South Carolina Law Review

Volume 69 | Issue 2 Article 5

Winter 2017

Autonomous Weapons and International Law

Christopher M. Ford

AUTONOMOUS WEAPONS AND INTERNATIONAL LAW

Lieutenant Colonel Christopher M. Ford*

---

* Lieutenant Colonel Christopher M. Ford, U.S. Army Judge Advocate General. This Article reflects research conducted while serving as a Military Professor at the Stockton Center for the Study of International Law at the United States Naval War College, and incorporates feedback and input from meetings with organizations and individuals, military and civilian, at the forefront of research and development in the field of autonomy. To this end, I would like to thank Professor Michael Schmitt, Jason Coats, Alan Schuller, Professor Peter Margulies, the U.S. Department of State, the U.S. Navy Autonomous Research Labs, the U.S. Defense Advanced Research Projects Agency, the United Nations Institute for Disarmament Research, the Geneva Academy, Human Rights Watch, Human Rights First, Professor Marco Sassóli, Professor Tim McCormack, Dustin Lewis, Rebecca Crootof, Heather Roff, Professor Leslie Kaelbling, Ben Price and colleagues at the Lincoln Laboratory, and Paul Scharre at the Center for a New American Security. The views expressed are the author's own and do not represent the U.S. Department of Defense or any other government entity.

I. INTRODUCTION

Autonomous weapons—those weapons that can select and engage targets without human involvement[1]—herald perhaps the most fundamental change in warfare in generations. The rise of autonomous weapons has sparked a robust international debate centered on one key question: Are the current laws of armed conflict sufficient to govern autonomous weapon systems? Positions span the spectrum from a call for preemptive prohibition[2] to arguments that current legal norms are adequate to regulate these future weapon systems.[3] While autonomy may give rise to circumstances in which the application of the law is rendered uncertain or difficult, the current normative legal framework is sufficient to regulate the new technology.

---

1. BONNIE DOCHERTY, HUMAN RIGHTS WATCH, LOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS (Steve Goose ed., 2012), https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots.

2. *E.g.*, *id.*

3. *E.g.*, 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/sites/default/files/uploads/documents/Anderson-Waxman_LawAndEthics_r2_FINAL.pdf.

The legality of such weapons has generated significant interest in both the public and private sectors. The International Committee for the Red Cross (ICRC) has hosted two meetings of experts,[4] and the United Nations, within the framework of the Convention on Certain Conventional Weapons (CCW), has convened three informal meetings of experts and recently recommended the establishment of a Group of Governmental Experts to study the issue.[5] State interest in autonomous weapons has been keen. Fourteen States have publicly called for a preemptive ban on autonomous weapons.[6] Other States have taken an active role in the debate, including notably the United States, the United Kingdom, China, Russia, India, Pakistan, Canada, France, Australia, the Netherlands, Belgium, and Germany.[7]

Numerous non-governmental organizations have been deeply involved in the issue, including the United Nations Institute for Disarmament Research (UNIDIR),[8] Geneva Academy,[9] Center for a New American Security,[10] Human Rights Watch and the Human Rights Clinic at Harvard Law School,[11] the Harvard Law School Program on International Law and

---

4. INT'L COMM. OF THE RED CROSS, EXPERT MEETING: AUTONOMOUS WEAPON SYSTEM: TECHNOLOGICAL, MILITARY, LEGAL, AND HUMANITARIAN ASPECTS 5 (2014).

5. Frank Sauer, *Arms Control for AWS: 2016 and Beyond*, INT'L COMM. FOR ROBOT ARMS CONTROL (Dec. 7, 2016), https://icrac.net/2016/12/arms-control-for-aws-2016-and-beyond/.

6. *Ban Support Grows, Process Goes Slow*, CAMPAIGN TO STOP KILLER ROBOTS (Apr. 15, 2016), https://www.stopkillerrobots.org/2016/04/thirdmtg/ (noting that the following countries have publicly endorsed a ban: Algeria, Bolivia, Chile, Costa Rica, Cuba, Ecuador, Egypt, Ghana, Holy See, Mexico, Nicaragua, Pakistan, State of Palestine, and Zimbabwe).

7. *See Statements from the 2016 CCW Meeting of Experts on Autonomous Weapons*, REACHING CRITICAL WILL (Apr. 12, 2016), http://www.reachingcriticalwill.org/disarmament-fora/ccw/2016/laws/statements.

8. *See, e.g.*, *Framing Discussions on the Weaponization of Increasingly Autonomous Technologies*, U.N. INST. FOR DISARMAMENT RESEARCH 1, 2 (2014), http://www.unidir.org/files/publications/pdfs/framing-discussions-on-the-weaponization-of-increasingly-autonomous-technologies-en-606.pdf.

9. *See, e.g.*, Nathalie Weizmann, *Autonomous Weapon Systems Under International Law Academy*, Briefing No. 8, GENEVA ACAD. OF INT'L HUMANITARIAN LAW AND HUM. RIGHTS (Nov. 2014), https://www.geneva-academy.ch/joomlatools-files/docman-files/Publications/Academy%20Briefings/Autonomous%20Weapon%20Systems%20under%20International%20Law_Academy%20Briefing%20No%208.pdf.

10. *See, e.g.*, Michael Horowitz & Paul Scharre, *An Introduction to Autonomy in Weapon Systems* 2 (Feb. 13, 2015) (working paper) (on file with Center for a New American Security), https://www.geneva-academy.ch/joomlatools-files/docman-files/Publications/Academy%20Briefings/Autonomous%20Weapon%20Systems%20under%20International%20Law_Academy%20Briefing%20No%208.pdf.

11. *See, e.g.*, DOCHERTY, *supra* note 1.

Armed Conflict,[12] the United States Army Strategic Studies Institute,[13] the International Institute of Humanitarian Law,[14] the Friedrich-Ebert-Stiftung,[15] the Stockholm International Peace Research Institute,[16] and the Advisory Council on International Affairs for the Netherlands.[17] Civil society organizations have likewise addressed the subject.[18]

This Article considers the entirety of positions and seeks to present a comprehensive, objective discussion of the relevant issues in international law. In the course of this examination, the Article reaches two broad conclusions. First, autonomy is less a technology as it is a capability comprised of multiple technologies. For this reason, this Article uses the term Autonomy Enabled Weapons (AEWs) rather than the more common Lethal Autonomous Weapons (LAWs). AEWs cannot be considered a homogeneous category of weapons that comply, or not, with the law of armed conflict. Rather, whether a given system meets the requirements of this body of law will depend on the system and the manner in which it is used. Second, this Article argues that the normative framework established by the law of armed conflict is sufficient to ensure the lawful operation of most types of weapons employing autonomous technologies.

This Article proceeds in four parts. Part II considers various approaches used by researchers trying to define autonomy, including the discussion in other fields such as aerospace and aeronautical engineering. Part III examines AEWs in the context of the law of armed conflict. In addition to consideration of the rules of distinction, proportionality, and precautions in

---

12. *See, e.g.*, DUSTIN A. LEWIS ET AL., HARVARD LAW SCH. PROGRAM ON INT'L LAW AND ARMED CONFLICT, WAR-ALGORITHM ACCOUNTABILITY (2016), https://pilac.law.harvard.edu/waa/.

13. *See, e.g.*, JEFFREY L. CANTON, U.S. ARMY WAR COLL. STRATEGIC STUDIES INST., AUTONOMOUS WEAPONS SYSTEMS: A BRIEF SURVEY OF DEVELOPMENTAL, OPERATIONAL, LEGAL, AND ETHICAL ISSUES xi (James G. Pierce ed., 2015).

14. *See, e.g.*, *Proceedings*, INT'L INST. OF HUMANITARIAN LAW, http://stage.iihl.org/proceedings/ (last visited Nov. 11, 2017).

15. *See, e.g.*, ROBIN GEISS, FRIEDRICH-EBERT-STIFTUNG, THE INTERNATIONAL-LAW DIMENSION OF AUTONOMOUS WEAPONS SYSTEMS 6 (2015), http://library.fes.de/pdf-files/id/ipa/11673.pdf.

16. *See, e.g.*, VINCENT BOULANIN, SIPRI INSIGHTS ON PEACE AND SECURITY NO. 2015/1, IMPLEMENTING ARTICLE 36 WEAPON REVIEWS IN THE LIGHT OF INCREASING AUTONOMY IN WEAPON SYSTEMS 1 (2015), https://www.sipri.org/sites/default/files/files/insight/SIPRIInsight1501.pdf.

17. *See, e.g.*, ADVIESRAAD INTERNATIONALE VRAAGSTUKKEN, AUTONOMOUS WEAPON SYSTEMS: THE NEED FOR MEANINGFUL HUMAN CONTROL (2016), http://aiv-advies.nl/8gr.

18. *See, e.g.*, *Ban Support Grows, Process Goes Slow*, *supra* note 6; *Introduction—Autonomous Weapons*, ARTICLE 36 (Apr. 12, 2016), http://www.article36.org/autonomous-weapons/introduction-autonomous-weapons/.

attack, this Part includes a discussion of the concept of "meaningful human control." Part IV addresses the requirement to test new weapons to ensure their compliance with international law. A discussion of accountability for the use of AEWs, including individual liability, command responsibility, and State responsibility concludes the piece.

II. DEFINING AUTONOMY

A common definition of an autonomous weapon used in the legal literature is a weapon that can select and engage a target without human involvement.[19] While this definition is commendably succinct, it raises significant questions: What does it mean to "select"? How and when can systems operate "without human involvement" if all systems are programmed by humans? Defining autonomy has proven vexing for all disciplines, particularly aeronautical and aerospace engineering, which have debated the issue for decades.[20] Diplomats and government arms control officials have struggled. At the 2016 CCW Informal Meeting of Experts, States failed to reach consensus over whether they should even attempt to define the term "autonomous."[21]

Where States and legal academics have tried to define relevant terms, they have adopted several distinct approaches.[22] These approaches are not necessarily contradictory, and at times commentators have adopted combinations thereof. This Part will briefly summarize each approach and then provide a suggested framework.

A. Defining Autonomy as a Discrete Concept

The term "autonomous" derives from the Greek words *autos* meaning "self" and *nomos* meaning "law."[23] The etymology of the word directly informs the common meaning of the term, often understood as something

---

19. *See, e.g.*, DOCHERTY, *supra* note 1.

20. *See generally* Chad R. Frost, *Challenges and Opportunities for Autonomous Systems in Space*, *in* FRONTIERS OF ENGINEERING: REPORTS ON LEADING-EDGE ENGINEERING FROM THE 2010 SYMPOSIUM 89–90 (showing the difficulty in defining the term "autonomy").

21. Chris Ford & Chris Jenks, *The International Discussion Continues: 2016 CCW Experts Meeting on Lethal Autonomous Weapons*, JUST SECURITY (Apr. 20, 2016), https://www.justsecurity.org/30682/2016-ccw-experts-meeting-laws/.

22. *See* discussion *infra* Sections II.A–II.D.

23. *Autonomous*, ONLINE ETYMOLOGY DICTIONARY, http://www.etymonline.com/index.php?term=autonomous&allowed_in_frame=0.

that is self-governing.[24] This construct does not surface in legal discourse, perhaps because of the legal term of art "govern." This formulation is, however, used on occasion in the engineering fields.[25]

In the legal space, most definitions tend to be a variation on the notion that the term denotes a weapon system that *can select and engage a target without human involvement*. Examples include definitions proffered by the U.S. Department of Defense,[26] the United Kingdom Ministry of Defense,[27] the Geneva Academy,[28] Center for a New American Security,[29] the Netherlands Advisory Council on International Affairs,[30] UN Special Rapporteur Christof Heyns,[31] Michael Schmitt and Jeff Thurnher,[32] and Rebecca Crootof.[33] Given the centrality of this definitional approach, it is useful to consider its parts: (1) *weapons*; that can (2) *select and engage a target*; (3) *without human involvement*.

1. Weapons

For this Article, "weapon" is read to mean those instruments intended to damage, destroy, or injure personnel or property.[34] This Article does not

---

24. *Autonomous*, MERRIAM-WEBSTER DICTIONARY (online ed. 2016) (autonomous is something "having the power or right to govern itself").

25. *See, e.g.*, FED. AGENCIES AD HOC AUTONOMY LEVELS FOR UNMANNED SYSTEMS WORKING GROUP PARTICIPANTS, AUTONOMY LEVELS FOR UNMANNED SYSTEMS (ALFUS) FRAMEWORK 8 (Hui-Min Huang ed., 2004) [hereinafter AD HOC AUTONOMY]; Panos J. Antsaklis et al., *An Introduction to Autonomous Control Systems*, in 5TH IEEE INT'L SYMPOSIUM ON INTELLIGENT CONTROL (1991).

26. U.S. DEP'T OF DEF., DIRECTIVE 3000.09, AUTONOMY IN WEAPON SYSTEMS 13–14 (Nov. 2012), http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf [hereinafter DOD DIRECTIVE].

27. U.K. MINISTRY OF DEFENSE, JOINT DOCTRINE NOTE 2/11: THE UK APPROACH TO UNMANNED AIRCRAFT SYSTEMS ¶ 203 at 2–3 (2011).

28. Weizmann, *supra* note 9, at 5.

29. Horowitz & Scharre, *supra* note 10, at 25.

30. ADVISORY COUNCIL ON INT'L AFFAIRS & ADVISORY COMM. ON ISSUES OF PUB. INT'L LAW, AUTONOMOUS WEAPON SYSTEMS: THE NEED FOR MEANINGFUL HUMAN CONTROL 11 (2015).

31. Christof Heyns (Special Rapporteur on Extrajudicial, Summary and Arbitrary Executions), Rep. to Human Rights Council, U.N. Doc. A/HRC/23/47 ¶ 38 (Apr. 9, 2013).

32. Michael N. Schmitt & Jeffrey S. Thurnher, *Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict*, 4 HARV. NAT'L SEC. J. 231, 235 (2013).

33. Rebecca Crootof, *The Killer Robots Are Here: Legal and Policy Implications*, 36 CARDOZO L. REV. 101, 106 (2015).

34. *See, e.g.*, U.S. DEP'T OF THE NAVY, NAVAL INSTRUCTION 5000.2E, ¶ 1.6.1.c; U.S. DEP'T OF THE AIR FORCE, AIR FORCE INSTRUCTION 51-402, WEAPONS REVIEW 1 (1994); Int'l Comm. of the Red Cross Geneva, *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977*, 88

distinguish between weapons designed to operate in a defensive posture and those intended for offensive applications. Nor is a distinction drawn between those weapons designed to kill or injure.

2. Select and Engage a Target

The second component of the definition refers to the ability of the system to "select and engage" targets.[35] The term "select" is widely understood to mean to "choose among" a group.[36] The ordinary meaning of "engage" in the military context—to "enter into combat or battle"[37]—requires some clarification. With regards to autonomous weapons systems, engage could refer to at least three different points in time: (1) when the system is activated; (2) when the system is operationally selecting targets; or (3) when the system is applying the instrument designed to kill, injure, or destroy its selected target. Most commentators—and this author—read engage to refer to the third meaning.[38]

Thus, by this approach, the machine is selecting among several targets and making a determination as to when and where to engage the chosen target with the weapon. This reading would exclude remotely controlled systems where a human operator determines which target to engage, and when, such as the General Atomics MQ-9 Reaper[39] and MQ-1 Predator.[40] The phrase "select and engage targets" also serves to exclude autonomous functions that are unrelated to targeting and engaging such as navigation.

INT'L REV. RED CROSS 864, 938 (2006) [hereinafter ICRC WEAPONS GUIDE]; INT'L GROUP OF EXPERTS, NATO COOPERATION CYBER DEF., TALLINN MANUAL ON THE INTERNATIONAL LAW APPLICABLE TO CYBER WARFARE 142 (Michael N. Schmitt ed., 2013) ("A weapon is generally understood as the aspect of the system used to cause damage or destruction to objects or to injure or death to persons.") [hereinafter TALLINN MANUAL]; THE PROGRAM ON HUMANITARIAN POLICY AND CONFLICT RESEARCH AT HARVARD UNIV., HPCR: MANUAL ON INTERNATIONAL LAW APPLICABLE TO AIR AND MISSILE WARFARE 49 (2010) ("The essence of a weapon is that it is an object used to cause (i) death of, or injury to, persons; or (ii) damage to, or destruction, of objects.").

35. *See, e.g.*, DOD DIRECTIVE, *supra* note 26, at 14; Weizmann, *supra* note 9, at 6.

36. *Select*, MERRIAM-WEBSTER DICTIONARY (11th ed. 2016).

37. *Engage*, MERRIAM-WEBSTER DICTIONARY (11th ed. 2016).

38. *See* DOD DIRECTIVE, *supra* note 26, at 14; Weizmann, *supra* note 9, at 6.

39. *MQ-9 Reaper*, U.S. AIR FORCE (Sept. 23, 2015), http://www.af.mil/AboutUs/FactSheets/Display/tabid/224/Article/104470/mq-9-reaper.aspx.

40. *MQ-1B Predator*, U.S. AIR FORCE (Sept. 23, 2015), http://www.af.mil/About-Us/Fact-Sheets/Display/Article/104469/mq-1b-predator/.

3. Without Human Involvement

The final component of the definition requires that the actions of the system occur "without human involvement."[41] The presence and degree of human participation in a weapons system are not always clear. Consider the Aegis Combat System, a ship-borne weapon that can identify threats and engage threats automatically, or allow a human operator to make the engagement decision.[42] Even in the latter case, it selects the targets, and the operator's decision to engage is based entirely on data provided by the system. In such a circumstance, is a human involved? Is the operator adding any value to the process or just pressing a button?

The difference between automaticity and autonomy is rather indistinct, since, as the ICRC has noted, both types of systems "have the capacity to independently select and attack targets within the bounds of their human-determined programming."[43] One is left wondering what degree of freedom in a system is sufficient such that the system is considered to be operating without human involvement.

The ICRC concludes that "[t]he difference appears only to be the degree of 'freedom' with which the weapon system can select and attack different targets."[44] Christof Heyns, the UN Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, draws a distinction in the nature of the system's operation: "Automatic systems, such as household appliances, operate within a structured and predictable environment. Autonomous systems can function in an open environment, under unstructured and dynamic circumstances."[45] This definition is unsatisfying since automatic systems can operate in unstructured and dynamic circumstances. Conversely, autonomous systems can function in structured and predictable environments.

The engineering literature provides a more robust discussion of definitions of autonomy.[46] In a paper examining autonomy in aerospace applications, a senior NASA researcher proposed the following:

---

41. *See, e.g.*, DOD DIRECTIVE, *supra* note 26, at 14; Weizmann, *supra* note 9, at 6.

42. Gary E. Marchant et al., *International Governance of Autonomous Military Robots*, 12 COLUM. SCI. & TECH. L. REV. 272, 273 (2011).

