

SETTING THE STAGE:

AUTONOMOUS LEGAL REASONING IN INTERNATIONAL HUMANITARIAN LAW

*Duncan B. Hollis**

Throughout history, law has operated as a quintessentially human enterprise; a set of social conventions formulated by—and applicable to—human beings and the institutions they create.¹ For example, international humanitarian law (IHL) and the international legal order of which it forms a part both exist as social conventions created (largely) by and for a particular type of human institution—the nation State.² States and their agents (e.g., militaries) operate and interact via human behavior. And it is humans who assess the international legality of that behavior.³

The rise of automation in technology (i.e., the ability to complete tasks without human control or supervision) poses well-documented challenges to law and legal reasoning.⁴ IHL is no exception: States and scholars are increasingly

* James E. Beasley Professor of Law, Temple University School of Law. I would like to thank Eriks Selga for his research assistance as well as participants in the workshop *Autonomous Legal Reasoning? Legal and Ethical Issues in the Technologies of Conflict* for helpful comments. The workshop was co-sponsored by Temple Law School and the International Committee of the Red Cross. Additional support for both the workshop and my research came from *Metanorm: A Multidisciplinary Approach to the Analysis and Evaluation of Norms and Models of Governance for Cyberspace*, an MIT-led project funded by the Office of Naval Research through the U.S. Department of Defense Minerva Research Initiative. All views expressed in this essay remain, however, entirely my own.

1. The definition of law itself has proved elusive, with deep philosophical divisions over the need for sovereign sanctions, the role of social “facts” in establishing authority, and law’s relationship with morality. *See, e.g.*, JOSEPH RAZ, *BETWEEN AUTHORITY AND INTERPRETATION: ON THE THEORY OF LAW AND PRACTICAL REASON* (2009); H.L.A. HART, *THE CONCEPT OF LAW* (3d ed. 2012); RONALD DWORKIN, *LAW’S EMPIRE* (1986); JOHN AUSTIN, *THE PROVINCE OF JURISPRUDENCE DETERMINED* 101 (Ashgate 1998) (1832).

2. I say “largely” because modern international law no longer consists solely of laws made by and for States; it reaches institutions that States create (e.g., international organizations or “IOs”) as well as human rights and responsibilities. *See, e.g.*, Duncan B. Hollis, *Why State Consent Still Matters – Non-State Actors, Treaties, and the Changing Sources of International Law*, 23 *BERKELEY J. INT’L L.* 137 (2005).

3. Individuals may do this in various capacities, whether as authoritative interpreters (e.g., international jurists), agents of States (e.g., diplomats, legal advisers), academics, or representatives of civil society.

4. *See, e.g.*, Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 *HARV. J. L. & TECH.* (forthcoming 2016). These challenges include—but are not limited to—those emerging from earlier technological innovations, such as the Internet. *See, e.g.*, Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 *CALIF. L. REV.* 513 (2015).

attentive to the international legal issues raised by the autonomous aspects of weapon systems.⁵ So far, this attention has emphasized *functional* issues, asking if existing IHL can serve its regulatory function given the capacities autonomous systems have (or may develop).⁶ But those same capacities require a *conceptual* inquiry as well. How do we conceive of law and legal reasoning if it occurs with limited (or no) human interactions? Under what circumstances (if any) would we qualify the acts of autonomous systems themselves as human and therefore subject to the requisite social conventions, including IHL?

The question of how to evaluate “human acts” has occupied scholars for centuries. Long before journalism students asked *who, what, where, when, how,* and *why*, Saint Thomas Aquinas’ *Summa Theologiae* used those same questions as the foundation for evaluating the circumstances of “human acts” according to natural (or divine) law.⁷ A similar analytic frame may help us evaluate the behavior of autonomous systems according to IHL.

