2014

Autonomous Weapons and Human Responsibilities

Jack M. Beard
University of Nebraska College of Law, jbeard2@unl.edu

45 Georgetown Journal of International Law 617 (2014)

ARTICLES

AUTONOMOUS WEAPONS AND HUMAN RESPONSIBILITIES

JACK M. BEARD*

ABSTRACT

Although remote-controlled robots flying over the Middle East and Central Asia now dominate reports on new military technologies, robots that are capable of detecting, identifying, and killing enemies on their own are quietly but steadily moving from the theoretical to the practical. The enormous difficulty in assigning responsibilities to humans and states for the actions of these machines grows with their increasing autonomy. These developments implicate serious legal, ethical, and societal concerns. This Article focuses on the accountability of states and underlying human responsibilities for autonomous weapons under International Humanitarian Law or the Law of Armed Conflict. After reviewing the evolution of autonomous weapon systems and diminishing human involvement in these systems along a continuum of autonomy, this Article argues that the elusive search for individual culpability for the actions of autonomous weapons foreshadows fundamental problems in assigning responsibility to states for the actions of these machines. It further argues that the central legal requirement relevant to determining accountability (especially for violation of the most important international legal obligations protecting the civilian population in armed conflicts) is human judgment. Access to effective human judgment already appears to be emerging as the deciding factor in establishing practical restrictions and framing legal concerns with respect to the deployment of the most advanced autonomous weapons.

I. INTRODUCTION 618
II. THE TECHNOLOGY OF AUTONOMOUS WEAPONS AND THE DIMINISHING ROLE OF HUMANS 622
A. *Autonomy, Artificial Intelligence, and the Inexorable March of "Smart" Machines* 622
B. *A Continuum of Autonomy in Military Technologies* 625
1. "Fire-and-Forget" Weapons 628

* Assistant Professor, University of Nebraska College of Law; former Associate Deputy General Counsel (International Affairs), U.S. Department of Defense. The author greatly appreciates helpful comments by Matthew Schaefer, Brian Lepard, and Richard Moberly on earlier drafts.

2. Autonomous Defense Systems 630
3. The Approaching Horizon: Autonomous Combatant Systems 632
III. THE SEARCH FOR A FRAMEWORK TO GOVERN AUTONOMOUS WEAPON SYSTEMS 634
IV. THE SEARCH FOR INDIVIDUAL RESPONSIBILITY FOR THE ACTIONS OF AUTONOMOUS WEAPONS 642
A. *The International Criminal Law Framework* 642
B. *Autonomous Weapon Systems and the Search for Criminal Culpability* 647
1. Manufacturers, Designers, Engineers, and Programmers 647
2. Military Personnel and the Employment of Complex Machines 652
3. The Question of Command Responsibility 655
4. Blame Everyone? Blame the Machine? 660
V. STATE RESPONSIBILITY FOR AUTONOMOUS MACHINES AND THE PROBLEMS OF DISTINCTION AND PROPORTIONALITY 663
A. *Human Judgment Along the Continuum of Autonomy* 666
B. *Higher End of the Continuum: The Uncharted Legal Territory of the Future* 670
C. *Human Judgment, Practical Controls, and Legal Accountability* 674
VI. CONCLUSION 678

I. INTRODUCTION

Although remote-controlled robots flying over Afghanistan, Pakistan, Yemen, and other countries may now dominate reports on new military technologies, robots that are capable of hunting and killing enemies on their own are quietly but steadily moving from the theoretical to the practical. Early versions of these weapon systems are already widely deployed by military forces around the world, and more advanced ones are on their way. U.S. military officials view such machines as a crucial part of their future fighting forces.[1] Well-funded efforts are thus underway in the United States and other countries to build a wide variety of robots that are designed to "think, see and react increasingly

1. *See* OFFICE OF THE SEC'Y OF DEF., UNMANNED-INTEGRATED SYSTEMS ROADMAP, 2013-2038 67 [hereinafter UNMANNED INTEGRATED SYSTEMS ROADMAP, 2013-2038] (stating that "[a]utonomy in unmanned systems will be critical to future conflicts that will be fought and won with technology").

like humans."[2] The results of these efforts are new generations of weapon systems that display greater and greater levels of autonomy.

The advent of autonomous war-fighting machines has raised various concerns in the international community and increasingly now generates objections from international and non-governmental organizations. In a Report to the U.N. Council on Human Rights, the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Mr. Christof Heyns, argued that the deployment of lethal autonomous robots (LARs) "may be unacceptable because no adequate system of legal accountability can be devised, and because robots should not have the power of life and death over human beings."[3]

Fearing a future in which "killer robots" threaten humanity, several non-governmental organizations have issued calls for autonomous weapon systems to be outlawed, with one group's leader stating that "[o]ur concern is that humans, not sensors, should make targeting decisions."[4] However, many existing weapon systems already have sensors that "make targeting decisions." As discussed below, robots under the sea already detect, identify, and fire torpedoes at enemy vessels on their own; on the surface of the sea and on the land, various sophisticated weapon systems autonomously engage missiles, aircraft, ships, and an increasing variety of other targets; in the skies above, unmanned combatant aerial vehicles are being tested with a goal to replace manned fighter bomber aircraft and ultimately conduct missions in an autonomous mode.

For clarity in assessing the implications of existing, emerging, and future military technologies, this Article focuses on lethal military machines and weapon systems that may be described as "autonomous" to the extent that they have the ability (in varying degrees) to sense, control, and act without external human intervention or control.[5] The official definition used by the U.S. Department of Defense (DoD)

---

2. Tim Weiner, *New Model Army Soldier Rolls Closer to Battle*, N.Y. TIMES, Feb. 16, 2005, at A1.

3. Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, U.N.G.A., Human Rights Council, 23d Sess., U.N. Doc. A/HRC/23/47, Summary (further recommending that "[s]tates establish national moratoria on aspects of LARs, and calls for the establishment of a high level panel on LARs to articulate a policy for the international community on the issue").

4. Paul Marks, *Anti-Landmine Campaigners Turn Sights on War Robots*, NEW SCIENTIST (Mar. 28, 2008), http://www.newscientist.com/article/dn13550-antilandmine-campaigners-turn-sights-on-war-robots.html (quoting Richard Moyes, Director of Policy and Research of Landmine Action, a London-based charity).

5. Along similar lines, the most threatening versions of futuristic autonomous weapons, sometimes referred to as "killer robots," are described as being able "to select and engage targets without human intervention." *See* HUMAN RIGHTS WATCH, LOSING HUMANITY: THE CASE AGAINST

refines this concept a bit further, stating that an autonomous weapon system is one that, "once activated, can select and engage targets without further intervention by a human operator."[6]

For a wide variety of reasons, autonomous weapon systems are the next logical and seemingly inevitable step in the continuing evolution of military technologies.[7] New and ever-more sophisticated versions of autonomous military machines are, in fact, being so rapidly developed and deployed that they risk outpacing efforts to evaluate interrelated legal, ethical, and societal concerns. Foremost among these concerns are questions about who or what will be accountable for the damaging actions of these weapons in armed conflicts, particularly when they harm civilians. This Article assesses these problems of accountability in the context of state responsibility and individual culpability for actions of autonomous weapons, particularly under the International Humanitarian Law (IHL) framework, also referred to as the Law of Armed Conflict and the Law of War.

Many of the risks, dangers, and challenges of future autonomous weapon systems are already present in existing, widely-deployed systems. Rather than a rampage by rogue robots in some futuristic Hollywood production, the real threat presented by these systems comes in the form of a slow, creeping, and continuous movement to autonomous war-fighting capabilities in increasingly complex technological conflicts. Machines and computers continue to take on more and more important roles in all aspects of these conflicts, and as they do, the precise level of human involvement or control over them continues to become more and more diminished and uncertain.

This diminishing level of human control will continue to raise increasingly difficult questions about both state and individual accountability for the actions of autonomous weapon systems. While state and individual accountability involve different legal regimes, they sometimes share some key components, particularly in applying the IHL framework to determine whether states, through their military forces and commanders, have violated key obligations designed to protect civilians and civilian objects. These similar but different dimensions of

---

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

6. DOD Directive 3000.09, Autonomy in Weapon Systems (Nov. 21, 2012) [hereinafter DOD AUTONOMOUS WEAPONS DIRECTIVE], Part II, at 13.

7. While "weapon systems" may comprise many different components, programs, and supporting technologies, the terms "autonomous weapon systems" and "autonomous weapons" are used interchangeably in this Article.

the IHL framework are particularly important in giving full meaning to the fundamental obligations incumbent on states and military commanders to distinguish between targets by attacking only military objectives and to attack in such a way as to avoid excessive civilian casualties.

Setting aside moral, ethical, and broad societal concerns to focus on the *legal* accountability of states and underlying human responsibilities for autonomous weapons, this Article argues that the central legal requirement relevant to determining such accountability for violating the most important IHL obligations protecting the civilian population (relating to discrimination and proportionality) is a meaningful connection to the effective exercise of *human judgment.* This piece further argues that the elusive search for individual culpability for the actions of autonomous weapons foreshadows fundamental problems in assigning responsibility to states for the actions of these machines, pointing inescapably to a legal requirement for human judgment in applying the most complex tests under the IHL framework. Lastly, it argues that access to human judgment already appears to be emerging as the deciding factor in establishing practical restrictions and framing legal concerns with respect to the deployment of the most advanced autonomous weapons.

Part II of this Article provides a brief overview of autonomous weapons, their underlying technologies, the inevitable growth of these systems and their missions, and the ever-diminishing role of humans across a spectrum or continuum of autonomy. Part III assesses the search for a framework to govern autonomous weapon systems and examines key issues related to the application of IHL obligations to these systems. Part IV explores the extraordinarily challenging search for individual (and human) responsibility for the actions of autonomous weapons. As noted, while state and individual human responsibilities involve some distinctly different issues, the search for criminal responsibility nonetheless provides important perspectives on key challenges associated with state responsibility for the actions of autonomous machines.

Part V explores the problem of state responsibility for ensuring that autonomous weapons comply with IHL obligations, focusing particularly on the observance of the cardinal IHL principle of distinction (or discrimination) and the related requirement of proportionality. This assessment illustrates how the legal requirement of human judgment—as distinct from ethical or moral requirements—underlies both individual and state responsibilities under the IHL framework for compliance with its most important principles. In making this assess-

ment, Part V draws on the examination of existing and future autonomous military technologies discussed in Part II in order to identify the legal connection, or lack thereof, between human judgment and lethal machines along a continuum of autonomy. Part VI concludes with an examination of the role that human judgment may continue to play in establishing practical, and ultimately legal, barriers to the deployment of autonomous weapons as they incrementally proceed to higher and higher levels of autonomy.

II. THE TECHNOLOGY OF AUTONOMOUS WEAPONS AND THE DIMINISHING ROLE OF HUMANS

A. *Autonomy, Artificial Intelligence, and the Inexorable March of "Smart" Machines*

To describe a machine as truly autonomous raises serious philosophical questions about the nature of humans and machines that lie beyond the scope of this work. The most developed definitions of autonomy encompass concepts such as complete self-governance and the ability to make decisions as a free and independent moral agent. In the narrower context of existing and emerging military weapon systems, autonomy can be said to describe "the capacity of a machine to operate in the real-world environment without any form of external (human) control, once it is activated."[8]

In spite of their limitations, machines with advanced capabilities (including various degrees or levels of autonomy) are often enthusiastically portrayed as possessing human competencies and characteristics. For example, although existing advanced weapon systems are not cognitive beings and cannot think like humans do, terms such as "smart" and "intelligent" are nonetheless repeatedly used to describe them. In the context of both civilian and military technology, the adjective "smart" has taken on an expansive definition. In the market place, sellers apply the term "smart" to almost anything, including diapers, shoes, and automobile tires. The term has become almost as ubiquitous in the military sphere, defying meaningful definition as military forces use it to describe bombs, weapons, equipment, or machines that do not think, but instead are in some way more accurate,

8. PATRICK LIN ET AL., AUTONOMOUS MILITARY ROBOTICS: RISK, ETHICS, AND DESIGN 4 (Cal. Polytechnic St. Univ. ed., 2008) [hereinafter AUTONOMOUS MILITARY ROBOTICS REPORT].

reliable, durable, or adaptable than previous conventional "dumb" variants.[9]

A more useful scientific definition of a "smart" machine or structure for purposes of this Article and contemporary autonomous military operations is: "a system containing multifunctional parts that can perform sensing, control, and actuation."[10] Systems that are able to perform these functions are often said to rely on "artificial intelligence," a term defined in the Oxford Dictionary as "the theory and development of computer systems able to perform tasks normally requiring human intelligence."[11] While some ambitious definitions of artificial intelligence focus on the actual assumption by computers of human cognitive functions such as learning, reasoning, adaptation, and self-correction, most instead emphasize the appearance or imitation by computers of human intelligence, not its actual duplication.[12]

Films, novels, and other forms of popular culture often portray a future in which genuinely smart machines think, work, and fight wars. It is unlikely, however, that military machines will soon (if ever) be able to "think" like humans, at least in the sense that they will be able to reason and achieve genuine "cognition."[13] However, even if machines are never able to achieve cognitive capabilities, sufficiently difficult and important legal problems are presented by existing, emerging, and near-future autonomous weapons that rely on advanced automation, programming, complex algorithms, and other increasingly sophisticated artificial intelligence applications. Such autonomous systems are not designed to replace or replicate all the intellectual capabilities of a human but instead are intended to provide sufficient intelligence for specific tasks (albeit increasingly complicated and dangerous ones).

The inexorable march of new autonomous military technologies and

---

9. *See, e.g., Intelligent Clothing Could Save U.S. Military Lives*, TECH NEWS DAILY (May 11, 2012), http://www.technewsdaily.com/5748-intelligent-clothing-military-lives.html (describing "smart" uniforms that include medical sensors built into their fabric); MAX BOOT, WAR MADE NEW: TECHNOLOGY, WARFARE AND THE COURSE OF HISTORY 347 (2006) (describing "dumb" bombs used in the 1991 Gulf War as being only a bit more accurate than those used in World War II, in contrast to modern guided "smart" bombs).

10. Wenwu Cao et al., *Smart Materials and Structures*, 104 PROC. NAT'L ACAD. SCI. 8830, 8830-31 (2009).

11. OXFORD DICTIONARY OF ENGLISH 89 (2d ed. 2003) [hereinafter ODE].

12. WEBSTER'S NEW UNIVERSAL UNABRIDGED DICTIONARY 119 (2003) ("[T]he capacity of a computer to perform operations *analogous* to learning and decision making in humans.") (emphasis added).

13. ODE, *supra* note 11, at 335 (defining cognition as "the mental action or process of acquiring knowledge and understanding through thought, experience, and the senses").

the ever-increasing dependence by military forces on them has not been occurring in a vacuum. Modern society is increasingly confronted with complicated tasks that involve large amounts of data, the need to rapidly process that data, the ever-present possibility of operator errors, and the inability of humans to fully monitor the entire state of increasingly complex software systems. In this context, modern society is willing—even compelled—to put more and more important decisions in the hands of computers.[14] Examples of this phenomenon include financial institutions with world-wide computer networks that can trade stocks and evaluate millions of transactions in an instant, and public utility companies with power grids dependent on the decisions made by computerized control systems.[15] Incidents such as the 2003 electrical blackout in the northeastern United States and the 2010 stock market "flash crash" have demonstrated the dangers of entrusting machines with such decision-making power.[16]

In the military context, both autonomous and remote-controlled capabilities now support unmanned systems that are intended to "free warfighters from the dull, dirty, and dangerous missions that might now be better executed robotically and enable entirely new design concepts unlimited by the endurance and performance of human crews."[17] In general, unmanned military systems provide extra-human endurance, flexibility for expanded missions, complete safety for human operators, fewer manning and training costs, and vast new attack capabilities.[18] However, there are inherent communication limitations related to operating remote-controlled weapon systems. These concerns have prompted U.S. military planners to conclude that "the

14. WENDELL WALLACH & COLIN ALLEN, MORAL MACHINES: TEACHING ROBOTS RIGHT FROM WRONG 3, 19 (2009) ("In a world where operator error is a fact of life, and humans are unable to monitor the entire state of system software, the pressures for increased automation will continue to mount.").

15. *Id.* at 3, 4.

16. U.S.-CAN. POWER SYS. OUTAGE TASK FORCE, FINAL REPORT ON THE AUGUST 14, 2003 BLACKOUT IN THE UNITED STATES AND CANADA: CAUSES AND RECOMMENDATIONS (2004) (discussing how automated systems were implicated in worsening the blackout crisis); Tom Lauricella et al., *How a Trading Algorithm Went Awry*, WALL ST. J., Oct. 2, 2010, at B1 (discussing how computerized trading programs sold stocks in response to a perceived crisis and sent the stock market reeling).

17. OFFICE OF THE SEC'Y OF DEF., UNMANNED SYSTEMS ROADMAP, 2007-2032 34 [hereinafter UNMANNED SYSTEMS ROADMAP, 2007-2032].

18. UNMANNED INTEGRATED SYSTEMS ROADMAP, 2013-2038, at 20 (noting that "[u]nmanned systems have proven they can enhance situational awareness, reduce human workload, improve mission performance, and minimize overall risk to both civilian and military personnel, and all at a reduce[d] cost . . . . Unmanned systems provide persistence, versatility, survivability, and reduced risk to human life . . . .").

unmanned systems community must wean itself from the telecommunication bandwidth. Autonomy will certainly be required in order to accomplish this goal."[19]

Other key factors relentlessly driving modern military forces to develop and rely on new autonomous capabilities include: the increasingly complex demands of modern military technologies, the speed of modern weapons and related need for immediate responses to threats, and the vast amount of information that must be quickly evaluated and acted on in modern combat situations. This drive to automation also reflects a natural extension of the pursuit by military forces of new robotic capabilities generally—a trend that is strongly supported by both commercial interests and the national legislatures of states seeking cost savings and greater efficiency in their defense budgets.[20]

B. *A Continuum of Autonomy in Military Technologies*

Autonomous systems currently support a wide and growing spectrum of military functions and operations. Many important reconnaissance missions are now performed by unmanned aerial vehicles (UAVs) that rely on sophisticated autonomous systems for take-off, landing, and many navigation functions.[21] New autonomous functions are continually being added to these systems.[22] However, current and future autonomous systems capable of employing deadly force present the most significant challenges for the IHL framework.

