of Incendiary Weapons to the Convention on Certain Conventional Weapons art. 2, *opened for signature* Apr. 10, 1981, 19 I.L.M. 1523 [hereinafter Protocol III]; Protocol on Prohibitions or Restrictions on the Use of Incendiary Weapons art. 1, Apr. 10, 1981, 19 I.L.M. 1524.

[158] BOOTHBY, *supra* note 154, at 53–54.

[159] *Id.* at 58.

[160] *Id.* at 49.

[161] Protocol on Non-Detectable Fragments to the Convention on Certain Conventional Weapons, *opened for signature* Apr. 10, 1981, 19 I.L.M 1523 [hereinafter Protocol I]; *see also ICRC Rule 70*, *supra* note 153 (noting examples of weapons that have been cited in practice as causing unnecessary suffering).

[162] *See* Int'l Comm. of the Red Cross, *Weapons That May Cause Unnecessary Suffering or Have*

Second, customary IHL prohibits weapons that are inherently indiscriminate.[163] This prohibition covers only those weapons that are incapable of being directed at a specific military objective or that have effects that cannot be contained as required by IHL.[164] Again, the key consideration is the design of the weapon rather than its effects in any given situation. Most weapons can lawfully be used in at least some circumstances.[165] Even inaccurate or "dumb" weapons are rarely prohibited per se, as they are capable of being employed lawfully in environments where few, if any, civilians are present.[166] Examples of inherently indiscriminate weapons include balloon bombs[167] and certain types of unguided rocket systems,[168] since they cannot be directed at a specific military target. Biological weapons are also considered indiscriminate since their effects cannot be contained.[169]

Treaties may prohibit other weapons that are not banned as a matter of customary IHL. The Ottawa Convention, for example, prohibits States Parties from using, developing, producing, acquiring, retaining or transferring anti-personnel mines.[170] The Oslo Convention similarly prohibits the use of cluster munitions by States Parties to that Convention.[171] Many major military powers, such as the United States, are not parties to these conventions, which limits their effectiveness.

Indiscriminate Effects: Report on the Work of Experts ¶¶ 21, 22, 147, 151 (1973).

[163]   *ICRC Rule 71*, *supra* note 154.

[164]   Additional Protocol I, *supra* note 155, art. 51(4)(b)–(c); *see also* BOOTHBY, *supra* note 154, at 67.

[165]   Stephen Townley, *Indiscriminate Attacks and the Past, Present, and Future of the Rules/Standards and Objective/Subjective Debates in International Humanitarian Law*, 50 VAND. J. TRANSNAT'L. 1223, 1226 (2017).

[166]   Schmitt, *supra* note 52, at 10 (noting that SCUD missiles, although inaccurate, were not unlawful per se because there are situations in which they can be deployed in a discriminate manner); *see also* Townley, *supra* note 165, at 1226 (stating that "most weapons and methods of attack are not inherently indiscriminate, because one could imagine a set of circumstances in which they could be used lawfully").

[167]   *ICRC Rule 71*, *supra* note 154. Balloon bombs date back to World War II, but they have also been employed in recent conflicts. During the unrest in the Gaza Strip in 2018, Palestinians attached gas-soaked rags to inflated condoms and flew them into Israeli territory. *See* Alisa Odenheimer, *Flaming Condoms, Kites, Balloons From Gaza Used to Set Fields Ablaze in Southern Israel*, CHI. TRIB. (June 21, 2018), https://www.chicagotribune.com/nation-world/ct-flaming-condoms-from-gaza-20180621-story.html.

[168]   BOOTHBY, *supra* note 154, at 220 n.6 (noting that the "VI and V2 rocket systems used by Germany during World War II have been cited as examples of weapons which today would breach the principle" of discrimination).

[169]   Int'l Committee of the Red Cross, *Rule 73. Biological Weapons*, IHL DATABASE, https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule73 (last visited Oct. 11, 2019) [hereinafter *ICRC Rule 73*] ("The use of biological weapons is prohibited.").

[170]   Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction art. 1, Sept. 18, 1997, 2056 U.N.T.S. 211. This treaty is more commonly referred to as the Ottawa Convention. *See Id.*

[171]   Convention on Cluster Munitions art. 1, May 30, 2008, 2688 U.N.T.S. 39. This treaty is often referred to as the Oslo Convention. *See id.*

There is no treaty or customary international law ban on the use of autonomous weapons per se.[172] Similarly, the use of autonomy is not subject to any specific regulation in IHL. As with all new weapon systems, States will need to undertake a rigorous legal review of each new autonomous weapon before deployment to determine whether it would be prohibited on the grounds that it is of a nature to cause unnecessary suffering or is inherently indiscriminate. Certain autonomous weapons may fail this legal review, but there is no basis to claim that all autonomous weapons would be prohibited by IHL.

Autonomy in weapon systems is not, by its nature, likely to cause unnecessary suffering.[173] The autonomous features of a weapon implicate the decision-making process for when and where force is deployed rather than the nature of the force itself.[174] While it is conceivable that an autonomous weapon could learn to systematically deploy force in a particularly cruel manner, such a weapon would be the rare exception rather than the norm.[175] There is nothing intrinsic to AI that suggests a weapon with autonomous functions would be of a nature to cause superfluous suffering.

Likewise, there is no ground to believe, as some NGOs suggest,[176] that autonomous weapons would be inherently indiscriminate. The presence of a human in or on the loop does not determine whether an autonomous weapon is capable of being used consistent with the principle of discrimination.[177] AI, combined with the development of sophisticated sensors, has made many weapons more discriminate. We can also imagine how autonomous weapons could be used in a discriminate manner. Autonomous submarine hunters programmed to search for certain hydrophonic signatures unique to enemy submarines would pose little risk to civilians or civilian objects. While it is certainly the case that autonomous weapons could be indiscriminate if appropriate safeguards are not in place, this observation is not unique to weapons with autonomy.

The fact that autonomous weapons are not per se prohibited by IHL does not end the inquiry. IHL rarely prohibits weapons per se, but it does restrict how

---

[172]   BOOTHBY, *supra* note 154, at 253.
[173]   Schmitt, *supra* note 52, at 9.
[174]   *See id.*
[175]   *Id.* at 35.
[176]   Docherty, *supra* note 143, at 938–39 (stating that fully autonomous weapons would be unable to comply with the rules of distinction).
[177]   Schmitt, *supra* note 52, at 13 ("a man in the loop is not a panacea during situations in which it may be difficult to distinguish civilians and civilian objects from combatants and military objectives.").

they can be used in armed conflict.[178] The development of autonomous weapons will challenge how IHL regulates human decision-making over the use of force. Part V examines some of these challenges.

IV. CHALLENGES IN APPLYING IHL TO AUTONOMOUS WEAPONS

Applying IHL to the use of autonomous weapons will present challenges not merely because AI and machine learning are novel and emerging technologies. Autonomy will significantly alter how war is waged and humans' role in hostilities. This transformation will upend key assumptions about warfare that have shaped the development of IHL. Section A discusses how the concept of the reasonable commander informs our understanding of combatants' obligations under IHL. Section B examines how autonomy will strain this core concept.

A. *The Reasonable Commander*

IHL is a uniquely human-centric legal framework in the sense that it focuses on the reasonableness of combatants' judgments and decisions rather on their effects. The provisions governing the conduct of hostilities generally do not impose concrete prohibitions or permissions that can be assessed based on observable evidence in the same way that the rules governing a game of soccer do. Many of IHL's provisions governing the conduct of hostilities are "standard-like" rather than "rule-like." Even those provisions that may appear rule-like "also contain standard-like elements."[179] Standards are more "indeterminate [than rules] in their application."[180] Standards seek to further a social value by promoting certain norms of behavior, but do not require particular outcomes given the importance of context. Rather, standards employ "'evaluative' criteria, such as reasonableness, good faith, or due care" in mandating norms of behavior.[181]

The "rule-like" and "standard-like" provisions of IHL require combatants to make certain factual *determinations* and evaluative *judgments* when using

---

[178]   *Id.* at 35.
[179]   Townley, *supra* note 165, at 1231. For example, combatants must take "all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss of civilian life." Additional Protocol I, *supra* note 154, art. 57(2)(a)(ii).
[180]   Russel D. Covey, *Rules, Standards, Sentencing and the Nature of Law*, 104 CAL. L. REV. 447, 459 (2016).
[181]   *Id.*

force.[182] A factual determination required by IHL may be whether an individual is an enemy combatant or "directly participating in hostilities," rendering him a legitimate target.[183] These factual determinations are often based on observable and objective evidence. An individual wearing an enemy uniform and carrying a weapon may easily be designated an enemy combatant. In other circumstances, however, this determination may not be so clear. Assessing whether an individual is directly participating in hostilities may require a more contextualized assessment of the individual's conduct. A shepherd carrying an AK-47 may not raise suspicions in countries where it is common to carry semi-automatic weapons. This shepherd may be deemed hostile though if he also carries a satellite phone and is positioned on a trail leading to a village occupied by enemy forces. A degree of professional intuition is often required in making these contextualized determinations.

The judgments required by IHL are often subjective in nature, especially when they require a balancing of competing values. The principle of proportionality, discussed in greater depth *infra*, requires a commander to determine whether an attack "may be expected" to cause collateral damage that would be "excessive to the military advantage anticipated."[184] This principle requires the balancing of two inapposite values—military advantage and civilian life—for which there is no agreed metric.[185]

Combatants' decisions and judgments are assessed based on a human-oriented standard: reasonableness.[186] The reasonableness standard is the

---

[182] DEP'T OF DEFENSE, DEPARTMENT OF DEFENSE LAW OF WAR MANUAL § 5.11.4 (2015) [hereinafter DoD LAW OF WAR MANUAL]. The specific determinations that a soldier is required to make will depend on his or her rank and role in the operation. *Id.*

[183] Townley, *supra* note 165, at 1255 n.174 (citation omitted).

[184] Additional Protocol I, *supra* note 154, art. 51(5)(b); Ian Henderson & Kate Reece, *Proportionality Under International Humanitarian Law: The "Reasonable Military Commander" Standard and Reverberating Effects*, 51 VAND. J. TRANSNAT'L L. 835, 836 (2018).

[185] MICHAEL BOTHE, KARL JOSEF PARTSCH & WALDERMAR A. SOLF, NEW RULES FOR VICTIMS OF ARMED CONFLICTS 310 (noting that proportionality assessment "requires a comparison of values which cannot be compared"); *see also* United Nations International Criminal Tribunal for the Former Yugoslavia, *Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia*, ¶ 50, 39 ILM 1257 (June 8, 2000) [hereinafter NATO Bombing Report] (describing the difficulties in assessing proportionality).

[186] *See, e.g.,* Prosecutor v. Galic, Case No. IT-98-28-T, Judgement, ¶ 58 (Int'l Crim. Trib. for the Former Yugoslavia Dec. 5, 2003) ("In determining whether an attack was proportionate it is necessary to examine whether a reasonably well-informed person in the circumstances of the actual perpetrator, making reasonable use of the information available to him or her, could have expected excessive casualties to result from the attack."); United Nations General Assembly, *Report of the United Nations Fact-Finding Mission on the Gaza Conflict*, A/HRC/12/48, Sept. 25, 2009 (applying the reasonable commander standard).

"touchstone of compliance with almost all [IHL] targeting rules."[187] Soldiers do not necessarily violate IHL if their determinations or judgments ultimately prove incorrect, as the fog of war makes absolute standards unrealistic. Instead, soldiers' determinations and judgments must be reasonable and made in good faith.[188]

The reasonableness standard has several significant implications for how we conceptualize IHL compliance. First, a combatant's knowledge and intent are generally more probative than the effects of his or her conduct. Civilian casualties do not necessarily mean that the attacking force violated IHL.[189] The question will be whether a commander or soldier *intentionally* targeted civilians or acted *recklessly* with respect to his obligation to discriminate between civilians and combatants and to avoid attacks that would result in disproportionate civilian casualties.