To be clear, I do not propose to resolve whether the acts of autonomous systems deserve the label “human.” My proposal is more modest. In asking how IHL responds to these six questions, I hope to set the stage for a broader conversation on the extent to which autonomous systems analogize to human behavior beyond the conclusory claims made to date.⁸ The Aquinas framework

5. See The United Nations Office at Geneva, *Background – Lethal Autonomous Weapons Systems*, [http://www.unog.ch/80256EE600585943/\(httpPages\)/8FA3C2562A60FF81C1257CE600393DF6?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/8FA3C2562A60FF81C1257CE600393DF6?OpenDocument) (listing papers and reports on IHL and lethal autonomous weapons systems); INT’L COMM. OF THE RED CROSS, REPORT OF THE ICRC EXPERT MEETING ON ‘AUTONOMOUS WEAPON SYSTEMS: TECHNICAL, MILITARY, LEGAL AND HUMANITARIAN ASPECTS’, 26–28 MARCH 2014, GENEVA (2014) (describing the challenges of autonomous weapons for government leaders and weapons experts) [hereinafter “ICRC Experts Meeting”]; Rebecca Crotoof, *The Killer Robots Are Here: Legal and Policy Implications*, 36 CARDOZO L. REV. 1837 (2015).

6. This emphasis is understandable given the divergent answers generated so far, especially in response to calls for a ban on fully autonomous lethal weapons. See *Losing Humanity: The Case Against Killer Robots*, HUMAN RTS. WATCH (Nov. 2012), http://www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf (arguing that autonomous weapons would not comply with current international humanitarian law); 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 INSTITUTION (2013), http://www.hoover.org/sites/default/files/uploads/documents/Anderson-Waxman_LawAndEthics_r2_FINAL.pdf (arguing against a ban on fully autonomous lethal weapons in favor of applying extant IHL); Michael A. Newton, *Back to the Future: Reflections on the Advent of Autonomous Weapons Systems*, 47 CASE WESTERN RES. J. INT’L L. 5 (2015) (exploring salutary effects of autonomous systems for IHL).

7. For Aquinas, the circumstances of “human” acts were defined by their voluntary character. THOMAS AQUINAS, II-I SUMMA THEOLOGIAE Q6 (Fathers of English Dominican Province trans., Christian Classics 1981) (1485). Aquinas proposed evaluating such acts according to seven questions: “*Who, what, where, by what aids, why, how, and when.*” *Id.* at Q7(3) (citing Cicero). Cicero offered similar categories, but not in the form of questions. D.W. Robertson, Jr., *A Note on the Classical Origin of “Circumstances” in the Medieval Confessional*, 43 STUDIES IN PHILOLOGY 6, 6–7 (1946). In this article, I focus on six questions, treating “by what aids” as a version of the “how” question.

8. For example, advocates of a ban on fully autonomous lethal weapons insist, without

provides a ready-made platform to pose new questions about the relationship between IHL and autonomous systems and to reformulate others that have already engendered scholarly attention. In doing so, this article (together with the others presented in this symposium) illustrates both the scope and limitations of current international legal thinking on autonomous systems. I conclude with a call for more crosscutting and interdisciplinary research on the ways in which autonomous systems relate to the “human” portion of IHL.

WHO? DEFINING THE SUBJECTS OF IHL IN THE AUTONOMOUS CONTEXT

Experts agree that there is “no doubt that the development and use of autonomous weapon systems in armed conflict is governed by international humanitarian law (IHL).”⁹ But *who* exactly is subject to IHL? Certainly, States are constrained and empowered by IHL’s terms, whether via treaty or customary international law.¹⁰ For example, States must give new weapons, including autonomous ones, a legal review to ensure that they are not unlawful per se (i.e., they are neither indiscriminate nor the cause of unnecessary suffering).¹¹

explanation, that IHL’s assessments and judgments require “uniquely human” behavior that autonomous systems can never achieve because they are not human. *See e.g.*, ICRC Experts Meeting, *supra* note 5, at 21; *Losing Humanity*, *supra* note 6, at Pt. IV. *See also* Vik Kanwar, *Review Essay: Post-Human Humanitarian Law: The Law of War in the Age of Robotic Weapons*, 2 HARV. NAT’L SEC. J. 616, 620 (2011) (With respect to IHL regulation of autonomous weapons, for “a series of partially coherent reasons, the ‘human element’ is seen as ‘indispensable’; for providing judgment, restraint, and ultimately responsibility for decisions.”).