Understanding the legal implications of autonomous military systems requires some degree of familiarity with modern military technologies, since many commonly used designations or descriptions of military technologies can be misleading. For example, so-called "smart" laser-guided PGMs (which first saw significant action near the end of

---

19. UNMANNED SYSTEMS ROADMAP, 2007-2032, *supra* note 17, at 49. Many of these systems are also dependent on communication satellites, a key vulnerability in some situations, particularly in any future conflict in which military-related assets in outer space could be threatened.

20. *See, e.g.*, National Defense Authorization Act for Fiscal Year 2001, Pub. L. No. 106-398, § 220(a), 114 Stat. 1654, 1654A-38 (2000) (mandating that one-third of designated U.S. military aircraft and ground combat vehicles be unmanned by 2010 and 2015, respectively); Noel Sharkey, *March of the Killer Robots*, TELEGRAPH (June 15, 2009), http://www.telegraph.co.uk/science/science-news/5543603/March-of-the-killer-robots.html (quoting a U.S. defense systems expert that "[u]nless we end war, end capitalism and end science, the use of robots will only grow . . . .").

21. *See, e.g.*, UNMANNED SYSTEMS ROADMAP 2007-2032, *supra* note 17, at 116 ("Global Hawk's airborne systems are designed to identify, isolate, and compensate for a wide range of possible system/subsystem failures and autonomously take actions to ensure system safety.").

22. *See, e.g.*, Jeremiah Gertler, *U.S. Unmanned Aerial Systems*, U.S. Congressional Research Service Report, Jan. 1, 2012, at 5 (discussing new autonomous air refueling capabilities).

the Vietnam conflict and gained prominence during the 1991 Persian Gulf conflict) are fundamentally lacking in "intelligence" or any autonomous capabilities.[23] Instead, they are best viewed as an extension of the human operator who optically guides these bombs to a target or the human who designates a target with a laser.

PGMs guided by humans therefore demonstrate key aspects of the continuing *dependence* of some advanced military technologies on human decisions and control: A human operator selects the types of targets to be attacked; senses and searches for the target (often with the assistance of several associated technologies); identifies the specific target to be attacked; assesses the value of the target and related risks of attack; initiates the attack; guides the bomb to its intended point of impact; and, in some cases, retains the ability to suspend the attack after launch by diverting the bomb (up to a point).

As machines replace human performance of each of the above functions, weapons reach successively higher levels of autonomy. In doing so, they raise fundamental questions about state and individual responsibilities for compliance with IHL obligations. These questions become even more pressing and complicated as autonomous weapons are given greater mobility, equipped with improved sensing capabilities, and assigned a wider range of missions and targets. An examination of the legal implications of autonomous weapons is thus best undertaken in the context of a continuum of autonomy (from least to greatest autonomy for the machine), taking into account not only the capabilities of the weapons and the nature of human involvement, but also the circumstances of their employment, missions, and intended targets.

The movement along this continuum appears to be both inevitable and incremental, presenting many practical difficulties in precisely defining diminishing levels of human control.[24] On the lower end of the continuum are machines that rely primarily on control by humans with only limited, brief, or supporting autonomous functions. Some of these machines may be classified as "semi-autonomous" systems, *e.g.*, those that, once activated, only engage individual targets selected by

23. BOOT, *supra* note 9, at 325-28 (discussing the rise of precision-guided weapons and the employment of laser-guided bombs).

24. 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. MONOGRAPH, Apr. 9, 2013, at 19, *available at* http://www.hoover.org/publications/monographs/144241 ("It seems far more likely that the evolution of weapons technology will be gradual, slowly and indistinctly eroding the role of the human in the firing loop.").

human operators.[25] On the higher end are groups of machines of the future that the military hopes will one day be "self-actualizing."[26] The U.S. military argues, however, that the decision to engage individual targets with lethal force will continue to be made by humans and not by machines "for the foreseeable future."[27] This simple and absolute statement is subject to increasing skepticism if one carefully surveys existing autonomous weapons systems and the roles played by their human operators.

The role of human involvement—whether it is described as control, delegation, supervision, or the possibility of intervention—is both fundamental and problematic in determining the operational and legal status of a weapon system along the continuum of autonomy. Although all machines that are able to sense and act based on their programming can be said to possess some degree or level of autonomy, it has been suggested that associated, diverse levels of autonomy can be divided into three categories: "Human-in-the-Loop;" "Human-on-the-Loop," and; "Human-out-of-the-Loop."[28]

"Human-in-the-Loop" refers to the lowest level of automation or autonomy, one in which machines (including many types characterized as "robots") can only "select targets and deliver force with a human command."[29] A second level of autonomy is "Human-on-the-Loop," which is described by the DoD as involving "human-supervised autonomous weapon systems" that "can select and engage targets without further human input after activation" and are "designed to allow human operators to override operation of the weapon system."[30] The third and final level approaching fully autonomous capabilities is described as "Human-out-of-the-Loop" and involves weapons that are "capable of selecting targets and delivering force without any human input or interaction."[31]

Yet clear distinctions between "levels of autonomy" may in practice be difficult to identify when different types of current and emerging military technologies are carefully assessed. Furthermore, the manner

25. DOD AUTONOMOUS WEAPONS DIRECTIVE, *supra* note 6, at 13.

26. *Id.*

27. U.S. DEPT. OF DEF., OFFICE of THE SEC'Y OF DEF., UNMANNED SYSTEMS INTEGRATED ROADMAP, FY2011-2036, at 17 (noting that such capabilities will instead "be retained under human control in unmanned systems").

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

29. *Id.*

30. DOD AUTONOMOUS WEAPONS DIRECTIVE, *supra* note 6, at 13-14.

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

in which autonomous weapon systems are deployed and the nature and complexity of their assigned missions may be highly significant in determining the legal significance of the limited human involvement in the operation of these systems. At each ascending level along the continuum of autonomy, there are various factors that may drive human operators to cede more responsibility to weapons, thereby diminishing human involvement, limiting access to human judgment, and complicating the legal assessment of accountability at each level.

To illustrate these points, it is useful to look at three broad categories of weapon systems that may be viewed as moving from lower to higher points along the continuum of autonomy. They are: "Fire-and-Forget" Weapons, Autonomous Defense Systems, and Autonomous Combatant Systems.

1. "Fire-and-Forget" Weapons

In pursuit of guidance systems for long-range missiles and various other types of projectiles and munitions, military forces in World War II embarked on a journey to develop autonomous capabilities for machines. Primitive autonomous guidance systems were developed for the first ballistic missiles (Nazi "V-2" rockets) and for torpedoes (U.S. "Mine Mark-24" torpedoes that were capable of seeking the acoustic signatures or sound pulses of ships).[32] Later in the 1950s, U.S. "Sidewinder" air-to-air missiles were equipped with guidance systems that allowed them to autonomously track targets based on heat emissions/infrared energy, followed in the 1960s by air-to-air missiles that were equipped with their own radar homing capabilities.[33]

Weapons designers later searched for weapons that could more reliably be used to attack targets "beyond visual range" (BVR) or "over the horizon" (OTH), leading to the development of new generations of sophisticated "fire-and-forget" weapons—so named because they did not require further operator guidance or involvement after launch.[34] Advanced versions of these weapons on the land, under the sea, and in

32. OXFORD COMPANION TO AMERICAN MILITARY HISTORY 13, 724 (John Whiteclay Chambers ed., 1999).

33. *Id.* at 311.

34. A BVR missile usually refers to an air-to-air missile, such as the U.S. AMRAAM missile, that is capable of engaging targets at ranges beyond 20-25 nautical miles. An OTH missile, such as the U.S. AGM-84D Harpoon anti-ship missile, has a sea-skimming cruise trajectory and a range in excess of 60 nautical miles. U.S. DEPT. OF DEF., DEFENSE INDUSTRIAL BASE CAPABILITIES STUDY: FORCE APPLICATION 35 (Oct. 2004); U.S. NAVY FACT FILE, HARPOON MISSILE, http://www.navy.mil/navydata/fact_display.asp?cid=2200&tid=200&ct=2 (last visited Oct. 22, 2010).

the air are now able to autonomously adjust their movements while tracking targets with a variety of guidance systems (or combinations of these systems) that include improved infrared optics, active radar guidance, computers with image processing algorithms, and perhaps most importantly, guidance systems with GPS receivers for satellite navigation.[35]

Increasingly sophisticated fire-and-forget weapons continue to limit direct human participation in war-fighting, making attacks more and more dependent on information that is programmed into weapons before launch (such as radar measurements, infrared images, and particularly, GPS coordinates). While some of these weapon systems can be connected to data links that allow them to receive additional information after launch or release, they may often operate with no further human involvement or "human in the loop." Humans still control the initial targeting and engagement decisions for these weapons, but they are less well positioned to evaluate the risks related to attacking specific targets when they lie "over the horizon." It is thus not surprising that the primary U.S. military regulation that establishes "guidelines designed to minimize the probability and consequences of failures in autonomous and semi-autonomous weapon systems that could lead to unintended engagements" is specifically made applicable to "guided munitions that can independently select and discriminate targets."[36]

Because a human makes the ultimate decision to launch advanced guided munitions at targets, these weapons are often excluded from many discussions of autonomous killing machines. Instead, "target recognition technology" is emphasized, along with a focus on "matching specific sensor information with predictive templates of the intended target."[37] Yet the appearance of ultimate human control here may be misleading. The realities of modern warfare are forcing humans to cede more and more responsibilities to targeting systems that can make their own rapid evaluations of data in chaotic combat conditions and also recognize and evade likely enemy countermeasures. The demands being placed on autonomous targeting sys-

35. JOHN ANDREAS OLSEN, A HISTORY OF AIR WARFARE 246 (2010) (noting how self-guided Joint Direct Attack Munitions (JDAMs) equipped with the Global Positioning System can "strike any static target by day or night, in all weathers, from heights up to forty thousand feet, with a 'near precision' accuracy of a few feet").

36. DOD AUTONOMOUS WEAPONS DIRECTIVE, *supra* note 6, at 2.a.(2).

37. U.S. DEPT. OF AIR FORCE, UNMANNED AIRCRAFT SYSTEMS FLIGHT PLAN 2009-2047 16 (May 18, 2009) [hereinafter UAS FLIGHT PLAN].

tems thus continue to grow and have led the U.S. military to conclude that, in the future, these weapons must be "capable of learning and exercising a spectrum of missions."[38]

2. Autonomous Defense Systems

Moving further along the continuum of autonomy are weapons that not only have their own sensory, movement, and attack capabilities after launch, but also have the power (once activated) to decide which targets will be attacked and then act on that "decision." Collectively referred to in this Article as "autonomous defense systems," they encompass a variety of guard, sentry, barrier, onboard, and point defense weapon systems, many of which are used in the static defense of manned platforms.

Although they are sometimes excluded from discussions about autonomous weapons, the most primitive autonomous defense weapons (and still among the most controversial) are anti-personnel landmines. Once these weapons are activated and put in place by humans, they have their own crude but effective sensors, identify and select the targets within their sensor range, and then attack those targets on their own (although humans still play a key role in target selection by placing and then activating the mines in a particular, stationary location). Other more sophisticated naval mine systems do not rely on direct contact, but instead are triggered by the magnetic, seismic, or pressure influence of a ship or submarine, and are designed to both detect and classify vessels.[39]

Other autonomous systems defend specific objectives such as ships or designated air space. For example, the U.S. "Phalanx Close-In Weapon System" is deployed on ships as a "fast reaction" defense capability against numerous threats.[40] The Phalanx consists of a radar-guided Gatling gun mounted on a swiveling base and is "capable of autonomously performing its own search, detect, evaluation, track,

38. *Id.*

39. MK 60 CAPTOR ASW MINE, 45 JANE'S AIR-LAUNCHED WEAPONS 614 (Mar. 2005) (noting how the U.S Mark 60 Encapsulated Torpedo (CAPTOR) can autonomously launch a torpedo when its sonar detects a "hostile" submarine").

40. U.S. NAVY FACT FILE, PHALANX CLOSE-IN WEAPON SYSTEM, http://ipv6.navy.mil/navydata/fact_display.asp?cid=2100&tid=487&ct=2 (last visited Oct. 22, 2013) (noting that the Phalanx provides an "inner layer point defense capability against Anti Ship Missiles (ASM), aircraft, and littoral warfare threats that have penetrated other fleet defenses").

engage and kill assessment functions."[41] A land-based variant of the Phalanx, called the "Centurion," is a "Counter-Rocket, Artillery and Mortar" system that is equipped with so-called "sense, warn and response capabilities."[42]

Although human operators of the Phalanx and Centurion systems have the option to intervene and make the decision to engage targets, the high speed of multiple incoming missiles and other threats will often make autonomous operation the preferred (if not unavoidable) option. This problem greatly complicates meaningful human oversight of many sophisticated autonomous weapon systems. For example, in discussing the operations of the Centurion, one author has noted:

> The human is certainly part of the decision making but mainly in the initial programming of the robot. During the actual operation of the machine, the operator really only exercises veto power, and a decision to override a robot's decision must be made in only half a second, with few willing to challenge what they view as the better judgment of the machine.[43]

Military forces have many incentives to deploy autonomous systems to counter threats posed by land-based intruders and attackers. Protecting large, important facilities or areas may require many human guards, which may be ineffective if the guards lose their focus, falter under severe weather conditions, fail to follow orders, or become fatigued, afraid, or otherwise indisposed. Such concerns were addressed by Samsung Techwin Company when it introduced another logical step in autonomous military systems: a robotic sentry called the SGR-1, intended for use by the Republic of Korea in guarding its side of the Demilitarized Zone (DMZ).[44] Similarly, the Israel Defense Forces

---

41. *Id.* Two sets of radars and sophisticated search and tracking subsystems allow the Phoenix to identify targets based on bearing, range, velocity, heading, and altitude information. A similar U.S. system, called the "SeaRam," relies on an 11-round missile launcher assembly instead of a Gatling gun.

42. *Lawmakers Approve $350 Million C-RAM Funding Boost for Afghanistan*, INSIDE THE ARMY, Oct. 5, 2009.

43. P.W. Singer, *War of the Machines: A Dramatic Growth in the Military Use of Robots Brings Evolution in Their Conception*, SCI. AM., July 2010, at 63.

44. Kim Tae-gyu, *Machine Gun-Armed Robots to Guard DMZ*, KOREA TIMES, June 24, 2010 (noting that the SGR-1 is equipped with state-of-the-art surveillance technology, heat and motion sensors, and a 5.57 millimeter machine gun; it can operate autonomously or on a remote-controlled basis; it is designed to recognize potential human targets; and it can be programmed to call on targeted individuals to declare a password or indicate surrender, and if necessary fire on

have deployed numerous stationary "robot gun-and-sensor installations" along Israel's borders with the Gaza Strip, which are now remotely controlled but are also capable of autonomous operation.[45]

In an attempt to move to a higher level of autonomous sentry capabilities, Samsung Techwin Company continues to work with a Korean government institute on a mobile variant of the SGR-1 that can perform patrolling functions along the DMZ.[46] In fact, prototypes of mobile robotic sentries are already patrolling ammunition depots in the United States, and airports and other facilities in Israel.[47] Military forces are evaluating the use of armed robots—initially only remote-controlled variants—for a wide range of perimeter security and sentry missions.[48] This trend is not confined to the land, as robotic unmanned surface vessels (USVs) are being developed for many different missions by the U.S. Navy and its foreign counterparts.[49]

3. The Approaching Horizon: Autonomous Combatant Systems

For its own planning purposes, the U.S. military recognizes that the most complex missions for future unmanned systems will require high levels of autonomy and that such "ultimate autonomy" will in turn require "some capabilities analogous to those of the human brain" (at least for specific missions).[50] As intensive efforts continue to capture more and more human capabilities in software algorithms and new

---

those targets with great accuracy.); Ed Pilkington, *'Killer Robots' Pose Threat to Peace and Should Be Banned, UN Warned*, GUARDIAN (U.K.), May 29, 2013 (noting that although the SGR-1 "is reported to have an automatic mode," it is currently configured to be operated by a human).

45. Lewis Page, *Israel Deploys Robo-Snipers on Gaza Border: Man Still in the Loop . . . to Begin With*, THE REGISTER (June 5, 2007), http://www.theregister.co.uk/2007/06/05/israel_robo_sniper_gaza/; Paul Koring, *Killer Robots Inching from Sci-Fi to Reality*, GLOBE & MAIL (Can.), Nov. 21, 2012.

46. Kim Tae-gyu, *supra* note 44.

47. Many systems are being tested in the United States, including one built by General Dynamics called the "Mobile Detection Assessment Response System" (MDARS); an Israeli version, called the Guardium, is used for border security operations and can be "fitted with machine guns" and "programmed to return fire if shot at." David Hambling, *Intruders Beware: Armed Robots on Patrol*, NEW SCIENTIST, Oct. 25, 2010, at 22.

48. *See, e.g.*, Erik Sofge, *America's Robot Army: Are Unmanned Fighters Ready for Combat?*, POP. MECHANICS, Dec. 18, 2009 ("Last June the Army deployed the first-ever armed UGVs. Three SWORDS (Special Weapons Observation Remote Direct-Action System) robots landed in Iraq, each equipped with an M249 light machine gun.").