Assessments of IHL compliance must be based on the facts reasonably available to the commander at the time of attack rather than on information obtained with the benefit of hindsight.[190] This principle is reflected in various articles of Additional Protocol I to the Geneva Conventions (API), as well as decisions of international criminal tribunals. Article 52 of API, for example, defines "military objective" based on whether an object's destruction would confer a definite military advantage "in the circumstances ruling at the time."[191] The obligation to take "feasible precautions" to protect civilians likewise requires an assessment of the circumstances ruling at the time.[192] The principle of proportionality requires a commander to determine the "expected" collateral

---

[187]   Geoffrey S. Corn, *Humanitarian Regulation of Hostilities: The Decisive Element of Context*, 51 VAND. J. TRANSNT'L L. 763, 768 (2018) (noting that "'reasonableness' is the touchstone of compliance with judgment-based LOAC targeting rules").

[188]   DOD LAW OF WAR MANUAL, *supra* note 182, at § 5.3 ("Assessing Information Under the Law of War").

[189]   *See* Charles P. Trumbull IV, *Re-Thinking the Principle of Proportionality Outside of Hot Battlefields*, 55 VA. J. INT'L L. 521, 541 (2015) ("IHL implicitly authorizes belligerents to kill civilians in the course of targeting military objectives, provided that the civilian casualties are not 'excessive'").

[190]   Prosecutor v. Gotovina, Case No. IT-06-90-A, Application and Proposed Amicus Curiae Brief, ¶ 9 (Int'l Crim. Trib. for the Former Yugoslavia Jan. 12, 2012) ("Both the laws and customs of war and the jurisprudence of international war crimes tribunals validate the conclusion that the legal standard for review of a targeting decision must be made based on the assessment of the situation confronting the commander at the time of the attack."); Stefan Oeter, *Methods and Means of Combat, in* THE HANDBOOK OF HUMANITARIAN LAW IN ARMED CONFLICT 179 (Dieter Fleck, ed., 1995).

[191]   NATO Bombing Report, *supra* note 185, ¶ 21 (noting that the "military worth of the target would need to be considered in relation to the circumstances prevailing at the time").

[192]   Convention on Prohibition of Restrictions on the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to Have Indiscriminate Effects, Protocol II, art. 3, Oct. 10, 1980, 1342 U.N.T.S. 137.

damage and "anticipated" military advantage that would result from an attack.[193] These assessments are "predictive in nature"[194] and thus necessarily dependent on the information reasonably available to the commander at the time of the attack rather than on information gained subsequent to an attack. Combatants in the heat of battle may make mistakes or need to act based on imperfect information. IHL demands only that soldiers act reasonably based on the available information.[195]

Second, the laws governing the conduct of hostilities are context dependent. Myriad factors may inform what steps a "reasonable commander" would take to mitigate risks to civilians in a particular situation, including the operational context and tempo, the risk to friendly forces, and the importance of the mission.[196] Given the importance of context, soldiers' specific obligations under IHL are often indiscernible in the abstract. A precaution designed to protect civilians (such as giving advance warning of an attack) that is feasible in one context may not be feasible in another.[197] In many situations, a commander can only determine what IHL permits or requires at the time of implementation, taking into account the circumstances at the time.

Third, combatants have a degree of discretion in making the determinations and judgments required by IHL.[198] In many cases, there will be no single "correct" course of action. Commanders may reasonably come to different conclusions on what precautions are feasible, the importance of destroying a military objective, or the value of a civilian life. These judgments may vary based on subjective elements, such as prior experiences, training, and culture. IHL accepts that commanders will make different decisions confronted with the same facts. A commander's judgment, however, must fall within a "range of

---

[193]   Additional Protocol I, *supra* note 154, art. 51(5)(b); Henderson & Reece, *supra* note 185, at 836.

[194]   Corn, *supra* note 188, at 773.

[195]   Prosecutor v. Gotovina, Case No. IT-06-90-A, Application and Proposed Amicus Curiae Brief, ¶ 8 (Int'l Crim. Trib. for the Former Yugoslavia Jan. 12, 2012) ("In short, the legal standard does not impose an obligation to always be right: it imposes and obligation to make a reasonable decision based on the information available at the time."); BOTHE, PARTSCH & SOLF, *supra* note 185, at 310 ("The best that can be expected of the decision maker is that he act honestly and competently.").

[196]   *See generally* DOD LAW OF WAR MANUAL, *supra* note 182, at § 5.

[197]   *Id.* at § 5.11 ("what precautions are feasible depends greatly on the context").

[198]   INT'L COMM. OF THE RED CROSS, COMMENTARY ON THE ADDITIONAL PROTOCOLS OF 8 JUNE 1977 TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949, at 684, ¶ 2210 (noting that the provision on feasible precautions "allows for a fairly broad margin of judgment …"); Frederick Schauer, *The Tyranny of Choice and the Rulification of Standards*, 14 J. CONTEMP. LEGAL ISSUES 803, 804 (2005) ("… standards leave most of the important choices to be made by the subject, the enforcer, or the interpreter, and leave them to be made at the moment of application.").

appreciation" of what a hypothetical reasonable commander would do in the same situation.[199]

B. *Autonomy, the Evolving Role of Humans in Warfare, and Reasonableness*

Autonomy will transform warfare in ways that directly implicate the traditional understandings and applications of IHL, including that of the reasonable commander. Autonomous systems will render obsolete underlying assumptions regarding humans' primary role in armed conflict, blurring the distinction between weapon and operator. IHL, however, only regulates human conduct. Machines do not have obligations under IHL.[200] Even if machines can ultimately make distinction and proportionality determinations, discussed in Part VI, they do not have an obligation to do so.[201] Targeting decisions that are currently regulated by IHL (because they are made by humans) will no longer be directly regulated if they are made by a weapon system, although IHL would still apply to the human decision to deploy the weapon. A submarine commander who attacks a fishing boat he confuses for an enemy ship can be held accountable for his mistake. An autonomous underwater vehicle that makes the same error would not itself violate IHL.

The increase in machine decision-making on the battlefield does not mean that humans can or will be mere observers in armed conflict, nor does it mean that humans will operate beyond the reach of IHL. "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."[202] Human decision-making on uses of force cannot be completely excised. Humans must decide which weapons to develop, how to program them, what tasks to assign to machines, and when and where to activate a weapon system.

Advances in autonomy, however, will significantly change how human judgment is exercised regarding uses of force. As discussed below, autonomy will allow (and potentially encourage) humans to deploy weapons with far less knowledge about the location, timing, or even ultimate target of the attack. Advances in autonomy and robotics may permit weapons to deploy for extended periods of time before engagement. Commanders will rely increasingly on information about the programming of the weapon system rather than on

---

[199] Corn, *supra* note 187, at 772 (emphasis omitted).

[200] *See* United States Working Paper, *Autonomy in Weapons Systems*, CCW/GGE.1/2017/WP.6, ¶¶ 10–13 (Nov. 10, 2017) [hereinafter *Autonomy in Weapons Systems*].

[201] *Id.*

[202] Schmitt, *supra* note 52, at 33.

knowledge of the particular circumstances where force is deployed. A relevant legal question today may be whether a commander is reasonably confident that a particular target is a military objective and not a civilian object. In the future, the inquiry may be whether the commander is reasonably confident that an autonomous weapon will be capable of determining that an object is a lawful target.

The delegation of greater decision-making to weapon systems will further complicate applications of human-oriented standards, such as reasonableness and good faith. First, autonomous weapons will be able to make certain targeting-related decisions after deployment, making the causal link between human decision-making and the ultimate deployment of force more attenuated. A commander may not be able to predict exactly how an autonomous weapon will react to unanticipated circumstances, and in some instances it may be impossible to reconstruct why an autonomous system acted as it did.[203] This attenuation will make it more difficult to assess whether a commander's decision to deploy the weapon was reasonable.

Second, the reasonableness standard requires a body of common experience among relevant professionals. The reasonableness inquiry entails assessing whether a commander's judgment falls "within a range of appreciation consistent with that of the hypothetical reasonable commander faced *with the same situation*."[204] It will take time to develop the professional norms necessary to inform what decisions and judgments a "reasonable commander" would make when deploying these weapons.[205] Consider an autonomous weapon system that starts to act erratically, but that does not pose an imminent threat to civilians. The commander may not know whether the weapon has malfunctioned or whether it has learned a legitimate (albeit unanticipated) behavior that will further the commander's ultimate objective. In case of the former, aborting the mission could reduce the risk to civilians by preventing an unintended engagement. If the latter, aborting the mission could nullify the military advantage anticipated and potentially expose the commander's troops to greater danger. How would a reasonable commander act in this situation? Until militaries gain more experience deploying autonomous weapons in combat,

---

[203]   Crootof, *supra* note 43, at 1373.

[204]   Corn, *supra* note 118, at 772.

[205]   Henderson & Reece, *supra* note 185, at 837, 839. Of course, certain decisions may be patently unreasonable, and the lack of a body of experience would not preclude such a determination in clear cases.

there may be little consensus on what constitutes a reasonable decision in a particular set of circumstances.[206]

Third, autonomy will further complicate the *ex ante* approach to determining IHL obligations and compliance by creating what this Article refers to as an "uncertainty problem." Commanders will deploy autonomous weapons to obtain a military effect, but may not know when, where, or how that effect will be achieved. This technology will accordingly create uncertainty on three levels: geographic, temporal, and functional. Loitering weapons may be deployed for hours, days, or even weeks before engaging the target. Similarly, search and destroy weapons may cover significant distance before hitting a target. In both cases, the human operator cannot have concrete knowledge of the circumstances at the moment force is deployed. Commanders may also deploy weapons despite having limited insight into their decision-making processes.[207] How can a commander determine that his intended target is a military objective and that the attack would not cause excessive collateral damage when he does not have knowledge of the timing or location of the ultimate attack? How can society assess whether that decision is reasonable?

Autonomy will inevitably alter the type and sources of information that a commander will need to consider in determining whether to deploy a weapon system in a particular operational context. The commander will likely focus less on information concerning the specific target and more on the weapon system's programming, capabilities, and operating environment in making the judgments required by IHL. Discrimination decisions, for example, will be based increasingly on knowledge of the weapon system's ability to attack only the intended targets in the circumstances in which it is deployed.

This evolution in the type of information that will inform legal judgments raises a separate challenge, which I refer to as the "distributed knowledge" problem. When authorizing attacks, a commander will rely on inputs from a

---

[206]   *See* Merel Ekelhof, *Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation*, 10 GLOB. POL. 343, 343–44 (2019). The reasonableness-standard employed to assess IHL compliance is already under significant pressure, particularly among those seeking greater accountability for civilian harm. The application of standards, the subjectivity of decision-making, the importance of context, the fog of war, the inherent secrecy of military operations, and the difficulty of proving mens rea pose challenges for accountability efforts. Even sophisticated observers cannot, in many cases, conclusively determine that an attack resulting in civilian casualties constitutes a violation of IHL. For this reason, there is an increasing desire to establish "rule-like" requirements in IHL that can be enforced based on empirical, publicly available, evidence. The emergence of autonomous weapons will further muddle this reasonableness inquiry, likely fueling calls to replace IHL standards with bright line rules.