43. INT'L COMM. OF THE RED CROSS, *supra* note 4, at 64.

44. *Id.*

45. Heyns, *supra* note 31, ¶ 42.

46. *See, e.g.*, Raja Parasuraman et al., *A Model for Types and Levels of Human Interaction with Automation*, 30 IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART A: SYSTEMS AND HUMANS 286, 287 (2000); AD HOC AUTONOMY, *supra* note 25, at 8; Antsaklis et al., *supra* note 25, at 5; BRUCE T. CLOUGH, AIR FORCE RESEARCH LIBRARY, METRICS, SCHMETRICS! HOW THE HECK DO YOU DETERMINE A UAV'S

> An *automated system* doesn't make choices for itself—it follows a script, albeit a potentially sophisticated script, in which all possible courses of action have already been made . . . . Thus for an automated system, choices have either already been made and encoded, or they must be made externally. By contrast, an *autonomous system* does make choices on its own. It tries to accomplish its objectives locally, without human intervention, even when encountering uncertainty or unanticipated events.[47]

Both autonomous and automated systems may be predictable with regards to the overall action of the system—say the destruction of a tank—but the particular, or component actions, of an autonomous system are not readily predictable. Thus, it seems that the distinction between automated and autonomous turns on the predictability of the system, specifically the component actions of the system.[48]

Plainly, what constitutes autonomous (e.g., without human involvement) and automated (e.g., with human involvement) is quite complex. This complexity points to the disadvantage of the discrete definitional approach. While such a definition is readily understood, it is too simplistic. Aside from the difficulties of parsing out automated from autonomous, consider what it means for a system to "select" a target? If the system has been programmed to select among two predetermined targets, has the machine selected anything? What if the selection is among 500 targets? Clearly, adding some nuance to the definition serves a useful function.

B. Defining Autonomy as a Spectrum

It is tempting to conceptualize autonomy as a binary proposition: either something is autonomous or not. It is more useful, however, to consider autonomy as a spectrum. Parsing out "highly" autonomous systems from "low-level" autonomous systems, however, requires the application of criteria against which one can judge the level of autonomy.

In a project designed to develop a "framework to facilitate characterizing and articulating autonomy for unmanned systems," the U.S. National Institutes of Standards and Technology proposed a three-part

---

AUTONOMY ANYWAY? (2002), http://www.dtic.mil/dtic/tr/fulltext/u2/a515926.pdf; WALT TRUSZKOWSKI ET AL., AUTONOMOUS AND AUTOMATIC SYSTEMS: WITH APPLICATIONS TO NASA INTELLIGENT SPACECRAFT OPERATIONS AND EXPLORATION SYSTEMS (2009).

47. Frost, *supra* note 20, at 2.

48. *See, e.g.*, U.K. of Gr. Brit. and N. Ir., Statement to the Informal Meeting of Experts on Lethal Autonomous Weapons Systems (Apr. 11–15, 2016).

framework that looked at human independence, mission complexity, and environmental complexity.[49] Other researchers have spoken of attributes of autonomy that include independence from operator interaction, adaptability to environment, and the machine's ability to adapt the means of achieving the ends.[50] Another researcher characterizes autonomy as a function of "the capacity to operate without outside intervention," the ability to choose actions and capacity to choose whether to choose.[51]

Pioneering researchers in the field of autonomy, Thomas Sheridan and William Verplank, developed a ten-level spectrum of autonomy, which, in its essence, is a spectrum of human involvement in a machine's process.[52] Similarly, in a document entitled *Unmanned Systems Integrated Roadmap FY 2011-2036*, the U.S. Department of Defense (DoD) expressed autonomy on a three-part scale that distinguishes between automated, semi-autonomous, and autonomous weapons—again depending on the level of human involvement.[53] The Center for a New American Security employs a similar scale.[54] The distinction between semi-autonomous and autonomous systems refers to whether a human operator selects the target.[55]

Considering autonomy as a spectrum provides some much-needed granularity to the discussion but fails to provide a satisfying definition. Looking to the DoD construct, whether a system is autonomous or semi-autonomous depends on whether the system "selects" the target. DoD Directive 3000.09 attempts to clarify this distinction by excluding from the definition of autonomous weapons those systems where "individual targets

---

49. Hui-Min Huang, Software & Mech. Eng'r with Nat'l Inst. of Standards & Tech., PowerPoint Presentation on Autonomy Levels for Unmanned Systems (July 20–21, 2005), http://www.nist.gov/el/isd/ks/upload/ALFUS-BG.pdf.

50. *See* Troy B. Jones & Mitch G. Leammukda, *Requirements-Driven Autonomous System Test Design: Building Trusting Relationships*, 15TH ANNUAL LIVE-VIRTUAL-CONSTRUCTIVE CONFERENCE INT'L TEST AND EVALUATION ASS'N 1, 6 (2011), https://www.researchgate.net/profile/Mitch_Leammukda/publication/228598990_Requirements-Driven_Autonomous_System_Test_Design_Building_Trusting_Relationships/links/00b49536d4800877cd000000/Requirements-Driven-Autonomous-System-Test-Design-Building-Trusting-Relationships.pdf.

51. O.G. Clark et al., *Mind and Autonomy in Engineered Biosystems*, 12 ENG'G APPLICATIONS OF ARTIFICIAL INTELLIGENCE 389, 397 (1999).

52. THOMAS B. SHERIDAN & WILLIAM L. VERPLANK, HUMAN AND COMPUTER CONTROL OF UNDERSEA TELEOPERATORS (1978), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.694.7165&rep=rep1&type=pdf.

53. U.S. DEP'T OF DEF., UNMANNED SYSTEMS INTEGRATED ROADMAP FY 2011-2036 43 (2013), https://fas.org/irp/program/collect/usroadmap2011.pdf. This scale is also seen in other DoD policy documents. *See, e.g.*, DOD DIRECTIVE, *supra* note 26, at 22.

54. Horowitz & Scharre, *supra* note 10, at 5–7.

55. *Id.* at 16.

or specific target groups [] have been selected by a human operator."[56] This language again begs the question of what is meant by "select." If the human operator instructs the machine to target threats (which have been preprogrammed), is the machine "selecting" when it attacks a particular object? This approach also fails to consider that a system may have varying levels of autonomy in its various subsystems.

C. Autonomy as a Loop

Another common method for conceptually organizing levels of autonomy is considering autonomy in the context of "the loop."[57] This approach is really a variation of the "autonomy as a spectrum" approach and has been used by Christof Heyns,[58] Human Rights Watch,[59] the Center for a New American Security,[60] Rebecca Crootof,[61] Markus Wagner,[62] and others. This definitional construct breaks autonomy into three categories based on the nature of human/machine interaction.[63]

In an influential paper, a group of computer scientists reviewed the cognitive psychology literature and suggested that all decision-making and action can be roughly simplified into four categories: information acquisition, information analysis, decision and action selection, and action implementation.[64] A popular formulation of this taxonomy is John Boyd's OODA loop, where the variables are described as observe, orient, decide, and act.[65] In this definitional framework, autonomy is an expression of the human's involvement (or lack thereof) in the "loop," which refers to the information processing cycle of the machine.[66] Autonomy then is expressed as a function of the "human in the loop," "human on the loop," or "human out of the loop."[67]

---

56. DOD DIRECTIVE, *supra* note 26, at 3.

57. Heyns, *supra* note 31, at 8; DOCHERTY, *supra* note 1, at 3; Horowitz & Scharre, *supra* note 10, at 8.

58. Heyns, *supra* note 31, at 8.

59. DOCHERTY, *supra* note 1, at 3.

60. Horowitz & Scharre, *supra* note 10, at 8.

61. Crootof, *supra* note 33, at 125.

62. Markus Wagner, *Taking Humans Out of the Loop: Implications for International Humanitarian Law*, 21(2) J.L. INFO. & SCI. 155, 155 (2011), http://www6.austlii.edu.au/cgi-bin/viewdoc/au/journals/JlLawInfoSci/2012/9.html.

63. *See* Heyns, *supra* note 31, at 2–3.

64. Parasuraman et al., *supra* note 46, at 288.

65. John Boyd, PowerPoint Presentation on The Essence of Winning and Losing (June 28, 1995), http://www.danford.net/boyd/essence.htm.

66. *See* DOCHERTY, *supra* note 1, at 2.

67. *Id.* at 2–3.

Human in the loop refers to a situation where the human operator plays an integral role in the operation of the machine—the machine cannot accomplish its task without human involvement.[68] A human on the loop system is one in which the human monitors the system and can intervene before the system takes action.[69] Finally, where the system is a human out of the loop, the human plays no role in the machine's execution of its task.[70]

There is also the possibility that a human could be "near the loop."[71] This phrase is a new concept that reflects situations where an autonomous system is deployed near humans. For instance, an autonomous system to provide logistical support. The system may be operating autonomously such that there is no one on or in the loop. The unit commander can, however, observe the battlefield and the operation of the system. If the commander sees something that would necessitate changes to the functioning of the autonomous system, the commander would have the ability to inform an operator who could then get on or in the loop and make the appropriate changes.

However one characterizes the loop, there are potential issues with this definitional framework. As some authors have cautioned, given the complexity of certain systems, it might be overly simplistic to characterize a particular system as an "in/on/out" system.[72] Further, a system may have several decisional loops regarding various aspects of its operations, some of which may (or may not) have human involvement. There is also an inherent difficulty in characterizing the precise nature of the human's role.

D. Information Processing Models and Scales of Autonomy

A final way to consider autonomy is to combine different frameworks such as informational processing models together with a scale of autonomy. A 2016 report from the U.S. Department of Defense Science Board (DSB), for example, uses the formulation of sense, think/decide, act, and team.[73] The report usefully further breaks down each factor into technology that is

68. *Id.*
69. *Id.* at 3.
70. *Id.*
71. *See id.*
72. *See, e.g.*, William C. Marra & Sonia K. McNeil, *Understanding "The Loop": Regulating the Next Generation of War Machines*, 36 HARV. J.L. & PUB. POL'Y 1139, 1179 (2012) [hereinafter *The Loop*].
73. DEF. SCIENCE BD., U.S. DEP'T OF DEF., TASK FORCE REPORT: SUMMER STUDY ON AUTONOMY 9 (2016).

available today, likely available in the near term, and may be available in the long term.[74]

Others, such as the Air Force Research Laboratory, have combined Sheridan and Verplank's scale of autonomy with Boyd's OODA loop to create a table of autonomous control levels.[75] This model allows one to express varying levels of autonomy for each part of the OODA loop. Thus, a system may be highly autonomous with regards to its ability to "observe" but completely lacking autonomy in its ability to "act." This construct allows commentators and policymakers a more precise mechanism for describing what is being discussed, regulated, or both.

There is perhaps no single best way to define autonomous weapons. It would seem, however, that there is utility in considering autonomous functions against an informational processing model such as the OODA loop. Such a formulation allows users to precisely describe the autonomous attributes of a given system, which may vary widely in a particular system. For example, a system might be highly autonomous in its ability to observe and orient itself but have a low level of autonomy in deciding and acting. This Article then suggests the following chart as a useful framework to consider the autonomous functions of a weapon system.

IV – Fully automated *– Operative without the need for human interaction in any function*				
Examples	**Observe**	**Orient**	**Decide**	**Act**
Level IV systems do not yet exist.	Computer gathers data without direction. Provides no information to a human controller. What is observed is unpredictable.	Computer analyzes data without any human input or report to a human controller. Computer analysis is unseen and unpredictable.	Computer ranks targets and determines where and when to engage. Targeting is unseen and unpredictable.	Computer decides when and where to execute. Actions are potentially unpredictable in time and space.

74. *Id.* at 11 tbl.1.
75. CLOUGH, *supra* note 46.

III – Largely automated *– Inoperative without human interaction in some functions*				
Examples	**Observe**	**Orient**	**Decide**	**Act**
Phalanx; Patriot in Automatic Mode; Israeli "Harpy" Counter-Radar System	Computer automatically gathers data based on previously established criteria. Observation is predictable. Information may be provided to human controller.	Computer analyzes data utilizing previously established criteria but without any contemporaneous human input. Analysis is predictable.	Computer ranks targets and determines where and when to engage within previously established parameters. Targeting is predictable.	Computer decides to execute based on previously established parameters. Actions are generally predictable in time and space.
II – Automated in many functions *– Inoperative without human interaction in most functions*				
Examples	**Observe**	**Orient**	**Decide**	**Act**
Phoenix over-the-horizon missile; Paladin artillery; Cruise missiles; MIA2 Abrams tank	Computer gathers data at human controller direction. Provides simultaneous information to human controller.	Computer analyzes data utilizing previously established criteria. Contemporaneous human input is present. Analysis is predictable.	Computer ranks targets, but human approval is required. Contemporaneous human input is present. Targeting is predictable.	Computer suggests execution and executes after human approval. Human is shadow for contingencies.

I – Little to no automation – Inoperative without human interaction in all functions				
Examples	**Orient**	**Orient**	**Decide**	**Act**
Short-range artillery (e.g., M114 155mm howitzer)	Information is primarily directed and gathered by human controller. Computer gathers raw information for human controller.	Human controller analyzes data with assistance from computer systems.	Human ranks targets and determines where and when to engage.	Computer executes on human command.

III. THE INTERNATIONAL LAW AND AUTONOMY

It is well accepted that the law of armed conflict applies to AEWs; therefore, the law imposes requirements on parties to a conflict on their use,[76] including the requirement that the attack is discriminate,[77] proportional,[78] and complies with requirements for precautions in attack.[79] However, the application of these concepts to autonomous weapons raises some questions. For example, can an autonomous system distinguish civilian from combatant? A civilian object from a military object? Can it do so in all environments? Or, in the context of proportionality, how does an autonomous system calculate and weigh anticipated military advantage and expected collateral damage?

Some commentators contend these issues are insoluble, arguing that AEWs cannot comply with the requirements because doing so requires

---

76. INTERNATIONAL HUMANITARIAN LAW AND THE CHALLENGES OF CONTEMPORARY ARMED CONFLICTS, INT'L COMM. OF THE RED CROSS, REP. NO. 31IC/11/5.1.2, at 36 (2011) ("There can be no doubt that IHL applies to new weaponry and to the employment in warfare of new technological developments . . . ."); Schmitt & Thurnher, *supra* note 32, at 243 ("There is universal consensus that the law of armed conflict applies to autonomous weapon systems.").

77. Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, art. 48, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter AP I].

78. *Id.* art. 51(5)(b); *id.* art. 57(2)(a)(iii); *see also* OFF. OF GEN. COUNS., U.S. DEP'T OF DEF., LAW OF WAR MANUAL ¶ 5.12 (2015).

79. AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 57.

inherently human judgments,[80] the technology is not sufficiently sophisticated,[81] or the current legal regime is insufficient to address autonomous weapons.[82] This Article concludes—as have others[83]—that the law of armed conflict is adequate to regulate autonomous weapons systems that currently exist or will likely exist in the near future.

That said, it is perilous to think of AEWs as a homogenous category that is either compliant with the law or not. Certain components of a given system may be highly autonomous, while other elements may not have autonomous features at all. Thus, asking whether autonomous weapons can comply with the law is fundamentally the wrong question, since an answer requires an impossible *ex-ante* judgment. To be clear, however, one could examine a particular autonomous weapons system and make that determination.

While this Section briefly discusses the impact of autonomy on the triggers for non-international armed conflicts (NIACs) and international armed conflicts (IACs), it does not address the *jus ad bellum* and law of neutrality issues that arise with the use of force by one State into the territory of another. Further, the Article draws no distinction between IAC and NIAC as the targeting provisions of the Additional Protocol I (API) to the Geneva Conventions are analogous and accepted as reflecting the customary international law applicable in both types of conflict.[84]

---

80. Peter Asaro, *On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making*, 94 INT'L REV. RED CROSS 687, 700–03 (2012).

81. *See* Noel Sharkey, *Grounds for Discrimination: Autonomous Robot Weapons*, RUSI DEFENCE SYSTEMS, Oct. 2008, at 86, 87; DOCHERTY, *supra* note 1, at 23; *Stop Killer Robots While We Still Can*, PAX (Feb. 26, 1014), https://www.paxforpeace.nl/stay-informed/news/stop-killer-robots-while-we-still-can.

82. Hin-Yan Liu, *Categorization and Legality of Autonomous and Remote Weapons Systems*, 94 INT'L REV. RED CROSS 627, 629 (2012).

83. *See* Kenneth Anderson et al., *Adapting the Law of Armed Conflict to Autonomous Weapon Systems*, 90 INT'L L. STUD. 386, 387 (2014); Markus Wagner, *The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapons Systems*, 47 VAND. J. TRANSNAT'L L. 1371, 1386 (2014); Peter Margulies, *Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts*, *in* RESEARCH HANDBOOK ON REMOTE WARFARE 1 (Jens David Ohlin ed., 2016).

84. *See* Michael N. Schmitt & Eric W. Widmar, *"On Target": Precision and Balance in the Contemporary Law of Targeting*, 7 J. NAT'L SEC. L. & POL'Y 379, 381 (The U.S., Israel, "and other non-party states consider nearly all the treaty's targeting provisions as reflective of customary international law."); Michael J. Matheson, Deputy Legal Adviser, Dep't of State, Remarks on the United States Position on the Relation of Customary International Law to the 1977 Protocols Additional to the 1949 Geneva Conventions at the Sixth Annual American Red Cross-Washington College of Law Conference on International Humanitarian Law (Jan. 22,

A. Circumstances Not Implicating LOAC

Arguments concerning autonomy and the law of armed conflict (LOAC) tend to focus on a limited subset of potential engagements, specifically those engagements involving fully autonomous systems using deadly force against a person in cluttered and complex battlespaces.[85] Unquestionably, conflicts will be fought in these environments and will raise significant issues regarding the employment of autonomous weapons systems. As discussed below, conflicts will also, however, be fought in ways and in environments that will not raise issues in the law of armed conflict because of the way autonomy is used and the context or manner in which it is used.

For example, systems may employ autonomy in *technologies* that do not implicate the law of armed conflict. The U.S. RQ-4 "Global Hawk" is an unarmed/unmanned surveillance aircraft that can autonomously refuel[86] and navigate[87]—functions that do not implicate the law of armed conflict.[88] Looking to the proposed framework definition of autonomy presented above, the concerns raised by opponents of AEWs are only potentially present in the "decide" and "act" aspects of machine decision making, and then only in the higher order systems (e.g., Tier IV and III).[89]

It is also possible to employ weapons that are fully autonomous but are employed in an *environment* that does not raise issues with some aspects of the law of armed conflict. For example, an autonomous weapon could be used in a very limited area (e.g., a remote desert battlefield) or in a very limited fashion (e.g., a weapon activates for a fraction of a second at a time when no civilians are present).[90] In such circumstances, aspects of the law of armed conflict, such as proportionality, may not be implicated.

1987), *in* 2 AM. U. J. INT'L L. & POL'Y 419, 426–27 (1987) (describing the Protocol I sections that deserve recognition as customary international law); *see generally* 1 JEAN-MARIE HENCKAERTS & LOUISE DOSWALD-BECK, INT'L COMM. OF THE RED CROSS, CUSTOMARY INTERNATIONAL HUMANITARIAN LAW (2005) (discussing certain elements of Protocol I that are accepted as customary international law).

85. *See e.g.*, DOCHERTY, *supra* note 1, at 30; Heyns, *supra* note 31, at 13.

86. *Autonomous High-Altitude Refueling*, DEF. ADVANCED RES. PROJECTS AGENCY, http://www.darpa.mil/about-us/timeline/autonomous-highaltitude-refueling (last visited Nov. 10, 2017).