49. Erik Sofge, *Robot Boats Hunt High-Tech Pirates on the High-Speed Seas*, POP. MECHANICS, Oct. 31, 2007.

50. OFFICE OF THE SEC'Y OF DEF., UNMANNED SYSTEMS ROADMAP, 2005-2030 52 (2005) (noting that "to achieve that level, machine processing will have to match that of the human brain in speed, memory, and quality of algorithms, or thinking patterns").

computational approaches, military technologies appear to be steadily moving along the continuum of autonomy to an ominous endpoint: the deployment of mobile, autonomous, combatants that will replace many humans on and above battlefields and at sea.

The first crossing of this threshold—to mobile, fully autonomous combatant systems—will probably occur in the air. The increasing complexity of aerial combat continues to highlight the limits of existing UAVs (as well as conventional, manned fighter aircraft) and has prompted senior U.S. military officials to comment that "[t]here's really no way that a system that is remote controlled can effectively operate in an offensive or defensive air combat environment. The requirement of that is a fully autonomous system."[51]

In addition to UAVs that now rely on sophisticated autonomous take-off, landing, and navigation systems, some UAVs already possess autonomous attack capabilities, at least against certain types of ground targets. For example, Israel has built and deployed a hunter-killer UAV variant (called the Harpy-2) that autonomously loiters over battlefields and is able to suppress enemy air defenses without human intervention by self-destructing into them.[52]

Efforts to build an unmanned combat aerial vehicle (UCAV) capable of autonomous operations continue in several countries, and prototype versions are already being tested.[53] Because UCAVs do not have to be "built around a human being," they can be "smaller, stealthier and more maneuverable."[54] The final step on the continuum envisioned by the U.S. Air Force involves the leveraging of new and fully autonomous UCAV capabilities (including automatic target engagement), swarming ("one pilot directing the actions of many multi-mission aircraft"),

---

51. Sharkey, *supra* note 20 (quoting the deputy director of the Pentagon's Unmanned Aerial Systems Task Force).

52. ANTHONY FINN & STEVE SCHEDING, DEVELOPMENTS AND CHALLENGES FOR AUTONOMOUS UNMANNED VEHICLES: AAA COMPENDIUM 178 (2010).

53. *See, e.g.*, Dominic Basulto, *The X-47B: Score One for the Away Team*, WASH. POST, Innovations, July 11, 2013 ("It appears that the future of warfare will revolve around unmanned drones that are armed to the teeth, fighting wars for their human masters based on computerized algorithms . . . . On July 10, the U.S. Navy landed the X-47B, the prototype for an unmanned, computer-controlled fighter plane, on the deck of a U.S. aircraft carrier located off the coast of Virginia . . . this was truly a robotic aircraft able to fly and land on its own, based on some earlier computerized algorithms.").

54. David Robertson, *Defence Groups Eye Drones for War and Policing*, TIMES, Aug. 17, 2009, at 39; *see also* UNMANNED SYSTEMS ROADMAP 2007-2032, *supra* note 17, at 19 (noting that "higher survivability, increased endurance, and the achievement of higher G-forces, as well as smaller sizes and thus signatures . . . are all made possible by removing the human from the aircraft").

and hypersonic technology, ultimately leading to "a revolution in the roles of humans in air warfare."[55]

Weapons designers continue to make great strides towards realizing a brave new future in which weapon systems will autonomously operate in complex air, ground, and maritime environments, and military planning documents stress the incredible transformational capabilities of such weapons.[56] The ambitious plans for future autonomous combatant systems found in many official U.S. Air Force documents do not, however, engage in any detailed discussion of the ethical or legal challenges that these systems present. Instead, the pursuit of greater autonomous combatant capabilities continues at a rapid pace, conditioned only in general terms on receiving the approval of lawyers and policymakers for the deployment of more advanced systems at some point in the future. In this regard, the U.S. Air Force notes that "[a]ssuming legal and policy decisions allow, technological advances in artificial intelligence will enable UAS to make and execute complex decisions required in this long-term phase of autonomy."[57]

Based on the ease with which many existing lethal and autonomous systems have been deployed, some weapons designers might anticipate few difficulties from legal reviews of future systems.[58] Yet the legal status of existing, emerging, and future autonomous military systems is complex and deserves careful examination.

III. THE SEARCH FOR A FRAMEWORK TO GOVERN AUTONOMOUS WEAPON SYSTEMS

No international conventions directly regulate autonomous weapons, other than agreements banning the precursors of modern advanced systems such as anti-personnel landmines and free-floating naval mines.[59] While specific convention-based prohibitions may be lacking, autonomous weapons must, like all weapons, be used in

55. UAS FLIGHT PLAN, *supra* note 37, at 50.

56. *Id.* at 16, 34 (noting that the U.S. Air Force further envisions manned and unmanned aircraft creating "a focused, relentless, and scaled attack").

57. *Id.* at 50.

58. Weiner, *supra* note 2, at A1 (quoting Dr. Gordon Johnson, a leader of "robotics efforts at the Joint Forces Command research center," as saying that "the lawyers tell me there are no prohibitions against robots making life-or-death decisions").

59. *See* Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction, Sept. 18, 1997, 2056 U.N.T.S. 211 [hereinafter the Anti-Personnel Land Mine Treaty]; Hague Convention (VIII) Relative to the Laying of Automatic Submarine Contact Mines, Oct. 18, 1907, 36 U.S.T. 541.

compliance with applicable customary international law as reflected in the IHL framework. States and individuals can thus be held responsible for violations of IHL obligations involving the use of any weapon, depending on the facts of a particular case. Another possibility is that certain weapons may be classified as illegitimate under the IHL framework. In the latter case, states are prohibited from employing such weapons under any circumstances because they are illegal *per se.*

Consistent with customary international law obligations, any state that studies, develops, acquires, or adopts a new weapon, means, or method of warfare is first required under Additional Protocol I to the Geneva Conventions of 1949 "to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party."[60]

Although the United States is not a party to Additional Protocol I, regulations applicable to each U.S. military department require new weapons to receive a legal review "to determine whether the weapons or weapon systems or their intended use in combat are consistent with the obligations assumed by the United States Government under all applicable treaties and with customary international law."[61] One possible result of such a review is a determination that a weapon cannot be adopted because its intrinsically unlawful nature renders it illegal *per se.*

Devices of a mechanical nature that are designed to automatically kill or seriously maim human beings have been relegated to a *per se* illegal status in many domestic legal systems. The laws of several states in the United States impose criminal and civil liability on persons setting traps and other devices with the intention of harming intruders or trespassers on their property.[62] Central to the reasoning behind such laws is the requirement that deadly force can only be used in

60. Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, art. 36, *opened for signature* Dec. 12, 1977, 1125 U.N.T.S. 3, *reprinted in* 16 I.L.M. 1391 (1977) [hereinafter "Additional Protocol I"].

61. Army Regulation 27-53, *Review of Legality of Weapons Under International Law* (1 January 1979), ¶ 4.e.(1) (assigning this responsibility in the U.S. Army to The Judge Advocate General). Each U.S. military department is required to promulgate such regulations pursuant to DoD Directive 5000.1, sec. E.1.15, which provides that "[t]he acquisition and procurement of DoD weapons and weapon systems shall be consistent with all applicable domestic law and treaties and international agreements . . . customary international law, and the law of armed conflict." *Id.*

62. *See generally* J.D. Perovich, *Use of Set Gun, Trap, or Similar Device on Defendant's Own Property,* 47 A.L.R.3d 646 (1973).

certain narrowly-defined circumstances, and that primitive mechanical devices act without regard to such requirements. Most domestic legal systems also restrict the use of lethal force by law enforcement personnel to only certain situations, while international human rights law also protects the right to life by placing limits on the use of lethal force by police officers.[63]

However, the specialized subset of public international law that governs armed conflict contains no broad prohibitions on the types of force that may be used by military entities. Highly destructive weapons intended to neutralize enemy combatants are not inherently illegal because they succeed by killing those combatants.[64] Instead, violations of the IHL framework related to weapons are usually based on their indiscriminate or otherwise unlawful use in a particular military engagement, not on their inherently unlawful nature. Unfortunately, there is a tendency among some critics of various weapons, including critics of autonomous weapon systems, to conflate or confuse the distinction between IHL prohibitions against inherently unlawful weapons and those prohibitions applicable to the unlawful use of otherwise lawful weapons.[65]

There is only a limited, narrow basis under international law to declare that a weapon itself is illegitimate. Rather than representing broad prohibitions on types of weapons, the criteria under IHL framework for determining whether a weapon itself is *per se* illegitimate focus on whether a weapon is specifically prohibited by an international convention or if it is intrinsically unable to comply with

---

63. *See, e.g.*, McCann v. United Kingdom, 21 Eur. Ct. H.R. 97 (1995) (requiring security forces responding to a suspected terrorist attack to minimize the threat to life); Eighth United Nations Congress on the Prevention of Crime and the Treatment of Offenders, Havana, Aug. 27-Sept. 7, 1990, *Basic Principles for the Use of Force and Firearms by Law Enforcement Officials*, ¶ 9, U.N. Doc. A/CONF.144/28/Rev.1 at 112 (requiring that "[l]aw enforcement officials shall not use firearms against persons except in self-defense or defense of others against the imminent threat of death or serious injury . . . . intentional lethal use of firearms may only be made when strictly unavoidable in order to protect life").

64. As succinctly stated by the U.S. Navy, "[a]ntipersonnel weapons are designed to kill or disable enemy combatants and are lawful notwithstanding the death, pain, and suffering they inflict." U.S. DEP'T OF THE NAVY, COMMANDER'S HANDBOOK ON THE LAW OF NAVAL OPERATIONS § 9.1.1 (2007) [hereinafter NAVAL COMMANDER'S HANDBOOK].

65. Michael N. Schmitt, *Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics*, HARV. NAT'L SEC. J. FEATURES (2013), *available at* http://harvardnsj.org/wp-content/uploads/2013/02/Schmitt-Autonomous-Weapon-Systems-and-IHL-Final.pdf (arguing that a recent report by Human Rights Watch entitled "Losing Humanity: The Case against Killer Robots," blurs this distinction).

the two "cardinal principles contained in the texts constituting the fabric of humanitarian law."[66]

The first is the principle of distinction (also referred to as discrimination) between combatants and civilians. As noted by the International Court of Justice (ICJ), "[s]tates must never make civilians the object of attack and must consequently never use weapons that are *incapable* of distinguishing between civilian and military targets."[67] Thus, a weapon may be used in such a way that it results in unintended harm to civilians, but this does not make it illegitimate under the IHL framework. The focus instead remains on whether a weapon is designed in such a way that it is incapable of complying with the principle of distinction. The ability of autonomous weapons to comply with the principle of distinction is assessed in detail in Part V below, along with the related principle of proportionality (which requires that, even if attacks are directed against legitimate military objectives, they must not cause injury or damage to civilian objects that are expected to be in excess of the concrete and direct military advantage to be gained).[68]

According to the second cardinal principle of the IHL framework, the prohibition against causing unnecessary suffering to combatants, it is "prohibited to use weapons causing them such harm or uselessly aggravating their suffering."[69] As a rule of customary international law, this fundamental prohibition traces its origins to the earliest and most important international conventions governing armed conflict. The modern restatement of this prohibition, found in Additional Protocol I, provides: "It is prohibited to employ weapons, projectiles and material and methods of warfare of a nature to cause superfluous injury or unnecessary suffering."[70]

However, suffering is "unnecessary" and an injury is "superfluous"

---

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

67. *Id.* (emphasis added). Similar statements are found in the military manuals of many countries. *See, e.g.*, NAVAL COMMANDER'S HANDBOOK, *supra* note 64 ("[W]eapons, which by their nature are incapable of being directed specifically against military objectives, and therefore that put civilians and noncombatants at equivalent risk, are forbidden due to their indiscriminate effect.").

68. Additional Protocol I, *supra* note 60, art. 51.5.b (defining attacks causing such excessive damage to civilian objects as a type of prohibited indiscriminate attack); *supra* art. 57.2.(a)(iii) (requiring those who plan or decide upon an attack to refrain from launching attacks which are expected to cause such excessive damage to civilian objects).

69. Nuclear Advisory Opinion, *supra* note 66, ¶ 78.

70. Additional Protocol I, *supra* note 60, art. 35(2).

only in a particular legal sense. While the ICJ has noted that the IHL framework has long prohibited certain types of weapons "because of the unnecessary suffering caused to combatants," it has defined such suffering as "harm greater than that avoidable to achieve legitimate military objectives."[71] For this reason, a weapon that is able to achieve such objectives in this manner is not banned for causing unnecessary suffering or superfluous injury, even if its effects are horrific.[72] Few weapon systems actually qualify as being inherently unable to achieve legitimate military objectives without causing, in every situation, unnecessary suffering or superfluous injury.[73] Autonomous weapons are not, based solely on their autonomous character, well suited for such a classification. The lethality of weapon systems operated by autonomous means does not in fact differ substantially from that of other weapon systems: their unique features do not involve abhorrent weapons themselves, but instead relate only to their autonomous control.[74]

Apart from the two cardinal principles of the IHL framework, various humanity- and morality-based arguments have been advanced as potential grounds for declaring weapons to be *per se* illegal. One seemingly attractive and broad basis for arguing that a weapon is intrinsically unlawful is that it violates "principles of humanity" or the "dictates of public conscience." While there is no consensus about the legality of giving autonomous machines the power to deliver lethal force, it has been suggested that many people find the idea "shocking and unacceptable," and that "states should take their perspective into account when determining the dictates of public conscience."[75] "Principles of humanity" and the "dictates of the public conscience," which were first set forth in the celebrated "Martens Clause" in the Hague Convention II of 1899, have undoubtedly been important in shaping the IHL

71. Nuclear Advisory Opinion, *supra* note 66, ¶ 78.

72. YORAM DINSTEIN, THE CONDUCT OF HOSTILITIES UNDER THE LAW OF INTERNATIONAL ARMED CONFLICT 59 (2004).

73. Unfortunately, as demonstrated by the inconclusive advisory opinion rendered by the ICJ on the legality of nuclear weapons, it can be difficult to establish that even an incredibly destructive, arguably indiscriminate weapon is unlawful *per se*. *See* Nuclear Advisory Opinion, *supra* note 66, ¶ 97 ("... the Court is led to observe that it cannot reach a definitive conclusion as to the legality or illegality of the use of nuclear weapons by a State in an extreme circumstance of self-defence, in which its very survival would be at stake.").

74. Jeffrey S. Thurnher, *No One at the Controls: Legal Implications of Fully Autonomous Targeting*, 67 JOINT FORCE Q. 77, 82 (Oct. 2012), *available at* http://www.ndu.edu/press/lib/pdf/jfq-67/JFQ-67_77-84_Thurnher.pdf.

75. LOSING HUMANITY, *supra* note 5, at 35.

framework itself.[76] They are not widely regarded, however, as constituting an independent basis or criterion for determining the legality of weapons.[77]

There are various morality-based arguments based loosely on concepts of fairness and sentiments similar to "principles of humanity" and the "dictates of public conscience" against the legality of new weapon systems. For example, it has been suggested that the very use of highly advanced weapons (particularly unmanned weapon systems) against other less capable adversaries is itself unfair or unjust.[78] Such arguments, however, find no basis under the IHL framework. As noted by Harold Koh, the former U.S. State Department Legal Advisor, "the rules that govern targeting do not turn on the type of weapon system used, and there is no prohibition under the laws of war on the use of technologically advanced weapon systems in armed conflict . . . so long as they are employed in conformity with applicable laws of war."[79]

Other morality-based arguments made by critics against the legality of autonomous weapons, although not based on any explicit prohibition found in the IHL framework, are based on the replacement of humans by machines. Such arguments posit that "[i]t is simply wrong per se to take the human moral agent entirely out of the firing loop" and that a machine capable of inflicting harm cannot replace a human since only a human is "possessed of a conscience and the faculty of moral judgment."[80]

Arguments about the morality of replacing men with machines are of course difficult to comprehensively address because they ultimately depend on the moral views of the individual evaluating the problem. However, many aspects of human involvement in autonomous military weapon systems, such as legal requirements linked to the exercise of human judgment or required levels of human control over specific targeting actions, involve issues that are distinct from moral concerns.

---

76. Hague Convention with Respect to the Laws and Customs of War on Land, Preamble, ¶ 8, July 29, 1899, 32 Stat. 1803.

77. DINSTEIN, *supra* note 72, at 56-57. *But see 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 INT'L REV. RED CROSS 931, 945 (2006) ("A weapon which is not covered by existing rules of international humanitarian law would be considered contrary to the Martens clause if it is determined *per se* to contravene the principles of humanity or the dictates of public conscience.").

78. Harold Hongju Koh, Legal Adviser, U.S. Dep't of State, The Obama Administration and International Law, Remarks at the Annual Meeting of the American Society of International Law (Wash., D.C., Mar. 25, 2010), *available at* http://www.state.gov/s/l/releases/remarks/139119.htm.