[207]   *See* SCHARRE, *supra* note 19, at 180–88.

potentially large number of disparate sources. To be sure, IHL does not prohibit soldiers from making targeting decisions based on second-hand information. In modern conflicts, soldiers are increasingly less likely to directly observe and engage their intended target. Rather, as technology allows militaries to deploy force from beyond the traditional front lines, operators and commanders must frequently rely on information from other sources to inform the judgments required by IHL.[208]

The use of autonomous weapons, however, will further distance the commander, geographically and temporally, from the kinetic action.[209] This means that the number of sources with information relevant to the targeting decision is also likely to increase, particularly when the timing and location of an attack are unknown. These sources may include computer programmers, the weapon's testers, intelligence units and friendly forces, satellite imagery, weather forecasters, and the weapon's sensors. The individuals (or machines) providing this data may likewise be acting on incomplete or second-hand information. Officials responsible for testing and reviewing the weapon, for example, may not be aware of all the circumstances in which it may be used or how it has evolved since it was initially tested. Similarly, "programs with millions of lines of code are written by teams of programmers; none of whom knows the entire program."[210]

The distributed knowledge effect may frustrate those who seek to determine individual accountability for civilian casualties. When using complex systems, minor mistakes can have a cascade effect that ends in catastrophe. Yet, no human may be held responsible for the ensuing harm if they acted in good faith.[211] It is a lamentable fact of war that civilians may be killed even when parties to the conflict strictly comply with their legal obligations. Seeking to hold commanders strictly liable for machine malfunction would create a significant military disadvantage and would not promote greater IHL compliance. As machines undertake roles traditionally performed by humans, however, there is

---

[208]   *See, e.g.,* Ekelhof, *supra* note 207, at 346 (describing the inputs an F-16 pilot will rely on in executing a deliberate strike against a military compound). As the Department of Defense Law of War Manual notes, "in a long-distance attack, a commander may rely on information obtained from aerial reconnaissance and intelligence units in determining whether to conduct an attack." DOD LAW OF WAR MANUAL, *supra* note 182, at § 5.4.3.

[209]   *See* Ford, *supra* note 29, at 445 (noting that autonomy may prolong the time between activation and engagement of the target).

[210]   Gary Marchant et al., *supra* note 50, at 284. As Marchant notes, the number of programmers involved means that "no individual can predict the effect of a given command with absolute certainty, since portions of large programs may interact in unexpected, untested ways." *Id.*

[211]   Crootof, *supra* note 43, at 1385.

a greater likelihood that no human will be held accountable when civilians are killed.

V.   APPLYING IHL TO THE USE OF AUTONOMOUS WEAPONS

The Group of Governmental Experts on LAWS concluded in 2018 that "'international humanitarian law continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems."[212] This is an important, albeit axiomatic, affirmation. Articulating how IHL would apply to the use of autonomous weapons in practice is substantially more complicated. This Part explains the IHL principles and rules that govern the development and use of weapons, including those with autonomous functions. Drawing on the discussion in Part V, it explains why the application of fundamental IHL principles to the use of autonomous weapons raises important questions and legal ambiguities that have not been adequately examined in the academic literature or the inter-State dialogue.

This Part focuses exclusively on how IHL may regulate the human judgments and decision-making regarding the use of autonomous weapons. The legality of an autonomous weapon system does not hinge, as some seem to suggest,[213] on whether the weapon system is capable of making the range of decisions or judgments that a human would make in combat. The technology required to fully replicate human decision-making across a range of tasks, known as "general AI," may never exist.[214] Moreover, States and combatants are responsible for implementing IHL. These obligations cannot be delegated to machines.[215] Accordingly, the important legal issues concern whether and how States and humans responsible for these weapons systems can act in accordance with IHL.

This Part focuses on three areas in which autonomy will significantly challenge traditional understandings of IHL. Section A discusses the first stages of the weapon life cycle: development, testing, and review. Section B examines how IHL might apply to the targeting process. Section C analyzes the challenges regarding accountability in the use of autonomous weapons.

---

[212]   *Report of the 2018 Session*, *supra* note 14, ¶ 21.

[213]   *See, e.g.,* Wagner, *supra* note 137, at 1387 ("AWS would have to be able to execute their combat operations in full compliance with these [IHL] rules.").

[214]   *See, e.g.,* Larry Lewis, *AI and Autonomy in War: Understanding and Mitigating Risks,* CNA 8 (2018) (stating that the existence of general AI for now and in the near future "can only be found in science fiction").

[215]   Neil Davison, *A Legal Perspective: Autonomous Weapons Systems Under International Humanitarian Law*, Perspectives on Lethal Autonomous Weapon Systems, UNODA Occasional Papers No. 30, at 7 (2017).

A. Testing and Review

The weapons review process is intended to ensure that States do not deploy new weapons that are prohibited by IHL. Article 36 of Additional Protocol I to the Geneva Conventions (API) states: "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."[216]

States have different procedures for weapons reviews, given their different treaty obligations. The United States practice is to determine, at a minimum, whether: (1) there is a specific law prohibiting or restricting the use of the weapon; (2) the weapon's intended use is calculated to cause superfluous injury; and (3) the weapon is inherently indiscriminate.[217] The focus of the review is on the planned or intended use of the weapon.[218] The attorney responsible for the review may also advise "whether other measures should be taken that would assist in ensuring compliance with law of war obligations related to the type of weapon being acquired or procured."[219] However, it is generally not the responsibility of the weapon reviewer to determine whether the use of a weapon would be legitimate "for a particular attack on a specified occasion."[220] The commander in the field, and his or her legal advisers, must make this determination.[221]

---

[216]   Additional Protocol I, *supra* note 154, art. 36. The United State is not a party to API and does not consider the text of Article 36 to reflect customary international law but has long required legal reviews of the intended procurement or acquisition of weapons or weapon systems as a matter of policy.

[217]   Meier, *supra* note 157, at 126.

[218]   *See* Department of the Navy Implementation and Operation of the Defense Acquisition System and the Joint Capabilities Integration and Development System, SECNAV INSTRUCTION 5000.2E, ¶ 1.6.1 (Sec. of the Navy, Sept. 1, 2011) (stating that weapons "shall be reviewed by the Judge Advocate General (JAG) of the Navy during the program decision process to ensure that the intended use of such weapons or systems is consistent with domestic and international law."); *see also* Netherlands and Switzerland Working Paper, *Weapons Review Mechanisms,* CCW/GGE.1/2017/WP.5, ¶ 3(d) (Nov. 7, 2017) ("However, it is of considerable importance that the review is limited to the normal, expected or intended use of the weapon. Almost all of the relevant sources identify the possibility of misuse or inventive abuse of any weapon and the review need not take all of those possible alternatives into consideration.").

[219]   DOD LAW OF WAR MANUAL, *supra* note 182, at § 6.2.2; *see also* Australia Working Paper, *Australia's System of Control and Applications for Autonomous Weapon Systems*, CCW/GGE.1/2019/WP.2, ¶ 17 (noting that this "review is a legal assessment to determine whether a weapon is (i) cleared, (ii) cleared subject to conditions, or (iii) not cleared for operational use.").

[220]   BOOTHBY, *supra* note 154, at 220.

[221]   *Id.*

Determining the reliability and predictability of a weapon system is critical to the weapon review process. Reliability refers to the rate at which a weapon system will perform in the same manner in a given set of circumstances. As Farrant and Ford note, "a weapon's systems reliability in performing as intended in its concept of employment will inform a great many of the substantive legal considerations."[222] In assessing whether a weapon is indiscriminate, for example, the reviewer will need to know its reliability in performing its intended function. A rocket launcher that overshoots its target fifty percent of the time will present greater concerns than one that misses only five percent of the time. The predictability of the weapon system—i.e., whether it performs as anticipated in different circumstances—similarly informs the legal considerations.[223] Even if a weapon system functions reliably in a controlled environment, its reviewer will need to be able to predict how the weapon will function in the various operational environments in which it is likely to be deployed in order to determine whether it can be used in a manner consistent with IHL.[224]

Establishing the reliability and predictability of weapons that have self-learning capabilities will present unique challenges.[225] As one author notes, "it is increasingly difficult for operators to predict with a high degree of probability how a system might actually perform against an adaptive adversary."[226] Unlike conventional weapons, such as landmines, weapons with self-learning capabilities constantly evolve.[227] These weapons will adjust their behavior, based on external inputs and an assessment of their performance in different environments, in order to maximize their reward function.[228] As the U.S.

---

[222]   Farrant & Ford *supra* note 103, at 410.

[223]   Schuller, *supra* note 17, at 408 (stressing the importance of being able to predict how autonomous weapons will operate once deployed).

[224]   Farrant & Ford *supra* note 103, at 411 ("a system that performed unpredictably would fail a weapon review if the reviewing lawyer could not be satisfied that the range of possible outcomes from a system's use was lawful."). *See* Lewis, *supra* note 214, at 14 (noting the importance of context and the need for weapon systems to be certified for use only in those contexts in which they can exercise discrimination effectively).

[225]   *See* Davison, *supra* note 164, at 10; *see also* Marchant et al., *supra* note 50, at 284 ("unpredictability in the behavior of complex robots is a major source of worry, especially if robots are to operate in unstructured environments …").

[226]   Hall, *supra* note 33, at 89; *see also* Paul Scharre, *Autonomous Weapons and Operational Risks*, CENTER FOR A NEW AMERICAN SECURITY 12 (2016) ("As a system becomes more complex, it becomes increasingly difficult for a human operator to predict precisely what the autonomous system might do in any given situation").

[227]   Artificial Intelligence will allow machines to develop tactics and strategies that humans cannot predict in advance. *Open AI Five*, OPENAI (June 25, 2018), https://blog.openai.com/openai-five/. To give one example, an AI bot team developed by OpenAI defeated human amateurs in Dota 2, a real time strategy game between two teams of five players. The AI team of five neural networks continuously improved its performance by playing 180 years' worth of games against itself every day, using reinforcement learning. *Id.*

[228]   Schuller, *supra* note 17, at 404 (describing how machines learn).

Defense Science Board Study on Autonomy acknowledges, self-learning software cannot be exhaustively tested because it "exhibits different behavior as it incorporates more data about its task, and learns to provide better results partly based on experience[.]"[229] Due to this self-learning capability, a system deployed in the field may not function in the same manner as when it was tested.[230] Moreover, self-learning machines may alter their behavior in ways that humans cannot foresee.

This inherent degree of unpredictability does not mean that autonomous weapons are unlawful. IHL does not require that weapons meet specific standards for reliability or predictability.[231] Prior attempts by international tribunals to establish fixed standards for weapons performance have proven controversial and unsuccessful.[232] As a number of IHL scholars wrote in an amicus brief to the International Criminal Tribunal for the former Yugoslavia (ICTY), a fixed standard could amount "to a strict liability standard of culpability" or an "implicit requirement that commanders be prepared to justify each and every effect produced by an attack."[233] This would be problematic because commanders cannot completely control the effects of an attack. As with the application of other IHL rules, context is important. Weapon performance is

---

[229]   DEF. SCI. BOARD, *supra* note 23, at 30.

[230]   Lewis, *supra* note 214, at 23.

[231]   *See* BOOTHBY, *supra* note 154, at 220 (noting that the "law as to the reliability of new weapons is limited to a voluntary assertion, based on best practice"); The Technical Annex to Protocol V of the Convention on Certain Convention Weapons, for example, encourages High Contracting Parties to take "best practice norms and operating procedures …" to enhance the reliability of explosive ordnance. These practices are not legally required, however, and there is no fixed legal standard for the reliability of weapons systems. *See* CCW Protocol V on Explosive Remnants of War, Technical Annex Art. 3 (2003).

[232]   Prosecutor v. Ante Gotovina, Case No. IT-06-90-A, Judgment, ¶¶ 58–61, 65 (Int'l Crim. Trib. for the Former Yugoslavia Nov. 16, 2012). The Gotovina case at the ICTY illustrates the complexities in seeking to establish fixed standards regarding weapon performance. Ante Gotovina, a General in the Croat Army, was accused of ordering unlawful artillery attacks against four cities in Croatia, with the purpose of removing the Serb population. The trial chamber determined, based on expert testimony, that the artillery used in the attack had a margin of error of 200 meters. Accordingly, based on an impact analysis, it concluded that shells landing more than 200 meters from a military objective were evidence of indiscriminate attacks. The Appeals Chamber reversed, finding that there was no basis for establishing a uniform margin of error for the artillery used in the attacks given the number of factors that could affect the artillery's reliability. After rejecting this margin of error, the Appeals Chamber concluded that it could not "exclude the possibility that all of the impact sites considered in the Trial Judgement were the result of shelling aimed at targets that the Trial Chamber considered to be legitimate." *Id.* ¶ 65. The Appeals Chamber decision has been criticized by some commentators, but it demonstrates the lack of international standards regarding the acceptable error rate of weapons, including conventional weapons such as artillery that have been in use for decades. *See, e.g.,* Jens David Ohlin, *Why the Gotovina Appeals Judgment Matters*, EJIL: TALK! (Dec. 21, 2012), https://www.ejiltalk.org/why-the-gotovina-appeals-judgment-matters/.