87. *See* Monroe Conner, *NASA Armstrong Fact Sheet: Global Hawk High-Altitude Long-Endurance Science Aircraft*, NASA, https://www.nasa.gov/centers/armstrong/news/FactSheets/FS-098-DFRC.html (last updated Aug. 4, 2017).

88. Though autonomous navigation would implicate other areas of the law such as sovereignty and the law of neutrality.

89. *See infra* Part III.

90. Schmitt & Thurnher, *supra* note 32, at 246 (noting that some systems might be lawful in some circumstances but not others).

Finally, an autonomous weapon could be employed that is highly autonomous but then used in a *manner* that does not implicate the law of armed conflict. For example, consider a hypothetical weapon directed to attack a particular building. Before activation, the system is programmed to *de*activate when any civilians are present. This process would be akin to an artillery round that can turn itself off when civilians are detected. Here, the autonomous feature of the system (e.g., the decision to deactivate) is being used in a manner that does not generate issues under the law of armed conflict, and in fact, can only cause enhanced compliance.

B. Armed Conflict Triggers and Autonomy

The existence of an IAC is determined based on the criteria established in Common Article 2, which applies the Conventions "to all cases of declared war or of any other armed conflict which may arise between two or more of the High Contracting Parties . . . ."[91] The official ICRC commentary notes that an armed conflict extends to "any difference arising between two States and leading to the intervention of members of the armed forces . . . . It makes no difference how long the conflict lasts, or how much slaughter takes place."[92] This view has found widespread[93] but not universal acceptance.[94]

Autonomous technologies are unlikely to affect the triggers for IACs. In the hands of a State, an autonomous weapon system, like a non-autonomous weapon system, is merely an instrument for the exercise of State authority. Even where the system is acting with extreme levels of autonomy, it is—at

---

91. Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, art. 2, Aug. 12, 1949, 75 U.N.T.S. 31 [hereinafter GC I]; Geneva Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea, art. 2, Aug. 12, 1949, 75 U.N.T.S. 85 [hereinafter GC II]; Geneva Convention (III) Relative to the Treatment of Prisoners of War, art. 2, Aug. 12, 1949, 75 U.N.T.S. 135 [hereinafter GC III]; Geneva Convention (IV) Relative to the Protection of Civilian Persons in Time of War, art. 2, Aug. 12, 1949, 75 U.N.T.S. 287 [hereinafter GC IV].

92. INT'L COMM. OF THE RED CROSS, COMMENTARY: GENEVA CONVENTION FOR THE AMELIORATION OF THE CONDITION OF THE WOUNDED AND SICK IN ARMED FORCES IN THE FIELD 32 (Jean S. Pictet ed., 1952).

93. *See, e.g.*, Prosecutor v. Tadić, Case No. IT-94-1-I, Decision on Defence Motion for Interlocutory Appeal on Jurisdiction, ¶ 70 (Int'l Crim. Trib. for the Former Yugoslavia Oct. 2, 1995) (endorsing the view that "an armed conflict exists whenever there is a resort to armed force between States").

94. *See, e.g.*, INT'L LAW ASS'N, FINAL REPORT ON THE MEANING OF ARMED CONFLICT IN INTERNATIONAL LAW 2 (2010) ("The Committee, however, found little evidence to support the view that the Conventions apply in the absence of fighting of some intensity.").

most—an organ or agent of the State whose actions are attributable to the State.[95] Actions will be attributable even where the system is acting in an entirely unpredictable manner and beyond the scope of the initial deployment.[96]

The effect of autonomy on NIACs, however, is somewhat different. Like IACs, there is no internationally accepted definition of an NIAC, though Additional Protocol II describes what does *not* constitute an NIAC; specifically, "internal disturbances and tensions, such as riots, isolated and sporadic acts of violence, and other acts of a similar nature."[97] The International Criminal Tribunal for the Former Yugoslavia's (ICTY) extensive treatment of the subject in the *Prosecutor v. Tadić* case is widely considered the definitive exposition on the issue.[98] In *Tadić*, the Appeals Chamber found that an NIAC exists when there is a situation of "protracted armed violence between governmental authorities and organized armed groups or between such groups within a State."[99]

The first prong of the *Tadić* test requires sufficiently intense violence.[100] While *Tadić* speaks to "protracted armed violence," in practice the intensity of the violence rather than the length of the violence has come to be regarded as the focus of the test.[101] In another case, the ICTY provided a list of "indicative factors" which can be used to evaluate the intensity criteria, including the following:

> the number, duration and intensity of individual confrontations; the type of weapons and other military equipment used; the number and calibre of munitions fired; the number of persons and type of forces partaking in the fighting; the number of casualties; the extent of

---

95. Int'l Law Comm'n on the Work of Its Fifty-Third Session, Responsibility of States for Internationally Wrongful Acts, U.N. Doc. A/56/83, at pt. 1, ch. II, art. 4 (2001) [hereinafter Articles].

96. *Id.* at pt. I, ch. II, art. 7.

97. Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of Non-International Armed Conflicts, art. 1(2), June 8, 1977, 1125 U.N.T.S. 609 [hereinafter AP II]; *see also* Rome Statute of the International Criminal Court, art. 8(2)(f), July 17, 1998, 2187 U.N.T.S. 90 [hereinafter Rome Statute] (NIACs exclude "situations of internal disturbance and tensions, such as riots, isolated and sporadic acts of violence or other acts of a similar nature.").

98. *Tadić*, Case No. IT-94-1-I, ¶ 70.

99. *Id.*

100. *Id.*

101. Prosecutor v. Haradinaj, Case No. IT-04-84-T, Judgment, ¶ 49 (Int'l Crim. Trib. for the Former Yugoslavia Apr. 3, 2008) ("The criterion of protracted armed violence has therefore been interpreted in practice, including by the *Tadić* Trial Chamber itself, as referring more to the intensity of the armed violence than to its duration.").

> material destruction; and the number of civilians fleeing combat zones.[102]

The organizational element of ICTY jurisprudence requires that the organized armed group that is a party to the conflict have a level of organization and a command structure capable of sustaining military operations.[103] The court relies on various criteria to determine whether a group is sufficiently organized.[104] In examining whether the Kosovo Liberation Army (KLA) was sufficiently organized, the ICTY utilized the following criteria:

> level of organization of the KLA: the existence of KLA headquarters and command structure; the existence of KLA disciplinary rules and mechanisms; territorial control exerted by the KLA; the ability of the KLA to gain access to weapons and other military equipment; to recruit members; to provide them with military training; to carry out military operations and use tactics and strategy; and to speak with one voice.[105]

Sasha Radin and Jason Coats have expressly addressed the question of how autonomy might impact the trigger for an NIAC.[106] In an analysis with which this author agrees, Radin and Coats argue that autonomous weapons have a direct bearing on the organization criteria in that they enable an armed group to inflict violence and control territory with a much smaller organizational structure than would be required using conventional weapons.[107] A single combatant could control dozens of autonomous weapons systems, which could replace hundreds or thousands of combatants. Further, autonomy simplifies an armed group's logistics and streamlines command and control,[108] allowing much smaller, more loosely organized

---

102. *Id.*

103. INT'L COMM. OF THE RED CROSS, HOW IS THE TERM "ARMED CONFLICT" DEFINED IN INTERNATIONAL HUMANITARIAN LAW? 3 (2008); *see also* Dapo Akande, *Classification of Armed Conflicts: Relevant Legal Concepts*, *in* INTERNATIONAL LAW AND THE CLASSIFICATION OF CONFLICTS 51 (Elizabeth Wilmshurst ed., 2012) ("In order to be a party to an armed conflict a non-state group must have a certain level of organization with a command structure.").

104. Akande, *supra* note 103, at 51–52.

105. *Haradinaj*, Case No. IT-04-84-T, ¶ 64.

106. Sasha Radin & Jason Coats, *Autonomous Weapons Systems and the Threshold of Non-International Armed Conflict*, 30 TEMP. INT'L & COMP. L.J. 133, 134 (2016).

107. *Id.* at 143–47.

108. *Id.* at 144–45.

groups to inflict levels of violence that previously would necessitate large, highly organized groups.[109] The court in *Tadić* likely did not consider the circumstance where a minuscule, potentially highly decentralized, non-state actor would have the ability to engage in violence on the scale of an armed conflict.

Regarding the intensity prong of the *Tadić* test, Radin and Coats note that "[a]s systems replace humans, the forms that armed clashes take could be altered, resulting in an increase in destruction and possible decrease in loss of life."[110] Thus, autonomous systems could be highly sophisticated and programmed to avoid civilian casualties. The resulting engagements could be extremely violent regarding damage to objects but result in no loss of life. Would such an engagement trigger the *Tadić* intensity criteria? While no clear answer exists, the question prompts consideration of the value ascribed to the various indicia of intensity. Where human lives are valued more than the destruction of property, a conflict with substantial loss of property could occur without triggering an armed conflict. Conversely, "if destruction were to be accorded the same weight as human life, [then this] could lead to an extreme situation where, for example, intense clashes between AWS could satisfy the intensity requirement, and thus trigger an armed conflict (if the organization criterion was also satisfied)" without any loss of human life.[111]

C. Distinction

Distinction requires a person conducting the attack to distinguish between lawful targets (combatants, civilians taking direct part in the hostilities, and military objectives) from unlawful targets (civilians, those *hors de combat*, civilian objects, and other protected persons and objects).[112] The International Court of Justice has described the principle of distinction as one of the two "cardinal principles" that constitute "the fabric of humanitarian law."[113] Together these principles reflect "intransgressible principles of international customary law."[114] Article 48 expresses this principle in Additional Protocol I:

---

109. *Id.* at 141.

110. *Id.* at 148.

111. *Id.* at 149.

112. AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 57(2)(a)(1) (The article requires the attacker "do everything feasible to verify that the objectives to be attacked are neither civilian nor civilian objects and are not subject to special protection but are military objectives.").

113. Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. Rep. 226, ¶ 78 (July 8, 1996).

114. *Id.* ¶ 79.

> In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives.[115]

Article 48 reflects customary international law, as does the relevant provision of Additional Protocol II[116] that establishes the same restrictions in NIACs.[117]

There are aspects of distinction that may prove particularly challenging for an autonomous weapons system. While distinguishing a uniformed combatant from a civilian is relatively easy, distinguishing a civilian from a civilian directly participating in hostilities, in many circumstances, may be very difficult. Determining when the direct participation begins and ends is more difficult still. Similarly, determining membership in an organized armed group could prove challenging for an autonomous weapons system.

1. Distinction with Respect to Persons

The civilian population is comprised of all persons[118] who are not members of the armed forces.[119] Parties to the conflict are enjoined from attacking[120] civilians "unless and for such time as they take a direct part in hostilities."[121] Combatants include members of the armed forces of a State;[122] members of a militia or volunteer corps that belong to a State;[123] and members of a *levée en masse*.[124] Additional Protocol I further extends

---

115. AP I, *supra* note 77, pt. IV, § 1, ch. I, art. 48.

116. AP II, *supra* note 97, art. 13.3.

117. HENCKAERTS & DOSWALD-BECK, *supra* note 84, at 3, 25.

118. AP I, *supra* note 77, pt. IV, § 1, ch. II, art. 50(2).

119. GC IV, *supra* note 91, art. 4; AP I, *supra* note 77, pt. IV, § 1, ch. II, art. 50(1); HENCKAERTS & DOSWALD-BECK, *supra* note 84, at 17.

120. AP I, *supra* note 77, pt. IV, § 1, ch. I, art. 48; MICHAEL BOTHE ET AL., NEW RULES FOR VICTIMS OF ARMED CONFLICTS 288 (1982).

121. AP I, *supra* note 77, pt. IV, § 1, ch. II, art. 51(3).

122. *Id.*, art. 43(2); OFF. OF GEN. COUNS., *supra* note 78, ¶ 4.3.3.

123. GC III, *supra* note 91, art. 4(2); OFF. OF GEN. COUNS., *supra* note 78, ¶ 4.3; *see also* INT'L COMM. OF THE RED CROSS, INTERPRETIVE GUIDANCE ON THE NOTION OF DIRECT PARTICIPATION UNDER INTERNATIONAL HUMANITARIAN LAW 22 (2009) ("[A]ll armed actors showing a sufficient degree of military organization and belonging to a party to the conflict must be regarded as part of the armed forces of that party.") [hereinafter INTERPRETIVE GUIDANCE].

124. GC III, *supra* note 91, art. 4(2); OFF. OF GEN. COUNS., *supra* note 78, ¶ 4.3.

the definition of combatants for States party to include members of organized armed groups that fulfill the criteria outlined in Additional Protocol I, Article 43.[125] Combatants may be targeted based solely on their status.[126] In both IACs and NIACs, members of organized armed groups are regarded as distinct from civilians for targeting purposes.[127] Which members of an organized armed group can be targeted remains a matter of some debate.[128]

As noted above, issues relating to distinction and autonomy only arise in a very specific subset of engagements—attacks involving autonomous technologies selecting and engaging targets where civilians or civilian objects, or those that are *hors de combat*, are potentially present. Broadly, autonomy implicates two issues related to the principle of distinction. The first arises from the length of time the system is deployed, whereas the second derives from the inherent technological sophistication of autonomous systems.

Autonomy allows systems to potentially be deployed for extended periods of time.[129] The U.S. "Global Hawk" unmanned surveillance aircraft, for example, can autonomously operate for more than thirty hours.[130] The U.S. Army is currently developing a system, the Persistent Aerial Reconnaissance and Communications System, which can stay aloft for an

---

125. AP I, *supra* note 77, pt. III, § 1, ch. II, art. 43(1) ("The armed forces of a Party to a conflict consist of all organized armed forces, groups and units which are under a command responsible to that Party for the conduct of its subordinates, even if that Party is represented by a government or an authority not recognized by an adverse Party.").

126. *See, e.g.*, AP I, *supra* note 77, pt. IV, § 1, ch. I, art. 48; OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.7.2.

127. Schmitt & Widmar, *supra* note 84, at 385 ("Consensus has emerged in the past decade as to another group of individuals who do not qualify as civilians *for the purpose of targeting*—members of 'organized armed groups.'"); *see also* OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.8.3 ("Like members of an enemy State's armed forces, individuals who are formally or functionally part of a non-State armed group that is engaged in hostilities may be made the object of attack because they likewise share in their group's hostile intent.").

128. *See* Michael N. Schmitt, *Deconstructing Direct Participation in Hostilities: The Constitutive Elements*, 42 N.Y.U. J. INT'L L. & POL. 697, 704 (2010) (arguing that all members of an organized armed group may be targeted regardless of their function in the group). *But see* INTERPRETIVE GUIDANCE, *supra* note 123, at 71 (noting that only members of the group who serve a "continuous combat function" can be targeted at any time).

129. In 2010, Boeing reported that an experimental unmanned aerial vehicle dubbed the Phantom Eye can loiter for up to ten days. Michael Barkoviak, *Boeing UAV Able to Loiter Above Target for 10 Days*, DAILY TECH (Aug. 3, 2010), http://www.dailytech.com/Boeing+UAV+Able+to+Loiter+Above+Target+For+10+Days/article19242.htm.

130. *Autonomous Global Hawk Unmanned Aircraft System Suppresses 200,000 Flight Hours*, NORTHROP GRUMMAN (July 26, 2016), http://news.northropgrumman.com/news/releases/autonomous-global-hawk-unmanned-aircraft-system-surpasses-200-000-flight-hours.

indefinite amount of time if it maintains a power supply connection with a ground station.[131] The longer the system operates, the greater the chance the environment will change. Changes can occur in the physical environment (e.g., atmospheric conditions, altitude, time of day, weather, etc.) and in the operational environment (e.g., the human element of the battlefield including the persons and human-made structures).

The interaction between environment and machine is critical, as the ability of a system to distinguish is a function of the sophistication of the system and the complexity of the environment. An increasingly complex environment requires an increasingly sophisticated system.[132]

Autonomy raises various issues with the rule of distinction in the context of attacks on persons.[133] It is not inconceivable that a system could be programmed to identify members of an armed force.[134] They are, after all, required to take measures to ensure that they are readily distinguishable from civilians.[135] An AEW could be easily programmed to identify a particular uniform or insignia. Identifying a civilian who is directly participating in hostilities may be more difficult since the civilian will likely not exhibit any outward indication of the civilian's status, and may, in fact, be attempting to hide his or her status. For these reasons, identifying a civilian directly participating is challenging for both humans and machines.

---

131. *The Future of High-Powered Commercial Drones*, CYPHY, http://cyphyworks.com/parc/ (last visited Nov. 11, 2017).

132. *See, e.g.*, U.S. NAVAL METEOROLOGY AND OCEANOGRAPHY PROF'L DEV. DETACHMENT ATLANTIC, ATMOSPHERIC EFFECTS ON EO SENSORS AND SYSTEMS (2005) (providing a detailed discussion of atmospheric effects on various types of electro-optical sensors); *see also* RICHARD C. SHIRKEY & BARBARA J. SAUTER, ARMY RESEARCH LAB. REP., WEATHER EFFECTS ON TARGET ACQUISITION PART I: SENSOR PERFORMANCE MODEL INFRARED ALGORITHMS 1 (2001) ("Detection and recognition ranges depend upon the target and background characteristics, atmospheric propagation, and senor performance."); *The UK Approach to Unmanned Aircraft Systems*, UK MINISTRY OF DEFENSE, DEVELOPMENT, CONCEPTS AND DOCTRINE CENTRE, JOINT DOCTRINE NOTE 2/11, 5–4 (Mar. 30, 2011) ("[F]or operating environments with easily distinguished targets in low clutter environments, a degree of autonomous operation is probably achievable now and data from programmes such as *Brimstone* and ALARM, for example, would have direct read-across.").

133. AP I, *supra* note 77, pt. IV, § 1, ch. II, art. 52(2).

134. *See* Matthew Rosenberg & John Markoff, *The Pentagon's "Terminator Conundrum": Robots That Could Kill on Their Own*, N.Y. TIMES (Oct. 25, 2016), https://www.nytimes.com/2016/10/26/us/pentagon-artificial-intelligence-terminator.html (describing recent U.S. military tests of an autonomous drone which "showed a spooky ability to discern soldier from civilian, and to fluidly shift course and move in on objects it could not quickly identify. Armed with a variation of human and facial recognition software used by American intelligence agencies, the drone adroitly tracked moving cars and picked out enemies hiding along walls").

135. Annex to the Convention No. IV Regulations Respecting the Laws and Customs of War on Land, Regulations, § I, ch. I, art. I, ¶ 2, Oct. 18, 1907, 36 Stat. 2227, T.S. 539.

In an attempt to provide more concrete guidance on civilians directly participating in hostilities, the ICRC has suggested a three-part cumulative test of "constitutive elements." This formulation has found widespread support[136] and holds that:

> Acts amounting to direct participation in hostilities must meet three cumulative requirements: (1) a threshold regarding the harm likely to result from the act, (2) a relationship of direct causation between the act and the expected harm, and (3) a belligerent nexus between the act and the hostilities conducted between the parties to an armed conflict.[137]

Protection against attack is lost for the period of time a civilian takes direct part in hostilities.[138]

Determining the contours of "for such time" has proven difficult. Most agree that the direct participation in hostilities extends for some point in time before and after the participation in hostilities, but identifying the moment direct participation begins and ends has proved elusive.[139] In the ICRC's interpretive guidance, the individual's direct participation extends to preparatory measures and deployment to and from the location of the act.[140] An alternative view holds that "for such time" should be extended "as far before and after a hostile action as a causal connection existed."[141] The United States' position is that persons taking direct part in hostilities are a legitimate target until "they have permanently ceased their participation" in hostilities.[142]

AEWs will likely face significant difficulty in identifying a person immediately before the act constituting a direct participation in hostilities. To be clear, this does not render "autonomous weapons" as a class of

---

136. Schmitt & Widmar, *supra* note 84, at 387 ("The ICRC's *Interpretive Guidance*, in an approach that has been widely accepted, suggests that acts of direct participation consist of three cumulative constitutive elements.").