79. *Id.*

80. Anderson & Waxman, *supra* note 24, at 15.

To argue that any "decision" by a machine to take a lethal action is always morally objectionable and necessarily *mala in se* is problematic. As noted above in Part II, machines in both the military and civilian sectors are incrementally taking on more and more functions that were formerly assigned to humans. In the military sphere, it has long been impossible (certainly since the advent of mines and booby-traps) to argue that only humans are permitted to "pull the trigger" or otherwise initiate a lethal action. The key issue remains whether or not a particular weapon system can be operated in compliance with IHL rules and obligations, not the presence or absence of a human moral agent. Nonetheless, a spirited debate continues between critics who suggest that only a human moral agent can be entrusted with weighty life and death decisions and proponents of artificial intelligence who suggest that machines may be better suited than humans in many situations to perform a variety of stressful and dangerous combat actions.[81]

Moving beyond purely moral questions about replacing humans with machines, other critics question whether programming and various forms of artificial intelligence will ever work well enough to field a weapon that can comply with IHL obligations, particularly when "decisions about what is proportionate often require fine distinctions and sophisticated judgment."[82] Such arguments flow not so much from the moral deficiencies of machines, but rather from the perceived limits of technological evolution. It is likely, however, that it will not be technology that limits the deployment of future autonomous weapons, but rather policy (and by implication, legal policy).[83]

Those who take a more optimistic vision of autonomous weapon systems suggest that critics who dismiss in advance the results of any future technological evolution (often in favor a complete ban on autonomous lethal weapons) would rule out many possible positive technological outcomes. Such outcomes could include the elimination of human failings from targeting decisions (such as fear, panic, vengeance, and other emotions), improved sensory capabilities that re-

81. *Id.* at 14 (noting arguments based on "beliefs about the special nature of human beings and their emotional and affective abilities on the battlefield that no machine could ever exhibit, even over the course of technological evolution").

82. *Unmanned Aerial Warfare: Flight of the Drones*, THE ECONOMIST, Oct. 8, 2011, *available at* http://www.economist.com/node/21531433.

83. *Id.* (quoting David Deptula, a retired general in charge of the U.S. Air Force's intelligence and surveillance operations, who said that "[t]echnologically, we can take [autonomy] pretty far, but it won't be technology that is the limiting factor, it will be policy").

duce the misidentification of targets, reduction of pressures that soldiers face in pursuit of their self-preservation, and of course the substitution of disposable robots for human lives.[84] Advocates of autonomous weapons further argue that it may be possible in the future to build a comprehensive architecture in autonomous weapons to serve as a sort of compliance mechanism—which might be described as an "ethical governor"—to ensure that they will operate in accordance with both IHL restrictions and the rules of engagement.[85]

While many of the possible future benefits or positive outcomes of autonomous weapons deserve careful evaluation, questions remain about whether such weapon systems (no matter how advanced they are) could replace humans for the purposes of meeting key legal standards in various combat situations. These questions, in their most problematic contexts, are assessed in detail in Parts IV and V below.

A final, philosophical line of reasoning for banning fully autonomous weapons is founded on both consequentialist and deontological arguments. This line of reasoning posits that if no one can be justly held responsible for the actions of such weapons, then their use would be unethical in war.[86] Some commentators have expanded on this concept to argue that, "since there is no fair and effective way to assign legal responsibility for unlawful acts committed by fully autonomous weapons," giving such weapons complete control over targeting decisions "would undermine yet another tool for promoting civilian protection."[87] It is argued, however, that there is no requirement under the IHL framework that "a human be held personally accountable for any mistakes or violations that may occur on the battlefield."[88] Nonetheless, the elusive search for individual accountability for IHL violations (as set forth in Part IV below) illuminates important issues that, together with a careful assessment of state responsibility for compliance with the principles of distinction and proportionality (as set forth in Part V below), help define the legally permissible limits of the most advanced autonomous weapons.

---

84. Anderson & Waxman, *supra* note 24, at 15.

85. RONALD C. ARKIN, GOVERNING LETHAL BEHAVIOR: EMBEDDING ETHICS IN A HYBRID DELIBERATIVE/REACTIVE ROBOT ARCHITECTURE, TECHNICAL REPORT GIT-GVU-07-11 20 (2007), *available at* http://www.cc.gatech.edu/ai/robot-lab/onlinepublicationsonlinepublicationsonline-publications onlinepublications/formalizationv35.pdf; RONALD C. ARKIN, GOVERNING LETHAL BEHAVIOR IN AUTONOMOUS ROBOTS 69 (2009).

86. Robert Sparrow, *Killer Robots*, 24 J. APPLIED PHIL. 62, 67 (2007).

87. LOSING HUMANITY, *supra* note 5, at 42.

88. Thurnher, *supra* note 74, at 82.

IV. THE SEARCH FOR INDIVIDUAL RESPONSIBILITY FOR THE ACTIONS OF AUTONOMOUS WEAPONS

A. *The International Criminal Law Framework*

In his summation for the prosecution at the Trial of the Major War Criminals at Nuremburg in 1946, Justice Robert Jackson mocked many of the defendants for "taking the position in effect that while there have been enormous crimes there are no criminals."[89] While it faced several challenges in establishing the individual criminal responsibility of the defendants before them, the Nuremburg Tribunal forcefully affirmed the central role that such responsibility must play in the enforcement of international law: "Crimes against international law are committed by men, not by abstract entities, and only by punishing individuals who commit such crimes can the provisions of international law be enforced."[90]

The principle of individual responsibility and punishment for crimes under international law is the enduring legacy of Nuremberg and has been described as the "cornerstone of international criminal law."[91] Contemporary international criminal tribunals and courts, such as the International Criminal Tribunal for the former Yugoslavia (ICTY), the International Criminal Tribunal for Rwanda (ICTR), the Special Court for Sierra Leone, and the International Criminal Court (ICC), all demonstrate the continuing importance of establishing individual criminal responsibility for war crimes and crimes against humanity.

The intertwined and sometimes competing purposes or aims of criminal law and penal sanctions in domestic legal systems are numerous, including retribution, condemnation, correction, special public control over persons disposed to commit crimes, rehabilitation, deterrence, promotion of respect for the law's authority, providing for an orderly society, and generally protecting citizens and public interests against harm.[92] These diverse goals can translate into varied concepts of culpability in domestic legal systems.

International criminal law operates in a different context, often with

---

89. 19 TRIAL OF THE MAJOR WAR CRIMINALS BEFORE THE INT'L MILITARY TRIBUNALS, NUREMBERG 419 (1948).

90. 1 TRIAL OF THE MAJOR WAR CRIMINALS BEFORE THE INT'L MILITARY TRIBUNALS, NUREMBERG 223 (1947).

91. Prosecutor v. Tadic, Case No. IT-94-1-T, Judgment, ¶¶ 664-66 (Int'l Crim. Trib. for the Former Yugoslavia May 7, 1997).

92. *See generally* Henry M. Hart, Jr., *The Aims of the Criminal Law*, 23 L. & CONTEMP. PROBS. 401 (1958).

divergent or narrower goals. While domestic criminal legal systems seek to address disparate social concerns in an effort to promote more productive, orderly, and safe societies, international criminal law operates in chaotic conditions of armed conflict or widespread violence and may directly relate to larger international security concerns. For example, the Rome Statute of the ICC recognizes that "grave crimes threaten the peace, security and well-being of the world" and that "the most serious crimes of concern to the international community as a whole must not go unpunished."[93]

Although the punishment of key offenders can promote justice in the international community and also serve international peace and security, the authority and legitimacy of international criminal law depends on its conformance to stringent requirements for the determination of legal—not moral—culpability. One fundamental requirement is personal responsibility, since "guilt by mere association is unjust."[94] While it has been suggested that an important requirement of *jus in bello* is ensuring that "someone can be held responsible for each death caused in war," establishing the identity of that person is often difficult.[95] This challenge becomes even more complicated with the advent of autonomous weapon systems.

Who then can be held accountable for unnecessary, excessive, or otherwise unjustified civilian deaths caused by autonomous weapons? Before turning to an assessment of specific, potentially liable entities and individuals below, two sets of overarching problems in evaluating international criminal law liability in the context of autonomous weapons must be noted. The first concerns the increasingly complex relationship between criminal law and modern technology generally, and the second concerns the framework of international criminal law. This first set of problems is not limited to military systems. As consumers in civilian society move from "using" technologies to "interacting" with new, ambient, and smart technologies, one scholar notes that "use [of]

93. Rome Statute of the International Criminal Court, preamble, *opened for signature* on July 17, 1998, 2187 U.N.T.S. 90 (entered into force July 1, 2002) [hereinafter Rome Statute of the ICC].

94. LUBAN, O'SULLIVAN, & STEWART, INTERNATIONAL AND TRANSNATIONAL CRIMINAL LAW 15 (2010); *see also* Uphaus v. Wyman, 360 U.S. 72, 79 (1959) (observing that guilt by association is a "thoroughly discredited doctrine").

95. Sparrow, *supra* note 86, at 67. *But see* Anderson & Waxman, *supra* note 24, at 17 ("But post-hoc judicial accountability in war is just one of many mechanisms for promoting and enforcing compliance with the laws of war . . . .").

context-specific rules to control the 'misuse' of technology continues to erode and ultimately vanishes."[96]

Problems of attribution and responsibility for the misuse of technology become even more acute in the case of machines that on some level can be described as autonomous. Furthermore, assessments of criminal liability now take place in a digital age in which the accelerating pace of technological developments continue to illustrate how the law often tends to lag—sometimes dramatically—behind those developments.[97] It is thus not surprising that lawbreakers (and those who see opportunities in obsolete legal structures) are among the first to make use of new technologies.[98]

A second set of problems looming over individual criminal responsibility in this area springs from the nature of international criminal law itself, which differs in several key respects from the criminal law of many domestic legal systems. These differences mean that it may be particularly difficult to find any theory of international criminal liability which will adequately hold human operators accountable for the actions of advanced autonomous weapons in the future.

One fundamental issue is the mental element necessary for establishing war crimes and the inability of negligence alone to meet this requirement. For example, the required *mens rea* for the offense of attacks on civilians requires that a perpetrator acted "willfully."[99] As noted by the ICTY, this means that the accused "must have acted consciously and with intent."[100] While the ICTY has held that the "the notion of 'willfully' incorporates the concept of recklessness," it excludes "mere negligence."[101] Different states of mental culpability may be required for different modes of participation for various offenses, but negligence alone is generally insufficient. Negligent conduct in the performance of duties is left by default to generally be addressed by domestic laws and applicable military regulations.

---

96. BRENNER, LAW IN AN ERA OF SMART TECHNOLOGY 149 (2007).

97. WALLACH & ALLEN, *supra* note 14, at 197.

98. BRENNER, *supra* note 96, at 13.

99. *See* Additional Protocol I, *supra* note 60, art. 85 (defining the act of making civilians the object of an attack as a "grave breach" when "committed willfully"); Prosecutor v. Galic, Case No. IT-98-29-T, Trial Judgment, ¶ 57 (Int'l Crim. Trib. for the Former Yugoslavia Dec. 5, 2003) (noting it is well established that "indiscriminate attacks, that is to say, attacks which strike civilians or civilian objects and military objectives without distinction, may qualify as direct attacks against civilians").

100. *Id.* ¶ 54.

101. *Id.* ("The perpetrator who recklessly attacks civilians acts 'wilfully.'").

Thus, the only argument for imposing strict or objective liability on persons responsible for the employment of autonomous weapons hinges on the classification of those weapons as illegal *per se*. As noted above, states are prohibited from using certain weapons under any circumstances. Consistent with this prohibition, the employment of illegitimate weapons can also give rise to individual criminal responsibilities. For example, the ICTY Statute specifically empowers the Tribunal to prosecute persons for violations of the law of war for the "employment of poisonous weapons or other weapons calculated to cause unnecessary suffering."[102]

Even more broadly, the jurisdiction of the ICC extends over war crimes related to the employment of "weapons, projectiles and material and methods of warfare which are of a nature to cause superfluous injury or unnecessary suffering or which are inherently indiscriminate in violation of the international law of armed conflict."[103] Unfortunately, the difficulty in determining which weapons are to be associated with this strict liability is demonstrated by the continuing failure of the States Parties to add a required annex to the Rome Statute that would definitively list those "weapons, projectiles and material and methods of warfare that are the subject of this comprehensive prohibition."[104]

Even when states are able to agree that certain weapons should be banned and enter into international conventions that clearly identify the subject matter of such prohibitions, the criminalization of specific actions related to the development, use, or transfer of these prohibited weapons may not be automatically required. In some cases, the imposition of related penal sanctions (as deemed appropriate by the state party) is required.[105] In other cases, agreements may not address this

---

102. Statute of the International Tribunal for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law Committed in the Territory of the Former Yugoslavia since 1991, art. 3(a), adopted by Security Council on May 25, 1993, U.N. Doc. S/RES/827 (May 25, 1993) [hereinafter ICTY Statute].

103. Rome Statute of the ICC, *supra* note 93, art. 8.2(b)xx. The statute also criminalizes the use of several specific weapons within this prohibition, including poisoned weapons, asphyxiating, poisonous or other gases, and bullets which expand or flatten easily in the human body. *Id.* arts. 8.2.(b)xvii-8.2(b)xix.

104. *Id.* art. 8.2.(b)xx (further providing that the required annex must be made by an amendment made in accordance with the relevant provisions set forth in articles 121 and 123).

105. *See, e.g.*, Anti-Personnel Land Mine Treaty, *supra* note 59, art. 9 (providing that "[e]ach State Party shall take all appropriate legal, administrative and other measures, including the imposition of penal sanctions, to prevent and suppress any activity prohibited to a State Party under this Convention undertaken by persons or on territory under its jurisdiction or control").

issue at all or may provide only a vague "necessary measures" requirement for implementing the agreement—measures that may fall short of providing an effective legal framework for holding individuals criminally responsible for violations of the terms of the agreement.[106] In any event, no existing international convention bans or regulates autonomous weapons, much less establishes an effective framework for imposing related criminal sanctions.

With no applicable weapons-related international agreement to rely upon, criminal penalties for war crimes associated with autonomous weapons remain dependent on the application of IHL obligations generally. As a first step, determining the liability of entities or individuals under international criminal law for such offenses requires navigation of its approaches to culpable modes of participation.

The statutes of the ICTY and ICTR (the Ad Hoc Tribunals) reflect approaches commonly found in many national systems by stating that a person "who planned, instigated, ordered, committed or otherwise aided and abetted in the planning, preparation or execution of a crime [within the jurisdiction of the tribunal] . . . shall be individually responsible for the crime."[107] As demonstrated by these statutes, accomplices as well as those who commit crimes are responsible for their conduct under international criminal law.[108]

With respect to group criminality, U.S. prosecutors often turn to conspiracy as the appropriate means of describing an individual's level of participation in a crime. The crime of conspiracy is not, however, recognized in many national legal systems and is not found in the statutes of Ad Hoc Tribunals, the ICC, or international criminal tribunals generally. While the Ad Hoc Tribunals have crafted a somewhat similar concept called "Joint Criminal Enterprise" (JCE) liability,

---

106. *See, e.g.*, Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on Their Destruction, art. IV, Apr. 10, 1972, 26 U.S.T.S. 583, 1015 U.N.T.S. 163 (providing that "[e]ach State Party to this Convention shall, in accordance with its constitutional processes, take any necessary measures to prohibit and prevent the development, production, stockpiling, acquisition, or retention of the agents, toxins, weapons, equipment and means of delivery specified in article I of the Convention, within the territory of such State, under its jurisdiction or under its control anywhere").

107. ICTY Statute, *supra* note 102, art. 7.1.; Statute of the International Criminal Tribunal for Rwanda, art. 6.1, adopted by U.N. Security Council on May 25, 1993, U.N. Doc. S/RES/955 (May 25, 1993) [hereinafter ICTR Statute].

108. *See also* Rome Statute of the ICC, *supra* note 93, art. 25.3(c) (imposing criminal responsibility on one who "aids, abets or otherwise assists" in the commission of genocide, war crimes or crimes against humanity).

this approach has not been without controversy.[109]

Along similar lines, the Rome Statute for the ICC criminalizes contributing to the commission of a crime "by a group of persons acting with a common purpose" if that contribution is intentional, "made with the aim of furthering the criminal activity or criminal purpose of the group," or "made in the knowledge of the intention of the group to commit the crime."[110] In the international criminal context, however, group criminality remains a limited concept with a relatively narrow scope of application.

B. *Autonomous Weapon Systems and the Search for Criminal Culpability*

1. Manufacturers, Designers, Engineers, and Programmers

It has been said that if a perpetrator of a crime cannot be prosecuted because it is a machine, such a possibility offends not only the rule of law but also "the more visceral human desire to find an individual accountable."[111] The desire to hold individuals (and companies) accountable for the failure of machines finds expression in many domestic legal systems in the form of civil lawsuits. Consumers in some countries, particularly the United States, enjoy the benefit of national product liability and safety laws that allow them to bring civil lawsuits against corporations for harm caused by articles manufactured or sold. These lawsuits are usually based on various types of negligence, including manufacturing and design defects, failure to take proper care or avoid foreseeable risks, failure to warn or provide reasonable instructions, etc.[112] Some commentators have also suggested similar civil lawsuits as an option for incentivizing autonomous weapons manufacturers to produce safer weapons.[113] However, this approach presents serious problems.

First, private weapons manufacturers are not generally punished for how individuals or governments use their weapons, particularly if the manufacturers are careful in disclosing (up front) to military purchasers any risks of malfunctions.[114] It is also highly unlikely that any

---

109. *See generally* David L. Nersessian, *Whoops, I Committed Genocide! The Anomaly of Constructive Liability for Serious International Crimes*, 30 FLETCHER F. WORLD AFF. 81 (2006).

110. Rome Statute of the ICC, *supra* note 93, art. 25.3.(d).

111. Darren Stewart, *New Technology and the Law of Armed Conflict: Technological Meteorites and Legal Dinosaurs?*, 87 INT'L L. STUD. 271, 275, (2011).

112. *See generally* AMERICAN LAW INSTITUTE, RESTATEMENT OF THE LAW THIRD, TORTS: PRODUCTS LIABILITY (1998).