[233]   Prosecutor v. Gotovina, Case No. IT-06-90-A, Application and Proposed Amicus Curiae Brief, ¶ 8 (Int'l Crim. Trib. for the Former Yugoslavia Jan. 12, 2012), ¶ 19.

often influenced by both operational factors (e.g., distance and timing) and external factors (e.g., wind speed, temperature, precipitation), making a single standard problematic. The operating environment will also influence what is considered an acceptable margin of error. Weapons fired against enemy formations in unpopulated areas may not need the same degree of precision and predictability as weapons fired in urban areas.

As humans assign greater decision-making authority to machines, and as the cost of weapon malfunction increases, we may question whether this legal indeterminacy is appropriate. Should there be a greater requirement on States to ensure that weapons systems will perform as intended? For example, concern with the humanitarian risks posed by unexploded cluster munitions prompted the U.S. Department of Defense to adopt a policy that it would only "procure cluster munitions containing submunitions or submunition warheads that do not result in more than one percent unexploded ordnance across the range of intended operational environments, or that possess advanced features to minimize the risks posed by unexploded submunitions."[234] In principle, similar performance requirements could be established for autonomous weapons.

Universal standards for autonomous weapons reliability and predictability are not the answer to the challenges inherent in testing and reviewing such weapons. As a practical matter, it is unrealistic that States could agree to specific standards for various types of autonomous weapons. Developing countries that lack the technology to develop these weapons would seek to impose high standards while major military powers would seek to preserve their flexibility. Even if States did agree on a single standard, it would be exceedingly difficult to enforce. States generally do not publicize the results of their weapons review processes (assuming such a review is undertaken). In addition, efforts to establish binding standards would be unwise. The appropriate reliability and predictability of a specific weapon will depend on its intended use and operational environment, making a single standard both under and over inclusive. It is also dubious whether a binding standard would address an extant problem. Unreliable and unpredictable weapons would be of little value to States. Such weapons cannot effectively be used to achieve desired military effects and may pose a risk to a State's own forces. In practice, States are seeking

---

[234] Memorandum from the Deputy Sec'y of Def. to the Sec'y of the Mil. Dep't, Chairman of the Joint Chiefs of Staff, Under Sec'y of Def., Commanders of the Combatant Commands, General Counsel of the DoD, Dir. of Cost Assessment and Program Evaluation (Nov. 30, 2017).

more precise and predictable weapons and discontinuing weapons that cannot reliably be used. [235]

A more effective approach would be to develop non-binding, good practices for improving autonomous weapon reliability, similar to those in the Technical Annex to the CCW's Protocol on Explosive Remnants of War (Protocol V). In that Annex, States agreed to apply "best practice norms and operating procedures" to "ensure the best possible long-term reliability of explosive ordnance."[236] These best practices include establishing certified quality control measures in production process, "live-fire testing over a range of conditions[,]" periodic testing of stockpiled munitions, and training of personnel in involved in handling, transport, and use of explosive ordinance.[237]

The U.S. Department of Defense has already codified a number of best practices in its Directive on Autonomous Weapons. This directive, which mandates a number of steps to ensure that semi-autonomous and autonomous weapons function as anticipated in realistic operational environments, could serve as the foundation for a soft-law instrument. The Directive requires, inter alia, that covered systems "go through rigorous hardware and software V&V and realistic system development and operational T&E, including analysis of unanticipated emergent behavior."[238] Further changes to the system require additional "V&V and T&E in order to ensure that critical safety features have not been degraded."[239] This includes a "regression test of the software … to validate critical safety features have not been degraded."[240] Weapons must also have "[s]ufficient safeties, anti-tamper mechanisms, and information assurance … to minimize the probability or consequences of failures that could lead to unintended engagements or to loss of control of the system."[241]

---

[235]   *See* Hall, *supra* note 33, 86, 89 (noting that forces in the field will not use systems that they cannot trust to achieve mission success); *see, e.g.,* SCHARRE, *supra* note 19, at 53–54. The Tomahawk Anti-Ship Missile (TASM) is one example of a military abandoning a weapon that it deemed too unpredictable. The TASM was designed to search for and engage presumed Soviet ships over a wide geographic area. The TASM initiated its search based on U.S. sensors that detected an enemy ship. But, the Navy could not control the TASM after it deployed, and the TASM could not discriminate between the intended enemy ship and other vessels that might be within its search pattern. The TASM was operational for twelve years but was retired in 1994. According to Paul Scharre's interview with Naval strategist Bryan McGrath, the TASM was removed from service because of the Navy's lack of confidence in its performance.

[236]   CCW Protocol V, Technical Annex, *supra* note 231, art. 3(a)–(e).

[237]   *Id.*

[238]   DoDD 3000.09, *supra* note 36, at 6.

[239]   *Id*.

[240]   *Id.*

[241]   *Id*. at 7.

B. *Targeting*

The conduct of hostilities is governed by, inter alia, the fundamental IHL principles of distinction, proportionality, and feasible precautions. This Section discusses how States might apply these principles to the use of autonomous weapons, and the interpretive challenges they will confront in doing so. Because the application of these principles requires humans to make certain legal judgments, including at different times of the life cycle of a weapon system, a recurring consideration will be whether commanders and operators are able to make meaningful judgments regarding the use of force.[242] This discussion requires consideration of hypothetical weapons and scenarios. States may not have, or ever develop, weapons with the capabilities discussed below. The hypotheticals in this Section are intended to generate discussion about how States might apply IHL to plausible uses of future weapons.[243]

1. *Distinction*

The principle of distinction, a rule of customary international law codified in Additional Protocol I, requires that parties to a 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."[244] Military objectives include only 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

---

[242] *See generally Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, CCW/GGE.1/2018/CRP.1, Annex III, (Oct. 23, 2018). The degree of human involvement required when using autonomous weapons will depend on a number of factors related to the operating environment in which the weapon is deployed (cluttered or uncluttered), the weapon's intended use (offensive or defensive; range and duration of mission; and type of munitions attached); and the sophistication and programming of the weapon system (safeguards and predictability). Importantly, the human judgment required over the use of force may be implemented throughout the lifecycle of the weapon system, including at the design, development, training, and deployment stages.

[243] These hypotheticals also presume the existence of an ongoing armed conflict, making IHL the applicable legal framework.

[244] Additional Protocol I, *supra* note 154, art. 48. As a matter of treaty law, API only applies to international armed conflicts. The principle of discrimination, however, is considered a rule of customary international law applicable to both IACs and NIACs. Furthermore, Article 13(2) of APII, which applies to NIACs, also makes clear that the "civilian population as such, as well as individual civilians, shall not be the object of attack." Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II) art. 13(2), *adopted on* June 8, 1977, 1125 U.N.T.S. 3 [hereinafter Additional Protocol II].

advantage."[245] In cases of doubt, individuals must be considered civilians rather than combatants.[246] Similarly, objects that are "normally dedicated to civilian purposes" should be presumed to be civilian objects in case of doubt regarding whether they are being used to make an effective contribution to military action.[247]

IHL also prohibits indiscriminate attacks, meaning attacks not directed at a specific military objective,[248] attacks which employ a method or means of combat which cannot be directed at a specific military objective, or attacks which employ a method or means of combat the effects of which cannot be limited.[249]

The challenges for compliance with the principle of distinction when deploying autonomous weapons will depend in part on the specific capabilities and programming of a weapon system.[250] Some autonomous weapons may be programmed to attack only pre-determined military objectives, like the Harpy that is used to search for and destroy radar installations.[251] These weapons have autonomous functions, but those functions only control how the weapon system tracks and engages its target.[252] Autonomous weapons that search for and engage specifically defined target sets do not pose significantly novel challenges in terms of compliance with the principle of distinction, as a human remains responsible for determining whether the programmed target set constitutes a military objective.[253] Although these weapon systems must "select"—in a narrowly defined meaning of the term—their targets, they do not "decide" what

---

[245] Additional Protocol I, *supra* note 154, art. 52(2).

[246] *Id*. art. 50(1).

[247] *Id*. art. 52(3).

[248] INT'L COMM. OF THE RED CROSS, COMMENTARY ON THE ADDITIONAL PROTOCOLS OF 8 JUNE 1977 TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949, ¶ 1946 (1987). The requirement to limit attacks to "specific" military objectives is best understood as limiting attacks to discrete military objectives rather than areas in which military objectives and civilians are intermixed. As the ICRC Commentary to Additional Protocol I notes, this provision was driven by "certain regrettable practices during the Second World War and subsequent armed conflicts … [where] too often the purpose of attacks was to destroy all life in a particular area or to raze a town to the ground[.]" Accordingly, this provision would not necessarily prohibit tactics, such as the laying of anti-vehicle mines, where the particular individual target is unidentified, so long as the attack is directed at military rather than civilian objectives.

[249] Additional Protocol I, *supra* note 154, art. 51(4). The United States is not a party to API but considers the Protocol's definition of "military objective" and its prohibitions on indiscriminate attacks to reflect customary international law applicable in all armed conflicts. *See* Brian Egan, *International Law, Legal Diplomacy, and the Counter-ISIL Campaign: Some Observations*, 92 INT'L L. STUD. 235 (2016).

[250] *See* Schmitt, *supra* note 52, at 8.

[251] *See e.g.*, SCHARRE, *supra* note 19, at 48.

[252] *See* Schmitt, *supra* note 52, at 6.

[253] *See id.*

constitutes a military objective in the same way that a human does.[254] The AEGIS, for example, does not consider whether the destruction of an incoming missile would provide a military advantage in the circumstances ruling at the time. Rather, it is programmed to search for and attack objects displaying certain characteristics, such as velocity and trajectory, which the human programmer determined would constitute military objectives (i.e., incoming missiles) under the circumstances in which the AEGIS would be deployed (i.e., in the open seas).

Future weapon systems may have greater capability to determine what constitutes a legitimate military target, perhaps within general parameters set by a human. Weapons systems may in the future use algorithms to identify enemy fighters, for example, similar to existing algorithms that assist police in predicting where crimes are likely to occur.[255] Ashley Deeks notes "[m]achine learning also facilitates pattern detection, which will help the military assess the activities of particular individuals and determine, based on those activities, whether they are members of enemy-organized armed groups."[256] As in the civilian sector, this technology will likely be used to recommend targets to human operators.[257] As algorithms become more sophisticated and accurate, however, the human may be cut from the loop.[258]

Consider a hypothetical unmanned aerial vehicle (UAV) programmed to identify and attack terrorists based on programmed "signatures"—unique characteristics or patterns of activity—that indicate terrorist activity.[259] The

---

[254]   *Id.*

[255]   *See, e.g.,* Randy Rieland, *Artificial Intelligence is Now Used to Predict Crime. But is it Biased?,* SMITHSONIAN MAG., Mar. 5, 2018 (describing how police departments use AI to identify patterns of criminal behavior).

[256]   Ashley Deeks, *Predicting Enemies*, 104 VA. L. REV. 1529, 1536 (2018).