137. INTERPRETIVE GUIDANCE, *supra* note 123, at 46.

138. AP I, *supra* note 77, pt. IV, § 1, ch. II, art. 51(3).

139. Michael N. Schmitt, *The Interpretive Guidance on the Notion of Direct Participation in Hostilities: A Critical Analysis*, 1 HARV. NAT'L SEC. J. 5, 36 (2010) [hereinafter Schmitt, *Interpretive*].

140. INTERPRETIVE GUIDANCE, *supra* note 123, at 65.

141. Schmitt, *Interpretive*, *supra* note 139, at 36–37 (citing Yoram Dinstein, *Distinction and the Loss of Civilian Protection in Armed Conflict*, *in* 84 INT'L L. STUD. 183, 189–90 (Michael D. Carsten ed., 2008)); *see also* Kenneth H. Watkin, *Controlling the Use of Force: A Role for Human Rights Norms in Contemporary Armed Conflict*, 98 AM. J. INT'L L. 1, 17 (2004).

142. OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.9.4.

weapons indiscriminate; rather, it limits the type and employment of weapons systems. Take for example a civilian who is emplacing an improvised explosive device (IED) along a road. An AEW could plausibly be deployed with radar that can detect a person digging along a road[143] and a spectrograph that can detect the chemical signature of explosives.[144] Assume the system has been programmed to engage a target when these detection criteria are met and no other persons are present.

Under the ICRC interpretative guidance,[145] the individual can be engaged as they deploy and redeploy from the act. Thus, an autonomous weapon could continue to track the person until no civilians are present and then engage. The ICRC guidance allows that such a person remains a direct participant in hostilities until they have "physically separated from the operation."[146] This determination "depends on a multitude of situational factors, which cannot be comprehensively described in abstract terms" and "must be made with utmost care and based on a reasonable evaluation of the prevailing circumstances."[147] This guidance is difficult to operationalize for commanders and soldiers on the ground, and even more so for programmers of an autonomous weapons system. The problem, however, is not insurmountable.

There are four control mechanisms that commanders could employ to ensure AEWs comply with the law of armed conflict. These mechanisms—briefly described as *Sophistication*, *Restriction*, *Updates*, and *Human Involvement*—address concerns about the use of force in unclear circumstances. In the context of direct participation in hostilities, the mechanisms would be applied as such:

- *Sophistication*: Deploying an AEW that is of such advanced technological sophistication that it can identify direct participants with reasonable certainty. In the above example, such a system could be

143. *See* Sevgi Zubeyde Gurbuz et al., *Comparison of Radar-Based Human Detection Techniques*, GA. TECH RESEARCH INST. (June 2010) (describing radar-based technology that can remotely identify humans), http://www.dtic.mil/dtic/tr/fulltext/u2/a523514.pdf.

144. *See* Ida Johnson, FOI, Swedish Def. Research Agency, Presentation on Stand-off Raman Spectroscopy for the Detection of Explosives, http://www.vtt.fi/files/newsletter/os/042013/Ida_Johansson.pdf (demonstrating the feasibility of a stand-off spectrometer that can detect the chemical signature of explosives); *see also* Ruth M. Doherty, U.S. Dep't of Homeland Security, Presentation on Science & Technology to Counter Improvised Explosive Devices (Apr. 28, 2010), http://www.dtic.mil/ndia/2010GlobalExplosive/Doherty.pdf (describing a U.S. program to develop technology to remotely detect explosives on a person).

145. INTERPRETIVE GUIDANCE, *supra* note 123, at 65.

146. *Id.* at 67.

147. *Id.* at 68.

programmed to determine when—using the ICRC language—a person is physically separated from the operation.

- *Restriction*: Limiting the AEW's geographical boundaries of operation, duration of the deployment, or target set/type such that the issue of direct participation will not arise. This is most easily accomplished by deploying the system for a discrete task or for a very short period.
- *Updates*: Updating the AEW with human-identified direct participants.
- *Human Involvement*: Retaining operator control or oversight of the AEW during deployment. This would include humans on, in, or near the loop.

Strictly speaking, one could argue that the latter two control mechanisms would render the system something other than autonomous, particularly where the control is significant. Most AEWs would likely use a combination of these control mechanisms. For example, a system might be deployed to a small operational area (*Restriction*) and be further programmed only to engage targets provided by a human operator (*Human Involvement*) based upon current intelligence (*Updates*).

2. *Distinction Between Objects*

Another aspect of distinction that autonomy potentially disrupts is the targeting of objects. Article 52(1) of Additional Protocol I prohibits targeting civilian objects, which are defined as all objects which "are not military objectives."[148] The article provides a two-part test for military objectives. First, they must "by their nature, location, purpose or use make an effective contribution to military action" and secondly, their "total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."[149] Effective contribution is a broad concept that "does not require a direct connection with combat operation[s]."[150]

---

148. AP I, *supra* note 77, pt. IV, § 1, ch. II, art. 52(1).

149. *Id.* at pt. IV, § 1, ch. II, art. 52(2).

150. OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.7.6.2 ("The object must make or be intended to make an effective contribution to military action; however, this contribution need not be 'direct' or 'proximate.'") (citing BOTHE ET AL., *supra* note 120, at 324) ("[A] civilian object may become a military objective and thereby lose its immunity from deliberate attack through use which is only indirectly related to combat action, but which nevertheless provides an effective contribution to the military phase of a Party's overall war effort.").

An object is a military objective by *nature* when its "intrinsic character" is military.[151] The ICRC commentary provides a non-exhaustive list including "all objects directly used by the armed forces: weapons, equipment, transports, fortifications, depots, buildings occupied by armed forces, staff headquarters, communications centres etc."[152] Autonomy likely does not have a significant impact on military objects by their nature since these objects are usually readily identifiable, can be programmed into an AEW, and are not likely to lose their status.

An object is a military objective by *location* when the location of the object provides an effective military contribution regardless of the use of the object.[153] A strategic bridge that affords enemy forces freedom of movement would be a valid military object by location regardless of how the bridge is used at the time of the attack. Again, autonomy has little impact on objects that are military by their location. As with intrinsically military objects, objects that are military by location can be identified and programmed into an autonomous system. Depending on the scope of autonomy, the system could then select targets from among the potential targets. Changes to the operational environment do not change the status of such objects—changes may, however, affect the military advantage calculation discussed below.

The military *purpose* of an object speaks to its future use.[154] The DoD Law of War Manual provides runways at civilian airports[155] as an example, and Yoram Dinstein uses civilian cruise ships that could be used as troop transports.[156] Designating an object a military objective by its purpose requires an understanding of the adversary's intent based on a knowledge of the enemy's tactics, techniques, and procedures (TTPs) and current intelligence showing enemy activity. This determination is more than supposition, and must be supported with reasonable certainty that the object will be converted to military use. As Yoram Dinstein cautions, often this evaluation is "crisply clear" while other situations "are not so easy to decipher."[157]

---

151. YORAM DINSTEIN, THE CONDUCT OF HOSTILITIES UNDER THE LAW OF INTERNATIONAL ARMED CONFLICT 96 (2d ed., Cambridge Univ. Press 2010).

152. Yves Sandoz et al., Commentary, *Additional Protocols of 8 June 1997 to the Geneva Conventions of 12 August 1949*, ¶ 2020 (1987).

153. *Id.* ¶ 2021 ("Clearly, there are objects which by their nature have no military function but which, by virtue of their location, make an effective contribution to military action.").

154. *Id.* ¶ 2022 ("The criterion of 'purpose' is concerned with the intended future use of an object, while that of 'use' is concerned with its present function.").

155. OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.7.6.1.

156. DINSTEIN, *supra* note 151, at 99–100.

157. *Id.*

AEWs may have difficulty making determinations as to when an object qualifies as a military objective by its purpose. Such a determination would require a system that can sense enemy activity and make targeting determinations about the enemy's future actions. Take a hypothetical where reliable intelligence indicates enemy forces are redirecting all aluminum to military purposes. Conceivably, an AEW could be sufficiently sophisticated to recognize all aluminum stock is being redirected to military installations. This level of sophistication is, however, unlikely given current technology. More likely, the system would have to be deployed with the information or provided an update regarding the status of the stock. Absent such updates, it is hard to conceive how an AEW could make sophisticated determinations regarding the future use of something.

An object is a valid military object by its *use* when the current function of a previously civilian object is now military in nature.[158] The object remains a valid military objective for such time as the object is used for military purposes.[159] This category of military object is analog to the principle that a civilian remains a civilian until and for such time that they directly participate in hostilities.[160] As with the targeting of civilians directly participating in hostilities, measures would have to be taken to ensure objects are attacked only during the time that they are military objectives.

Determining the point in time when a civilian object becomes a military object, and when it regains its civilian status, may prove difficult depending on the circumstances. It is relatively clear, for instance, when a school is being used as a fighting position. Conversely, it is relatively *unclear* when a school is being used as a military headquarters. Whether an autonomous system can determine when an object is no longer being used for military purposes is a technical question. In some circumstances, this determination might be simple and could be made with current technology. More complex scenarios would necessitate more sophisticated systems or programming the system such that it is not permitted to make these determinations.

3. *Doubt as to Status of the Target*

Article 50(1) of Additional Protocol I holds that where there is "doubt whether a person is a civilian, that person shall be considered to be a

158. OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.7.6.1. ("'Use' refers to the object's present function.").

159. HENCKAERTS & DOSWALD-BECK, *supra* note 84, at 35; *see also* OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.7.6.1.

160. DINSTEIN, *supra* note 151, at 98.

civilian."[161] Article 52(3) provides an analog provision in situations where there is doubt as to whether an object that is "normally dedicated to civilian purposes . . . is being used to make an effective contribution to military action . . . ."[162] It is not clear whether these provisions reflect customary international law.[163] Though the United States rejects the customary nature of these provisions,[164] the DoD Law of War Manual requires decisions to be made in good faith based on something more than "merely hypothetical or speculative considerations."[165]

Can machines be programmed to account for doubt?[166] In a lengthy report commissioned by the DoD, noted roboticist Ronald Arkin argues that systems could be programmed with an "ethical governor" bounding the actions of a system within predetermined limits.[167] Such bounds could be programmed to consider uncertainty, which can be expressed "in a variety of ways: discrete (e.g., binary: absent or present; categorical: absent, weak, medium, strong) or it can be real valued and continuous."[168] How then does one quantify doubt on the battlefield? Clearly, the easiest circumstance is where the machine is programmed to consider doubt as a binary function: either the system is 100% certain and attacks, or it is less than 100% certain and refrains from attacking. The law, however, requires only reasonable, not absolute, certainty.[169] Quantifying reasonable certainty is inherently more subjective; mathematically, what constitutes "reasonable certainty" is unclear.

Compliance with distinction will depend on the complexity of the environment in which the AEWs are operating. Systems displaying a large amount of autonomy are today employed in uncluttered environments (e.g., the open sea) against readily identified targets (e.g., an incoming missile). Difficulties arise when AEWs operate in dynamic circumstances or in situations that require contextual decisions. As others have correctly noted, "it is conceivable that the battlefield situation might be too cluttered for the system to accurately distinguish between military objectives and civilian

161. AP I, *supra* note 77, pt. IV, § 1, ch. II, art. 50(1).

162. *Id.* at pt. IV, § 1, ch. III, art. 52(3).

163. *See* Schmitt & Widmar, *supra* note 84, at 197 ("Under customary international law, no legal presumption of civilian status exists for persons or objects.").

164. OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.1.2.

165. *Id.* ¶ 5.5.3.2.

166. Schmitt & Thurnher, *supra* note 32, at 263.

167. RONALD C. ARKIN, GOVERNING LETHAL BEHAVIOR IN AUTONOMOUS ROBOTS 127 (CRC Press ed., 2009).

168. *Id.* at 59.

169. *See* OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.5.3.2.

objects or between combatants and the civilian population. In those cases, an autonomous weapons system would be unlawful to use."[170]

D. Proportionality

The object of an attack must be not only a legitimate object of attack, but the attack itself must comply with the rule of proportionality, which prohibits an "attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated."[171] Proportionality is widely considered to be a norm of customary international law in both IACs and NIACs.[172] Determining anticipated military advantage and the expected collateral damage and then weighing these unlike values against one another is both subjective and contextual and therefore can be difficult for the most skilled humans, let alone computer systems.[173]

Proportionality operates only to protect civilians and civilian objects.[174] Where there is no danger of collateral damage, the principle is not implicated.[175] Similarly, proportionality is read to apply to loss of life, injury, and damage to civilian objects and thus would not extend to prohibit attacks that harass or inconvenience.[176] It is equally important to consider

---

170. Jeffrey S. Thurnher, *Means and Methods of the Future: Autonomous Systems*, *in* TARGETING: THE CHALLENGES OF MODERN WARFARE 188 (Paul A.L. Ducheine, Michael N. Schmitt, & Frans P.B. Osinga eds., 2016); *see also* Marco Sassóli, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified*, 90 INT'L L. STUD. 308, 320 (2014) ("If it is technically not feasible to respect certain requirements of IHL with autonomous weapons, this is not a sufficient reason for abandoning those requirements. The use of autonomous weapons in such cases is simply unlawful.").

171. AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 57(2)(iii); *see also id.* at pt. IV, § 1, ch. IV, art. 51(5)(b) ("[A]n 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.").

172. HENCKAERTS & DOSWALD-BECK, *supra* note 84, Rule 14, at 46.

173. *See generally* Thurnher, *supra* note 170.

174. AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 51(5)(b).

175. OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.12.1 ("In conducting attacks, the proportionality rule only need be applied when civilians or civilian objects are at risk of harm from attacks on military objectives. It would not apply when civilians or civilian objects are not at risk."); DINSTEIN, *supra* note 151, at 129 ("Proportionality has nothing to do with injury to combatants or damage to military objectives.").

176. OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.12.2 (citing Yoram Dinstein, *Distinction and Loss of Civilian Protection in International Armed Conflicts*, 84 INT'L L. STUD. 183, 186 (2008)); *cf.* WILLIAM H. BOOTHBY, THE LAW OF TARGETING 370 (2012).

that the principle does not require an equitable balancing between military advantage and collateral damage; rather, an attack would violate this section only where the collateral damage is "excessive" relative to the "concrete and direct military advantage anticipated."[177]

In this context, "concrete and direct" indicates that the military advantage should be "substantial and relatively close" to the attack.[178] It need not be instantaneous,[179] though a "remote advantage to be gained at some unknown time in the future would not be a proper consideration to weigh against civilian losses."[180] The law does not require that the decision to attack be made with a perfect awareness of the direct or indirect and immediate or long-term consequences of the attack.[181] Rather,

> [i]n determining whether an attack was proportionate it is necessary to examine whether a reasonably well-informed person in the circumstances of the actual perpetrator, making reasonable use of the information available to him or her, could have expected excessive civilian casualties to result from the attack.[182]

What then constitutes a "reasonably well-informed person"? The U.S. military's joint targeting doctrine suggests considering "a mix of empirical data, probability, historical observations, and complex modeling" to estimate collateral damage.[183] Computer systems (autonomous or conventional) are especially suited for this analysis. Computers can process large volumes of data relevant to a proportionality analysis, including the nature and destructive effects of various weapons systems, the composition and durability of buildings near the target, the probability of civilian presence based on historical data, and more.[184] Thus, the collateral damage aspect is unlikely to cause significant issues for an AEW.

---

177. AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 57(2)(a)(iii).

178. Sandoz et al., *supra* note 152, ¶ 2209.

179. OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.12.5 ("There is no requirement that the military advantage be 'immediate.'").

180. BOTHE ET AL., *supra* note 120, at 365.

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

182. *Id.*

183. CHAIRMAN OF THE JOINT CHIEFS OF STAFF, U.S. DEP'T OF DEF., CHAIRMAN OF THE JOINT CHIEFS OF STAFF INSTRUCTION D-1 (2009), https://www.aclu.org/files/dronefoia/dod/drone_dod_3160_01.pdf.

184. *Id.* at D-2 ("[T]he CDM's science and art provide essential information that the commander uses in context with other factors and sound judgment . . . .").

Determining military advantage, however, is particularly challenging for AEWs because the evaluation is contextual and dynamic and does not lend itself to a mathematically precise calculation.[185] Military advantage must be calculated "in the circumstances ruling at the time."[186] This calculation requires an understanding of the military value of the target, including the contribution the item is making to the enemy in the circumstances of the time, and the benefit that will accrue from its neutralization or damage.

Autonomy (potentially) expands the time between activation and engagement of the target, allowing for extended loiter times. During this period, it is possible for the military advantage to change. For example, on a large battlefield, the destruction of a single unarmed bridging vehicle would provide little military advantage. If, however, that bridging vehicle is being used, or is about to be used, to facilitate the advance of hundreds of enemy vehicles, then the military advantage of its destruction increases over time.

Marco Sassóli has identified this area of the law as "the most serious [international humanitarian law] argument against the even theoretical possibility of deploying weapons that remain fully autonomous over considerable periods of time."[187] Sassóli suggests that autonomous weapons could not apply proportionality unless "constantly updated about military operations and plans."[188] This suggestion is insightful and deserves additional consideration.

If one employs a weapon that is "fully autonomous" for a "considerable" period of time, a mechanism for receiving updates on changes to the military advantage will likely be needed. This does not necessarily mean "constant" updates. Even on today's modern, fast-moving battlefield, the military advantage of some targets remains fairly static. The military advantage of an enemy's headquarters, for instance, will likely remain fairly constant throughout a battle. An AEW could likely be deployed against such an objective without significant proportionality concerns. Where the circumstances are dynamic, however, the challenges become more acute.

Consider a hypothetical battle involving three enemy tank battalions (A, B, and C) each with fifty-eight tanks. Before the engagement, a friendly forces commander will develop an operational framework that is a mechanism to frame their "concept of operations in time, space, purpose,

185. *Id.*
186. AP I, *supra* note 77, pt. IV, § 1, ch. III, art. 52(2).
187. Sassóli, *supra* note 170, at 332.
188. *Id.*

and resources."[189] One operational framework methodology is to designate primary and secondary efforts.[190] In this example, the commander designates the destruction of Battalion A as the primary effort, and the destruction of Battalions B and C as supporting efforts. Target tanks in enemy Battalion A would then have a greater military advantage than tanks in B and C. The AEWs participating in the battle (AEW A, AEW B, and AEW C) would be programmed with the value of tanks in each battalion. This example assumes individual tanks in each battalion are fungible and would carry the same military advantage value as another tank but for the commander's designation to the contrary.