9. ICRC Experts Meeting, *supra* note 5, at 8.

10. *See, e.g.*, Geneva Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, Aug. 12, 1949, 75 U.N.T.S. 31; JEAN-MARIE HENCKAERTS & LOUISE DOSWALD-BECK, 1 CUSTOMARY INTERNATIONAL HUMANITARIAN LAW (2005). When it comes to new technologies, moreover, States regularly wrestle with *which* IHL rules apply, including persistent questions about the role of the Martens Clause. ICRC, *Background Paper*, in ICRC Experts Meeting, *supra* note 5, at 92–94. Named after Russia’s Friedrich Martens, the Martens Clause provides that “the principles of humanity” and the “dictates of public conscience” provide IHL protections even in the absence of specific treaty provisions. *See, e.g.*, Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflict (Protocol I), June 8, 1977, 1125 U.N.T.S. 3, art. 1(2) [hereinafter “AP I”]. Although the Clause appears in numerous IHL treaties and the International Court of Justice has endorsed it as customary international law, scholars disagree on how the Martens Clause can or should regulate autonomous weapon systems. *See* ICRC, *Background Paper*, in ICRC Experts Meeting, *supra* note 5, at 92–94 (“[T]he exact interpretation of the Martens Clause remains subject to significant variation among experts.”) [hereinafter “ICRC Background Paper”]; *see also* Michael Schmitt, *Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics*, HARV. NAT’L SEC. J. FEATURES 31–32 (2013); *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, 1996 ICJ Rep. ¶ 78(h) (Jul. 8).

11. AP I, *supra* note 10, art. 35(1) (prohibiting weapons “of a nature to cause superfluous injury or unnecessary suffering”); art. 51(4)(b)-(c) (prohibiting “indiscriminate” weapons). For more on IHL and weapons reviews in the autonomous context see (in this volume) Michael W. Meier, *Lethal Autonomous Weapons Systems (LAWS): Conducting a Comprehensive Weapons Review*, 30 TEMP. INT’L & COMP. L. J. (2016).

IHL also clearly subjects individuals to its regulations, whether as agents of States (e.g., combatants) or in some other role (e.g., children, journalists).¹² Assuming a lawful weapon, IHL mandates that its human operators employ it with discrimination and precautions in a manner proportionate to military objectives.¹³ The capacity of humans to do this as weapons become more autonomous has become *the* central question for IHL in this context.¹⁴ The dominant typology thus differentiates among technologies where there is a “human-in-the-loop” (semi-autonomous systems where a human controls the technology as it operates); a “human-on-the-loop” (human supervised autonomous systems where a human can intervene and alter or terminate operations); and a “human-out-of-the-loop” (fully autonomous systems that operate independent of human control).¹⁵ Of course, saying that individuals must comply with IHL in designing and using autonomous weapons does not mean this task is an easy one. The greater a system’s autonomy, the more ambiguities, tensions, and novel questions it raises about IHL’s meaning vis-à-vis individuals. Nonetheless, much of the existing IHL literature focuses on these questions from the human perspective, including, most notably, questions about a requirement of “meaningful human control” over autonomous weapon systems.¹⁶

We could, however, address the “who” question quite differently. Instead of limiting IHL’s subjects to social institutions like States and the humans who represent them, why not interpret it to reach autonomous systems directly?¹⁷

12. See, e.g., AP I, *supra* note 10, art. 43(2) (right of combatants to participate in hostilities); art. 44(3) (duty of combatants to distinguish themselves from the civilian population); art. 77 (protection of children); art. 79 (measures for protection of journalists). The advent of international criminal law, moreover, signals the growing reach of IHL over individual behavior in armed conflicts. See *generally* Rome Statute of the International Criminal Court, July 17, 1998, 2187 U.N.T.S. 3.

13. See ICRC Background Paper, *supra* note 10, at 77 (finding that IHL rules of distinction, proportionality and precautions apply); see also AP I, *supra* note 10, art. 48 (distinction); art. 51(5) (proportionality); art. 57 (precaution).

14. See *generally* ICRC Experts Meeting, *supra* note 5, at 8 (describing current debates regarding the importance of human control over autonomous weapons).

15. Paul Scharre & Michael C. Horowitz, *An Introduction to Autonomy in Weapon Systems* 8 (CTR. FOR A NEW AM. SEC., Working Paper No. 021015, 2015), http://www.cnas.org/sites/default/files/publications-pdf/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf; *Losing Humanity*, *supra* note 6, at 2 (providing “human-in-the-loop,” “human-on-the-loop,” and “human-out-of-the-loop” definitions while acknowledging the nature of the categories is debated); ICRC Experts Meeting, *supra* note 5, at 63 (defining and differentiating among autonomous weapon systems, human supervised autonomous weapon systems, and semi-autonomous weapon systems).