[257]   *See* Jennifer Kite-Powell, *Making Facial Recognition Smarter with Artificial Intelligence,* FORBES (Sep. 30, 2018, 2:56 PM), https://www.forbes.com/sites/jenniferhicks/2018/09/30/making-facial-recognition-smarter-with-artificial-intelligence/#3e607b80c8f1. Security companies are currently using AI and machine learning to identify patterns of behavior or objects that suggest a potential threat. This allows guards to focus their attention on a smaller subset of individuals. *Id.*

[258]   *See* Taylor Owen, *The Violence of Algorithms*, FOREIGN AFF. (May 25, 2015), https://www. foreignaffairs.com/articles/2015-05-25/violence-algorithms ("Increasingly, such tools, and the algorithms that power them, are being used to automate violence.").

[259]   DAN KLAIDMAN, KILL OR CAPTURE: THE WAR ON TERROR AND THE SOUL OF THE OBAMA PRESIDENCY 41 (2012). The United States reportedly carried out "signature strikes" against terrorist groups in Pakistan, Yemen, and Somalia. These strikes target "groups of men who bear certain signatures, or defining characteristics associated with terrorist activity, but whose identities aren't known." For a discussion of signature strikes, see MICAH ZENKO, COUNCIL OF FOREIGN RELATIONS, REFORMING U.S. DRONE STRIKE POLICIES 12 (Special Report No. 65, 2013); Dan De Luce & Paul Mcleary, *Obama's Most Dangerous Drone Tactic is Here to Stay*, FOREIGN POL'Y (Apr. 5, 2016), https://foreignpolicy.com/2016/04/05/obamas-most-dangerous-drone-tactic-is-here-to-

UAV navigates to a region where terrorist camps are suspected, observes events on the ground, identifies patterns of life, assesses whether such patterns indicate terrorist activity, and without further human intervention engages targets that it determines are terrorists. The commander launching the autonomous UAV is responsible for complying with the principle of distinction but is not able to assess the information that the weapon system relies upon when deciding to engage a target. How can that commander determine that deploying this weapon would be consistent with the principle of distinction?

In answering this question, we should first consider what the principle of distinction does and does not require, and how militaries comply with distinction when using advanced weapons. Distinction does not require that a human independently confirm the identity of each target. Militaries currently rely on semi-autonomous and human supervised weapons to select, and sometimes engage, certain targets that humans cannot independently verify to be military objectives. The complex sensors and computer systems in an F/A-18 Hornet may lock in on targets beyond visual range for its pilot to shoot and then will guide the missile to its target. Although the pilot must press the button to deploy the munition (also known as "pickling"), he will likely not question whether the object identified by the plane's sensors is a military objective, especially if the computer indicates that it poses an imminent threat. The pilot's decision to engage the identified target is based almost entirely on his understanding of the mission, the pre-planned strike coordinates, knowledge of the aircraft's computer systems, and the operating environment.[260]

Similar to the Hornet pilot, a commander deploying an autonomous weapon will need to consider both the operating environment and the weapon system's capabilities in order to ensure the attack is directed against lawful targets. Relevant information about the operating environment may include the likelihood that enemy combatants are within range of the weapon, the likelihood that civilians or civilian objects are in the proximity, and the likelihood that both combatants and civilians are intermingled in such a manner that could result in unintended engagements. A commander deploying an autonomous weapon in an uninhabited region would have greater confidence that its use would be discriminate, even though she cannot independently verify the ultimate target. Similarly, the area, duration, and objective of the mission will be relevant in making the discrimination assessment. Weapons deployed for extended periods of time or over large areas may require more sophisticated discrimination

---

stay/.
[260] Ekelhof, *supra* note 206, at 335–36 (describing the pilot's role in the targeting cycle).

capabilities in order to compensate for the operator's inability to foresee the circumstances at the time of kinetic action.

The capability of the autonomous weapon system to discriminate between intended targets and unlawful targets will be a critical consideration. Some autonomous weapons may be as good, or better, than humans in distinguishing combatants and military objectives from civilians and civilian objects, at least when given narrowly defined missions.[261] A weapon system, for example, could have facial recognition technology that allows it to positively identify known enemy combatants.[262] Some weapons may be used solely to engage specific enemy objects, such as tanks or radar installations. These weapons may be "equipped with sensors that are designed to detect specific 'signatures'—unique, identifying characteristics that would be specific to a military objective, such as frequencies of electromagnetic radiation that are generally not found naturally or among civilian objects."[263]

By contrast, autonomous weapons may never reach human capabilities in making determinations that require contextualized judgments, such as when targeting objects that have both civilian and military uses. Whether an object constitutes a legitimate military objective depends on whether it makes an "effective contribution to military action" and whether its "total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."[264] A pick-up truck parked at Walmart would almost certainly be a protected civilian object. This same truck parked near a known terrorist camp, however, could be a legitimate military objective. Similarly, a bridge used predominantly for civilian purposes may become a military objective if it would be a viable escape route for enemy forces in an upcoming assault. Determining whether dual use objects can be attacked requires contextualized judgments, based on a strategic or tactical understanding of the specific operating environment, that machines may never be capable of making.

---

[261]   NATO Bombing Report, *supra* note 185. As the NATO Report on Kosovo makes clear, human operators can often make good faith mistakes in selecting targets that result in significant civilian casualties.

[262]   Schuller, *supra* note 17, at 394. Facial recognition software has improved significantly over the past few years and is now being used by government agencies like DHS at airports to screen international passengers. *See* U.S. DEP'T HOMELAND SECURITY, FISCAL YEAR 2018 ENTRY/EXIT OVERSTAY REPORT (2018).

[263]   *Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies, supra* note 79, ¶ 15.

[264]   Additional Protocol I, *supra* note 154, art. 52(2); *see also* DoD LAW OF WAR MANUAL, *supra* note 182, at § 5.6.3.

The fact that an autonomous weapon cannot replicate human decision-making does not render its use unlawful. The key question for the commander is not whether the weapon system can discriminate among lawful targets in *all* circumstances, but whether it can adequately do so under the *specific* circumstances of the assigned mission.[265] In the autonomous UAV hypothetical above, a commander would need to assess whether the UAV's algorithms can reliably distinguish between terrorist fighters and civilians based on observed evidence and patterns of activity, such as a "person's behavior, location and appearance in relation to other circumstances[.]"[266] This assessment may require consideration of the sophistication of the weapon's algorithms, the quantity and quality of data on which it was trained, and the weapon's ability to collect sufficient data on the intended targets prior to engagement.

Even if a commander is confident in the capability of an autonomous weapon to perform the assigned mission, self-learning weapons will have a degree of unpredictability.[267] This raises the question of the degree of certainty that operators must have that the weapon will attack combatants or military objectives rather than civilians and civilian objects. Additional Protocol I states that individuals and objects should not be attacked in "case of doubt" regarding their status.[268] Can a commander comply with this requirement if she does not know when, where, or even what the autonomous weapon will attack?

API does not provide clear guidance on the degree of certainty required when authorizing an attack,[269] but IHL does not require absolute confidence that the intended target is a lawful one.[270] According to one respected commentary, the critical issue is that the commander not act on "the basis of mere speculation."[271] Similarly, the ICTY has interpreted this requirement to prohibit attacks "when it is not reasonable to believe, in the circumstances of the person

---

[265]   Shane R. Reeves & Jeffrey S. Thurnher, *Are We Reaching a Tipping Point? How Contemporary Challenges are Affecting the Military Necessity-Humanity Balance*, HARV. NAT'L SEC. J., June 24, 2013, https://harvardnsj.org/2013/06/are-we-reaching-a-tipping-point-how-contemporary-challenges-are-affecting-the-military-necessity-humanity-balance/ ("There may be situations, such as battles that occur in remote regions … where an autonomous weapon might be lawful despite having virtually no ability to distinguish between civilian and military objectives.").

[266]   BOTHE, PATSCH, & SOLF, *supra* note 186, at 336, 337.

[267]   *See supra* notes 225–30.

[268]   Additional Protocol I *supra* note 154, art. 50, 52.

[269]   *See, e.g.,* Kevin Heller, '*One Hell of a Killing Machine': Signature Strikes and International Law*, 11 J. INT'L CRIM JUSTICE 89 (2013).

[270]   Schmitt, *supra* note 52, at 16 ("The mere existence of some doubt does not bring the presumption into operation").

[271]   BOTHE, PATSCH, & SOLF, *supra* note 185, at 336; *see also* Schmitt, *supra* note 52, at 16 (stating that the test is whether the degree of doubt would cause a reasonable commander to hesitate before attacking).

contemplating the attack, including information available to the latter," that the target is a lawful military objective.[272]

As the preceding discussion makes clear, autonomy in weapon systems will not obviate the need for human judgment, but it will transform the type of information that commanders must consider in exercising such judgment. As information about specific targets decreases, a commander will require greater knowledge of the weapon system (including its reliability and predictability) and operating environment in order to comply with the principle of distinction.[273] To this end, DoD is taking steps to ensure that operators have the appropriate familiarity with autonomous weapons systems under their control. The DoD Directive on Autonomous Weapons instructs the military departments to "[d]esign human-machine interfaces for autonomous and semi-autonomous weapon systems to be readily understandable to trained operators" and to "[c]ertify that operators of autonomous and semi-autonomous weapons systems have been trained in system capabilities, doctrine, [tactics, techniques and procedures] in order to exercise appropriate levels of human judgment in the use of force and employ systems with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable ROE."[274]

Lastly, some commentators have suggested that autonomous weapons could not be used consistent with IHL if they are unable able to recognize attempts to surrender.[275] Paul Scharre writes, "[t]o employ weapons that could not recognize when soldiers are hors de combat would not only violate the modern laws of war, but would trespass on millennia-old norms of warfare."[276] This assertion raises significant ethical concerns, but arguably errs in anthropomorphizing autonomous weapons and assigning them obligations that apply only to humans.

Combatants are considered hors de combat, and thus protected from attack, if their surrender is genuine, clear and unconditional, and made under circumstances where it is feasible for the opposing party to accept the

---

[272]   Prosecutor v. Galić, Case No. IT-98-29-T, Trial Judgement (Int'l Crim. Trib. for the Former Yugoslavia Dec. 5, 2003).

[273]   It is always good practice for commanders and operators to understand the weapons they deploy. For conventional weapons, obtaining this familiarity may be relatively easy. Understanding how a Howitzer operates, however, is simpler than understanding how an autonomous weapon with self-learning capabilities is likely to behave in different environments.

[274]   DoDD 3000.09, *supra* note 36, at Enclosure 4, 4(a)(4)–(5).

[275]   Heyns, *supra* note 8, at ¶ 67 ("It would be difficult for robots to establish, for example, whether someone is wounded and hors de combat, and also whether soldiers are in the process of surrendering.").

[276]   SCHARRE, *supra* note 19, at 371.

surrender.[277] There is no requirement that weapons be capable of recognizing the surrender of enemy forces. Soldiers, for example, cannot surrender to an inbound missile. In order for a surrender to be valid, it must be "practical and safe for the opposing force to take custody of the surrendering persons in the circumstances."[278] Even if an autonomous weapon could recognize an attempt to surrender, it is unlikely that it would be capable of taking custody of the surrendering individuals, at least in the absence of ground forces in the vicinity. At the same time, to the extent that autonomous weapons eventually undertake missions traditionally performed by ground forces, there is a legitimate question as to whether IHL would or should impose an obligation on the attacking force to take measures to recognize and accept surrender.

2. *Proportionality*

The principle of proportionality, codified in Article 51(5)(b) of Additional Protocol I, places additional restraints on attacks directed at military objectives. It prohibits any "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."[279] There is no metric for determining whether collateral damage is excessive, as the proportionality assessment requires "the balancing of two dissimilar values- military advantage and civilian life[.]"[280] A commander's proportionality assessment will necessarily have a subjective element, but it must be reasonable in light of the information reasonably available at the time of the attack.[281] For example, bombing an entire apartment

---

[277] DoD LAW OF WAR MANUAL, *supra* note 182, at § 5.9.3; *see also* U.S. DEP'T OF DEF. OF DEFENSE, FINAL REPORT TO CONGRESS ON THE CONDUCT OF THE PERSIAN GULF WAR 641 (1992) (stating that the communication of surrender "must be made at a time when it can be received and properly acted upon—an attempt at surrender in the midst of a hard-fought battle is neither easily communicated nor received.").