Tanks will be destroyed as the battle progresses. Consequentially, the military advantage of a single tank would increase as the overall number of tanks decreases. How then, could AEW A, which is targeting Battalion A, know that the number of tanks in Battalion B have decreased, thus increasing the relative value of each tank in Battalion A? Here again, consider the control mechanisms of the autonomous weapons system.

- *Sophistication of the System*: In this example, the AEW systems could detect and react to changes in the military advantage. Thus, AEW A/B/C would be connected to each other or could observe attrition rates across the battlefield and calculate changes to the military advantage.
- *Restrictions on Use*: AEW A could be deployed for a short period of time or with restricted operational parameters (e.g., destroy five tanks and return). The military advantage is unlikely to change in these limited circumstances.
- *Availability of Updates*: AEW A would be updated by human operators regarding the status of the tanks in Battalion B and C.
- *Human Control*: Retaining operator control or oversight of the system during deployment. This would include humans on, in, or near the loop.

Proportionality is unquestionably a potential challenge for the lawful operation of autonomous weapons, particularly in dynamic circumstances or where the systems are deployed for long periods of time where the military advantage is likely to change. Lawful use of autonomous systems in such complex situations would require careful consideration of how systems

189. DEP'T OF THE ARMY, ARMY DOCTRINE PUBLICATION 3-0, UNIFIED LAND OPERATIONS ¶ 47, at 11 (2011).

190. *See id.* ¶¶ 58–60.

would account for changes in the military advantage. The control mechanisms set forth above provide a framework through which operators could ensure the lawful employment of autonomous weapons systems.

E. Precautions in Attack

Persons conducting attacks with autonomous weapons systems must take feasible precautions to reduce the risk of harm to civilians and other protected persons and objects.[191] For States party to Additional Protocol I, this requirement appears in Article 57, which requires "constant care" to be taken to "spare the civilian population, civilians and civilian objects."[192] The Additional Protocol requirement applies to "[t]hose who plan or decide upon an attack,"[193] including commanders who make the decision to employ a weapon system on the battlefield and those staff officers who plan the employment of the weapon system."[194]

A distinction should be made here between echelons of command. Armed conflict is typically conducted at three levels of war—strategic, operational, and tactical.[195] Strategic operations synchronize instruments of power to achieve overall objectives, while operational-level operations plan and implement strategies and campaigns designed to employ tactical forces to achieve strategic objectives.[196] Tactical operations concern the employment of forces on the battlefield.[197] The duty to take constant care attaches to tactical-level commanders and planners as they have the means to control the application of force, and the intelligence to determine if constant care is being taken during the course of an operation.[198] In the context of

---

191. *See* OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.11.

192. AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 57(a)(1).

193. Id. at pt. IV, § 1, ch. IV, art. 57(a)(2).

194. BOTHE ET AL., *supra* note 120, at 362 ("[Article 57] imposes three distinct duties on commanders who decide upon attacks and staff officers who plan an attack . . . ."); *see also* TALLINN MANUAL, *supra* note 34, at 166 ("[T]he duty of care requires commanders and all others involved in the operations to be continuously sensitive . . . ."); PROGRAM ON HUMANITARIAN POLICY AND CONFLICT RESEARCH, MANUAL ON INTERNATIONAL LAW APPLICABLE TO AIR AND MISSILE WARFARE Rule 34 (2009) ("Constant care must be taken by all those involved in planning, ordering and executing air or missile combat operations to spare the civilian population, civilians and civilian objects.").

195. CHAIRMAN OF THE JOINT CHIEFS OF STAFF, U.S. DEP'T OF DEF., JOINT PUBLICATION 3-0, JOINT OPERATIONS I-12 (2011) [hereinafter JOINT PUBLICATION 3-0].

196. *Id.* at I-13 to I-14.

197. *Id.* at I-14.

198. BOTHE ET AL., *supra* note 120, at 363.

AEWs, this would include the commander who orders the activation of the AEW and the planners and staff who execute the commander's order.[199]

Neither the Protocol nor ICRC commentary to Article 57 defines constant care, but by its plain meaning, it creates something more than a one-time obligation.[200] That is to say, it would be insufficient to take constant care when the weapon is deployed but ignore the weapon as it loiters for months. In the cyber context, the *Tallinn Manual* notes that the duty of care "requires commanders and all others involved in the operations to be continuously sensitive to the effects of their activities on the civilian population and civilian objects, and to seek to avoid any unnecessary effects thereon."[201] An analog obligation should be read into the employment of AEWs.

Article 57(2) further requires that those who "plan or decide upon an attack . . . [d]o everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects and are not subject to special protection but are military objectives . . . ."[202] In this context, the interpretation of what is "feasible" should be a "matter of common sense and good faith."[203]

In discussing precautions in attack, the military targeting process provides a useful framework in which to consider the requirements. Modern military practice distinguishes between preplanned targets and dynamic or

---

199. *See* William Henry Boothby, *Autonomous Attack—Opportunity or Spectre?*, *in* 16 YEARBOOK OF INT'L HUMANITARIAN LAW 71, 81 (Terry D. Gill et al. eds., 2013) ("[T]hose who decide that a sortie involving automated/autonomous attack technology shall be initiated have, for these purposes, 'decided upon' the attacks that the weapon system undertakes in accordance with the instructions that are fed into the mission control system at the commencement of the sortie. Those decision-makers therefore have a responsibility to satisfy themselves before the sortie commences that the scope of possible decisions that the weapon system is being permitted to make properly reflects the Article 57(2)(a)(i) obligations.").

200. *Constant*, OXFORD ENGLISH DICTIONARY, https://en.oxforddictionaries.com/definition/constant (last visited Nov. 11, 2017); *Constant*, MERRIAM-WEBSTER DICTIONARY, https://www.merriam-webster.com/dictionary/constant (last visited Nov. 11, 2017) (defining "constant" as "happening all the time or very often over a period of time"); *see also* TALLINN MANUAL, *supra* note 34, at 166 ("Use of the word 'constant' denotes that the duty to take care to protect civilians and civilian objects is of a continuing nature throughout all cyber operations.").

201. TALLINN MANUAL, *supra* note 34, at 166; *see also* Eric Talbot Jensen, *Cyber Attacks: Proportionality and Precautions in Attack*, 89 INT'L L. STUD. 198, 202–03 (2013) ("[C]onstant care would likely require a commander to maintain situational awareness at all times . . . [in order] to adjust operations if the tool or operation began to have effects that the commander determined would have an illegal impact on civilians.").

202. AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 57(a)(2)(i).

203. Sandoz et al., *supra* note 152, ¶ 2189.

emerging targets.[204] Preplanned targets are "known to exist in the operational environment."[205] They are typically static, as in the case of buildings or military facilities.[206] Dynamic targets are those that emerge during an armed conflict;[207] they are targets of opportunity. A tank that suddenly emerges from a dense forest or a group of personnel engaging friendly forces exemplifies dynamic targets. Autonomous weapons systems could be used against dynamic or preplanned targets. Here, this Section will focus on dynamic targets since they present the most challenging issues of distinction, and one could argue that any system that is directed to engage a particular target is, by definition, not autonomous.

In United States doctrine, targeting utilizes a six-step process: Find, Fix, Track, Target, Engage, Assess.[208] First, intelligence collection identifies a potential target.[209] The fix step of the process refers to actions taken to confirm the nature and location of the target.[210] These first two steps can occur simultaneously.[211] Once a target's location has been fixed, the target is then tracked until engaged.[212] The targeting step refers to the allocation of resources against the target, a risk assessment, deconfliction with other friendly assets, and target validation.[213] During the validation, the operator asks a series of questions designed to verify the validity of the target.[214] In the final step of the process, an assessment is made as to whether the mission was a success or failure.[215]

---

204. *See* CHAIRMAN OF THE JOINT CHIEFS OF STAFF, U.S. DEP'T OF DEF., JOINT PUBLICATION 3-60, JOINT TARGETING II-2 to II-3 (2012) (noting U.S. doctrine categorizes targets as either deliberate or dynamic, planned, or targets of opportunity) [hereinafter JOINT PUBLICATION 3-60].

205. *Id.* at II-2.

206. *See id.*

207. *See id.* at II-2.

208. *Id.* at II-23 fig.II-10.

209. *Id.*

210. *Id.*

211. *Id.* at II-24.

212. *See id.* at II-26 to II-27.

213. *See id.* at II-29.

214. DEP'T OF THE ARMY, ARMY TECHNICAL PUBLICATION 3-60, ¶ 2-56 (May 7, 2015) ("*Target validation asks such questions as*: . . . Is engaging the target lawful? . . . Does the target contribute to the enemy capability and will to wage war? Is the target (still) operational? Is it (still) a viable element of a target system? . . . Are there any facilities or targets on the no-strike list or restricted target list collocated with the target being validated? What is the relative potential for collateral damage or collateral effects, to include casualties? . . . Would engaging the target generate significant environmental impacts or arouse environmental sensitivities?").

215. *See* JOINT PUBLICATION 3-60, *supra* note 204, at II-31.

The obligation to comply with proportionality extends from the planning of the attack through the execution of the attack. Article 57(2)(b) requires an attack be:

> cancelled or suspended if it becomes apparent that the objective is not a military one or is subject to special protection or that the attack may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated.[216]

Thus, the requirement to cancel or suspend can be implicated by a change in either the military advantage or the anticipated collateral damage. This requirement applies to those who have the authority to cancel or suspend, those at higher echelons of command who possess information that would necessitate cancellation or suspension.[217]

As discussed in the context of distinction and proportionality, autonomy has the potential to allow the activation of a weapons system long before targets are engaged. A prolonged engagement begs the question of when precautions in attack should be taken: When the system is deployed, when it is activated, when it is about to engage, or throughout the process? Precautions in attack are continuous in nature and run from the activation of the system to the engagement.[218] Technology has the potential to enhance these requirements. Thus, if a cruise missile has a video feed and the ability to abort, the operator would be obligated to monitor the feed and abort the missile should the proportionality calculation change significantly. Autonomy then raises the possibility that a system could be sophisticated enough to take continual precautions in attack. If an AEW possessed such a

---

216. AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 57(2)(b).

217. BOTHE ET AL., *supra* note 120, at 366 ("The Committee expressed the obligation in the passive voice so that it would apply to all commanders who have the authority to cancel or suspend attacks, including those at higher echelons who frequently have better intelligence sources than those actually engaged. But it also applies to the commander in the military organizations actually engaged in combat."); *see* OFF. OF GEN. COUNS., *supra* note 78, ¶ 5.1.2 ("[I]nternational obligations would only apply to those persons belonging to the party's forces with the domestic authority to make the decisions necessary to implement those obligations.") (citing United Kingdom, Statement on Ratification of AP I, Jan. 28, 1998, 2020 U.N.T.S. 75, 78 ("The United Kingdom understands that the obligation to comply with [art. 57] paragraph 2(b) only extends to those who have the authority and practical possibility to cancel or suspend the attack.")).

218. *See* BOTHE ET AL., *supra* note 120, at 363.

capability, a commander could rely on the system if he or she were confident the system could conduct the precautions analysis with reasonable certainty.

In sum, the requirements for precautions in attack are continuing obligations which affix to commanders and planners and all others who have the requisite information and ability to cancel or suspend an attack if necessary. Autonomy creates additional complexities in that the autonomous weapon system itself may possess the capability to conduct the feasibility analysis. There is nothing legally objectionable with this possibility, assuming the system is of sufficient sophistication that the commander employing the system is reasonably certain the system will comply with the obligations to take feasible precautions in the attack.

F. Control Over Weapons

1. Meaningful Human Control

Weapons used in armed conflict should be controlled by their users. The law of armed conflict is predicated on the idea of distinction, and thus "method[s] or means of combat which cannot be directed at a specific military objective" are unlawful.[219] Here "directed" is synonymous with "controlled."[220] Similarly, the ability to cancel or suspend an attack and to take feasible precautions in attack necessarily requires some degree of control over the employment of the system. As a matter of practice, militaries and commanders spend considerable time and money to maximize control over their weapons systems. Indeed, control is arguably the very essence of a military—whether control of troops, units, weapons, or munitions.

In the debate surrounding autonomous weapons, the concept of control has manifested itself in the phrase "meaningful human control."[221] Meaningful human control holds, in short, that humans should exert some

219. AP I, *supra* note 77, pt. IV, § 1, ch. II, art. 51(4)(b).

220. *Directed*, OXFORD ENGLISH DICTIONARY, http://www.oxforddictionaries.com/us/definition/american_english/direct?q=directed#direct__19 (last visited Nov. 11, 2017) ("directed" means to "control the operations of").

221. *See e.g.*, Horowitz & Scharre, *supra* note 10, at n.54; U.N. INST. FOR DISARMAMENT RESEARCH, THE WEAPONIZATION OF INCREASINGLY AUTONOMOUS TECHNOLOGIES: CONSIDERING HOW MEANINGFUL HUMAN CONTROL MIGHT MOVE THE DISCUSSION FORWARD (2014), http://www.unidir.org/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf.

level of control over AEWs.[222] While all acknowledge AEWs should be controlled, there is considerable debate over the necessary level of control, how that control is exerted, and upon what the control is exerted—all matters of meaningful discussion.[223] There further appears to be some confusion about the requirement of control: Is it required by existing international law or is it a policy imperative? This Section addresses the issues of international law implicated by control—or lack thereof—of weapons systems and attempts to provide some granularity to the discussion.

While the concept of meaningful human control has found currency with some States,[224] the United States has instead adopted the phrase "appropriate human judgment."[225] At the 2016 CCW Informal Meeting of Experts on Lethal Autonomous Systems, the United States explained its rationale behind this phrase.[226] After noting its discomfort with the subjectivity and lack of clear meaning, the United States stated:

> We view the optimization of the human/machine relationship as a primary technical challenge to developing lethal autonomous weapon systems . . . . Because this human/machine relationship extends throughout the development and employment of a system and is not limited to the moment of a decision to engage a target, we consider it more useful to talk about "appropriate levels of human judgment."[227]

This formulation echoes a statement made by the United Kingdom at the same meeting, where it expressed dissatisfaction with the phrase meaningful

---

222. The phrase was first introduced by Article 36, a British NGO. *See* ARTICLE 36, *KILLER ROBOTS: UK GOVERNMENT POLICY ON FULLY AUTONOMOUS WEAPONS* (2013), http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf.

223. *See* General Statement by Ger. at the 2015 U.N. Convention on Conventional Weapons Informal Meeting of Experts on Lethal Autonomous Weapons Systems (Apr. 13–17, 2014) (At the 2015 U.N. Convention on Conventional Weapons Informal Meeting of Experts on Lethal Autonomous Weapons Systems, the chair of the meeting, Germany, noted that the 2014 meeting produced "a common understanding regarding the necessity to exercise appropriate levels of human control over the use of force.").

224. Countries making public statements on the concept include Croatia, Denmark, Germany, Ireland, Japan, the Netherlands, South Africa, South Korea, and Sweden.

225. DOD DIRECTIVE, *supra* note 26.

226. Michael W. Meier, Statement of the U.S. Delegation to the Convention on Certain Conventional Weapons Informal Meeting of Experts on Lethal Autonomous Systems (Apr. 11, 2016).

227. *Id.*

human control for the same reasons voiced by the United States.[228] As an alternative, the United Kingdom proposed the concept of an "intelligent partnership" between human and machine that holds that a fully autonomous system is impractical if not impossible.[229] Instead, the United Kingdom acknowledged that computers and humans have different strengths and weaknesses, which necessitates a partnership between the two.[230]

2. *Mechanisms of Control*

The mechanism of control can be exercised through physical or technological means.[231] Historically, weapons were controlled through physical means. A human operator physically manipulates the weapon through positioning and manual activation of the weapon system (e.g., a rifle that is manually aimed and physically triggered by the user). Control through technological means is control that is manifest in the software of a system, such as an air-to-surface missile that, once launched, cannot be controlled by the operator but has been programmed to target a particular object. Control is exerted through the programming of the missile.

Control can be manifest across either or both vectors. By way of example, a 120mm mortar is a weapons system that is entirely mechanical and is controlled only by physical control.[232] A Paladin M109A6 155mm artillery system, by contrast, is a highly automated artillery system comprised of a 155mm artillery gun mounted on a tracked vehicle and controlled by a sophisticated computer control system.[233] The Paladin is controlled through a combination of physical control (e.g., where the operator drives the vehicle) and technological control (computation of firing data and automatic aiming of the gun tube). A cyber weapon, which has no physical manifestation, would be entirely controlled by the technological means, that is, through the programming of the weapon.

The relationship between physical and technological control is particularly interesting in the context of highly automated technologies. As

---

228. Statement of the U.K. of Gr. Brit. and N. Ir. to the Informal Meeting of Experts on Lethal Autonomous Systems 6 (Apr. 11–15, 2016).

229. *Id.* at 5–6.

230. *Id.* at 3.

231. *See* Sassóli, *supra* note 170, at 320 ("What counts is that either the system itself through technical means, or the human beings using it, are able to acquire information indicating that the attack must be interrupted . . . .").

232. *See generally* U.S. ARMY ACQUISITION CORPS, WEAPON SYSTEMS HANDBOOK 284–85 (2016), https://www.army.mil/e2/c/downloads/431298.pdf (providing a description of Mortar Systems).

233. *See generally id.* at 72.

weapons increase in technological sophistication, the requisite skill of the operator also increases; increasingly complex weapons require increasingly technologically savvy operators. At some point, however, the level of technology and automation becomes sufficiently high that a sophisticated operator may not be necessary. When computers were first introduced, they were operated exclusively by highly knowledgeable users who physically controlled the systems.[234] As the sophistication of technology increased, the requisite skill needed to operate the systems diminished and control was increasingly manifest through technology (software) rather than physical manipulation of punch cards and memory tapes. To be clear, control over the system (computer or otherwise) does not diminish with increasing sophistication, but rather, the nature of control (physical vice technological) changes.

As a matter of law, there is no distinction between effecting control through physical manipulation of a weapon (e.g., a wire-guided missile) and effecting control through a computer program (e.g., a Tomahawk Cruise Missile), so long as the requisite span of control can be expressed through computer programming, and the program or system will operate with a reasonable degree of certainty in a given environment.[235] The Phalanx provides a useful illustration. A human operator programs certain parameters into the system regarding when and what the system will engage. Based on his or her knowledge of the system and the environment in which it is operating, the human operator can be reasonably certain the weapon will function in a particular manner. When an object—an enemy missile for instance—triggers the Phalanx engagement criteria, the system will automatically engage the missile. What and when the system engages is entirely controlled by the human operator, albeit through a pre-programmed set of instructions.

---

234. *See Invention of the PC*, HISTORY, http://www.history.com/topics/inventions/invention-of-the-pc (last visited Nov. 11, 2017) ("The earliest electronic computers were . . . enormous and hugely expensive, and they required a team of engineers and other specialists to keep them running.").

235. *See* WILLIAM BOOTHBY, WEAPONS AND THE LAW OF ARMED CONFLICT 233 (2009) (noting that precautionary requirements can be met by having a person "in the loop" or by advance programming which controls "the timing, location, objective, and means of . . . attack").