16. See, e.g., Rebecca Crootof, *The Meaning of “Meaningful Human Control,”* 30 TEMP. INT’L & COMP. L. J. (2016); Bryant Walker Smith, *Controlling Humans and Machines*, 30 TEMP. INT’L & COMP. L. J. (2016); Michael C. Horowitz & Paul Scharre, *Meaningful Human Control in Weapon Systems: A Primer*, (CTR. FOR A NEW AM. SEC., Working Paper No. 031315, 2015), http://www.cnas.org/sites/default/files/publications-pdf/Ethical_Autonomy_Working_Paper_031315.pdf.

17. See Patrick Lin et al., *Autonomous Military Robotics: Risk, Ethics, and Design*, 55 et seq. CAL. POLYTECHNIC ST. U. (Dec. 2008), <http://www.unog.ch/80256EDD006B8954>

Indeed, science fiction fans will recall that the author Isaac Asimov defined his career by introducing the “Laws of Robotics”¹⁸—where robots themselves (rather than their designers) were the subjects of the rules.¹⁹

Some will object on the grounds that these systems are not (or cannot) become the subjects of IHL.²⁰ But humanity already has a long history of constituting and regulating all sorts of “legal” persons whether States and international organizations under international law, or corporations and trusts under domestic law. Of course, these “legal” persons are still social institutions subject to human control in ways that will be limited (or perhaps at some point absent) in the autonomous context. Moreover, subjecting autonomous systems directly to the requirements of, say, proportionality, would complicate already difficult questions of responsibility for the harms such systems cause.²¹

That said, nothing about law generally (or IHL in particular) precludes holding multiple actors responsible for the same conduct. Autonomous systems *could* be governed by specific IHL rules alongside the rules IHL has for States and human actors just like a corporation may be held accountable alongside—and in addition to—the fiduciary duties of its directors and officers.²² The fact that

(/httpAssets/A70E329DE7B5C6BCC1257CC20041E226/\$file/Autonomous+Military+Robotics+Risk,+Ethics,+and+Design_lin+bekey+abney.pdf (discussing the ethical, as opposed to legal, perspective of IHL reaching autonomous systems directly).

18. The “Laws of Robotics” first appeared in Asimov’s fourth robot short story entitled “Runaround.” Roger Clarke, *Asimov’s Laws of Robotics: Implications for Information Technology, Part I*, 26 IEEE Computer 53-61 (1993). The “Laws of Robotics” are programmed into the robot’s “platinum-iridium positronic brains.” ISAAC ASIMOV, RUNAROUND 33–51 (1942).

19. See Clarke, *supra* note 18, at 53 (assessing the practical potential of Asimov’s work in modern technology).

20. See Marco Sassoli, *Can Autonomous Weapon Systems Respect the Principles of Distinction, Proportionality and Precaution?*, in ICRC Experts Meeting, *supra* note 5, at 42 (“The feasibility of precautions must be understood to refer to what would be feasible for human beings using the machine, not to the options available to the machine.”); *id.* at 43 (opining that “all rules of IHL are addressed only to human beings”); *id.* at 41 (suggesting that only humans can be inhumane and can deliberately choose not to comply with the rules they were given).

21. See, e.g., *Mind the Gap: The Lack of Accountability for Killer Robots*, HUMAN RTS. WATCH (Apr. 9, 2015), <https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots> (noting that proportionality analysis involves a highly contextual, subjective determination and the difficulty in programming an autonomous weapon to be able to respond to every possible scenario); Crotoft, *supra* note 5, at 1878–84 (discussing the current inability of robots to weigh the anticipatable harm with the military advantage in military situations); Sassoli, *supra* note 20, at 42 (explaining that the plans and development of military operations on both sides of a conflict constantly change and preclude a robot from adequately applying the proportionality principle); Schmitt, *supra* note 10, at 18–21 (considering the human judgment and psychological process needed to evaluate proportionality).