[278] DoD LAW OF WAR MANUAL, *supra* note 182, at § 5.9.3.3.

[279] Additional Protocol I, *supra* note 154, art. 51(5)(b) (emphasis added). The United States accepts this as CIL, see DoD LAW OF WAR MANUAL, *supra* note 182, § 5.4.2.

[280] Charles P Trumbull IV, *Re-Thinking the Principle of Proportionality Outside of Hot Battlefields*, 55 VA. J. INT'L L. 521, 542 (2015); *see also* BOTHE, PARTSCH & SOLF, *supra* note 185, at 351 (noting that proportionality assessment "requires a comparison of values which cannot be compared").

[281] DoD LAW OF WAR MANUAL, *supra* note 182, at § 5.10.2.2 ("The commander's decisions on proportionality must be reasonable"); *see also id.* at § 5.10.2.3 (stating that the "assessment of whether a decision-maker has complied with the legal requirements must be based on the information available to that person at that time"); BOTHE, PARTSCH & SOLF, *supra* note 185, at 351; Understanding, Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, June 6, 1991, 1643 U.N.T.S. 473 ("In relation to Articles 51 to 58 inclusive it is the understanding of Australia that military commanders and others responsible for planning, deciding upon, or executing attacks, necessarily have to reach their decision on the basis of their assessment of the information

building in order to kill a single sniper would almost certainly violate this principle because the collateral damage would be excessive in relation to the military advantage gained.

The proportionality principle presupposes that the commander will be in a position to assess both the military advantage and the collateral damage expected to result from an attack. Autonomous weapons that can select and engage targets, and deploy for extended periods of time, will challenge this assumption and raise two critical questions regarding the interpretation and application of the principle.

a. Timing of the "Attack"

The legal judgments and determinations required by IHL, including the balancing required by the principle of proportionality, generally must be made at the time of attack. The use of autonomous weapons, however, may generate uncertainty as to when a specific attack commences and ends. Does the attack commence when the commander deploys the autonomous weapon, even if the anticipated use of force may not occur for days or weeks? Or, does it occur at some point closer to the kinetic action, for example, at the last point in which the operator can exercise control over the weapon system? Does an autonomous weapon that can strike multiple targets commit multiple attacks, or are multiple strikes the continuation of a single attack?

Neither the text of API nor its Commentary answers this question. Article 49(1) defines "attacks" as "acts of violence against the adversary, whether in offence or defence."[282] The ICRC's 1987 Commentary adds that the term "refers simply to the use of armed force to carry out a military operation at the beginning of or during the course of armed conflict."[283] During the negotiations of Article 49, the ICRC explained that it understood the term "attack" to be "related to only one specific military operation, limited in space and time."[284] Bothe, Partsch, and Solf add that the "ICRC had intended to refer generally to the coordinated acts of violence against the adversary by a specific military formation engaged in a specific military operation."[285] During the negotiation of this Article, however, a question arose that foreshadows our current inquiry: At what point does the use of mines constitute an attack? Does the attack commence when the

---

from all sources, which is available to them at the relevant time.").

[282]   Additional Protocol I, *supra* note 154, art. 49(1).
[283]   INT'L COMM. OF THE RED CROSS, *supra* note 198, at 603.
[284]   BOTHE, PARTSCH, & SOLF, *supra* note 185, at 288.
[285]   *Id.*

mine is laid, armed, or detonated? The general consensus was that the attack started "whenever a person is directly endangered by a mine laid."[286]

A close reading of related provisions in Additional Protocol I sheds a little more light on the temporal dimension of an attack. The obligation in Article 51(2)(b) to cancel or suspend an attack if it becomes apparent that the objective is not a military one or that the attack would cause excessive collateral damage suggests that the "attack" may commence before the moment of kinetic activity, such as when the weapon or troops are heading towards the intended target.[287] The obligation in Article 52 to limit "attacks" to military objectives, as defined in that Article, indicates that the determination of what constitutes a military objective must be made in close temporal proximity to the use of force. In particular, Article 52 defines military objectives as "those objects … whose total or partial destruction, capture, or neutralization, *in the circumstances ruling at the time*, offers a definite military advantage."[288] This requires an assessment that an object is a military objective at the time of its destruction or capture, rather than at the time a weapon system is deployed.

Until recently, there has been little need to clarify the ambiguity on when an attack occurs. With most conventional weapons, the time between the deployment and impact is negligible, such that there is little possibility of intervening events that would affect the proportionality assessment.[289] Advances in weapons technology, however, have generated some debate regarding when an "attack" commences.[290] Certain precision-guided munitions (PGMs), for example, can be re-directed in flight.[291] This practice, referred to as a "shift

---

[286]   INT'L COMM. OF THE RED CROSS, *supra* note 198, at 603. *But see* BOTHE, PARTSCH, & SOLF, *supra* note 185, at 350. Bothe, Partsch, and Solf, by contrast, appear to take the position that the attack occurs when the mines are laid, but the attack is directed against an area of land rather than against individuals. They note further that there "is nothing in Art. 50(1) which excludes a delayed act of violence from the definition" of attack. *Id.* at 350 n. 26.

[287]   INT'L COMM. OF THE RED CROSS, *supra* note 198, at 534, 677, 686.

[288]   Additional Protocol I, *supra* note 154, art. 52 (emphasis added).

[289]   This is not always the case. High altitude bombs, for example, may fall for a minute or more before detonation. This time lag between the decision to "attack" and the moment of impact creates the possibility that civilians will enter the blast radius. Unforeseen intervening events generally do not affect whether the decision to "attack" was lawful, as this decision must be based on the facts known at the moment of attack. A school bus that unexpectedly enters a militarily strategic bridge moments after a bomber deploys its ordnance, for example, does not render the decision to attack that bridge unlawful.

[290]   *See generally* Jonathan B. Tucker, *The Future of Chemical Weapons*, 26 NEW ATLANTIS 3–29 (2010); Darran Anderson, *The Grim Future of Urban Warfare*, ATLANTIC (Dec. 11, 2018), https://www.theatlantic.com/technology/archive/2018/12/technology-will-make-war-even-worse/577723/.

[291]   Michael Schmitt & Lt. Col Matthew King, *The "Shift Cold" Military Tactic and International Humanitarian Law*, JUST SECURITY, (Feb. 20, 2018), https://www.justsecurity.org/52198/shift-cold-tactic-international-humanitarian-law/ (noting that only certain precision guided munitions, such as laser-guided

cold," is primarily used to mitigate or avoid civilian harm during air strikes.[292] An operator may implement this tactic, for example, when civilians unexpectedly come within range of a missile after its launch.[293] As Schmitt and King note, if a commander is capable and authorized to divert a PGM when he becomes aware that its programmed trajectory would result in excessive civilian harm, "it is difficult to fashion an argument that there would not be an obligation to do so."[294] Schmitt and King consider the diversion of the missile as a precautionary measure that is part of "a continuation of an ongoing attack rather than a separate attack" because it involves the "[s]ame aircraft, same weapon, and same personnel in control of the weapon[.]"[295] Under this view, a new proportionality assessment would not necessarily be required at the time in which the munition is re-directed.[296] Adil Haque, by contrast, takes the view that the re-direction of a missile towards a different target is a new "attack" for the purposes of Article 51, requiring a fresh discrimination and proportionality assessment.[297] Alternatively, he argues that if the redirection of a missile is considered the continuation of an ongoing attack, "then the principles of distinction and proportionality continue to apply as well, until the *final* target is selected and the *final* balance is struck between anticipated military advantage and expected civilian harm."[298]

This debate regarding what constitutes an "attack" has little practical significance for the use of existing weapons, as the circumstances in which operators can execute a shift cold are rare.[299] Only a limited set of weapons are capable of such diversion, and the short time between launch and detonation makes it unlikely that intervening events will alter the initial proportionality assessment.[300] This question is of significant consequence for autonomous weapons. Given that the endurance and range of autonomous weapons is likely

---

weapons, can be re-directed; number of practical considerations, such as flight time and blast radius, will determine whether such re-direction is feasible under the circumstances).

[292]   *Id.*

[293]   *Id.*

[294]   *Id.*

[295]   *Id.*

[296]   *Id.* When a weapon is re-directed away from a military objective, it could also be considered a "cancellation or suspension of an attack." Adil Haque, *The "Shift Cold" Military Tactic: Finding Room Under International Law*, JUST SECURITY (Feb. 20, 2018), https://www.justsecurity.org/52713/shift-cold-military-tactic-finding-room-under-international-law/.

[297]   *Id.*

[298]   *Id.*

[299]   Schmitt & King, *supra* note 291.

[300]   *Id.* Even if an unforeseen event occurs, the commander may not have sufficient time to execute the shift cold.

to far exceed the capabilities of existing conventional weapons, there is a greater chance that autonomous weapons will encounter circumstances that the commander did not foresee when deciding to deploy the weapon system.

States and practitioners will need to address this definitional ambiguity. Determining when the "attack" occurs will be critical for understanding a commander's legal obligations when using autonomous weapons. There may not be a single test, applicable to all weapons, at which we can definitively say that an attack has commenced. Yet, for the principle of proportionality to have any consequence in an era of autonomous weapons, we must consider the "attack" to begin—and thus the obligation to undertake the proportionality assessment—at a time in which the commander can reasonably predict the consequences of the use of force and thus make the balancing judgment required by IHL.[301] Thus, for example, an attack might commence at the moment of deployment for a weapon that is intended to strike a target within a limited period of time. When using weapons that can deploy for days or weeks, the "attack" may commence at a later point, such as when the commander has the opportunity to communicate with the weapon and re-assess the mission.[302] Alternatively, as with mines, the attack could commence whenever a person is endangered by the weapon.[303] The critical point is that States must resolve the definitional ambiguity in such a way that preserves the essential role for humans to make informed judgments over uses of force.

b. Uncertainty in Attacks

The use of autonomous weapons raises the related question of how commanders can reasonably determine the collateral damage and military advantage that "may be expected" to result from an attack. This assessment is already fraught with difficulty given the lack of agreed metrics, the fog of war, and the need to act based on incomplete or unverified information.[304] The use of autonomous weapons will further complicate this determination due to potential uncertainty regarding the location and timing of the engagement.[305]

---

[301]   *See generally* Emanuela-Chiara Gillard, *Proportionality in the Conduct of Hostilities: The Incident Harm Side of the Assessment*, 2018 CHATHAM HOUSE 9.

[302]   While it may not always be possible to maintain a communication link between the commander and the autonomous weapon, such a link should be established whenever feasible to help ensure that new information regarding potential collateral damage is taken into account.

[303]   *See generally* INT'L COMM. OF THE RED CROSS, *supra* note 198, at 603.

[304]   Corn, *supra* note 188, at 770. These difficulties have led some academics, international courts, and organizations to call for an effects-based test for determining IHL compliance.

[305]   *See generally* Gillard, *supra* note 301.

The uncertainty introduced by autonomy is most likely to affect the collateral damage side of the balancing equation.[306] Because commanders will deploy autonomous weapons to achieve some desired tactical, operational, or strategic effect,[307] they will presumably be able to predict the military advantage resulting from the attack. The extent of anticipated collateral damage, however, will vary depending on when and where the target is attacked.[308] This prediction will become more difficult as the number of potential engagement scenarios increases.

Consider the following example: A State engaged in hostilities against a terrorist organization deploys an autonomous UAV with facial recognition capabilities to search for and engage a known, mid-level military leader of that group. The UAV is programmed to operate over a geographic area where the target is known to maneuver. Once the UAV confirms the target's identity, it is programmed to deploy a Hellfire missile. The target knows he is on a kill list, and generally stays in hiding. The only time he appears in the open is when he travels once a week by car along a desert road to a mosque. Video surveillance of the target indicates that he generally travels with just one bodyguard, who is also an enemy combatant. But once a month the target takes his three children to the mosque. Killing the terrorist and his bodyguard would be lawful under IHL as the attack would not involve any civilian casualties. Killing the terrorist and his children, on the other hand, would cause disproportionate civilian harm. Would the deployment of the UAV be consistent with the principle of proportionality?