3. *Control Predicates*

As discussed below, control of a weapon is manifest across three dimensions: time, space, and effects.[236] All weapons systems, conventional or autonomous, are controlled through these dimensions.[237] A dumb bomb dropped from an aircraft is controlled by time (when it is dropped), space (where it is dropped), and weapons effects (what size bomb is dropped). The ability to exert control over each dimension, however, is contingent on two predicates: the skill of the operator and an effective interface that allows the user to effect control over the system. An AEW cannot be controlled unless these two predicates are first met. This Section discusses first these predicates of control (skill and effective interface) and then the dimensions of control (time, space, and effects).

a. Operator Skill as a Control Predicate

The first predicate requires that the AEW operator possess the skill to operate the system. There is a direct relationship between the skill of the operator and the ability to control a weapon. A RAND Corporation study on the effect of personnel quality on the performance of the Patriot Air-Defense System provides a useful illustration. The RAND report found "considerable evidence that [standardized test scores have] a direct and consistent effect on the outcomes of air battles . . . ."[238] Perhaps unsurprisingly, the study concluded that "soldiers with higher [standardized test scores] can be expected to suffer significantly less asset damage, destroy more hostile aircraft, and be more effective in missile conservation."[239] Needless to say, an air-defense system operator who is more efficient with regards to missile conservation and enemy aircraft identification is less likely to engage a civilian aircraft.

Operator skill includes the skills necessary to operate the AEW, including activation, initiation, execution, maintenance, deactivation, and a basic understanding of the operational characteristics of the AEW, including its operational characteristics (e.g., how far it can fire, what munitions it

---

236. *See id.*

237. This is not a novel argument. *See, e.g.*, Peter Asaro, *On Banning Autonomous Weapons Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making*, 94 INT'L REV. RED CROSS 687, 695 (2012).

238. BRUCE R. ORVIS ET AL., RAND ARROYO CENTER, EFFECT OF PERSONNEL QUALITY ON THE PERFORMANCE OF PATRIOT AIR DEFENSE SYSTEM OPERATORS vi (1992), https://www.rand.org/content/dam/rand/pubs/reports/2006/R3901.pdf.

239. *Id.*

fires, etc.) as well as its operational reliability (e.g., what level of predictability does the system exhibit, how accurate is the system, etc.). An operator with a greater degree of skill will have a better ability to exert control vice an operator with a low degree of skill. For example, if a person with no military experience were to attempt to operate a 120mm mortar, the person would have effectively no control over the system, whereas a skilled operator would have the ability to control the mortar.

b. Effective Interface as a Control Predicate

The second control predicate requires an operator/system interaction such that the operator can exert control over the function of the AEW. In other words, an operator must have both the necessary skill and system understanding, but also an ability to actually effect control. If the AEW is out of communication, broken, or physically too distant, the human user cannot exert control. Not only must the AEW be capable of being controlled, the means to effect control includes a readily understood user interface.

In the *2016 Summer Study on Autonomy*, the U.S. Defense Science Board discussed the issue of user interface at some length.[240] The report grounds the discussion of control in the concept of trust, such that the machine will do what the user expects and intends the machine to do.[241] As the report concludes, "[e]stablishing trustworthiness . . . and providing adequate indicator capabilities so that inevitable context-based variations in operational trustworthiness can be assessed and dealt with at run-time is essential, not only for the operator and the Commander, but also for designers, testers, policy and lawmakers, and the American public."[242] The report lists several "barriers to trust" including "ineffective interfaces."[243]

When considering control over technologically sophisticated weapons systems, a corollary to an effective user interface is the requirement of mutual understanding between human operator and machine. An AEW "may have different sensors and data sources than any of its human teammates," and thus "may be operating on different contextual assumptions of the operational environment."[244] Similarly, machines and human operators need to have a shared understanding of the goals the operator wishes to achieve. By way of example, the *Summer Study* notes many of the aviation accidents of the 1990s where "the flight crew had one goal (e.g., staying on the glide

240. DEF. SCIENCE BD., *supra* note 73.
241. *Id.* at 14.
242. *Id.*
243. *Id.* at 15.
244. *Id.* at 14.

slope during an approach) and the flight management computer had another (e.g., executing a go-around)."[245]

4. Control Through Effects, Space, and Time

Assuming the operator satisfies both predicates (skill and effective interface), he or she can exert control across three dimensions. First, control may be asserted through the effects of weapon selection: A targeting officer selects a 500-pound dumb bomb for a given mission, rather than a 1,000-pound dumb bomb. Thus, before the munition has left the aircraft, the operator has taken actions to control the weapon.[246] In an advanced weapon system with autonomous features, the process would be the same. The type of weapon selected for the engagement and the features activated would be determined based on the nature of the target.

The second vector of control is spatial control. Conventionally, this is expressed by where the user points the weapon. For example, an artillery round is geographically controlled by where the operator aims the gun tube. Most conventional weapons are activated and employed in the same geographic locale, which provides the user with an understanding of where and how the weapon is affecting the battlefield.

AEWs differ in that they might be activated in one geographic area, but their effects manifest elsewhere. That is not to say they cannot be controlled; it is just that the nature of the control changes. To use a conventional example, consider a cruise missile that travels hundreds of miles. Here, the spatial control is perhaps less apparent but is no less considerable. Spatial control is exerted by programming the location of where the missile will strike. Thus, a cruise missile fired 1,000 miles from the target may be more accurate than an artillery round fired ten miles from the target. The same control could be exerted over an AEW where its operation could be easily geographically bounded.

The third dimension of control is time. Temporal restrictions on AEWs could restrict operations during a specified period, or prohibit operations for a given period. An artillery round, for example, is controlled in time by choice of the operator when to engage the system. AEWs change this aspect

245. *Id.* at 15.

246. *See* U.S. DEP'T OF THE AIR FORCE, AIR FORCE PAMPHLET NO. 14-210, USAF INTELLIGENCE TARGETING GUIDE 56 (1998). This process is commonly referred to as weaponeering, which is defined as "the process of estimating the quantity of a specific type weapon required to achieve a specific level of damage to a given target, considering target vulnerability, weapon effects, munition delivery errors, damage criteria, probability of kill, weapon reliability, etc." *Id.*

of control in that they have the ability to operate for extended periods of time, turn on and off in a given period, or both. Depending on the nature of the system, this could result in increased *or* decreased temporal control. An AEW deployed for an extended period is unconstrained from a temporal perspective. Conversely, that same system could be programmed to only operate in short periods of predetermined time, which would enable a high degree of temporal control.

IV. WEAPON REVIEWS

A. Generally

Before a commander uses any weapon in combat, he or she will (or should) demand assurances that the weapon will act in accordance with the specifications provided and in a lawful manner. The weapons review process generates this information.[247] The legal lodestar for this obligation is Article 35 of Additional Protocol I, which reaffirms the longstanding proposition that the methods and means of warfare are not unlimited.[248] The weapons review process ensures weapons are not unlawful.[249] For States party, Article 36 of Additional Protocol I operationalizes the weapons review requirement, which imposes an obligation to review new weapons.[250] The mechanism for this obligation is the weapons review process.[251] While the Additional Protocol does not mandate the form of the weapons review, it is widely accepted that a review should consider both the weapon itself and the planned and normal circumstances of the weapon's use.[252]

International humanitarian law prohibits two broad categories of weapons as unlawful per se: those that cause superfluous injury or unnecessary suffering;[253] and those that are inherently indiscriminate, including weapons that cannot be aimed or whose effects cannot be

247. Sandoz et al., *supra* note 152, ¶ 1469.

248. AP I, *supra* note 77, pt. III, § 1, art. 35.

249. OFF. OF GEN. COUNS., *supra* note 78, ¶ 6.2.2.

250. AP I, *supra* note 77, pt. III, § 1, art. 36.

251. OFF. OF GEN. COUNS., *supra* note 78, ¶ 6.2.2.

252. Sandoz et al., *supra* note 152, ¶ 1469; OFF. OF GEN. COUNS., *supra* note 78, ¶ 6.2.2 (the U.S. considers the "weapon's intended use" to determine whether the weapon is "calculated to cause superfluous injury"); *see also* ICRC WEAPONS GUIDE, *supra* note 34, at 938 ("A weapon or means of warfare cannot be assessed in isolation from the method of warfare by which it is to be used. It follows that the legality of a weapon does not depend solely on its design or intended purpose, but also on the manner in which it is expected to be used on the battlefield.").

253. AP I, *supra* note 77, pt. III, § 1, art. 35.

controlled.[254] Both categories of prohibitions reflect customary international law.[255] There is nothing inherent in an autonomous weapons system that raises unique issues in this regard, as the technology of autonomy does not, in and of itself, create superfluous injury or unnecessary suffering. In the same way, the technology of autonomy does not create indistinction. Weapons reviews should, of course, consider these per se prohibitions, though they are unlikely to be violated by the autonomous aspect of the weapon. Determining the lawfulness of an autonomous weapons system in its normal and expected circumstances of use is more challenging.

B. Challenges with Testing Advanced Technology

While not unique to autonomous weapons, the implicit technological sophistication of autonomous systems demands increasingly sophisticated means of testing the systems. Addressing the issue of testing technologically complex weapons, Alan Backstrom and Ian Henderson write, "[t]he use of a guided weapon with an autonomous firing option requires an understanding of the legal parameters; the engineering design, production, and testing (or validation) methods; and the way in which the weapon might be employed on the battlefield."[256] More than that, advancing technologies will require new approaches to test technologically sophisticated weapons.

The software that animates AEWs presents a particularly challenging area for testing. Many AEWs will rely on machine learning algorithms, which enable the system to "iteratively learn from data," so that the system will "produce reliable, repeatable decisions and results."[257] There are several methods used to train machine learning algorithms, including supervised and unsupervised learning models.[258] The most common learning method, supervised learning, requires that the learning algorithm be fed training data to learn what the user desires the machine to learn. Thus, an algorithm

---

254. AP I, *supra* note 77, pt. IV, § 1, ch. 2, art. 51(4)(b)–(c).

255. HENCKAERTS & DOSWALD-BECK, *supra* note 84, at 237, 244–45; *see generally* BOOTHBY, *supra* note 235, at 46–73 (discussing the historical development of prohibition on weapons that cause superfluous injury and unnecessary suffering).

256. Alan Backstrom & Ian Henderson, *New Capabilities in Warfare: An Overview of Contemporary Technological Developments and the Associated Legal and Engineering Issues in Article 36 Weapons Reviews*, 94 INT'L REV. RED CROSS 483, 484 (2012).

257. *Machine Learning: What It Is and Why It Matters*, SAS, https://www.sas.com/en_us/insights/analytics/machine-learning.html (last visited Nov. 11, 2017).

258. Jason Brownlee, *A Tour of Machine Learning Algorithms*, MACHINE LEARNING MASTERY (Nov. 25, 2013), http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/.

designed to identify enemy tanks would be provided millions of images of tanks and the system would self-adjust as it became more accurate.

In the context of AEWs, there are two broad concerns with machine learning. First, a machine-learning algorithm is only as good as the training it receives. A supervised learning process for an AEW designed to identify and destroy enemy tanks might include feeding the algorithm images labeled as tanks, and then later asking it to identify a tank among a group of vehicle images. Corrupt data will lead to corrupt results. Thus, if the algorithm is provided pictures of a truck and told that it is a tank, the algorithm will have an impaired ability to identify tanks (and trucks). The second problem occurs when the training data does not fully replicate the environment in which the system is designed to operate. If then, an algorithm is trained to identify tanks by being provided images of tanks taken in a wooded environment, the algorithm may come to define a tank as a tank-like object plus a wooded environment. In a desert or mountainous environment, the algorithm might be unable to identify a tank. Where, as in these examples, the algorithm cannot identify its target, it is not malfunctioning; rather, the problem is with the training.

C. Scope of Testing

The importance of testing weapons in the circumstances of their expected use exists too with regards to conventional weapons, but the importance is particularly acute when considering autonomous weapons and other highly sophisticated weapons. A bullet, for instance, will perform the same in the daytime as in the nighttime. Where a weapon relies on a suite of sophisticated sensors, however, weather and time of day are critical. By way of example, consider the Israeli HARPY system. The HARPY is an unmanned aircraft that loiters above a battlefield until it detects enemy radar, wherein it will engage the target in a kamikaze-style attack.[259] Say, hypothetically, that during testing it was determined the HARPY could distinguish the military objective from civilian objects 98% of the time in dry, sunny weather. This conclusion is relevant to the use of the HARPY in dry, sunny weather, but it is irreverent with regards to other circumstances of use (e.g., at night, in the rain, in the fog, etc.).

The parallels to an armed conflict scenario utilizing autonomous weapons systems are evident. Systems may be trained with incorrect data or data that is unconsciously biased. Weapons testing will only identify such

259. *Harpy Air Defense Suppression System*, DEFENSE UPDATE INT'L ONLINE DEFENSE MAGAZINE, http://defense-update.com/directory/harpy.htm#cont (last updated Mar. 4, 2006).

issues where the training environment accurately reflects the context in which the system will be used. In all reviews, certain best practices should be considered. These include the following:

- The weapons review should either be a multi-disciplinary process or include attorneys who have the technical expertise to understand the nature and results of the testing process.
- The review should delineate the circumstances of use for which the weapon was approved.
- The review should provide a clear delineation of human and system responsibilities. Who will do what in a given circumstance?
- Optimally, the review should occur at three points in time. First, when the proposal is made to transition a weapon from research to development. Second, before the weapon is fielded.[260] Finally, AEWs should be re-reviewed periodically based upon feedback on how the weapon is functioning. This suggestion would necessitate the establishment of a clear feedback loop that provides information from the developer to the reviewer to the user, and back again. This suggestion is perhaps not unique to AEWs, but it is of particular importance given the adaptability of autonomous weapons systems.

The review should also address the learning capacity of the AEW—are all sister systems trained exactly the same? Does it learn *in situ* in the operational environment?

In short, there are certainly aspects of weapons reviews in the context of autonomous weapons that need to be carefully considered. A robust weapons review is fundamental to ensuring autonomous weapons systems are used consistent with international humanitarian law.

V. ACCOUNTABILITY

If AEWs are used in armed conflict, there exists the possibility that they might be involved in a violation of the law of armed conflict. This prospect raises unique issues regarding the allocation of responsibility for the violation. Can a commander be held accountable for the actions of an AEW? Can the developer of the AEW be held accountable? Can the AEW itself be

260. This two-step review is the process adopted by the U.S. Department of Defense Directive on autonomous weapons. DOD DIRECTIVE, *supra* note 26, at 7.

held accountable? These are truly challenging issues of law, which have generated significant scholarship and necessitate careful consideration before the deployment of autonomous systems. This Part considers the requirement and mechanisms for accountability under international law. Here, consideration is given to both criminal and civil liability from the perspective of the individual, the commander, and the State.

A. Requirement to Hold Accountable

Treaty[261] and customary international law[262] obligate States to hold accountable those that seriously violate the law of armed conflict.[263] Accountability under international law includes individual responsibility in the form of direct[264] and command responsibility,[265] as well as State

---

261. GC I, *supra* note 91, art. 49; GC II, *supra* note 91, art. 50; GC III, *supra* note 91, art. 129; GC IV, *supra* note 91, art. 146; AP I, *supra* note 77, art. 86 (requires States party to "repress grave breaches, and take measures to suppress all other breaches, of the Conventions or of this Protocol . . . ."); *see also* Rome Statute, *supra* note 97, pmbl.; Convention on the Prevention and Punishment of the Crime of Genocide, art. VI, Dec. 9, 1948, 102 Stat. 3045, 78 U.N.T.S. 277; Second Protocol to the Hague Convention of 1954 for the Protection of Cultural Property in the Event of Armed Conflict, art. 28, Mar. 26, 1999, 2253 U.N.T.S. 212; Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, art. 7, Dec. 10, 1984, 1465 U.N.T.S. 85; Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction, art. VII(1), Jan. 13, 1993, 1974 U.N.T.S. 45; Amended Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices, art. 14, May 3, 1996, 2048 U.N.T.S. 93; Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction, art. 9, Sept. 17, 1997, 2056 U.N.T.S. 211.

262. Other Cruel, Inhuman or Degrading Treatment or Punishment, art. 7(1), Dec. 10, 1984, 1465 U.N.T.S. 112; Convention on the Prevention and Punishment of the Crime of Genocide, art. VI, Dec. 9, 1948, 78 U.N.T.S. 277; GC I, *supra* note 91, art. 49; GC II, *supra* note 91, art. 50; GC III, *supra* note 91, art. 129; GC IV, *supra* note 91, art. 146.

263. OFF. OF GEN. COUNS., *supra* note 78, ¶ 18.3.1 ("Each member of the armed forces has a duty to comply with the law of war in good faith.").

264. *See generally* ANTONIO CASSESE ET AL., CASSESE'S INTERNATIONAL CRIMINAL LAW 65 (3rd ed. 2013) ("War crimes are *serious violations* of customary or treaty rules belonging to international humanitarian law . . . .") [hereinafter CASSESE'S INTERNATIONAL LAW]; DINSTEIN, *supra* note 151, at 263 (citing DEP'T OF THE ARMY, FM-27-10: DEPARTMENT OF THE ARMY: FIELD MANUAL: THE LAW OF LAND WARFARE ch. 8 (1956), http://www.aschq.army.mil/gc/files/FM27-10.pdf ("In the past, it was frequently contended that '[e]very violation of the law of war is a war crime.' But such assertions have never elicited support in actual State practice.")).

265. Rome Statute, *supra* note 97, pmbl.

responsibility.[266] International Human Rights Law also, arguably, contains requirements for accountability.[267]

B. Criminal Liability

1. Threshold Issues

Individual criminal responsibility for a violation of the law of armed conflict necessitates the existence of an armed conflict.[268] This seemingly perfunctory threshold raises significant problems in the context of developers of autonomous weapons. As Tim McCormack and Tim McFarland point out, autonomous weapon development will often occur before the armed conflict commences.[269] This fact generates novel issues of accountability. How does one hold accountable, for example, an engineer who in peacetime develops an AEW that is designed to commit war crimes, and does, in fact, do so during a later armed conflict? McCormack and McFarland were "unable to identify any extant jurisprudence from international war crimes trials to support the notion of individual criminal

266. Rome Statute, *supra* note 97, art. 28; AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 86(1), 87(1); OFF. OF GEN. COUNS., *supra* note 78, ¶ 18.4 ("Military commanders have a duty to take appropriate measures as are within their power to control the forces under their command for the prevention of violations of the law of war.").

267. *E.g.*, International Covenant on Civil and Political Rights, art. 2(3), Dec. 1966, 99 U.N.T.S. 171 (requiring each State Party to ensure that "any person whose rights or freedom . . . are violated shall have an effective remedy . . . ."); Convention for the Protection of Human Rights and Fundamental Freedoms as Amended by Protocols No. 11 and No. 14, art. 13, Nov. 4, 1950, 213 U.N.T.S. 222 ("Everyone whose rights and freedoms . . . are violated shall have an effective remedy . . . .") (amended June 1, 2010); Basic Principles and Guidelines on the Right to a Remedy and Reparation for Victims of Gross Violations of International Human Rights Law and Serious Violations of International Humanitarian Law, U.N. Doc A/RES/60/147 (Dec. 16, 2005) (stating that States have a duty to "[t]ake appropriate legislative and administrative and other appropriate measures to prevent violations . . . [and] where appropriate, take action against those allegedly responsible in accordance with domestic and international law").