22. Exploring the application of IHL to autonomous systems requires close attention to their actual capacities—we have yet to (and may never) see a system matching Hollywood images of artificial intelligence replicating the intellectual capacities of a human. *Part I: Summary Report by the International Committee of the Red Cross*, in ICRC Experts Meeting *supra* note 5, at 7. Nonetheless, to the extent that weapon systems gain increasing (even if not full) autonomy in

autonomous system behaviors will be unpredictable (and maybe even emergent) strengthens the case for rules that go beyond system designers and operators and speak to the systems themselves.²³ At a minimum, it would seem worthwhile to expand existing research to inquire as to the relative costs and benefits of having IHL rules (whether those in existence or some more precisely tailored ones) regulating autonomous weapon systems directly in lieu of the conventional wisdom that would leave IHL to human subjects who create or operate these systems. The answer may be that truly autonomous systems are not yet sufficiently realistic to warrant the coverage, but it remains a question worth asking.

WHAT? WHAT AUTONOMOUS SYSTEMS DOES IHL NEED TO ADDRESS?

What is an autonomous weapons system? Although definitional debates continue, the U.S. Department of Defense definition provides a useful starting point: “a weapon system that, once activated, can select and engage targets without further intervention by a human operator.”²⁴ Without further qualifications, this definition would appear to cover a wide range of systems. Yet, the relevant IHL discussions have (mostly) treated the issue more narrowly by focusing on lethal systems, and, then, only those of a particular type (i.e., kinetic systems).²⁵ The focus on technologies that independently target and kill humans is easy to appreciate given popular attention to the prospect of “killer robots,” even if no such fully autonomous systems yet exist. The lethal capacity of remotely piloted aircraft (or “drones” to use the colloquial term) has further stirred that pot, even though their weaponry is separated from their semi-autonomous flight functions.²⁶

their operations, there is utility in giving more attention to how the systems themselves may understand and apply specific parameters for their operations.

23. Experts have noted, for example, that “programs with millions of lines of code are written by teams of programmers, none of whom knows the entire program; hence, no individual can predict the effect of a given command with absolute certainty, since portions of large programs may interact in unexpected, untested ways.” Gary E. Marchant et al., *International Governance of Autonomous Military Robots*, 12 COLUM. SCI. & TECH. L. REV. 272, 284 (2011). Of course, such unpredictability may extend to how an autonomous system interprets any IHL rules that cover it directly. *Id.* But that variation seems less risky assuming IHL continues to also regulate programmers and operators on the theory that two regulatory levers may be more effective than one. *Id.*

24. See U.S. DEP'T OF DEF., DIR. 3000.09, AUTONOMY IN WEAPONS SYSTEMS 13–14 (Nov. 21, 2012) [hereinafter DoD DIRECTIVE 3000.09] (“This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation”); See *Part I: Summary Report by the International Committee of the Red Cross*, in ICRC Experts Meeting *supra* note 5, at 7 (noting that “[t]here is no internationally agreed definition of autonomous weapon systems. For the purposes of the meeting, ‘autonomous weapon systems’ were defined as weapons that can independently select and attack targets, i.e. with autonomy in the ‘critical functions’ of acquiring, tracking, selecting and attacking targets.”).

25. See, e.g., *Part I: Summary Report by the International Committee of the Red Cross*, in ICRC Experts Meeting *supra* note 5, at 7; ICRC Background Paper, *supra* note 10, at 64.

26. See, e.g., ICRC Background Paper, *supra* note 10, at 64 (describing the ability of remotely piloted aircraft to engage weaponry via remotely located human operators).

More importantly, almost all the discussion of IHL and autonomous weapon systems has occurred in *kinetic* contexts (involving robotic systems, whether those on the ground, at sea, or in the air). Thus, States have used the Conventional Weapons Convention (CWC) as the primary forum for discussing this technology.²⁷ These conversations appear to presume that the autonomous characteristics of kinetic weapons, such as those of Israel's Iron Dome missile defense system, are what IHL needs to address most.²⁸ Indeed, the United States has even cabined its policy on autonomy in weapon systems to cover kinetic systems, while excluding "autonomous or semi-autonomous cyberspace systems for cyberspace operations."²⁹

It may be a mistake, however, to ignore cyber capabilities in asking what IHL regulates in terms of autonomy. As Eric Messinger has noted, the increasing speed of cyber operations will likely require more fully autonomous systems in both defensive *and* offensive scenarios such that a call to remove autonomy from the equation "is arguably tantamount to advocating a ban on cyberwarfare altogether."³⁰ At the same time, existing malware may offer examples of more fully autonomous systems than the kinetic context, where such a capacity is usually still described as unavailable. Jason Healy labeled Stuxnet, for example, as "the first autonomous weapon with an algorithm, not a human hand, pulling the trigger."³¹

27. See, e.g., Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, Mar. 11, 2015, *Revised Annotated Programme of Work for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems Geneva, 13–17 April 2015*, CCW/MSP/2015/WP.1/Rev.1.