One might argue that Article 51(5)(b), by its express terms, would not prohibit the attack. The risk of killing the children is only twenty-five percent, and therefore not "expected."[309] The more likely, or expected, scenario is that the autonomous weapon would kill only the objective and his bodyguard. This argument finds some support in the negotiating history of Additional Protocol I. According to the ICRC Commentary, some delegations preferred the broader

---

[306]   *See generally id.*
[307]   Corn, *supra* note 188, at 776, 782
[308]   Gillard, *supra* note 301. We can imagine scenarios in which the collateral damage would not exceed a certain level given pre-programmed constraints on the autonomous weapon's ability to engage a target. For example, a weapon system could be programmed not to engage a military objective if it detects any humans within the blast radius.
[309]   *Expect*, MERIAM-WEBSTER'S DICTIONARY (2019), https://www.merriam-webster.com/dictionary/expected (Meriam-Webster's dictionary defines the word "expect" as "to consider probable or certain.").

formulation "which risks causing" but the Committee ultimately adopted the wording "which may be expected to cause."[310]

A literal interpretation of the principle of proportionality, as codified in Article 51, that permits an attack so long as no single scenario among a range of potential outcomes is "expected" to result in excessive collateral damage cannot be correct, given the object and purpose of the Protocol.[311] Such an interpretation would incentivize ignorance in attacks and would be manifestly inconsistent with the stated purpose of Article 51, which is to give effect to the rule that "civilians shall enjoy general protection against dangers arising from military operations."[312] It would also conflict with other IHL provisions that impose a duty at all stages of the targeting cycle to ensure that only military objectives are targeted and that civilians are not made the object of an attack,[313] to minimize civilian casualties,[314] and to refrain from or cancel attacks expected to cause disproportionate harm to civilians or civilian objects.[315] A commander must have sufficient knowledge of the circumstances and context in which force will be used in order to comply with these requirements.[316] Ignorance regarding the circumstances in which force is deployed should not excuse what would otherwise be a violation of IHL, at least if it is feasible to acquire the information that would be relevant to the proportionality assessment.

The more logical interpretation of Article 51 is that the likelihood of civilian casualties must be considered as part of the "weight to be assigned to the harm in proportionality assessments."[317] In other words, harm that is possible, but not expected, must still be taken into account in the proportionality assessment.

---

[310]   INT'L COMM. OF THE RED CROSS, *supra* note 198, at 684. Similarly, the UK delegation proposed the formulation "and that the risks to civilians lives and objects are not disproportionate to the military advantage anticipated."

[311]   *Id*. at 1450.

[312]   *Id.* at 613.

[313]   *Id*. at 677.

[314]   *Id.*

[315]   *Id.*

[316]   Corn, *supra* note 188, at 776 ("Identifying the best approach to implementing LOAC obligations, and assessing what qualifies as a lawful attack decision, must obviously start with an understanding of this type of operational context."); Charles J. Dunlap, Jr., *Accountability and Autonomous Weapons: Much Ado About Nothing?,* 30 TEMPLE INT'L & COMP. L. J. 63, 69 (2016) ("[A] commander must have a reasonable understanding of the AWS and how it will work before deploying it in a particular situation."). The judgments required by a commander in deploying autonomous weapons are in some ways similar to the judgments required in maneuver operations. These operations involve decentralized decision-making by soldiers based on an understanding of the commander's intent. In such circumstances, a commander "must anticipate the *nature* of the attack decisions that subordinates will be required to make in such a dynamic decision-making context, and provide them with the tools they need to make these decisions consistent with LOAC obligations." Corn, *supra* note 188, at 775.

[317]   Gillard, *supra* note 301, ¶ 60.

Harms that are less likely to occur, however, may be given lesser weight than harms that are probable.[318] In the above hypothetical, the operation may still be lawful, provided that the commander reasonably determines that a twenty-five percent chance of killing three civilians is not excessive in relation to the advantage of killing the terrorist.

c. Feasible Precautions

Parties to a conflict have an obligation to take feasible precautions to protect the civilian population.[319] This includes taking "constant care" to spare the civilian population, doing everything feasible to verify that objectives of attack or neither civilians nor civilian objects, taking feasible precautions in the choice of means and methods of attack to minimize civilian casualties, and providing advance warning in attacks that may affect the civilian population.[320] The obligation to take feasible precautions does not mean that States must do anything and everything possible to protect civilians. Feasible precautions are understood to mean those measures "which are practicable or practically possible, taking into account all circumstances ruling at the time, including humanitarian and military considerations."[321] This means that if a precaution, while technically feasible, would jeopardize troops or undermine the mission, it would not be legally required. While the specific precautions that are mandated will depend on the operational context, there is "a continuing obligation to assign a high priority to the collection, collation, evaluation, and dissemination of timely target intelligence."[322]

The importance of taking precautionary measures cannot be overstated. As Professor Corn notes, "precautionary measures hold, from an operational implementation perspective, potentially greater potential [than other IHL principles] for producing this [civilian risk mitigation] outcome."[323] For this reason, the "precautions step" is perhaps the most important in the targeting process. At this stage, the commander considers whether "civilian risk can be mitigated by adjusting the timing of the attack, by issuing a prior warning, or by

---

[318]   *Id.* ¶ 71.

[319]   Corn, *supra* note 188, at 764 n.3 (noting that the obligation to take precautionary measures is a core LOAC principle that has arguably the greatest potential to reduce civilian casualties).

[320]   INT'L COMM. OF THE RED CROSS, *supra* note 198, at 713.

[321]   Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices as Amended on 3 May 1996 (Protocol II to the 1980 Convention as Amended on 3 May 1996), art. 3(10).

[322]   BOTHE, PARTSCH, & SOLF, *supra* note 185, at 405.

[323]   Geoffrey S. Corn, *War, Law, and Precautionary Measures: Broadening the Perspective of This Vital Risk Mitigation Principle*, 42 PEPP. L. REV. 419, 424 (2015).

selecting an alternate means or method for the attack."[324] An integral component of this inquiry is the "weaponeering" analysis, in which the command staff evaluates the available munitions and their likely effects.[325] In conducting this analysis, the command staff should favor an autonomous weapon over other munitions only if it "achieve[s] the desired effect while minimizing collateral damage."[326]

The obligation to take precautionary measures when using autonomous weapons is not disputed. Regular validation and verification, for example, is one clear example of a feasible precaution.[327] There are, however, questions as to who is responsible for taking such precautionary measures and how they could be implemented. As stated in Additional Protocol I, the obligation to take precautionary measures applies to those who "plan or decide upon an attack," which would generally mean the commander in charge of the operation.[328] When using conventional weapons, the focus on person who "plan or decide" the attack is most logical, as they are in the best position to decide which precautionary measures are feasible under the circumstances. This may not be the case with respect to autonomous weapons. In many circumstances, weapons developers and programmers may have the greatest capability to ensure that precautionary measures are implemented.[329] For example, a weapons developer may build in a communications link to allow the commander to abort a mission in the event of malfunction, or program code to require a weapon system to take multiple steps to confirm the target's identity before engaging.[330] This allocation in capabilities raises the following questions: Do weapons programmers have an obligation to take precautionary measures, even if they are not responsible for planning or deciding a specific attack? How can a commander comply with the obligation to take precautionary measures when using autonomous weapons, especially if the commander is not aware of the specific circumstances of the attack?

---

[324]   *Id.* at 435–36.

[325]   *Id*. at 435.

[326]   Jennifer M. O'Connor, Gen. Counsel of Dep't of Def., Applying the Law of Targeting to the Modern Battlefield (Nov. 28, 2016).

[327]   *See*, *e.g.*, Corn, *supra* note 323, at 435.

[328]   Additional Protocol I, *supra* note 154, art. 57(2)(a)(i-iii).

[329]   *See Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects*, INT'L COMM. OF THE RED CROSS, 23, (2014), https://www.icrc.org/en/document/report-icrc-meeting-autonomous-weapon-systems-26-28-march-2014.

[330]   *Id.*

C. Accountability

States and individuals are jointly responsible for complying with IHL, and both can be held responsible for violations.[331] States are liable for breaches of international law under the principle of State responsibility.[332] A fundamental premise of international law is that "[e]very internationally wrongful act of a State entails the international responsibility of that State."[333] The conduct of government officials, including soldiers, is attributed to the State "provided they are acting in their official capacity."[334] Soldiers and civilians can also be held individually liable, under both domestic and international law, for violations of IHL.[335] As explained in Part V, the crucial factor in assessing individual criminal liability will be whether a human operator intentionally or recklessly disregarded his obligations under IHL.[336]

The use of autonomous weapons should not, at least in theory, create the accountability gap that many commentators envision.[337] The GGE on LAWS concluded in 2018 that "[a]ccountability for developing, deploying and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law" and that "[h]uman responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines."[338] Accountability will be relatively straight-forward with respect to intentional violations of IHL. A weapons programmer who intentionally instructs an autonomous weapon to attack civilians can be held accountable for war crimes.[339] Similarly, a commander who deploys a weapon with the knowledge that it will engage in indiscriminate attacks violates IHL.[340] The autonomous capabilities of a weapon

---

[331] Draft Articles on Responsibility of States for Internationally Wrongful Acts, with Commentaries, 2001, art. 1 [hereinafter Draft Articles].

[332] *Id.* at 46.

[333] *Id.*; *see also* Int'l Comm. of the Red Cross, *Rule 149. Responsibility for Violations of International Humanitarian Law*, https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule149 (last visited Feb. 7, 2020) ("It is a long-standing rule of customary international law … that a State is responsible for 'all acts committed by persons forming part of its armed forces.'").

[334] Draft Articles, *supra* note 331, art. 4 cmt. 7; *see also Autonomy in Weapons Systems*, *supra* note 200.

[335] *See* Crootof, *supra* note 43, at 1361.

[336] *Id.* at 1375–76 (describing the mental elements for war crimes)

[337] *See* Wagner, *supra* note 137, at 1399–1409 *(*for a discussion of accountability challenges).

[338] *Report of the 2018 Session*, *supra* note 14.

[339] Tim McFarland & Tim McCormack, *Mind the Gap: Can Developers of Autonomous Weapons Systems Be Liable for War Crimes?*, 90 INT'L L. STUD. 361, 374 (2014).

[340] *Id.* at 380–81.

system may make it more difficult to identify the source of malfeasance,[341] but intentionally causing harm to civilians remains a criminal offense.[342]

The more difficult and likely more common situation will be determining accountability in the absence of intent to harm civilians or civilian objects.[343] For some criminal offenses, recklessness (rather than intent) is sufficient to establish criminal liability under international law.[344] Determining what constitutes recklessness in the use of autonomous weapons will raise significant evidentiary and analytical challenges.[345] As discussed in Part V, recklessness is assessed by comparison to the reasonable commander, but the "the standard of care or regard that is due in conducting military operations with regard to the protection of civilians is a complex question to which the law of war does not provide a simple answer."[346]

The standard of care required when deploying autonomous weapons is even murkier given the lack of a common body of experience with such weapons. Moreover, the inherent uncertainty regarding how autonomous weapons will behave, the distributed knowledge problem discussed in Part V.B, the inability in many cases to reconstruct why an autonomous weapon acted as it did, and the secrecy of military systems and operations, will make it difficult to establish recklessness in many cases. States will need to develop norms of professional behavior, as well as national rules of engagement (ROE) and doctrine in order to inform what can be expected of a reasonable commander.

---

[341]   *Id*. at 384 (noting evidentiary problems).

[342]   *See id*. at 381.