268. Prosecutor v. Tadić, Case No. IT-94-1-I, Decision on Defence Motion for Interlocutory Appeal on Jurisdiction, ¶ 67 (Int'l Crim. Trib. for the Former Yugoslavia Oct. 2 1995), http://www.icty.org/x/cases/tadic/acdec/en/51002.htm; *compare* Rome Statute, *supra* note 97, art. 8 (defining war crimes), *with* Definition of War Crimes, U.N. OFFICE ON GENOCIDE PREVENTION AND THE RESPONSIBILITY TO PROTECT, http://www.un.org/en/genocideprevention/war-crimes.html (last visited Nov. 11, 2017) (stating the requirement that the alleged "conduct took place in the context of and was associated with . . . armed conflict").

269. 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).

liability for war crimes where an accused's acts have occurred prior to the commencement of an armed conflict."[270]

The International Criminal Tribunal for the Former Yugoslavia's (ICTY) jurisprudence also calls into question the ability to prosecute a person for a war crime where the conduct occurs *before* the armed conflict.[271] In *Tadić*, the tribunal concluded "each of the four Geneva Conventions contains language intimating that their application may extend beyond the cessation of fighting."[272] The tribunal provides no consideration to the idea that the conventions may extend *before* the armed conflict. Indeed, *Tadić* is particularly notable for its studied consideration of when an armed conflict commences, and by extension, when criminal liability attaches.[273]

2. *Direct Individual Criminal Responsibility*

Direct physical perpetration refers to "physically and personally perpetrating a crime or engendering a culpable omission in violation of a rule of criminal law."[274] While addressing Article 3 of the ICTY Charter, the appeals chamber in *Tadić* articulated a cumulative four-part test for war crimes that has found broad support.[275] To constitute a war crime, the tribunal found the following elements must all be satisfied:

i. the violation must constitute an infringement of a rule of international humanitarian law;

---

270. *Id.* at 377.

271. Prosecutor v. Tadić, Case No. IT-94-1-I, Decision on Defence Motion for Interlocutory Appeal on Jurisdiction, ¶ 67 (Int'l Crim. Trib. for the Former Yugoslavia Oct. 2 1995).

272. *Id.*

273. *Id.* ¶¶ 67, 70; *see* Prosecutor v. Haradinaj, Case No. IT-04-84-T, Judgment, ¶¶ 40, 51, 135 (Int'l Crim. Trib. for the Former Yugoslavia Apr. 3, 2008), http://www.icty.org/x/cases/haradinaj/tjug/en/080403.pdf.

274. *See* Prosecutor v. Naletić, Case No. IT-98-34-T, Judgment, ¶ 62 (Int'l Crim. Trib. for the Former Yugoslavia Mar. 31, 2003), http://www.icty.org/x/cases/naletilic_martinovic/tjug/en/nal-tj030331-e.pdf (defining "committing" with the same language); Prosecutor v. Tadić, Case No. IT-94-1-A, Appeal Judgment, ¶¶ 187–88 (Int'l Crim. Trib. for the Former Yugoslavia July 15, 1999), http://www.icty.org/x/cases/tadic/acjug/en/tad-aj990715e.pdf (using similar language to describe conduct creating liability).

275. *See* Robert Cryer, *Individual Liability in International Law*, *in* ROUTLEDGE HANDBOOK OF THE LAW OF ARMED CONFLICT 538, 541 (Rain Liivoja & Tim McCormack eds., 2016) ("The current, most influential, statement of the conditions for a violation of IHL to be considered a war crime was given by the ICTY in its seminal *Tadić* decision.").

ii. the rule must be customary in nature or, if it belongs to treaty law, the required conditions must be met . . . ;

iii. the violation must be "serious", that is to say, it must constitute a breach of a rule protecting important values, and the breach must involve grave consequences for the victim . . . ;

iv. the violation of the rule must entail, under customary or conventional law, the individual criminal responsibility of the person breaching the rule.[276]

A full examination of the elements of this test is beyond the scope of this Article. Consider instead the last element of the test that requires "individual criminal responsibility." This element encompasses both objective (*actus rea*) and subjective (*mens rea*) elements. The *actus rea* for a war crime is the same regardless of whether the crime is committed by an autonomous weapons system or a human.

In the context of individual criminal responsibility, mere negligence is insufficient to satisfy the *mens rea*.[277] Absent intent, criminal liability will attach only when an individual's negligence "reaches the threshold of gross or culpable negligence."[278] Antonio Cassese describes this mental state as evidenced by "conduct that is blatantly at odds with the prescribed standards."[279] Conduct that "falls short of the standard of precautions" would constitute simple negligence, and thus not satisfy the *mens rea* requirement.[280] It is also worth noting that the Rome Statute appears to impose a higher culpability to certain crimes such as "willful killing,"[281] "intentionally directing attacks against" civilians,[282] and other crimes.[283]

---

276. Prosecutor v. Tadić, Case No. IT-94-1-I, Decision on Defence Motion for Interlocutory Appeal on Jurisdiction, ¶ 94 (Int'l Crim. Trib. for the Former Yugoslavia Oct. 2 1995).

277. *See* THE OXFORD COMPANION TO INTERNATIONAL CRIMINAL JUSTICE 413 (Antonio Cassese et al. eds., 2009).

278. CASSESE'S INTERNATIONAL LAW, *supra* note 264, at 53.

279. *Id.* at 52.

280. *Id.* at 52–53.

281. Rome Statute, *supra* note 97, art. 8(2)(a)(1).

282. *Id.* art. 8(2)(b)(i), 8(2)(e)(i).

283. DINSTEIN, *supra* note 151, at 279 ("Article 8(2)(b)(xi) of the Rome Statute sets a higher standard of intent by insisting that killing or wounding will be perpetrated 'treacherously.'"). *But see* Johan D. Van der Vyver, *The International Criminal Court and the Concept of Mens Rea in International Criminal Law*, 12 U. MIAMI INT'L & COMP. L. REV. 57, 71, 111–12 (2005) (citing Michael Bothe, *War Crimes*, *in* THE ROME STATUTE OF THE INTERNATIONAL CRIMINAL COURT: A COMMENTARY 370, 389 (Antonio Cassese, et al. eds., 2002)) (rejecting the assertion made by Michael Bothe and others that the modifiers such as "willful," "treacherously," and "intentionally" convert these crimes to special intent crimes).

In the conventional sense, direct individual criminal responsibility is the soldier who willfully kills a civilian. Whether the soldier does so with his or her hands or rifle, the soldier directly perpetrates the crime. Article 25(3)(a) of the Rome Statute sets forth various modes of principal liability for the International Criminal Court, including direct perpetration where a person commits a crime as an individual.[284] Article 7 of the ICTY Statute contains similar language to Article 25, which establishes individual criminal responsibility for "[a] person who planned, instigated, ordered, committed or otherwise aided and abetted in the planning, preparation or execution of a crime . . . ."[285]

Where, however, a machine with some degree of autonomy commits the willful killing, one may fairly question whether the operator who activated the system—but does not control it—has directly perpetrated the crime. Making this determination requires an examination of the elements of the substantive crime, including the *actus reus* and *mens rea*. If for example, an autonomous machine attacks a civilian during armed conflict, the question arises as to who would be liable for the war crime of "willful killing." The *actus reus* of the crime requires that the "perpetrator killed one or more persons."[286] The Rome Statute explains "[t]he term 'killed' is interchangeable with the term 'caused death.'"[287] The degree of autonomy exhibited by the autonomous weapons system would dictate whether the person activating the system could be said to have caused the death of the victim.

Where there is a low level of autonomy, the causal link between the death and activation is evident. For example, an operator activates a surface to air missile that destroys an unequivocally civilian aircraft ten seconds later. It can be said that the operator caused the deaths of the civilians on the plane. Demonstrating a causal link for a more autonomous system could prove challenging. Consider a hypothetical AEW designed to target enemy aircraft. If the system is activated in January and engages a civilian aircraft in April, did the person who activated the system in January cause the civilian deaths in April? If the AEW was programmed to shoot down the first civilian airliner it encounters, then yes. If the AEW selected the civilian

284. Rome Statute, *supra* note 97, art. 25(3)(a).

285. Updated Statute of the International Criminal Tribunal for the Former Yugoslavia art. 7(1) (2009), http://www.icty.org/x/file/Legal%20Library/Statute/statute_sept09_en.pdf (last visited Nov. 11, 2017) [hereinafter ICTY Statute].

286. *Elements of Crimes*, INTERNATIONAL CRIMINAL COURT, https://www.icc-cpi.int/resourcelibrary/official-journal/elements-of-crimes.aspx (last visited Nov. 11, 2017).

287. *Id.* at 13 n.31 (explaining in the explanatory note that the structure of the elements mirrors the structure of the corresponding articles in the Rome Statute).

airliner from among many targets, then it becomes a question of the programming of the system. Did the system have the ability to distinguish between military and civilian? If so, how accurate was the software? Did the commander understand the system's limitations? Was the system tested in the same environment in which it was utilized?

Assuming the *actus reus* element can be satisfied, the *mens rea* of this crime requires that the perpetrator was aware of the victim's protected status. The perpetrator in this example is arguably the AEW, which has no awareness of the victim's status. The perpetrator could also be the commander who ordered the AEW activated. Again, however, with sufficient levels of autonomy, the commander who ordered the system will be unaware of any specific victims until after the engagement, let alone their status. In order to satisfy the *mens rea* requirement for direct responsibility, the commander would have to be acting with culpable negligence; that is to say, the commander acted "blatantly at odds with the prescribed standards."[288] This highlights the importance of weapons testing. Knowing the results of the testing process helps inform the question as to whether the commander's decision was reasonable.

3. *Co-Perpetrator Criminal Responsibility*

Two additional modes of liability are co-perpetration, where a person commits a crime "jointly with another,"[289] and indirect perpetration, where a person commits a crime "through another person."[290] In a series of recent decisions, the ICC has also applied a doctrine of indirect co-perpetration.[291]

---

288. CASSESE'S INTERNATIONAL LAW, *supra* note 264, at 52.

289. Rome Statute, *supra* note 97, art. 25(3)(a).

290. *Id.*

291. Prosecutor v. Muthaura, Case No. ICC-01/09-02/11-382-Red, Decision on the Confirmation of Charges pursuant to Article 61(7)(a)–(b) of the Rome Statute, ¶ 298 (Jan. 23, 2012), https://www.icc-cpi.int/CourtRecords/CR2012_01006.PDF (last visited Nov. 11, 2017); Prosecutor v. Al Bashir, Case No. ICC-02/05-01/09-3, Warrant of Arrest for Omar Hassan Ahmad Al Bashir, ¶ 210 (Mar. 4, 2009), https://www.icc-cpi.int/CourtRecords/CR2007_02360.PDF (last visited Nov. 11, 2017) (citing Prosecutor v. Dyilo, Case No. ICC-01/04-01/06-803-tEN, Decision on Confirmation of Charges, ¶¶ 326–28 (Jan. 29, 2007)); Prosecutor v. Katanga, Case No. ICC-01/04-01-07-717, Decision on the Confirmation of Charges, ¶¶ 533–34 (Sept. 30, 2008), https://www.icc-cpi.int/CourtRecords/CR2008_05172.PDF (last visited Nov. 11, 2017). *But see* Prosecutor v. Gombo, Case No. ICC-01/05-01/08-424, Decision Pursuant to Article 61(7)(a)–(b) of the Rome Statute on the Charges of the Prosecutor against Jean-Pierre Bemba Gombo, ¶ 345 (June 15, 2009), https://www.icc-cpi.int/CourtRecords/CR2009_04528.PDF (last visited Nov. 11, 2017) (stating analysis of indirect co-participation is not applicable since the Prosecutor put forth precisely the elements of co-participation).

Article 25(3)(b) further contemplates several accessorial modes of liability where one commits a crime when one "[o]rders, solicits, or induces the commission" of a crime within the jurisdiction of the court.[292] Finally, Article 25(3)(c) criminalizes the facilitation of a crime, including those acts which aid, abet, or assist the commission of the crime.[293]

The ICTY has developed its own variant of co-perpetrator liability known as Joint Criminal Enterprise (JCE).[294] While not expressly mentioned in Article 7 of the ICTY Statute, the ICTY has found that the article "does not exclude those modes of participating in the commission of crimes which occur where several persons having a common purpose embark on criminal activity that is then carried out either jointly or by some members of this plurality of persons."[295] The ICTY has actually developed three variants of JCE, the details of which are not relevant to the instant discussion.[296] For all three variants, the *mens rea* element requires that "those who take part in a common criminal act are aware of its purpose and share its requisite criminal intent . . . ."[297] Thus, JCE would be an inapplicable mode of liability for the development and use of autonomous weapons unless the design and use were intended "to perpetrate a certain crime" or "further the criminal activity or the criminal purpose of a group . . . ."[298]

Co-perpetration is the "division of essential tasks for the purpose of committing a crime between two or more persons acting in a concerted manner."[299] Co-perpetration necessitates a plan with "an element of criminality . . . ."[300] The court further clarified that this would encompass co-perpetrators that "(a) are aware of the risk that implementing the common plan (which is specifically directed at the achievement of a non-criminal goal) will result in the commission of the crime, and (b) accept such an

---

292. Rome Statute, *supra* note 97, art. 25(3)(b).

293. *Id.* art. 25(3)(c).

294. ROBERT CRYER ET AL., AN INTRODUCTION TO INTERNATIONAL CRIMINAL LAW AND PROCEDURE 368 (2d ed. 2010).

295. Prosecutor v. Tadić, Case No. IT-94-1-A, Appeal Judgment, ¶¶ 189–90 (Int'l Crim. Trib. for the Former Yugoslavia July 15, 1999), http://www.icty.org/x/cases/tadic/acjug/en/tad-aj990715e.pdf (last visited Nov. 11, 2017).

296. *See generally* CASSESE'S INTERNATIONAL LAW, *supra* note 264, at 163–72 (describing the three variants of JCE).

297. *Id.* at 263; *see also Tadić*, IT-94-1-A ¶ 228 (describing the three *mens rea* elements).

298. *Tadić*, IT-94-1-A ¶ 228.

299. Prosecutor v. Dyilo, Case No. ICC-01/04-01/06-803-tEN, Decision on Confirmation of Charges, ¶ 342 (Jan. 29, 2007), https://www.icc-cpi.int/CourtRecords/CR2007_02360.PDF (last visited Nov. 11, 2017).

300. *Id.* ¶ 344.

outcome."[301] Even setting aside the unlikely scenario where an autonomous weapons system is specifically developed to commit war crimes, this mode of liability could still be of relevance to those deploying autonomous systems. For example, military planners and commanders may coordinate to employ an AEW despite the fact that for the given circumstance or environment, it cannot distinguish between civilians and combatants with reasonable certainty. If commanders and planners know this, arguably they could be considered co-perpetrators (with one another) if the autonomous system commits a war crime.

Indirect co-perpetration is a hybrid mode of liability that combines an indirect perpetrator (e.g., an individual who exercises control over the person who commits the crime) with a co-perpetrator (e.g., an individual who exercises control over the person who sets up a common plan to commit a crime).[302] This mode of liability concerns a circumstance in which leaders act together to commit criminal acts, with each leader using an organization under their control.[303] Indirect co-perpetration allows for the attribution of the crimes to all leaders involved in the plan, regardless of whether the leader had control over all the subordinate forces committing the war crimes.[304] This mode of liability is likely inapplicable to the context of autonomous weapons.

Aiding and abetting is a mode of liability found in the Rome Statute,[305] as well as the Statutes of the ICTY,[306] the International Criminal Tribunal for Rwanda (ICTR),[307] and the Special Court for Sierra Leone (SCSL).[308] Aiding and abetting contemplates those acts "specifically directed to assist,

---

301. *Id.*

302. CASSESE'S INTERNATIONAL LAW, *supra* note 265, at 178–79; *see* Prosecutor v. Al Bashir, Case No. ICC-02/05-01-09-3, Warrant of Arrest for Omar Hassan Ahmad Al Bashir, ¶ 213 (Mar. 4, 2009), https://www.icc-cpi.int/CourtRecords/CR2009_01514.PDF (last visited Nov. 11, 2017); Prosecutor v. Katanga, Case No. ICC-01/04-01-07-717, Decision on the Confirmation of Charges, ¶¶ 490, 492–94 (Sept. 30, 2008), https://www.icc-cpi.int/CourtRecords/CR2008_05172.PDF (last visited Nov. 11, 2017). *But see* Prosecutor v. Gombo, Case No. ICC-01/05-01/08-424, Decision Pursuant to Article 61(7)(a)–(b) of the Rome Statute on the Charges of the Prosecutor against Jean-Pierre Bemba Gombo, ¶ 351 (June 15, 2009), https://www.icc-cpi.int/CourtRecords/CR2009_04528.PDF (last visited Nov. 11, 2017) (describing the elements for liability as a co-perpetrator).

303. CASSESE'S INTERNATIONAL LAW, *supra* note 264, at 178–79.

304. *Id.*

305. Rome Statute, *supra* note 97, art. 25(3)(c).

306. ICTY Statute, *supra* note 285, art. 7(1).

307. Statute of the Tribunal of the International Criminal Tribunal for Rwanda art. 6(1) (1994), http://ictr-archive09.library.cornell.edu/ENGLISH/basicdocs/statute/2007.pdf (last visited Nov. 11, 2017) [hereinafter ICTR Statute].

308. Statute of the Special Court for Sierra Leone art. 6(1) (2002), http://www.rscsl.org/Documents/scsl-statute.pdf (last visited Nov. 11, 2017).

encourage or lend moral support to the perpetration of a certain specific crime . . . ."[309] "[A]iding and abetting includes all acts of assistance by words or acts that lend encouragement or support, as long as the requisite intent is present."[310] Much like JCE, the intent, or *mens rea*, of aiding and abetting entails "knowledge that assistance aids the commission of criminal acts, along with awareness of the essential elements of these crimes."[311] Thus, unless the developer or commander employing the autonomous system had knowledge that the system could commit criminal acts, they cannot be said to have aided or abetted the crimes, or both.

4. Command Responsibility[312]

Command responsibility "provides for a mode of liability, through which superiors may be held criminally responsible for crimes within the jurisdiction of the Court committed by his or her subordinates . . . ."[313] Command responsibility is *sui generis*;[314] the superior is not being held

309. Prosecutor v. Tadić, Case No. IT-94-1-A, Appeal Judgment, ¶ 229 (Int'l Crim. Trib. for the Former Yugoslavia July 15, 1999), http://www.icty.org/x/cases/tadic/acjug/en/tad-aj990715e.pdf (last visited Nov. 11, 2017).

310. Prosecutor v. Tadić, Case No. IT-94-1-T, Opinion and Judgment by Judges Stephen & Vohrah, ¶ 689 (Int'l Crim. Trib. for the Former Yugoslavia May 7, 1997), http://www.icty.org/x/cases/tadic/tjug/en/tad-tsj70507JT2-e.pdf (last visited Nov. 11, 2017).