113. LOSING HUMANITY, *supra* note 5, at 44.

114. *Id.*

company would produce and sell weapons, which are inherently dangerous, knowing the firm could be held strictly liable for any use that violates international humanitarian law. Second, product liability involves civil suits, which require victims to pursue appropriate legal actions. It is unrealistic to expect, however, that the disadvantaged, impoverished, and geographically displaced civilian victims of war will be in a position to sue for relief in a foreign court against a manufacturer, even if there are legal rules in place that would allow them to recover damages.[115]

Even larger issues loom over any attempt to impose international criminal liability on companies, as opposed to civil liability. Corporations are not generally subject to prosecution before international criminal courts, and international law rarely imposes criminal responsibility on corporations.[116] While civil lawsuits against corporations for product-related injuries are common in many national justice systems and some states also permit criminal proceedings in their courts against corporations themselves, many states object to the prosecution of legal entities. The philosophical differences behind these objections are reflected in the ICTY, ICTR, and ICC Statutes, all of which provide jurisdiction over natural persons only.

International law can, however, impose individual criminal responsibility on corporate executives, as demonstrated by the post-WWII trials of Nazi industrialists.[117] Most of these industrialists were tried as principals for their role in the planning and execution of the Nazi slave labor program and the plunder of private and public property.[118] However, the top officials of the firm that manufactured and sold Zyklon B for use in the Nazi gas chambers were convicted by a British military court as accessories to war crimes.[119] The Nuremberg Charter had recognized this possibility by imposing individual responsibility on

---

115. *Id.*

116. Doug Cassel, *Corporate Aiding and Abetting of Human Rights Violations: Confusion in the Courts*, 6 NW. J. INT'L HUM. RTS. 304, 306 (2008).

117. *See* Matthew Lippmann, *War Crimes Trials of German Industrialists: The Other Schindlers*, 9 TEMP. INT'L & COMP. L.J. 173, 267 (1995) (observing that the trials of these industrialists "stand as the most prominent prosecutions of corporate executives for international crimes" and that the resulting judgments "affirmed that private individuals possess obligations which transcend the dictates of domestic doctrine").

118. *Id.* at 183-85, 214-21, 239-49 (discussing how the industrialists were convicted of various spoliation, plunder and slave labor charges but that charges that they prepared, planned, initiated, waged, and conspired to engage in Crimes against Peace were dismissed).

119. *See* Trial of Bruno Tesch and Two Others (The Zyklon B Case), 1 Law Reports of Trials of War Crim. 93 (Brit. Mil. Ct., Hamburg Mar. 1-8, 1946).

"accomplices participating in the formulation or execution of a common plan or conspiracy to commit" a crime enumerated within the Charter.[120]

The U.S. Supreme Court famously summarized the fundamental *actus reas* and *mens rea* elements of American criminal law by noting that crime is a "compound concept . . . generally constituted only from concurrence of an evil-meaning mind with an evil-doing hand."[121] The "evil-doing hand" or *actus reus* requirement of international criminal law is uncontroversial: as stated by the ICTY, it consists of rendering "practical assistance, encouragement, or moral support which has a substantial effect on the perpetration of the crime."[122]

The *mens rea* requirement under international criminal law for the offense of aiding and abetting is less clear. The ICTY adopted a knowledge test: "The *mens rea* required is the knowledge that these acts assist in the commission of the offence."[123] On the other hand, a purpose test was adopted in the Rome Statute of the ICC for most cases of aiding and abetting, criminalizing the conduct of one who, "[f]or the purpose of facilitating the commission of such a crime, aids, abets, or otherwise assists in its commission or its attempted commission, including providing the means for its commission."[124] In spite of this, a strong argument can be made that customary international law as reflected in most of the post-World War II cases, the case law of the both the ICTY and the ICTR, and the International Law Commission's Draft Code on Crimes and Offences Against Mankind recognizes that the required *mens rea* for those who aid and abet is having the *knowledge* that they are assisting criminal conduct.[125]

Whether the *mens rea* requirement for the international crime of aiding and abetting is purpose or knowledge, it can be difficult to apply it to officers of corporations that develop armaments and weapon systems—at least when the products they manufactured are not illegal *per se* or when they can be as unambiguously and knowingly linked to crimes against humanity and genocide as Zyklon B was connected with

120. Nuremberg Charter, Agreement for the Prosecution and Punishment of the Major War Criminals of the European Axis, and Charter of the International Military Tribunal, Charter II, art. 6, *entered into force* Aug. 8, 1945, 82 U.N.T.S. 280.

121. Morissette v. United States, 342 U.S. 246, 251 (1952).

122. Prosecutor v. Furundzija, Case No. IT-95-17/1-T, Judgment, ¶ 249 (Int'l Crim. Trib. for the Former Yugoslavia Dec. 10, 1998).

123. *Id.* ¶ 274.

124. Rome Statute of the ICC, *supra* note 93, art. 25 (3)(c).

125. Cassel, *supra* note 116, at 314.

mass murder in Nazi Germany. While it is undeniable that weapons may be used in violation of the IHL framework by military personnel, this does not necessarily mean that the developers of those weapons incur any related criminal liability. Even weapons that are capable of causing horrific injuries if used against civilians may nonetheless have legitimate applications against certain military objectives. For example, the use of incendiary weapons against civilians, such as napalm, is prohibited under both customary and convention-based international law, yet they remain highly effective for their designed purpose of attacking hardened military targets (located outside concentrations of civilians).[126]

Questions about the way a particular weapon will be used and the environment into which it will be introduced present significant obstacles to determining the knowledge or purpose of the developers, designers, engineers, and programmers of that weapon, especially in the case of autonomous weapons. The possibility that the most advanced autonomous weapons will make choices that are not "preprogrammed" means that their decision-making may be very difficult to predict. The "Intelligent Decision-making Techniques" (IDTs) of these machines may in fact be so complex that their projected actions are indeterminable in some environments.[127] The most sophisticated autonomous machines may also make reproducible testing difficult, raising serious questions about holding a developer or designer legally responsible for the actions of an IDT that cannot be meaningfully tested.[128] These problems will become even more pronounced if a system is someday designed, as the military anticipates it will, to "learn" from the mistakes it makes.

To the extent that the law links the knowledge or purpose of alleged aiders and abettors to the predictable behavior of the weapons they are accused of developing and putting on the battlefield, autonomous systems present herculean challenges. Even if advanced autonomous weapons are designed and programmed to be directed against only military objectives, the inherent risks and unpredictability of these

126. *See* Protocol on Prohibitions or Restrictions on the Use of Incendiary Weapons (Protocol III), art. 2, Dec. 2 1983, 1342 U.N.T.S. 171 [hereinafter Protocol III] (providing that "[i]t is prohibited in all circumstances to make any military objective located within a concentration of civilians the object of attack by air-delivered incendiary weapons").

127. FINN & SCHEDING, *supra* note 52, at 183.

128. *Id.* (noting that "the more intelligent the decision-making process, the harder it is likely to be to functionally establish or fully test the response of the system to repeatable or verifiable system stimuli").

systems grow with their increasing autonomy. Some unpredictability may even be desired by the military: As autonomous weapons continue to assume important new roles and their operations are observed by enemy forces, military planners are likely to want these systems to be designed to be unpredictable (to at least some degree) in order to deal with predictable counter-measures.[129]

Regardless of the specifics of the design process, autonomy presents its own challenges for establishing individual responsibilities with respect to programmers. The essence of autonomy in machines means that inputted data and programming prior to a particular operation will not necessarily "result in a specific outcome in response to any given set of circumstances" and that, absent establishing the aberrant behavior of command programmers, "it would be almost impossible to attribute the autonomous system's behavior *directly* to a particular human."[130]

Furthermore, the culpability of individual designers, engineers, programmers, or developers of specific components of autonomous weapon systems is problematically linked (for criminal law purposes) to a collaborative and advanced technological achievement. The complex and polymorphic nature of such systems is likely to defy attempts to assign clear responsibility for overall system failures. Decision-making by these machines will probably be distributed across several programs and processors, results will be derived from data input that originates from many different types of sensors, and no single agent may be able to be identified as the "decision-maker" in a complicated system of concurrently interacting human and machine components.[131] These machines will be further characterized by interactions between many different programs that have been developed by multiple programmers (perhaps previously unknown to each other), on different processors and operating systems, possibly stretching across a network of different autonomous platforms.

The environment in which autonomous weapons are placed, the missions that they are assigned, and the context in which they are used may also make it inappropriate to attribute responsibility to designers, engineers, and programmers who cannot limit the potential uses or harms that these weapons cause. The risks inherent in these systems

129. *Id.* at 184.

130. Stewart, *supra* note 111, at 290.

131. FINN & SCHEDING, *supra* note 52, at 183.

may thus place special responsibilities on those that choose to employ them.[132]

Finally, as is the case with all complicity crimes, responsibility as an accomplice for offenses related to the employment of autonomous weapons is derivative in nature—it remains dependent on the commission of a crime by other parties. Turning to the battlefield itself, the search for such criminal conduct by a principal (in most situations, the military officer who orders and oversees the employment of these systems) raises even more challenging questions.

2. Military Personnel and the Employment of Complex Machines

It is not controversial that military personnel must comply with the IHL framework in their employment of any weapons, including autonomous systems. Such systems, however, present considerable challenges to the traditional framework for assessing individual responsibilities under the IHL framework for several reasons.

First, unlike other weapons, they do not have "operators" in the conventional meaning of that term because once these weapons are placed in the autonomous mode, they are no longer under the direct control of a human (although, in some cases, a human may in theory be able to intervene). For the most advanced systems on the higher end of the continuum that autonomously sense, identify, and attack targets, this focuses attention on the individuals (in most cases, the military officers responsible for these weapon systems and supporting personnel) who order or authorize their employment. This is in contrast to the individual operators of less advanced weapon systems who select a particular target and then "pull the trigger," launch a missile, or release a bomb.

Second, assuming that a particular autonomous weapon is not illegal *per se*, the officer who authorizes its use in an attack that causes excessive civilian casualties is not subject to any form of objective or strict liability. Instead, the officer must be judged by the IHL standards applicable to all individuals charged with war crimes: he or she must not only be causally linked to the *actus reus* element of the crime (in this case, targeting civilians or causing excessive civilian casualties) but must also have the necessary mental state of culpability. The *mens rea* of officers in these situations is fundamentally linked to their relationship

132. AUTONOMOUS MILITARY ROBOTICS REPORT, *supra* note 8, at 57 (noting that in light of the risks inherent in the use of robots, "at some point the users may be judged by a court to have knowingly assumed these risks in the very act of choosing to use a robot").

with, and understanding of, the autonomous machines for which they are responsible.

As acknowledged by the DoD, new types of autonomous operations "will be considered as confidence in the reliability of autonomous systems and their components grows."[133] Human-robot teams are currently described by the DoD as presenting "a unique challenge," since technologies supporting unmanned systems must be developed which will "enable the human to predict, collaborate, and develop trust with the unmanned system."[134] At present, such statements appear to counsel great caution in assigning missions to autonomous combatant systems if there is a significant possibility that they could threaten civilians and cannot be "trusted" to attack only military objectives.

Until such time as autonomous combatant systems can be trusted to "pull the trigger" on their own, a related policy holds that human control should be maintained with a "human-in-the-loop." As discussed above, however, this will increasingly not be practical or desirable once many of these systems have been activated and deployed and meaningful human control is diminished.[135] What, then, is a "reasonable person" to do with such a system if he or she is a military officer planning an attack and seeking to ensure compliance with the principles of necessity and proportionality? The answer appears to be that the officer must make a complicated and subjective decision, one that requires a judgment about the capabilities of the system, the circumstances in which it is to be deployed, and the nature and type of operations in which the system can be expected to function appropriately.

Since "virtuous robots" with ethical programs that ensure compliance with IHL obligations are not yet (and may never be) a reality, it appears that the use of autonomous combatants in armed conflicts in the foreseeable future will have to be subject to carefully designed operational constraints based on their capabilities. Some proposed constraints, described as "workaround policy solutions," include: limiting the deployment of lethal robots to a specific "kill box"; designing a robot to target only other machines or weapons; not giving robots a

133. UNMANNED SYSTEMS ROADMAP, 2007-2032, *supra* note 17, at 54.

134. *Id.* at 51.

135. *See* Singer, *supra* note 43, at 63 (noting that "[w]ith each step in the growing lethality and intelligence of robotics, the role of the "man in the loop" of decision making in war has begun to diminish").

self-defense mechanism; or creating robots with only less-than-lethal strike capabilities.[136]

These so-called workaround policy solutions are, of course, only partial solutions to questions about the employment of autonomous systems and are unlikely to be suitable for many combat situations. However, they do reflect the reality that an autonomous military robot will, certainly for the foreseeable future, have at best only a partially virtuous or law-abiding character, limited by specific circumstances, conditions, and missions. It will thus be up to military officers responsible for these systems to ensure that such a robot "*only enters situations in which its character can function appropriately.*"[137] Stated in the terms of the IHL framework, the obligation to "take all feasible precautions in attack" might be said to encompass, at a minimum, an obligation for officers who "plan or decide upon an attack" not to introduce autonomous combatant systems into situations where they threaten noncombatants but cannot be expected to function appropriately.[138] This obligation leaves a military officer who is responsible for these systems with a difficult decision, upon which his or her individual criminal responsibility may depend (and, as discussed below in Part V, upon which state responsibility may depend as well).

An officer who makes the decision to employ an autonomous combatant system in an attack incurs responsibility for that attack but has not committed a war crime unless he or she is found to have had the necessary culpable state of mind in relation to the criminal conduct associated with that attack. To argue that an officer willingly targeted civilians or willingly caused excessive civilian casualties by ordering an attack with such a system requires that he or she had considerable familiarity with the system involved, understood its capabilities, and made the decision to deploy it in an inappropriate situation with a conscious or reckless disregard of its known or apparent limitations under such circumstances.

In an age of increasingly sophisticated machines intruding into every aspect of military operations (as well as civilian life), it is not hard to imagine military officers who lack a comprehensive or even accurate understanding of precisely how some complicated pieces of equipment

---

136. AUTONOMOUS MILITARY ROBOTICS REPORT, *supra* note 8, at 91.

137. *Id.* at 41.

138. All those "who plan or decide upon an attack," are required to "take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss or civilian life, injury to civilians and damage to civilian objects." Additional Protocol I, *supra* note 60, art. 57.2.

work, including a detailed knowledge of the inner workings of their computer laptops, cell phones, and other modern electronic appliances. Instruction manuals for such common items may already defy understanding and certainly do not attempt to explain the underlying science or the manner in which these devices make "decisions."

It may, of course, be difficult even for the specialists who are responsible for designing, engineering, and programming complex autonomous combatant systems to discern their technical limitations in advance of actual battlefield situations. Such technical uncertainties seem likely to complicate any efforts to establish the necessary *mens rea* for holding military officers criminally responsible for the civilian deaths caused by these machines. It thus can be argued that at some point of advanced stages of weapon autonomy and complexity, it may be both legally difficult and morally unfair to hold a military officer responsible for the actions of an autonomous machine—yet the situation will also be quite "unfair" to the potential civilian casualties.[139]

Military officers will still, however, remain responsible for making the key determination of whether a particular autonomous weapon system is suited for a specific combat situation. The circumstances surrounding such a decision do not preclude the possibility that a court could find that apparent dangers or technical limitations were recklessly (and willfully) disregarded by that officer, but such a finding in most cases would probably be quite difficult to reach. While it may be very difficult for the officer in question to see these apparent dangers or limitations based on his or her own technical expertise, the decision to employ these weapons may also be the subject of relevant directives, policies, and guidance received from his or her commanding officer.

3. The Question of Command Responsibility

In contrast to a soldier or military officer who incurs criminal responsibility as the principal offender or as an aider and abettor (by directly providing some form of practical assistance, support, or encouragement that substantially effects the commission of a crime by the principal offender), another form of liability—the doctrine of superior or command responsibility—can attach to superiors who make no such

139. Sparrow, *supra* note 86, at 74-75 (further noting that "the more autonomous these systems become, the less it will be possible to properly hold those who designed them or ordered their use responsible for their actions").

direct contribution to the *actus reus* of the crime.[140] These commanders incur a *sui generis* form of "liability of omission" by affecting the *environment* in which a crime is committed by a subordinate.[141] Command responsibility dictates that the superior is responsible for his failure to take acts required of him by international law—acts that created or contributed to the circumstances in which a subordinate was able to commit a crime (through a failure to prevent) or in which the subordinate was able to go unpunished (through a failure to punish).[142] However, there are several requirements that must be met in order to successfully apply this doctrine and impose criminal responsibility on military commanders.

First, although the doctrine of command responsibility is a special type of liability, it is not a form of strict liability for the crimes of subordinates. As observed by the Appeals Chamber of the ICTY, "[i]t cannot be overemphasized that, where superior responsibility is concerned, an accused is not charged with the crimes of his subordinates but with his failure to carry out his duty as a superior to exercise control."[143] This observation is in accordance with the foundational tenet of modern international criminal law—the principal of personal culpability—which holds that the officer cannot be held "criminally responsible for acts or transactions in which he has not personally engaged or in some other way participated."[144]

Second, a commander does not incur criminal liability under the command responsibility doctrine based on a mere "breach of duty." Such situations, whereby a commander fails to fulfill the responsibilities that are expected of his rank, are usually instead dealt with through some form of disciplinary action within national command structures,

---

140. GUÉNAËL METTRAUX, THE LAW OF COMMAND RESPONSIBILITY 43 (2009) (noting that "[w]hilst in the case of aiding and abetting the contribution is made to the underlying crime itself, to its *actus reus*, with superior responsibility the causal contribution is more indirect . . . .").

141. *Id.* at 41-43.

142. *Id.* at 45 (noting that the liability of a superior is incurred "for a personal failure to perform an act required of him by international law, namely, to take necessary and reasonable measures to prevent or punish crimes of subordinates"); United States v. Wilhelm List (The Hostages Case), 11 Trials of War Criminals Before the Nuremberg Tribunals Under Control Council Law No. 10 1230, 1298 (1948) (holding that a superior commander is responsible for criminal acts "if he orders, permits or acquiesces in such criminal conduct").