[343]   Jens David Ohlin, *The Combatants Stance: Autonomous Weapons on the Battlefield*, 92 INT'L L. STUD. 1, 21 (2016) ("The far more frequent occurrence is one where the commander deploys the [autonomous weapon system] for military operations and the AWS violates a core prohibition of IHL[.]").

[344]   Int'l Comm. of the Red Cross, *Rule 156. Definition of War Crimes*, https://ihl-databases.icrc.org/ customary-ihl/eng/docs/v1_rul_rule156 (last visited Feb. 7, 2020) ("International case-law had indicated that war crimes are violations that are committed willfully, i.e., either intentionally (*dolus directus*) or recklessly (*dolus eventualis*). The exact mental element varies depending on the crime concerned."); Knut Dormann, *5.2. Elements of Specific Crimes Under Art. 8(2)(a) ICC Statute, in* ELEMENTS OF WAR CRIMES UNDER THE ROME STATUTE OF THE INTERNATIONAL CRIMINAL COURT: SOURCES AND COMMENTARY 38, 43 (2003) ("It may be concluded from the cases rendered by the ad hoc tribunals that the notion 'willful' includes 'intent' and 'recklessness', but excludes ordinary negligence."); Ryan Goodman, *Explainer: What Mental State is Requires to Commit a War Crime*, JUST SECURITY (Sept. 1, 2016), https://www.justsecurity.org/32644/explainer-mental-state-required-commit-war-crime/.

[345]   Ohlin, *supra* note 343, at 3. ("[T]here is still on jurisprudential area where international criminal law is ill-suited to prosecuting AWS cases, and that involves the mental state of recklessness."); Wagner, *supra* note 137, at 1404 ("Responsibility for negligence could only be established while the system is not designed to learn independently from past behavior, or in situations where designers acted negligently in supervising the development of AWS software when it comes to discretionary decision making.").

[346]   *Autonomy in Weapon Systems*, *supra* note 200.

In many cases, accidents that result in civilian casualties may not involve intentional or reckless conduct.[347] Human error, accidents, and equipment failures are inevitable in war, and may not constitute war crimes, "even if civilians are killed or injured as a result of those malfunctions."[348] The fact that a human is not prosecuted for a war crime in the event of unintended engagements, however, does not mean that there is an accountability gap. Prosecutions for war crimes receive significant public and scholarly attention, but they play a relatively minor role in promoting compliance with the law of war.[349] "[P]ost-hoc judicial accountability … is just one of many mechanisms for promoting and enforcing compliance with the laws of war[.]"[350]

States have an incentive to hold soldiers to high standards of conduct, to enforce compliance with ROEs, and to limit civilian casualties.[351] As the UK noted in a submission to the GGE on LAWS, "Responsibility is discharged through the military Chain of Command, and accountability measures are set out clearly in the orders, directives, and Standard Operating Procedures that are enforced by all responsible militaries engaged in the conduct of operations."[352] Incidents involving civilian casualties generally trigger investigations.[353] Soldiers may be subject to administrative discipline, or criminal sanction within the State's military justice system, even though they did not commit crimes under international law.[354] Dereliction in the performance of duties, for example, is an offense subject to U.S. court martial.[355] Investigations of suspected IHL violations may also focus more broadly on the conduct of governmental entities

[347]   Laura Dickinson, *Lethal Autonomous Weapons Systems: The Overlooked Importance of Administrative Accountability*, *in* THE IMPACT OF EMERGING TECHNOLOGIES ON THE LAW OF ARMED CONFLICT (Eric Talbot Jensen & Ronald Alcala eds., 2018 Forthcoming).

[348]   *Autonomy in Weapon Systems*, *supra* note 200, ¶ 30.

[349]   Crootof, *supra* note 43, at 1364 ("[T]he tendency of some to treat individual criminal accountability as the sole remedy to violations of [IHL] is a mistake.").

[350]   Kenneth Anderson & Matthew Waxman, *Law and Ethics for Robot Soldiers*, HOOVER INST. (Dec. 1, 2012), https://www.hoover.org/research/law-and-ethics-robot-soldiers.

[351]   Exec. Order No. 13732, 3 C.F.R. § 13732 (2017) ("The protection of civilians is fundamentally consistent with the effective, efficient and decisive use of force in pursuit of U.S. national interests.").

[352]   UK Working Paper, *Human Machine Touchpoints*, *supra* note 73.

[353]   Exec. Order No.13732, *supra* note 351 (requiring relevant agencies to "review or investigate incidents involving civilian casualties, including by considering relevant and credible information from all available sources[.]").

[354]   Dickinson, *supra* note 347, at 28 (noting that governments "invoke a variety of administrative procedures and mechanisms" when there is a suspected violation of IHL, which may result in "demotion, firing, reassignment, or financial penalties.").

[355]   10 U.S.C. § 892 (2014).

or organizations, and may result in recommendations for institutional reform to prevent similar conduct in the future.[356]

The Department of Defense's response to the tragic bombing of the Medecins Sans Frontiers (MSF) trauma center in Kunduz Province, Afghanistan, demonstrates how militaries can promote accountability in the absence of criminal conduct.[357] An extensive investigation into the causes of the mistaken attack found that personnel "misidentified and struck the MSF Trauma Center" due to a "combination of human errors, compounded by process and equipment failures."[358] Although the investigation concluded that no war crime had been committed, it identified "sixteen U.S. servicemembers whose conduct warranted consideration for appropriate administrative or disciplinary action, including a general officer."[359] The resulting administrative actions included "suspension and removal from command, letters of reprimand, formal counseling, and extensive retraining."[360] This type of administrative investigation and discipline could similarly be used in the event that human negligence or error leads to unintended engagements by autonomous weapons.

The doctrine of State responsibility also plays a critical role in preventing an accountability gap. Despite recent focus on international criminal law, State responsibility has traditionally been the primary vehicle for addressing violations of IHL.[361] This mode of responsibility may again assume a more prominent role with respect to the use of autonomous weapons, as proving individual accountability becomes more difficult.[362] As the United States observed in a submission to the GGE on LAWS: "States are responsible for the acts of persons forming part of their armed forces. It follows that States are responsible for the [use of autonomous weapons] by persons forming part of their armed forces as well as other such acts that may be attributable to a State under the law of State responsibility."[363] The legal consequences for a State that has committed an internationally wrongful act include "offer[ing] appropriate assurances and guarantees of non-repetition, if circumstances so require," and

---

[356]   Dickinson, *supra* note 347, at 27.

[357]   *CENTCOM Releases Investigation Into Airstrike on Doctors Without Borders Trauma Center*, U.S. CENT. COMMAND (Apr. 29, 2016), http://www.centcom.mil/MEDIA/PRESS-RELEASES/Press-Release-View/Article/904574/april-29-centcom-releases-investigation-into-airstrike-on-doctors-without-borde/.

[358]   *Id.*

[359]   *Id.*

[360]   *Id.*

[361]   *See* Crootof, *supra* note 43, at 1354–58.

[362]   Heyns, *supra* note 8, ¶ 14 ("In general, a stronger emphasis on State as opposed to individual responsibility may be called for, except in respect of its use by non-state actors.").

[363]   *Autonomy in Weapon Systems*, *supra* note 200, at ¶ 26.

"mak[ing] full reparation for the injury caused by the internationally wrongful act.[364] The rise of autonomous weapons may thus warrant a shift in collective attention from individual criminal responsibility back to State responsibility.

Focusing on State responsibility, rather than individual responsibility, is a rational response to the use of weapons systems that involve disparate individuals acting in furtherance of a governmental policy. Unintended engagements by autonomous weapons will likely be the result of systematic problems rather than individual malfeasance. Nevertheless, the doctrine of State responsibility does not fully resolve the potential accountability gap. State responsibility requires an internationally wrongful act (in this context a violation of IHL) and not simply harm to civilians or civilian objects. Given the focus over the past several decades on international criminal law, violations of IHL are generally viewed through a criminal law lens, requiring intent or recklessness. In many cases, the delta between war crime and a non-criminal violation of IHL is non-existent or at least unclear. For this reason, Rebecca Crootof has called for the establishment of a "war tort" regime, under which States could be held liable for harms caused by autonomous weapons that do not constitute war crimes.[365] States may resist establishing such a mode of legal liability, but focusing on State responsibility nevertheless provides a more useful framework for accountability than individual criminal law. Even when States deny legal wrongdoing, they may face strong political or diplomatic pressures to provide remedies to victims.[366] Further work on approaches and mechanisms for holding States accountable, and providing redress to victims, for autonomous weapons malfunctions is needed.

In sum, the use of autonomous weapons will not necessarily create an accountability gap, although it may alter how we understand and seek

---

[364] Draft Articles, *supra* note 331, arts. 30–31.

[365] Crootof, *supra* note 43.

[366] In practice, States have provided remedies to victims in the form of ex gratia payments, even when they deny legal responsibility for civilian harm. Executive Order 13732, for example, requires relevant agencies, as appropriate and consistent with mission objectives and applicable law, to "acknowledge U.S. Government responsibility for civilian casualties and offer condolences, including *ex gratia* payments, to civilians who are injured or to the families of civilians who are killed." Exec. Order No. 13732, *supra* note 351, sec. 2(b)(ii). This practice is not new. In 1988, the USS Vincennes shot down an Iranian commercial airliner over the Persian Gulf, killing 290 civilians, after confusing it for a hostile F-14 fighter jet. Aerial Incident of 3 July 1988 (Iran v. U.S.), Order, 1989 I.C.J. Rep. 132, at 132–33 (Dec. 13). The United States denied legal wrongdoing after a formal investigation concluded that the shootdown was reasonable in light of the circumstances and information known to the commander at the time. *Id.* at 12. Nevertheless, following a suit brought by the Iranian Government at the International Court of Justice, the United States agreed to settle by paying compensation to the families of the victims. David K. Linnan, *Iran Air Flight 655 and Beyond: Free Passage, Mistaken Self-Defense, and State Responsibility*, 16 YALE J. INT'L L. 245, 258 (1991).

accountability. Given the gray areas concerning standards of care in the use of emerging technologies, the accountability pendulum may again shift away from individual criminal liability to State responsibility (in either the legal or diplomatic sense). States have robust tools for encoring military discipline and compliance with ROE and can likewise be held responsible under international law for violations of IHL committed by members of their military.

CONCLUSION

Autonomous weapons will be increasingly prevalent in future conflicts, but it is premature to say whether this is a cause for concern. As with other technological developments in warfare, autonomy can be used in a manner that enhances military effectiveness while also limiting risk to civilians. It can also, if developed or used irresponsibly, present grave dangers to the civilian population.

States have made clear that IHL applies to the use of autonomous weapons. This axiomatic assertion, however, only raises more questions regarding how IHL will apply in practice. We should not be sanguine about the challenges that States and academics will face in applying a legal framework that was designed for conventional warfare to a fundamentally new type of weapon.

Autonomy will transform warfare in ways that directly implicate the traditional understandings and applications of the law of war. IHL's focus on human decision-making will need to account for machine decision-making. Human-oriented standards, such as the reasonable commander, will come under greater strain as decisions to employ force will hinge to a greater extent on knowledge of the capabilities of the weapon system, potentially leading to greater focus on the effects of an attack. The importance of context in assessing IHL requirements will be questioned as commanders are increasingly removed, both geographically and temporally, from the use of force. The human judgments required by IHL will be further complicated as knowledge of the weapon system and operations is distributed among a larger number of individuals.

This is not to say that new international law is warranted at this time. IHL has proved sufficiently flexible to regulate the use of modern weapons with modern technologies. Autonomy in weapons, however, presents unique challenges because it has the potential to blur the distinction between weapon and operator. The ongoing debates regarding whether to preemptively ban autonomous weapons will likely have little effect on the major military powers

that are committed to developing this technology. It is time to shift the focus of this debate to how IHL can and should be applied in practice. This Article has sought to further that conversation by examining a number of conceptual and interpretive challenges that States, practitioners and the legal academy will confront with the emergence of such weapons.