311. Prosecutor v. Perišić, Case No. IT-04-81-A, Judgment, ¶ 48 (Int'l Crim. Trib. for the Former Yugoslavia Feb. 28, 2013) (citing Prosecutor v. Mrkšić, Case No. IT-95-13/1-A, Appeals Judgment, ¶ 159 (Int'l Crim. Trib. for the Former Yugoslavia May 5, 2009)), http://www.icty.org/case/mrksic/4 (last visited Nov. 11, 2017); Prosecutor v. Blaškić, Case No. IT-95-14-A, Appeal Judgment, ¶ 49 (Int'l Crim. Trib. for the Former Yugoslavia July 29, 2004), http://www.icty.org/x/cases/perisic/acjug/en/130228_judgement.pdf (last visited Nov. 11, 2017).

312. The terms "command responsibility" and "superior responsibility" will be used interchangeably for the purposes of this Article. For a thorough discussion on command responsibility in the context of autonomous weapons, see Peter Margulies, *Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts*, *in* RESEARCH HANDBOOK ON REMOTE WARFARE (Jens David Ohlin ed., 2017).

313. Prosecutor v. Gombo, Case No. ICC-01/05-01/08-3343, Judgment Pursuant to Article 74 of the Statute, ¶ 171 (Mar. 21, 2016), https://www.icc-cpi.int/CourtRecords/CR2009_04528.PDF (last visited Nov. 11, 2017) (citing Prosecutor v. Gombo, Case No. ICC-01/05-01/08-424, Decision Pursuant to Article 61(7)(a)–(b) of the Rome Statute on the Charges of the Prosecutor against Jean-Pierre Bemba Gombo, ¶ 341 (June 15, 2009)), https://www.icc-cpi.int/CourtRecords/CR2016_02238.PDF (last visited Nov. 11, 2017)).

314. *Id.* ¶ 174.

responsible for the acts of the subordinate,[315] but rather the superior's own acts or omissions.[316] Command responsibility is not a form of strict liability.[317]

The Statutes of the ICTY,[318] the ICTR,[319] and the corresponding jurisprudence,[320] reflect customary law. The most comprehensive discussion of the issue came in the *Čelebići* case,[321] which articulated the requirements under customary law and ICTY/ICTR statute:

> (1) The superior must exercise direct and/or indirect command or control whether *de jure* and/or *de facto*, over the subordinates who commit serious violations of international humanitarian law, and/or their superiors.
>
> (2) The superior must know or have reason to know, which includes ignorance resulting from the superior's failure to properly supervise his subordinates, that these acts were about to be committed, or had been committed, even before he assumed command and control.
>
> (3) The superior must fail to take the reasonable and necessary measures, that are within his power, or at his disposal in the circumstances, to prevent or punish these subordinates for these offences.[322]

As with direct responsibility, *mens rea* must be established to hold a person accountable under the theory.[323] The second prong of the *Čelebići*

---

315. Prosecutor v. Krnojelac, Case No. IT-97-25-A, Judgment, ¶ 171 (Int'l Crim. Trib. for the Former Yugoslavia Sept. 17, 2003), http://www.icty.org/x/cases/krnojelac/acjug/en/krn-aj030917e.pdf (last visited on Nov. 11, 2017).

316. Prosecutor v. Delalić, Case No. IT-96-21-T, Judgment, ¶ 333 (Int'l Crim. Trib. for the Former Yugoslavia Nov. 16 1998), http://www.icty.org/x/cases/mucic/tjug/en/981116_judg_en.pdf (last visited Nov. 11, 2017).

317. *Id.* ¶ 383 ("The doctrine of superior responsibility does not establish a standard of strict liability for superiors for failing to prevent or punish the crimes committed by their subordinates.").

318. ICTY Statute, *supra* note 285.

319. *Id.*

320. Prosecutor v. Blaškić, Case No. IT-95-14-A, Appeal Judgment (Int'l Crim. Trib. for the Former Yugoslavia July 29, 2004), http://www.icty.org/x/cases/blaskic/acjug/en/bla-aj040729e.pdf (last visited Nov. 11, 2017).

321. *See* Prosecutor v. Delalić, Case No. IT-96-21-T, Judgment, ¶ 333 (Int'l Crim. Trib. for the Former Yugoslavia Nov. 16 1998), http://www.icty.org/x/cases/mucic/tjug/en/981116_judg_en.pdf (last visited Nov. 11, 2017).

322. *Id.* ¶ 344.

323. *Id.* ¶ 327.

test articulates the *mens rea* requirement for command responsibility.[324] Thus, commanders can be held liable under the doctrine where they have actual knowledge of the unlawful acts, or had reason to know of the unlawful acts. As Antonio Cassese has written, "knowledge is required in most cases of command responsibility."[325] In cases where actual knowledge does not exist, gross negligence is sufficient to establish the *mens rea*.[326] The clearest expression of this is the "should have known" standard found in the Additional Protocol,[327] Rome Statute,[328] and the ad hoc tribunals.[329]

Article 28 of the Rome Statute deviates from customary international law in two regards. In the first, the Rome Statute establishes different *mens rea* standards for military vice civilian superiors.[330] Military commanders are liable where they "either knew or, owing to the circumstances at the time, should have known that the forces were committing or about to commit such crimes,"[331] whereas the *mens rea* for civilian superiors is the more relaxed standard that they "either knew, or consciously disregarded information which clearly indicated, that the subordinates were committing or about to commit" crimes.[332] Secondly, the Rome Statute adds the requirement of causation, that is, a requirement that the subordinate's crimes occurred "as a result of . . . [the superior's] failure to exercise control."[333]

Unsurprisingly, all formulations of the doctrine of command responsibility envision "subordinates" committing the underlying crime.[334] In common parlance, the term subordinate connotes a person.[335] The dictionary definition, however, supposes a broader definition. The *Oxford*

---

324. *Id.*

325. CASSESE'S INTERNATIONAL LAW, *supra* note 264, at 51.

326. *See id.* at 53.

327. *See, e.g.*, AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 86(2) ("The fact that a breach of the Conventions or of this Protocol was committed by a subordinate does not absolve his superiors from penal or disciplinary responsibility . . . if they knew, or had information which should have enabled them to conclude . . . that he was committing or was going to commit such a breach . . . .").

328. Rome Statute, *supra* note 97, art. 28(a)(i) ("That military commander or person either knew or, owing to the circumstances at the time, should have known that the forces were committing or about to commit such crimes.").

329. *Delalić*, Case No. IT-96-21-T, ¶ 344; Prosecutor v. Halilović, Case No. IT-01-48-T, Judgment, ¶ 67 (Int'l Crim. Trib. for the Former Yugoslavia Nov. 16, 2005).

330. Rome Statute, *supra* note 97, art. 28.

331. *Id.* art. 28(a)(i).

332. *Id.* art. 28(b)(i).

333. *Id.* art. 28(a); *see also* Cryer, *supra* note 275, at 396–97.

334. *See, e.g.*, Rome Statute, *supra* note 97, art. 28(b).

335. *See Subordinate*, MERRIAM-WEBSTER DICTIONARY, https://www.merriam-webster.com/dictionary/subordinate (last visited Nov. 11, 2017) (noting that "subordinate" can mean "[i]nferior: a subordinate officer").

English Dictionary defines subordinate as "dependent upon, subservient to . . . something which is subordinate; a subordinate thing . . . ."[336] Case law from the ad hoc tribunals seems to support the idea that command responsibility can exist over a "thing" such as a military unit, rather than a specific subordinate person.[337] Citing ICTY decisions, Guénaël Mettraux writes:

> The prosecution would not necessarily be required to identify [the subordinates] by name, if it can be established, as a minimum, that the perpetrators were part of a unit, organ or structure over which the accused had authority and that the accused was able to exercise effective control over the members of that body or group, including those who committed the crimes.[338]

When then does the commander exercise "direct and/or indirect command or control" over an autonomous system? Culpability under the doctrine of superior responsibility requires that the superior must have had "effective control" over those subordinates who have committed the underlying crime.[339] Effective control, in turn, has been interpreted to mean the commander can "prevent [subordinates] from committing crimes or punish them after they committed the crimes."[340] As Mettraux has written, "[i]t is a relationship of authority which goes almost unquestioned between its two poles: one side orders; the other obeys."[341] Sufficient levels of autonomy strain this "unquestioned" relationship. At some point along the spectrum of autonomy, there exists the possibility that the system may not obey. Deploying such a system raises issues of individual responsibility but would sever the supervisor/subordinate relationship needed under the doctrine of command responsibility.

Knowledge of the existence of some risk of a future violation of the law is not sufficient. As noted by the ICTY Appeals Chamber in *Prosecutor v. Blaškić*, "[t]he knowledge of any kind of risk, however low, does not suffice

---

336. *Subordinate*, OXFORD ENGLISH DICTIONARY, https://en.oxforddictionaries.com/definition/subordinate (last visited Nov. 11, 2017); *see also id.* (defining "subordinate" to include something "submissive to be controlled by authority").

337. *See, e.g.*, Rome Statute, *supra* note 97, art. 28(a) (The statute uses terminology such as "forces" when referring to subordinates, indicating that it is not always referring to a specific person.).

338. GUÉNAËL METTRAUX, THE LAW OF COMMAND RESPONSIBILITY 159 (2009).

339. Prosecutor v. Delalić, Case No. IT-96-21-T, Appeals Judgment, ¶ 197 (Int'l Crim. Trib. for the Former Yugoslavia Feb. 20, 2001).

340. *Id.* ¶ 198.

341. METTRAUX, *supra* note 338, at 157.

for the imposition of criminal responsibility for serious violations of international humanitarian law."[342] Knowledge of previous breaches is relevant.[343] The knowledge may be either actual or constructive.[344] If constructive, the ad hoc tribunals have applied the standard that the superior "had reason to know."[345] By way of example, the Appeals Chamber in the *Čelebići* case noted that a commander would have the requisite constructive knowledge if they had "received information that some of the soldiers under his command have a violent or unstable character, or have been drinking prior to being sent on a mission . . . ."[346]

Thus, it would constitute constructive knowledge if the commander was informed that an autonomous system could not, for instance, reliably comply with the requirement for distinction. Here, again, is where the importance of weapons testing rises to the fore, as it is the weapons testing process that provides knowledge to the commander as to when a given system should be employed. A robust weapons testing process also serves as a measure a supervisor can take to comply with the requirement to "take measures necessary to suppress . . . breaches [of the law]."[347] Reasonable measures could also encompass a requirement to monitor the activity of the autonomous system and take measures to address malfunctions or actions that violate the law of armed conflict. Testing and refining autonomous systems in a closed environment may also be considered a measure to reduce breaches of the law of armed conflict.

Article 87 of Additional Protocol I further imposes on commanders an obligation "to ensure that members of the armed forces under their command are aware of their obligations . . . ."[348] The commentary notes that the commander satisfies this obligation by ensuring that the "unit gets proper training."[349] While a machine is not a subordinate within the meaning of the Conventions or Protocols, some autonomous systems can be trained as one

342. Prosecutor v. Blaškić, Case No. IT-95-14-A, Appeal Judgment, ¶ 41 (Int'l Crim. Trib. for the Former Yugoslavia July 29, 2004), http://www.icty.org/x/cases/perisic/acjug/en/130228_judgement.pdf (last visited Nov. 11, 2017).

343. Sandoz et al., *supra* note 152, ¶ 3545.

344. *See id.* ¶ 3546 ("[T]aking into account the circumstances, a knowledge of breaches committed by subordinates could be presumed.").

345. *Delalić*, Case No. IT-96-21-T, ¶ 238; AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 86(2) ("[H]ad information which should have enabled them to conclude in the circumstances at the time, that [the subordinate] was committing or was going to commit such a breach . . . .").

346. *Delalić*, Case No. IT-96-21-T, ¶ 197.

347. AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 86(1).

348. AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 87(2).

349. Sandoz et al., *supra* note 152, ¶ 3558.

would train a subordinate.[350] Such a capability may give rise to a parallel obligation to train the autonomous systems.

C. *State Responsibility*

The International Law Commission's Articles on State Responsibility (the Articles) articulate several well-accepted[351] general principles of State responsibility, which hold that first, "[e]very internationally wrongful act of a State entails the international responsibility of that State."[352] Second, "an internationally wrongful act" exists where, by act or omission, the conduct "(a) is attributable to the State under international law; and (b) constitutes a breach of an international obligation of the State."[353] Article 12 then defines an "internationally wrongful act" as an act of a State "not in conformity with what is required of it by that obligation, regardless of its origin or character."[354] As the commentary to the Draft Articles notes, "[i]nternational obligations may be established by a customary rule of international law, by a treaty or by a general principle applicable within the international legal order."[355]

As detailed elsewhere in this Article, there are any number of international obligations incumbent upon States, including the obligation to review weapons to ensure their compliance with international law,[356] and the obligation to exercise distinction,[357] proportionality,[358] and precautions in the attack.[359] Further, as discussed immediately above, Additional Protocol I

---

350. *See, e.g.*, STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 693–767 (3d. ed. 2010) (detailing various methods of machine learning).

351. *See* James R. Crawford, *State Responsibility*, MAX PLANCK ENCYCLOPEDIA OF PUBLIC INTERNATIONAL LAW ¶ 17 (Sept. 2006), http://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e1093?prd=EPIL ("These are well established, even axiomatic.").

352. Articles, *supra* note 95, art. 1.

353. *Id.* art. 2.

354. *Id.* art. 12.

355. *Id.* art. 12 cmt. 3.

356. AP I, *supra* note 77, pt. III, § 1, art. 35; *see also* TALLINN MANUAL, *supra* note 34, art. 48 ("All states are required to ensure that the cyber means of warfare that they acquire or use comply with the rules of the law of armed conflict that bind the State."); PROGRAM ON HUMANITARIAN POLICY AND CONFLICT RESEARCH AT HARVARD UNIVERSITY, MANUAL ON INTERNATIONAL LAW APPLICABLE TO AIR AND MISSILE WARFARE § C(9) (2010) ("States are obligated to assess the legality of weapons before fielding them.").

357. *See* AP I, *supra* note 77, pt. IV, § 1, ch. I, art. 48; HENCKAERTS & DOSWALD-BECK, *supra* note 84, art. 1, 7.

358. *See* AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 51(5)(b), art. 57(a)(iii); HENCKAERTS & DOSWALD-BECK, *supra* note 84, art. 14.

359. *See* AP I, *supra* note 77, pt. IV, § 1, ch. IV, art. 57(1).

obligates States to "repress grave breaches, and take measures necessary to suppress all other breaches, of the Conventions or of this Protocol which result from a failure to act when under a duty to do so."[360] Where an autonomous system employed by a State breaches these or other international obligations, State responsibility may incur. Before a State can be held responsible for such violations, the actions of the autonomous system must be attributable to the State.

Chapter II (Articles 4–11) of the Articles address attribution: when an action should be attributable as a matter of law to a State. Here, the Articles attribute to the State the conduct of the organs of the State[361] and entities exercising elements of governmental authority.[362] The Articles do not distinguish between different elements of the government (e.g., between military and intelligence agencies).

The Articles do not, of course, consider the potentiality of autonomous weapons systems. Article 5 does, however, consider a situation in which "a person or entity empowered by the law of that State to exercise elements of the governmental authority shall be considered an act of the State under international law."[363] This article might be regarded as autonomous weapons, which are systems that have been "empowered" by the State (e.g., activated) in order to "exercise elements of the governmental authority" (e.g., conduct combat operations).

If this article is read to trigger State responsibility for an autonomous weapons system activated by a State, the attendant liability will extend to acts both anticipated and unanticipated. Article 7 makes clear that State responsibility lies even where the agent of the State is acting *ultra vires*.[364] The commentary makes clear that a State will be responsible "even if the organ or entity acted in excess of authority or contrary to instructions."[365]

Even where a State is found to have violated an international obligation, there are circumstances precluding the wrongfulness of the act of the State. Chapter V of the Articles lays out six circumstances precluding the wrongfulness of conduct, to wit: consent,[366] self-defense,[367] countermeasures,[368] *force majeure*,[369] distress,[370] and necessity.[371] A

360. AP I, *supra* note 77, pt. V, § 1, ch. I, art. 86(1).
361. *See* Articles, *supra* note 95, art. 4.
362. *See id.* art. 5.
363. *Id.*
364. *Id.* art. 7.
365. *Id.* art. 7, cmt. 1.
366. *Id.* art. 20.
367. *Id.* art. 21.
368. *Id.* art. 22.
369. *Id.* art. 23.

detailed discussion of the substance of these circumstances is beyond the scope of this Article. It is sufficient to acknowledge the existence and full application of these circumstances to situations where States employ autonomous weapons.

Where there are no applicable circumstances precluding the wrongfulness of the act of the State, a breach of an international obligation generates obligations for the breaching State and rights for the injured State.[372] The obligations on the breaching State include the obligation to "cease the wrongful conduct [and] make full reparation for the injury caused by the internationally wrongful act."[373] Additional consequences follow if the internationally wrongful act "constitutes a serious breach by the State of an obligation arising under a peremptory norm of general international law."[374]

Thus, where a State violates an international obligation through the operation of an autonomous weapon and there is no circumstance precluding the wrongfulness of the act, then there exist several fora in which the State may be held liable. The Draft Rules also incorporate a robust regime for countermeasures which can be taken by the injured State.[375] Judicial fora include the International Court of Justice and domestic courts; though accountability through both systems would be hindered by issues of jurisdiction.[376]

VI. CONCLUSION

Autonomy will undoubtedly have an enormous impact on the conduct of hostilities. The newness and inherent complexity of the technology underlying autonomy creates a great deal of uncertainty about how the law of armed conflict applies to autonomy. As with any weapon system, the employment of autonomous weapons systems requires an understanding of

---

370. *Id.* art. 24.

371. *Id.* art. 25.

372. *See generally* Daniel Bodanksy & John R. Crook, *Symposium: The ILC's State Responsibility Articles*, 96 AM. J. INT'L L. 773, 785 (2002) ("The breach of an international obligation entails two types of legal consequences: it creates new obligations for the breaching state, principally, duties of cessation and nonrepetition (Article 30), and a duty to make full reparation (Article 31); and it creates new rights for injured states . . . .").

373. Articles, *supra* note 95, art. 28, cmt. 2 (referencing arts. 30 & 21).

374. *Id.* art. 28, cmt. 2 (referencing Articles 40 & 41).

375. *Id.* arts. 49–53.

376. *See generally* Daniel N. Hammond, *Autonomous Weapons and the Problem of State Accountability*, 15 CHICAGO J. INT'L L. 654 (2015).

the system and ability to control the system. The nature and reliability of the system's operation is informed through the weapons testing process.

Effectively testing autonomous weapons requires sophisticated, and possibly novel, means of testing. Tests must be designed to replicate the environment in which the system is sought to be used, and weapons reviews should reflect the scope of the testing protocols. An understanding of a system, however, is irrelevant if the system cannot be controlled. Control—expressed in time, space, and effects—can be exercised through the physical manipulation of a system or through the programming of the system. Effective control requires a skilled operator who can effect control over the system.

It is clear that the law of armed conflict applies and provides an effective normative framework to ensure the lawful employment of autonomous weapons. It is equally clear that not all autonomous weapons are the same. Some systems will be lawful and others will not. Indeed, a given system might be lawfully employed in one circumstance but not another. Simply put, some systems employing autonomous technologies will be lawful in some circumstances, while other systems will not. In all circumstances, development and use of autonomous weapons systems should carefully consider the uniqueness of the technology and the novel ways in which this technology affects the function of the weapons system.