143. Prosecutor v. Krnojelac, Case No. IT-97-25-A, Appeal Judgment, ¶ 171 (Int'l Crim. Trib. for the Former Yugoslavia Sept. 17, 2003).

144. Prosecutor v. Tadic, Case No. IT-94-1-A, Appeal Judgment, ¶ 186 (Int'l Crim. Trib. for the Former Yugoslavia July 15, 1999).

and not under the auspices of international criminal law.[145]

Third, although command responsibility is not a form of aiding and abetting, it still involves complicity: a crime must have been committed by subordinates, one in which the commander acquiesced and in so doing assumed legal responsibility for his own failure to act.

It is thus first necessary to establish criminal conduct on the part of subordinate military personnel for the actions of an autonomous system. If the subordinate's conduct was in fact criminal, the doctrine of command responsibility then requires that the commander had knowledge of that conduct, the power to prevent or punish it, and the duty to do so.[146]

Identifying those duties that, if breached, would incur individual criminal responsibility is not always an easy task. The ICTY Statute provides that a superior is not relieved of criminal responsibility for the acts of subordinates if, assuming requisite knowledge requirements are met, the superior "failed to take the *necessary and reasonable measures* to prevent such acts or to punish the perpetrators thereof."[147] What constitutes "necessary and reasonable measures" for preventing and suppressing IHL violations is not defined in any international agreement. Aside from requiring that a superior must be under "a specific, and pre-existing, legal duty and had formal legal competence to adopt a particular measure," international law does not provide much guidance as to what a commander is concretely required to do, leaving much of this task to domestic law and regulations.[148] This interplay between international and domestic law may thus place special importance on any state or military regulations or directives governing the employment of autonomous weapon systems and the related assignment of command responsibilities regarding the employment of these systems in situations where there is a likelihood of civilian casualties.

Assuming that necessary and reasonable measures can be identified, the next step in finding a commander criminally liable in connection with the acts of subordinates employing autonomous weapons may

145. Jamie Allan Williamson, *Some Considerations on Command Responsibility and Criminal Liability*, 90 INT'L REV. OF THE RED CROSS 303, 303 (2008).

146. Prosecutor v. Halilović, Case No. IT-01-48-T, Trial Judgment, ¶ 38 (Int'l Crim. Trib. for the Former Yugoslavia Nov. 16, 2005) (noting that "[i]t is thus the 'failure to act when under a duty to do so' which is the essence of this form of [superior] responsibility").

147. ICTY Statute, *supra* note 102, art. 7.3 (emphasis added); ICTR Statute, *supra* note 107, art. 6.3 (using identical language); *see also* Additional Protocol I, *supra* note 60, art. 86 (requiring superiors to "take all feasible measures within their power to prevent or repress the breach").

148. METTRAUX, *supra* note 140, at 49.

pose particularly daunting challenges: establishing the requisite state of mental culpability. At the outset, it should be noted that all weapon systems involve some degree of risk. The inherent complexities of autonomous systems may leave a commander with little more than a general appreciation of the many risks that may be present in various situations in which his or her subordinates introduce these weapons.

If an accused commander is charged with a failure to prevent war crimes by subordinates employing autonomous weapons, a mere awareness of a risk that crimes might be committed by those subordinates is not a sufficiently culpable state of mind to attract criminal liability. As noted by the ICTY Appellate Chamber, "[t]he knowledge of any kind of risk, however low, does not suffice for the imposition of criminal responsibility for serious violations of international humanitarian law."[149] Instead, the risk must be "clear, strong, real and serious" and must rise to a level of "a real and concrete likelihood" that is sufficient for establishing the required culpable state of mind.[150] Given the inherent uncertainties and complexities of advanced autonomous weapon systems, this is a formidable requirement.

Knowledge on the part of a commander of the actual or impending commission of crimes by his subordinates is fundamental to the concept of individual criminal responsibility. For this reason, the Ad Hoc Tribunal statutes and Additional Protocol I provide that it is appropriate to hold commanders criminally liable for their failure to prevent or suppress breaches only if, respectively, they "had reason to know" or "had information which should have enabled them to conclude in the circumstances at the time" that a subordinate was committing or was going to commit such a breach.[151]

The "had reason to know" and "had information" standards have not been interpreted by international criminal tribunals as establishing any form of vicarious liability. Instead, these standards refer to information that was in fact available to the commander at the time in question and should not generally be read as imposing a "should have known" standard.[152] In fact, a "should have known" *mens rea* requirement for

149. Prosecutor v. Blaškić, Case No. IT-95-14-T, Appeal Judgment, ¶ 41 (Int'l Crim. Trib. for the Former Yugoslavia July 29, 2004).

150. METTRAUX, *supra* note 140, at 204.

151. ICTY Statute, *supra* note 102, art. 7.3. Identical "had reason to know" language is found in the ICTR Statute, *supra* note 107, art. 6.3; *see also* Additional Protocol I, *supra* note 60, art. 86, § 1.

152. Prosecutor v. Delalic, et al. (the Celebici Case), Case No. IT-96-21-A, Appeal Judgment, ¶ 241 (Int'l Crim. Trib. for the Former Yugoslavia Feb. 20, 2001) ("[A] superior will be criminally

military commanders has been explicitly rejected by the Ad Hoc Tribunals as being inconsistent with customary international law.[153]

In the case of autonomous weapon systems, what is it that a commander accused of acquiescing in war crimes "had reason to know?" What information did the commander have that should have allowed him to conclude that war crimes were being committed or about to occur? Prior to an attack, a multi-volume set of technical specifications for these complex systems and how they work is unlikely to be useful information to the commander in overseeing the conduct of subordinates responsible for these systems. During or after an attack, information that civilians have been killed may be hard to explain or link to specific system failure and related personnel misconduct.

These uncertainties point to the advantage—and necessity—of military organizations putting clear rules and regulations in place to govern the employment of autonomous weapons in specific types of operations. Without such regulations available to evaluate the performance of a commander in overseeing the actions of subordinates, it will be difficult if not impossible to establish that a commander had sufficient knowledge of the misuse of complex autonomous weapon systems to justify the imposition of criminal liability for his or her failure to prevent or suppress violations of the IHL framework. These problems are only magnified by the *mens rea* requirements of international criminal law that demand more than a mere breach of duty, simple negligence, or an awareness of risks.

Yet the development of clear, unambiguous, and effective regulations governing the employment of autonomous weapons may be an elusive goal. A DoD directive approved in 2012 illustrates this problem by providing only a broad and vague mandate, which requires that "[p]ersons who authorize the use of, direct the use of, or operate autonomous and semi-autonomous weapon systems must do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement."[154] As both the complexity and range of missions for autono-

---

responsible through the principles of superior responsibility only if information was available to him which would have put him on notice of offences committed by subordinates.").

153. *Id.* ¶¶ 235, 241. The Ad Hoc Tribunals have continued to affirm this principle, notwithstanding art. 28 of the Rome Statute of the ICC, which codifies the "should have known" standard for some cases of command responsibility.

154. DOD AUTONOMOUS WEAPONS DIRECTIVE, *supra* note 6, at 4.b. This directive has been criticized; *see* Mark Gubrud, *US Killer Robot Policy: Full Speed Ahead*, BULL. ATOMIC SCIENTISTS (Sept. 20, 2013), http://thebulletin.org/us-killer-robot-policy-full-speed-ahead (arguing that "[a]

mous weapon systems grow, commanders will unquestionably struggle (as will their responsible governments) with what constitutes "appropriate care."

While other provisions in DoD regulations set forth some basic procedures to better ensure that autonomous weapon systems are operated in accordance with IHL obligations, ultimately the unique problems presented by these systems—including their unpredictability, complexity, and inherent human oversight limitations—make it difficult to clearly identify many of their operators' specific responsibilities. Such difficulties are magnified when assessing state responsibility for compliance with IHL obligations that place the most complicated demands on human judgment, particularly the principle of distinction (as discussed below in Part V). These problems further illustrate the clear linkage between the elusive concept of individual operator responsibility and the requirement that states be held responsible for corresponding violations of IHL obligations.

Finally, even if it is possible to establish knowledge on the part of a commander that subordinates have committed or are about to commit a war crime through their employment of autonomous weapons, that is not a sufficient state of mental culpability to incur criminal liability under the doctrine of command responsibility. To be held liable, the superior must also have consciously failed to perform his duties "either by deliberately failing to perform them or by culpably or willfully disregarding them."[155] Expressed differently, the negligence of the commander in preventing or punishing war crimes "must be so serious that it is tantamount to malicious intent, apart from any link between the conduct in question and the damage that took place."[156]

4. Blame Everyone? Blame the Machine?

A final basis for imposing criminal sanctions on an individual (including a commander) for IHL violations is JCE liability. As previously

---

careful reading of the directive finds that it lists some broad and imprecise criteria and requires senior officials to certify that these criteria have been met if systems are intended to target and kill people by machine decision alone . . . . Far from applying the brakes, the policy in effect overrides longstanding resistance within the military, establishes a framework for managing legal, ethical, and technical concerns, and signals to developers and vendors that the Pentagon is serious about autonomous weapons.").

155. Prosecutor v. Bagilishema, Case No. ICTR 95-1A-A, Appeal Judgment (Reasons), ¶ 35 (July 3, 2002).

156. INT'L COMM. OF THE RED CROSS, COMMENTARY ON THE ADDITIONAL PROTOCOLS OF 8 JUNE 1977 TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949, at 1012 (Yves Sandoz et al. eds., 1987).

noted, this form of liability is not explicitly found in the statutes of the Ad Hoc Tribunals but was instead developed chiefly by the judges and prosecutors of the ICTY to impose criminal responsibility on individuals responsible for furthering the commission of mass murder, "ethnic cleansing," and other atrocities in the former Yugoslavia. The development of JCE liability has been criticized as greatly diminishing the requirements of individual criminal responsibility, earning the scornful nickname "Just Convict Everyone."[157] Nicknames notwithstanding, JCE liability is not well-suited as a mechanism for imposing criminal liability on military personnel who may be responsible for complex weapon systems. It would be appropriate only in cases of the most egregious, intentional, and systematic misuse of those systems, where such actions could also be described as being part of a "common criminal purpose," "common concerted design," or "joint criminal enterprise."[158]

There are three forms of JCE liability. In each, the prosecution must show (1) "[a] plurality of persons"; (2) "[t]he existence of a common plan, design or purpose which amounts to or involves the commission of a crime provided for in the [ICTY] Statute"; and (3) "[p]articipation of the accused in the common design."[159] For each type of JCE liability, the group in question must have a common criminal purpose and individuals must have one of three required culpable mental states.

In the basic form of JCE liability, the prosecution must prove that the perpetrator acted with "the intent to perpetrate a certain crime."[160] In the second or "systemic" form of JCE liability (which relates to "systems of ill-treatment," primarily concentration camps, and in which the perpetrator holds a position of authority in a military or administrative unit), the accused must have "personal knowledge of the system of ill-treatment" and "inten[d] to further this common concerted system of ill-treatment."[161] In the extended and most controversial form of JCE liability, the accused must have "a criminal intention to participate in a common criminal design" while "criminal acts other than those

157. MARK A. DRUMBL, ATROCITY, PUNISHMENT AND INTERNATIONAL LAW 39 (2007).

158. These and other terms were used interchangeably by ICTY in describing JCE liability in the first case in which it was applied. *See* Prosecutor v. Tadic, *supra* note 144, ¶¶ 187, 203, 220.

159. *Id.* ¶ 227.

160. *Id.* ¶ 196 (noting that the prosecutor must show that "the accused . . . voluntarily participate[d] in one aspect of the common design" and "the accused, even if not personally effecting the [crime] . . . intend[ed] this result").

161. *Id.* ¶ 228.

envisaged in the common criminal design are likely to be committed by other participants in the common design."[162]

If autonomous weapon systems were to be intentionally used in a joint criminal enterprise aimed at attacking civilians, the situation would not be unlike the misuse of other technologies in the deliberate commission of the most serious war crimes against civilian populations. Such misuse is not, however, a common scenario for most types of military engagements. Whatever its merits in addressing organized efforts to systematically murder and terrorize civilians in the former Yugoslavia, JCE liability would be a basis for the imposition of criminal liability on those responsible for autonomous weapons only in those unusual cases in which a common criminal purpose is present and an intention to systematically target civilians can be established.

As artificial intelligence systems and autonomous weapons assume more and more functions analogous to humans, some commentators have raised the question of the responsibility of these machines themselves. Existing legal structures drawing a clear distinction between persons and machines are likely to be challenged by increasingly sophisticated artificial intelligence systems, even to the point that someday there may be calls for such machines to be held legally accountable for their actions.[163]

The assignment of moral and legal accountability to machines of course risks absolving humans of any related moral or legal responsibilities. The resulting temptation for humans to simply pass the buck to machines when things go wrong has left many ethicists and philosophers of technology unwilling to attribute such moral agency to machines, regardless of technological advances in the field of artificial intelligence.[164]

While ethicists, lawyers, and philosophers continue to examine the barriers to the legal accountability of machines, it is safe to say that there currently is an enormous gap between any technologies that meet the conditions necessary for "moral personhood" and any existing, planned, or even contemplated autonomous weapon system.[165]

---

162. *Id.* ¶¶ 204, 206.

163. WALLACH & ALLEN, *supra* note 14, at 204-05.

164. *Id.* at 199, 201.

165. Sparrow, *supra* note 86, at 73. *But see* Peter M. Asaro, *How Just Could a Robot War Be?*, *in* CURRENT ISSUES IN COMPUTING AND PHILOSOPHY 51 (A. Briggle et al. eds., 2008) (assessing a "continuum of autonomy and moral responsibility" and suggesting that "[a]s robotic technologies advance, it is possible that they will acquire moral capacities that imitate or replicate human moral capacities").

Because international criminal law (and criminal law generally) is fundamentally linked to concepts of punishment and deterrence, it generally presumes that criminal actions are performed by moral agents that are ultimately capable of being punished.[166] It is not yet possible, and perhaps unlikely to ever be possible, to meaningfully punish, shame, praise, or reward a machine. Moreover, even if such emotions were to be regarded as valuable restraining influences and it were feasible to replicate them in machines, some have argued that introducing emotions into robots is "a virtual Pandora's box filled with both benefits and ethical challenges."[167]

One solution to future artificial intelligence accountability issues might be to give advanced autonomous weapon systems a quasi or "legal" person status similar to that given to limited liability corporations in many legal systems.[168] While such a proposal might be useful in addressing certain product liability issues in national legal systems, it would conspicuously fail to meet the requirements of an international legal system intended to suppress war crimes and promote compliance with the IHL framework.

Holding a robot accountable as a "legal" person for war crimes as if it were a human appears to be impractical on many levels. Such an approach assumes that machines can possess both key human reasoning and judgment capabilities, as well as sufficiently human attributes to be subject to meaningful punishment. As unlikely and unworkable as this proposition may be, it has echoes in efforts to assign states responsibility for the actions of autonomous machines as if those machines can be viewed as "reasonable persons," performing legal functions for states under the IHL framework, as discussed below.

V. STATE RESPONSIBILITY FOR AUTONOMOUS MACHINES AND THE PROBLEMS OF DISTINCTION AND PROPORTIONALITY

No problem highlights the challenges of applying existing legal frameworks to autonomous weapon systems more than efforts to ensure compliance by military forces with the cardinal principle of distinction (or discrimination) and its corollary, the principle of propor-

166. *See* Morissette v. United States, 342 U.S. 246, 250 n.4 (1952) (quoting Roscoe Pound that "[h]istorically, our substantive criminal law is based upon a theory of punishing the vicious will. It postulates a free agent confronted with a choice between doing right and doing wrong and choosing freely to do wrong.").

167. WALLACH & ALLEN, *supra* note 14, at 196.

168. *Id.* at 204.

tionality.[169] The capabilities, actions, and decisions of individual human commanders and weapons operators are inescapably linked with state IHL obligations in this area. Thus, key problems that make the search for individual criminal responsibility elusive for the actions of autonomous weapons also haunt efforts to impose responsibility on states for violations of IHL obligations.

Proponents of autonomy argue that there is a "possibility that autonomous weapon systems could in the long run be more discriminating and ethically preferable to alternatives."[170] It may seem incongruous to describe a weapon that is specifically designed to attack only military objectives as inherently indiscriminate, yet autonomy raises a question not yet clearly confronted by the IHL framework: What level or quality of *human* judgment is ultimately required to apply the principles of distinction and the related principle of proportionality? Scientists and weapons designers currently face enormous challenges in building autonomous weapons that are able to consistently distinguish civilians from combatants. An even greater task awaits them with respect to ensuring observance of the principle of proportionality.

The inherently subjective and complex decisions required for applying the proportionality principle so as to avoid "excessive" civilian casualties—decisions that often challenge human judgment—seem certain to defy easy resolution by formulas, the most advanced algorithms, and any current or contemplated artificial intelligence systems. Yet regardless of the capabilities of existing and future autonomous systems, legal requirements appear to dictate the presence of the human mind (a "reasonable" one) for making decisions relating to key IHL obligations.

Legal requirements linked to the need for human judgment must be clearly separated for purposes of analysis from concepts of morality. To the limited extent that it has been considered by states and the international community as a whole, the subject of non-human decision making about human life and death has never been easy to reconcile with either legal or moral principles. Unfortunately, these two different sets of principles have often been blurred or conflated by courts and commentators.

---

169. As now codified in its most widely accepted formulation, the proportionality principle requires that an attack be cancelled or suspended if it "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 . . . ." Additional Protocol I, *supra* note 60, art. 57.2.b.

170. Anderson & Waxman, *supra* note 24, at 15.

In cases before courts of domestic legal systems that have addressed the use of deadly mechanical traps, and in debates at the first international conferences addressing the legality of mines, there are objections to both the indiscriminate nature of the earliest autonomous killing machines as well as an aversion to their non-human character.[171] Decisions by U.S. courts have referred to various lethal mechanical traps and spring guns as "silent instrumentalities of death" and "secret implements of destruction," and have opined that "[t]he taking of human life by such means is brutally savage and inhuman."[172] Commentary on such U.S. cases has also emphasized that "[w]here the actor is present, there is always the possibility he will realize that deadly force is not necessary, but deadly mechanical devices are without mercy or discretion."[173]

While it may be relatively easy for domestic courts to conclude that simple traps applying lethal force lack the necessary reasoning ability and discretion to do so, the autonomous capabilities of modern military technologies present much greater challenges. Yet even if a machine is able to clearly distinguish combatants from non-combatants (a significant assumption), the balancing tests encompassed within the proportionality principle yield questions that demand human judgment related to the value of human life in the context of both specific, immediate facts and larger, long-term goals. These questions include: "What are the relative values to be assigned to the military advantage gained and the injury to non-combatants and/or the damage to civilian objects? . . . What do you include or exclude in totaling your sums? . . . What is the standard of measurement in time or space?"[174]

As noted, autonomous weapons do not appear to meet established criteria required for banning a weapon as illegal *per se*. These weapons are unlikely to be calculated to cause unnecessary suffering and are not inherently indiscriminate (at least in sense that current versions cannot be sufficiently controlled to be directed at military objectives). However, this Article argues that a key factor in applying the principles of

171. *See, e.g.*, JAMES J. BUSUTTIL, NAVAL WEAPONS SYSTEMS AND THE CONTEMPORARY LAW 14 (1998) (noting the arguments of the Colombian delegate at the Second Peace Conference of 1907 that "naval mines are the most treacherous method of modern war" and that they present the disturbing prospect of "courageous sailors destroyed by a murderous device set by an absent enemy").

172. People v. Ceballos, 526 P.2d 241, 244 (Cal. 1974).

173. Perovich, *supra* note 62.

174. ICTY, FINAL REPORT TO THE PROSECUTOR BY THE COMMITTEE ESTABLISHED TO REVIEW THE NATO BOMBING CAMPAIGN AGAINST THE FEDERAL REPUBLIC OF YUGOSLAVIA (June 8, 2000), *reprinted in* 39 ILM 1257, 1271 (2000) [hereinafter ICTY REPORT ON NATO BOMBING CAMPAIGN].

distinction and proportionality to autonomous systems deserves special consideration: the *legal* requirement of a proximate and significant level of *human judgment.* A reassessment of existing, emerging, and future planned autonomous weapon systems helps demonstrate the important role this requirement plays.

A. *Human Judgment Along the Continuum of Autonomy*

The necessary connection between technology, the principles of distinction and proportionality, and human judgment is illustrated by a brief revisit of weapons found along the continuum of autonomy that was discussed in Part II. At the lower end of the continuum of autonomy, fire-and-forget weapons use a variety of guidance systems (or combinations of these systems) to autonomously seek targets after they have been launched by human operators. Because these weapons are capable of autonomously tracking and attacking targets that are "over the horizon" or "beyond visual range," civilian objects may unintentionally be hit. These capabilities point to the need for special care in target discrimination but do not necessarily make such weapons unlawful.[175]

The earliest long-range missiles, such as the German V-1 and V-2 rockets of World War II, are cited as examples of "weapons that are incapable of being directed at a military objective" and are thus "forbidden as being indiscriminate in their effect."[176] Modern autonomous guidance systems, however, have so dramatically improved the accuracy of long range missiles and other fire-and-forget weapons that they cannot be classified as illegal *per se* under the IHL framework due to their indiscriminate nature. On the contrary, all major military forces now rely on these weapon systems, particularly for anti-ship and anti-aircraft attacks, precisely because of their accuracy.[177] According to U.S. military regulations, these weapons are "lawful provided they are equipped with sensors or are employed in conjunction with external sources of targeting data that are sufficient to ensure effective target discrimination."[178] As a general matter, effective target discrimination hinges on the question of whether weapons "are not, or cannot be, directed against a specific military objective."[179] This standard

175. SAN REMO MANUAL ON INTERNATIONAL LAW APPLICABLE TO ARMED CONFLICTS AT SEA 117 (2205).

176. NAVAL COMMANDER'S HANDBOOK, *supra* note 64, § 9.10.

177. JAMES J. BUSUTTIL, NAVAL WEAPONS SYSTEMS AND THE CONTEMPORARY LAW 191 (1998).

178. NAVAL COMMANDER'S HANDBOOK, *supra* note 64, § 9.10.

179. SAN REMO MANUAL, *supra* note 175, ¶ 78.4

appears to be easily met by modern autonomous guidance systems, particularly when state practice has demonstrated that even "unguided bombs are not as such indiscriminate by nature."[180]

Given that "dumb" or unguided bombs aimed by humans at targets can easily satisfy the discrimination requirement (even if significant unintended damage to civilian objects occurs), weapons with advanced autonomous guidance systems seem unlikely to fail that requirement. The latter type of weapons may, however, attack targets far "over the horizon," beyond any human operator's vision. Based on the widespread acceptance of these weapons by states, the element of human judgment necessary to ensure discrimination and apply the proportionality balancing test appears to be sufficiently present in the key acts of target selection and launching that are performed by human operators.

Moving on to the mid-range of the continuum of autonomy, a variety of autonomous defense systems are already widely deployed by military forces, including close-in defense systems on ships, air defense systems, and land based systems intended to counter rockets, mortars, and artillery. These weapons are undoubtedly capable of causing civilian casualties by autonomously attacking civilian planes, ships, and other civilian objects. However, the many states that have adopted variants of these weapon systems and the accompanying lack of any legal objections by other states now make it difficult to argue that they are illegitimate, indiscriminate weapons under customary international law.

The failure of states to express concern regarding the legality of existing autonomous defense systems is problematic for those groups now broadly advocating bans on all "killer robots." As discussed above, some existing autonomous systems have already crossed a fundamental human-machine threshold: Once activated, they have the ability to identify and track targets and autonomously "decide" to use force in a way that can cause death or injury to non-combatants.

One possible explanation for the lack of criticism of most existing autonomous defense systems is their specialized or immobile nature—and the way this means that their placement and their use arguably

180. MANUAL ON INTERNATIONAL LAW APPLICABLE TO AIR AND MISSILE WARFARE, Program on Humanitarian Policy and Conflict Research at Harvard University (2013), at 61 (further noting that, "[i]n many circumstances, they can be delivered against enemy combatants with little harm to civilians or civilian objects").

remain linked to human judgment.[181] Close-in weapon systems are just that: they generally pose risks only to civilian aircraft and vessels that venture past outer defenses and deep into combat zones. These risks greatly increase, of course, in the case of naval engagements that spread to busy shipping lanes, ports, or other areas with a significant civilian presence or, in the case of land-based military activities, when they encroach on densely populated urban areas. In many cases, the intended targets of specialized defense systems (missiles, military aircraft, vehicles, etc.) can often be destroyed with little or no accompanying risk of civilian casualties.

Unlike existing autonomous anti-missile, anti-aircraft, and sea-based Phalanx systems, robotic guard or sentry systems designed to perform primarily anti-personnel missions on a fully autonomous basis are not yet widely deployed, although many remote-controlled precursors are now being evaluated in the field. Stationary sentry systems that are able function autonomously, such as the South Korean SGR-1, raise many of the same concerns that led to the conclusion of a convention banning anti-personnel landmines. For this reason, some anti-mine activists are seeking a similar ban on autonomous weapon systems.[182] In fact, some NGOs suggest that landmines and fully autonomous robots "share essential elements of DNA" and that they should thus both be logically included behind "a red line" banning "fully autonomous targeting."[183]

A central criticism of anti-personnel landmines by their opponents is that these weapons should be regarded as inherently indiscriminate owing to their inability to distinguish combatants from non-combatants. Although autonomous guard systems are more dangerous in the sense that they can fire at multiple targets at a considerable distance, they may offer several advantages over comparatively primitive landmines, such as improved identification capabilities and the ability to broadcast a warning (and even demand surrender) before

181. It is also argued that the risk many of these weapons pose to civilians and civilian objects is further reduced because they are primarily designed by humans "to destroy munitions, not launch offensive attacks." LOSING HUMANITY, *supra* note 5, at 12.

182. Marks, *supra* note 4 ("It is the first time a high profile non-governmental organisation has campaigned against such technology. London-based charity Landmine Action wants autonomous robots capable of killing people banned under the same kind of treaty that has outlawed landmines in over 150 countries.").

183. Matthew Bolton et al., *Ban Autonomous Armed Robots*, ARTICLE 36 (Mar. 5, 2012), http://www.article36.org/statements/ban-autonomous-armed-robots/ (The authors are members of the U.K.-based NGO "Article 36" who argue that "[l]andmines and fully autonomous weapons . . . all provide a capacity to respond with force to an incoming 'signal'" and thus both represent "killing by machine" that should be banned.).

firing. In addition, long-buried and forgotten anti-personnel landmines are indiscriminate killers even in a temporal sense, often threatening civilians long after a particular armed conflict has ended.

While they share some similarities with landmines, autonomous guard systems do not fall within the narrow legal scope of exploding devices banned by the landmine convention, and they are not regulated by any other international convention.[184] Unless disguised in such a way as to constitute a prohibited deceptive practice, they are also unlikely to be characterized as perfidious under the IHL framework. Although they lack the ability of human guards to distinguish combatants from non-combatants, autonomous machines with sophisticated sensors and warning/identification capabilities are an improvement over simple mechanical explosive devices. Furthermore, since the practice of states indicates that the landmine ban is not yet customary international law, it is difficult to argue that immobile, autonomous guard systems are unlawful on a similar basis.[185]

Ultimately, the restricted geography and immobile nature of landmines and various autonomous guard systems ensure that human judgment is able to play an important role in their operation through the act of placement, and whatever additional precautions and warnings can be devised for the defined area in which these systems are allowed to employ lethal force. However, the most advanced of these systems now in existence are increasingly called upon to make more and more lethal "decisions" with very limited human "supervision."

Current DoD regulations implicitly recognize the unique challenges and risks associated with the high level of autonomy found in existing and emerging defensive systems that employ lethal force. Even if a human is theoretically able to intervene, the pace and complexity of many defensive operations against multiple targets and the huge amount of data that must be quickly processed means that the human operator must increasingly trust the machine to do its job. Operators in such situations can even experience what is called "automation bias," which is "the tendency to trust an automated system, in spite of

184. *See* Anti-Personnel Land Mine Treaty, *supra* note 59, art. 2.1 (prohibiting mines which are "designed to be exploded by the presence, proximity or contact of a person and that will incapacitate, injure or kill one or more persons").

185. JEAN-MARIE HENCKAERTS & LOUIS DOSWALD-BECK, CUSTOMARY INTERNATIONAL HUMANITARIAN LAW, VOLUME I: RULES 282 (2005) (noting that in light of current state practice, "it cannot be said at this stage that the use of anti-personnel landmines is prohibited under customary international law").

evidence that the system is unreliable, or wrong in a particular case."[186] At some point, although theoretical human intervention is possible, human involvement in these advanced autonomous defensive systems comes to resemble a version of the so called "human-out-of-the loop."

Because of the limitations, protections, and restrictions that are provided, through the exercise of human judgment in the placing and employment of advanced autonomous defensive weapons for only the "static defense of manned installations" and "the onboard defense of manned platforms," DoD regulations make an exception allowing for these deadly killer robots. These regulations provide that "[h]uman-supervised autonomous weapon systems may be used to select and engage targets, with the exception of selecting humans as targets, for local defense to intercept attempted time-critical or saturation attacks."[187]

In a clear expression of concern about releasing "human-supervised autonomous weapon systems" into environments and operations that are *not* restricted by the static defense of manned installations or the onboard defense of manned platforms, DoD regulations require that other types of missions intended for other autonomous or semi-autonomous weapon systems must be specifically approved by the Under Secretary of Defense for Policy, the Under Secretary of Defense for Acquisition, Technology, and Logistics, and the Chairman of the Joint Chiefs of Staff "before formal development and again before fielding."[188] Thus, beyond the controls and restrictions afforded by human judgment through the placement and employment of advanced autonomous weapons in static positions and on manned platforms (often in wide swaths of empty oceans) lies a great and dangerously challenging zone for ensuring their compliance with the principles of distinction and proportionality.

B. *Higher End of the Continuum: The Uncharted Legal Territory of the Future*

Unlike existing point defense and stationary autonomous guard/barrier systems that can to some extent reflect human judgment and be directed at military objectives by their placement, mobile, autonomous

186. Peter M. Asaro, *Modeling the Moral User*, IEEE TECH. & SOC'Y MAG., Spring 2009, at 22, *available at* http://peterasaro.org/writing/Asaro%20Modeling%20Moral%20User.pdf.

187. DOD AUTONOMOUS WEAPONS DIRECTIVE, *supra* note 6, at 4.c.(2).

188. *Id.* at 4.d (further noting that this approval must be made "in accordance with applicable guidelines, policies and issuances").

combatant systems pose a far greater risk of causing civilian casualties as a result of their mobility, expanded range, and missions. Semi-autonomous variants of these weapons with anti-personnel attack capabilities are being tested on the ground, and the first unmanned combatant vehicles with autonomous capabilities are already being tested in the air. In air combat situations far removed from civilian air traffic, autonomous attack capabilities may pose little risk to non-combatants. In other air, ground, and sea scenarios, the risks are considerably greater and call into question the ability of mobile, autonomous combatant weapon systems (with distinct anti-personnel functions) to comply with rules of engagement and IHL obligations without external human input. It is such planned and near-future "killer robots" that raise the most pointed legal, ethical, and societal issues and present the most serious problems for regulation under the existing IHL framework.

The U.S. government's response to concerns about the future deployment of autonomous, mobile, multiple-mission killer robots appears to be two-fold. The first response is the affirmation that some continuing level of direct human supervision is necessary and present. For example, the DoD notes that, "[f]or a significant period into the future, the decision to pull the trigger or launch a missile from an unmanned system will not be fully automated, but it will remain under the full control of a human operator."[189]

However, maintaining this so-called "human in the loop" often runs against the very forces that are driving the creation of these systems in the first place. As a result of improvements in artificial intelligence, the speed of attacking missiles and planes, the need to quickly process vast quantities of data, and the ever-increasing tempo of modern combat generally, human operators are already increasingly proven to be not only redundant, but also disadvantageous in the functioning of these systems.[190]

Even if human involvement in an autonomous weapon system is deemed desirable, the communications link necessary to provide human-in-the-loop oversight is increasingly threatened by bandwidth limitations (and may also be difficult to maintain in some combat conditions).[191] These and other factors have compelled the U.S. Air

189. OFFICE OF THE SEC'Y OF DEF., UNMANNED SYSTEMS ROADMAP, 2009-2034 10 (2009) [hereinafter UNMANNED SYSTEMS ROADMAP 2009-2034].

190. Sparrow, *supra* note 86, at 68.

191. *See* UAS FLIGHT PLAN, *supra* note 37, at 43 (further noting that not only is it difficult to maintain "security in a contested environment," but bandwidth requirements "become more

Force to regard UCAVs as the next logical step in fighter aircraft. Tellingly, official U.S. Air Force planning documents concede the inherent difficulty of keeping a "man in the loop" in future autonomous systems, preferring instead to use their own terminology—"man on the loop"—to describe a less direct form of human involvement that focuses on human monitoring of the "execution of certain decisions."[192] Although it is not uncommon for media reports to note that the "official policy" of states is to not concede decision-making over the use of lethal force to machines or at least preserve the possibility of human intervention,[193] existing military technologies are already calling such "official policies" into question.

The second U.S. government response to concerns about the future deployment of autonomous, mobile, multiple-mission killer robots is a focus on the possibility of equipping such future systems with ethical programming that could make them capable of complying with IHL obligations. For example, recognizing that autonomous combatant systems are no longer a thing of fiction, the U.S. Navy recently commissioned "a preliminary investigation" into the risks and ethics related to autonomous military systems, with a particular focus on battlefield robotics.[194] Surveying work in this field, the authors of the Navy study envisioned a future role for "artificial moral agents," ethical programs, and other types of software architecture that may someday constrain an autonomous military robot's lethal actions and allow it to conform with IHL obligations.

According to some leading experts, there is reason to hope that military robots could someday "behave more ethically on the battlefield than humans currently can," unhindered by human emotions, weaknesses, and vulnerabilities.[195] Unfortunately, numerous technological thresholds must be crossed before the development of "virtuous ro-

---

demanding for stealthy operations such as cooperative engagements that require low probability-of-intercept or detection").

192. *Id.* at 41 ("Increasingly humans will no longer be 'in the loop' but rather 'on the loop'—monitoring the execution of certain decisions.").

193. *See, e.g.*, Jason Palmer, *Call for Debate on Killer Robots*, BBC NEWS (Aug. 3, 2009), http://news.bbc.co.uk/2/hi/8182003.stm ("Current rules of engagement to which the UK subscribes prohibit the use of lethal force without human intervention.").

194. AUTONOMOUS MILITARY ROBOTICS REPORT, *supra* note 8, at iii.

195. *See, e.g.*, Ronald C. Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*, GA. INST. OF TECH. C. OF COMPUTING 1 (2007), *available at* http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf (Professor Arkin suggests that correctly designed software architecture could someday provide "an ethical control and reasoning system potentially suitable for constraining lethal actions in an autono-

bots" become a serious possibility.[196] In the meantime, new autonomous weapon systems are being developed at such a rapid pace that they risk outstripping the ability of engineers and designers to develop supporting programs to even ensure their safety, let alone make certain that they are capable of complying with IHL obligations.[197] This problem is particularly serious and challenging with respect to ensuring the observance of the principle of distinction, as some scientists have suggested that it is "simply too difficult to design a machine that can distinguish between a combatant and a non-combatant, particularly as insurgents pose as civilians, as required for the LOW [law of war] and ROE [rules of engagement]."[198]

Although autonomous combatant systems may be designed and programmed to attack only military objectives, the more removed they become from human control, the less capable they are of "being directed" at specific military targets (at least by human operators). Existing autonomous weapon systems at the lower-to-mid range of the continuum still have human operators that perform key functions in activating, placing, targeting, or operating these weapons. By identifying specific targets and by initiating and overseeing attacks—either by direct involvement, programming, or placement—it can be argued that humans ultimately remain in control of these weapons and that their operation remains linked to human decision-making.

However, lethal military machines are continuing to progress along a continuum to a point where they have higher and higher degrees of autonomy, mobility, new combatant capabilities, and eventually may no longer be restricted to specialized missions. At this point of diminished human control and greater risk of civilian casualties, advanced autonomous weapons will directly confront international law (and humanity) with the question of whether decisions involving adherence to the cardinal principles of international humanitarian law can ultimately ever be made by machines—or at least with a removed, limited, or much lower level of human participation and, more specifically, human judgment.

---

mous robotic system so that they fall within the bounds prescribed by the Laws of War and the Rules of Engagement.").

196. AUTONOMOUS MILITARY ROBOTICS REPORT, *supra* note 8, at 40.

197. *Id.* at 41.

198. *Id.* at 76 (further noting that "the challenge of creating a robot that can properly discriminate among targets is one of the most urgent, particularly if one believes that the (increased) deployment of war robots is inevitable").

C. *Human Judgment, Practical Controls, and Legal Accountability*

Whatever promise "ethical governors" and other ethical programming may offer in the future for autonomous machines, advanced robots will never be humans and will never possess human judgment *per se* (no matter how close machines may ever come to achieving cognition). Recognizing such limitations, the DoD has required that autonomous and semi-autonomous weapon systems "be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force."[199]

Efforts to ensure appropriate levels of human judgment over the use of force by autonomous systems include requirements that such systems "go through rigorous hardware and software verification and validation (V&V) and realistic system developmental and operational test and evaluation."[200] Furthermore, DoD regulations require that training, doctrine, tactics, techniques, and procedures be established to ensure that autonomous and semi-autonomous weapon systems:

> (a) Function as anticipated in realistic operational environments against adaptive adversaries.
>
> (b) Complete engagements in a timeframe consistent with commander and operator intentions and, if unable to do so, terminate engagements or seek additional human operator input before continuing the engagement.
>
> (c) Are sufficiently robust to minimize failures that could lead to unintended engagements or to loss of control of the system to unauthorized parties.[201]

In addition, in order to avoid "the potential consequences of an unintended engagement or loss of control of the system to unauthorized parties," DoD requires that physical hardware and software be "designed with appropriate . . . [s]afeties, anti-tamper mechanisms, and information assurance and [h]uman-machine interfaces."[202] Finally, in order for the operators to "make informed and appropriate decisions in engaging targets," DoD regulations provide that "the interface between people and machines for autonomous and semi-autonomous weapon systems shall: (a) Be readily understandable to

199. DOD AUTONOMOUS WEAPONS DIRECTIVE, *supra* note 6, at 4.a.

200. *Id.* at 4.a.(1).

201. *Id.* at 4.a.(1)(a)-(c).

202. *Id.* at 4.a.(2).

trained operators; (b) Provide traceable feedback on system status; [and] (c) Provide clear procedures for trained operators to activate and deactivate system functions."[203]

While the above regulations have been explicitly imposed in order to allow commanders and operators to exercise appropriate levels of human judgment over the use of force by autonomous systems, these limits are implicitly recognized in those same regulations as insufficient except for systems involving the static defense of manned installations and the onboard defense of manned platforms.[204] All other types of lethal "human-supervised autonomous weapon systems" require, as noted, special approvals from multiple authorities. It thus appears that human judgment is considered too attenuated (for purposes of pre-approval) with respect to other advanced autonomous systems with broader missions that may pose a serious threat to civilian objects.

The emerging, explicit focus on the role of human judgment in DoD regulations governing autonomous weapons is not surprising. Human judgment—in different states or conditions—underlies legal obligations generally, including those found in the IHL framework. In determining whether an actor has complied with the IHL requirements of distinction and proportionality, a court must inquire into human perception and reasoning in the context of specific, often complex sets of immediate facts and longer-term goals relating to the choice of targets to be attacked, the value of those targets, and the incidental costs resulting from their destruction.[205]

Legal and political accountability flows from the judgments of commanders who must evaluate the direct military benefits of launching an attack against a military object in relation to the expected loss of innocent civilian lives. Not only is it necessary that a human make this judgment, but increasingly (under the applicable policies of states), a human at the highest levels of authority must make this judgment, depending on how many innocent civilian lives are at risk.[206] Thus, in

---

203. *Id.* at 4.a.(3).

204. *See supra* discussion at page 40.

205. ICTY REPORT ON NATO BOMBING CAMPAIGN, *supra* note 174, at 1271 (noting that "[i]t is much easier to formulate the principle of proportionality in general terms than it is to apply it to a particular set of circumstances because the comparison is often between unlike quantities and values. One cannot easily assess the value of innocent human lives as opposed to capturing a particular military objective.").

206. U.S. policy appears to mandate that if a given planned attack involves anticipated civilian casualties above a maximum allowable number, that attack must be approved at a high level of military and political authority. *See, e.g.*, Bradley Graham, *U.S. Moved Early for Air Supremacy*, WASH. POST, July 20, 2003, at A26 (noting that the former commander of U.S. Central Command

addition to their other limitations in replacing humans, robots seem particularly ill-suited to serve as elected officials or cabinet secretaries who must make the ultimate decisions about the value of civilian lives to be lost in pursuit of national security objectives.

The law is also clear about the type of person that must serve as the standard for making decisions about the true meaning of the principles of distinction and proportionality. The ICTY has concluded that, whenever there is doubt about the status of persons to be attacked, "the Prosecution must show that in the given circumstances a *reasonable person* could not have believed that the individual he or she attacked was a combatant."[207]

The need for a reasonable person (or a reasonable military commander) to make complex choices becomes particularly pressing in the case of applying the principle of proportionality when civilian casualties are clearly expected.[208] Although the focus of courts in relevant cases is on what a reasonable person would do in a particular situation with the information available to him or her, there is an implicit understanding in their findings that the "proportionality determination equates to a judgment call, which has always belonged to a human."[209] If the person required to make this judgment call does so on the basis of imperfect or contradictory information, courts may explore more aspects of his or her humanity by requiring that an

---

Air Forces confirmed that, in the early stages of Operation Iraqi Freedom, approval to bomb targets—if the deaths of thirty or more civilians might result—had to be reserved for Secretary of Defense Rumsfeld). Although the precise number of anticipated casualties triggering this review remains classified, it is reportedly dropping. *See, e.g.*, Mark Benjamin, *When Is an Accidental Civilian Death Not an Accident?*, SALON.COM (July 30, 2007) (quoting a journalist, who has interviewed intelligence analysts familiar with the policy, as saying that the days of the "magic number" of 30 are over).

207. Prosecutor v. Galić, Case No. IT-98-29-T, ¶ 55 (Int'l Crim. Trib. for the Former Yugoslavia Dec. 5, 2003) (emphasis added).

208. ICTY REPORT ON NATO BOMBING CAMPAIGN, *supra* note 174, at 1271 (noting that the answers to questions raised by the application of proportionality "are not simple" and that "[i]t may be necessary to resolve them on a case by case basis, and the answers may differ depending on the background and values of the decision maker . . . it is unlikely that military commanders with different doctrinal backgrounds and differing degrees of combat experience or national military histories would always agree in close cases. It is suggested that the determination of relative values must be that of the 'reasonable military commander'").

209. Thurnher, *supra* note 74, at 81 (further noting that this judgment call has "traditionally . . . been compared against what a 'reasonable person' or a 'reasonable commander' would do in such a situation.").

"honest judgment" be made with the information available.[210]

In contrast to artificial intelligence systems that are designed to support specific tasks or functions, a system endowed with sufficient human attributes to serve as a "reasonable person" under the law is highly unlikely to ever be a reality. Enormous obstacles confront efforts to build a machine with artificial reasoning that can genuinely replicate human judgment and cognition. Scientists, philosophers, and ethicists continue to study and debate the difficulties that are inherent in creating an unambiguous set of "top-down" ethical rules for any machine to follow.[211] More importantly, no matter what its capabilities, it will not be a human, nor will it be a commander, elected official, or other "person" in authority. It is also highly unlikely that a "reasonable robot" standard will emerge in the foreseeable future to replace existing norms governing human commanders under the IHL framework and international criminal law.

The challenge confronting military forces in assigning missions to fully autonomous combatant systems will thus involve, particularly for purposes of ensuring distinction and proportionality, limiting the level of risk that missions will pose to non-combatants by allowing the machine to perform only certain actions within some sort of restricted spectrum; one that is subject to pre-established controls reflecting human decision-making. For other missions involving greater risk to non-combatants, fundamental questions must be confronted. While some roboticists argue that machines can perform more reliably and better comply with IHL obligations than humans in combat, the question remains whether a machine can ever perform such functions *for purposes of the law.*

In situations where the risk of civilian casualties is high and difficult decisions about the application of distinction and proportionality are present, the IHL framework demands that a reasonable person in a position of authority, rather than a machine, be able to make key judgments and evaluations. To the extent that existing autonomous

---

210. *See, e.g.*, United States v. Wilhelm List (The Hostages Case), 11 TRIALS OF WAR CRIMINALS BEFORE THE NURENBERG TRIBUNALS UNDER CONTROL COUNCIL LAW NO. 10, 1297 (1950) (The Tribunal held that "[w]e are concerned with the question whether the defendant at the time of its occurrence acted within the limits of honest judgment on the basis of the conditions prevailing at the time. The course of a military operation by the enemy is loaded with uncertainties . . . . It is our considered opinion that the conditions, as they appeared to the defendant at the time were sufficient upon which he could honestly conclude that urgent military necessity warranted the decision made. This being true, the defendant may have erred in the exercise of his judgment but he was guilty of no criminal act.").

211. WALLACH & ALLEN, *supra* note 14, at 84, 97.

weapon systems operate with very limited human involvement, they do so only when employed on specialized missions that remain restricted in key respects by some form of human judgment.

In the future, more advanced, mobile autonomous combatant systems may intrude into complex decision making roles that are now performed only by humans. While it may be difficult to classify these weapons as illegal *per se* if they can perform properly in very restricted conditions, their employment would unquestionably be problematic where civilians are put at risk and the principles of distinction and proportionality must be carefully applied. In such scenarios, human decision making takes on a legal significance that overshadows the practical capabilities of any machine.

VI. CONCLUSION

Proponents of autonomous weapon systems suggest that proposals to ban these systems would preclude many possible positive technological outcomes (particularly outcomes that could better ensure protection of the civilian population in armed conflicts) and that "radical skepticism" about the capabilities of future autonomous technologies is unjustified.[212] Continued research into improving the discriminating capabilities of defensive autonomous systems with restricted missions could indeed be valuable. Whatever the capabilities of future robots, however, they will not qualify as "reasonable persons" with the human judgment necessary to make key decisions about obligations under the existing IHL framework. The inevitable creep of military weapons, on an incremental basis, along the continuum of autonomy will continue to bring this problem into focus.

The incremental creep of autonomy in weapon systems may make their effective regulation almost impossible, especially in the context of arms control, where it may be impossible to determine "what constitutes the tipping point into impermissible autonomy given that the automation of weapons' functions is likely to occur in incremental steps."[213] Yet the incremental march of machines to greater autonomy may already be beginning to carve out important legal defining bound-

212. Anderson & Waxman, *supra* note 24, at 15.

213. *Id.* at 16 (noting further that "even states and groups inclined to support treaty prohibitions or limitations will find it difficult to reach agreement on scope or definitions because lethal autonomy will be introduced incrementally—as battlefield machines become smarter and faster, and the real-time human role in controlling them gradually recedes, agreeing on what constitutes a prohibited autonomous weapon will be unattainable").

aries, founded on the role that human judgment plays in applying the most challenging legal tests required by the IHL framework.

These developments are perhaps best reflected in the struggle evident within the DoD as it attempts to reconcile the ever-increasing autonomy of lethal defensive systems (and their continuing widespread deployment) with proposals for new, more dangerous autonomous combatant systems. DoD officials have no choice but to focus their regulations on the connection such future systems must have with the most important controlling legal component: human judgment. For this reason, only those advanced autonomous systems that perform restricted defensive missions (within the parameters of their defined placement) are made available for pre-approved, conditional deployment.

It is no coincidence that the most dangerous and advanced autonomous weapon systems now in use are defensive ones, constrained by placement and missions that allow human judgment to play a key role in their operation. Future systems that move beyond purely defensive missions will require severe constraints to allow human judgment to play such a prominent role, limiting them in the near future (and perhaps forever) to unpopulated areas in the air and at sea where their operations will pose only a minimal risk to civilians.

The operation of autonomous weapon systems beyond such limited, constrained missions will lead both individuals and states into a fundamental conflict with the legal requirement for human judgment. The great difficulty in imposing individual criminal responsibility on operators and commanders (or on any humans) for the failure of autonomous weapons is inescapably linked to fundamental problems associated with imposing liability on states for violation of key IHL obligations that depend on human judgment. In turn, while machines may indeed perform many military functions more reliably than humans, the IHL framework requires states to put "reasonable commanders" in positions in which they are able to exercise their judgment in complying with that state's key IHL obligations, in particular the principles of distinction and proportionality.

International law does not now demand that a human be in charge of all of the targeting functions of weapons. No international conventions directly address the issue, and customary international law remains largely silent as well. Although relevant state practice in this area is currently quite limited, it does strongly support the lawfulness of existing, widely deployed autonomous defensive systems. Such support can be directly linked to the issues discussed in this Article regarding the legal significance of their restricted missions and the role that

human judgment is able to play in the operation of these systems.

It is useful, however, for purposes of assessing the early stages of the formation of rules of customary international law, to draw attention to the stated positions of the most developed military powers with respect to their intended use of autonomous weapon systems. For example, at least one official U.S. government publication notes that, although many aspects of targeting and firing sequences will someday be fully automated, the "decision to fire will not likely be fully automated until legal, rules of engagement, and safety concerns have all been thoroughly examined and resolved."[214] While this is a welcome statement, the qualifier "likely" gives the reader reason to pause.

In a recent parliamentary debate in the United Kingdom, U.K. Minister for Counter Proliferation, Alistair Burt, stressed that the "UK does not possess fully autonomous weapon systems and has no intention of developing them . . . . As a matter of policy . . . the operation of our weapons will always be under human control as an absolute guarantee of human oversight and authority and of accountability for weapons usage."[215] Although not speaking on behalf of the U.K. government, Nia Griffith, Vice-Chair of the All Party Parliamentary Group on Weapons and Protection of Civilians, further noted, "robots may never be able to meet the requirements of international humanitarian law, as its rules of distinction and proportionality require the distinctively human ability to understand context and to make subjective estimates of value."[216]

Cumulatively, statements by government officials eschewing lethal autonomous military operations until safety, reliability, and legal concerns are fully resolved (when made by the most advanced military powers able to field such systems) can have legal significance.[217] Clear

---

214. UNMANNED SYSTEMS ROADMAP, 2009-2034, *supra* note 189, at 10.

215. 17 June 2013, PARL. DEB., H.C. (2013) 731 (U.K.), *available at* http://www.publications.parliament.uk/pa/cm201314/cmhansrd/cm130617/debtext/130617-0004.htm#130617-0004.htm_spnew1.

216. *Id.* (noting that "[t]he open-endedness of the rule of proportionality in particular, combined with complex circumstances on a battlefield, could result in undesired and unexpected behaviour by LARs. It is clear that existing law was not written to deal with LARs.").

217. The second criterion for establishing a rule of customary international law requires a general practice by states that is followed out of "a sense of legal obligation." RESTATEMENT (THIRD) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 102(2) (1987). This subjective requirement can be problematic but may often attribute considerable importance to official statements and policies of states explaining the motives behind their conduct when they are expressed in terms other than courtesy, practical considerations, policy, indifference, or mere political expediency.

and unambiguous versions of such statements, together with reluctance by technologically advanced states to actually deploy combatant systems in a fully autonomous mode—and widespread objections by other states to any such deployments—could ultimately serve as evidence of an emerging rule of customary international law banning the use of some of these systems.

Such a result, however, is far from inevitable. Another possibility is the creeping and unregulated expansion of autonomous weapons, as increasingly diverse types of lethal military technologies are activated in the autonomous mode and little or no objections are raised by states. Effective legal restraints also seem less likely in light of the ambiguous official statements of government officials in which only present "intentions" are expressed and self-imposed bans on these weapons are explicitly made only as a "matter of policy" and not as a matter of law.

Yet the complexity and unpredictability of emerging autonomous weapons are clearly giving states, including the United States, reason to carefully evaluate their design and deployment. Foremost among serious concerns are the difficult situations in which military officers and their commanders will be forced to perform their missions, as they attempt to determine if new autonomous weapons can be expected to "function appropriately" in dynamic battlefield environments, especially where civilians may be at risk.

While state responsibility under the IHL framework and the individual culpability of operators and commanders involve different concepts in many instances, the link between these two regimes for purposes of autonomous weapons is clear: a meaningful connection with the effective exercise of human judgment. Many dangerous missions envisioned for autonomous weapon systems are unlikely to permit the effective exercise of such judgment by a reasonable person. The future contours of customary international law in this area are unknown. However, the difficulties that states now face in deploying autonomous weapons systems beyond special or restricted missions will continue to form a boundary requiring, at a minimum, increasing attention to be focused on the practical and legal significance of human judgment in controlling lethal machines.