28. The Iron Dome works by firing intercepting missiles to destroy incoming rockets before they can reach their target. Michael Martinez & Josh Levs, *How Iron Dome Blocks Rockets from Gaza, Protects Israelis*, CNN (July 9, 2014, 10:23 AM), <http://www.cnn.com/2014/07/09/world/meast/israel-palestinians-iron-dome/index.html>. The Iron Dome operates automatically according to pre-programmed parameters, subject to human operator intervention. Schmitt, *supra* note 10, at 4 n.11.

29. DoD DIRECTIVE 3000.09, *supra* note 244, at 2(b) (also excluding unarmed, unmanned platforms; unguided munitions; munitions manually guided by the operator (e.g., laser- or wire-guided munitions); mines; or unexploded explosive ordnance). Given that the United States is one of the first (and few) States to have any public policy on autonomous weapons, this distinction may prove influential as other States begin to devise their own policies in this area.

30. See Eric Messinger, *Is It Possible to Ban Autonomous Weapons in Cyberwar?*, JUST SECURITY (Jan. 15, 2015, 9:27 AM), <https://www.justsecurity.org/19119/ban-autonomous-weapons-cyberwar/>.

31. See Jason Healy, *Stuxnet and the Dawn of Algorithmic Warfare*, HUFFPOST TECH BLOG (Apr. 16, 2013), http://www.huffingtonpost.com/jason-healey/stuxnet-cyberwarfare_b_3091274.html. Paul Walker, in contrast, views Stuxnet as a semi-autonomous weapon at most, as he explains in this volume. See Paul A. Walker, *Military Necessity in the Digital Age*, 30 TEMP. INT'L & COMP. L. J. (2016). Stuxnet was a form of malware that infected industrial control systems globally and then executed a payload on just one specific system (i.e., the one at Iran's Natanz nuclear enrichment facility), instructing centrifuges there to run at various—and unsustainable—speeds, causing over 1000 to break in the process. KIM ZETTER, COUNTDOWN TO ZERO DAY: STUXNET AND THE LAUNCH OF THE WORLD'S FIRST DIGITAL WEAPON 341-42 (2014).

Of course, Stuxnet's compliance with IHL has proved a thorny subject.³² Nonetheless, the potential for cyber operations to "select and engage targets" with limited or no human intervention suggests a need to broaden existing conversations on what autonomous systems IHL regulates beyond the conventional weapons context.³³

WHERE? LOCATING THE LEGAL PROCESSES FOR AUTONOMOUS SYSTEMS

One might frame the "where?" question in jurisdictional terms—asking where geographically IHL regulates autonomous weapon systems. Doing so would implicate larger debates over IHL's geographic scope, albeit not necessarily in any way that creates issues unique to the circumstances of autonomous weapons.³⁴ But the "where?" question can also be raised in more novel terms by asking where the discourse on IHL's operations vis-à-vis autonomous weapon systems should occur.

In recent years, both international law and international relations scholarship have explored the emergence of different governance regimes, and the various settings in which legal processes can occur.³⁵ That literature has revealed just how much the locale for autonomous legal reasoning matters. The movement to "ban killer robots" favors a multilateral treaty conference, presumably because that terrain proved favorable to the attainment of the Land Mines Convention.³⁶ At the other end of the spectrum, some States may prefer the Westphalian system's default terrain: auto-interpretation, where, in the absence of agreement, each State is free to interpret what IHL requires for itself.³⁷ This may lead different States to

32. See, e.g., TALLINN MANUAL ON THE INTERNATIONAL LAW APPLICABLE TO CYBER WARFARE 58, 83–84 (Michael N. Schmitt ed., 2013) (noting how the International Group of Experts that authored the *Tallinn Manual* was divided over the characterization of Stuxnet for IHL purposes).

33. Crootof, *supra* note 5, at 1840.

34. See, e.g., Laurie R. Blank, *Debates and Dichotomies: Exploring the Presumptions Underlying Contentions about the Geography of Armed Conflict*, in 16 YEARBOOK OF INTERNATIONAL HUMANITARIAN LAW 297 (Terry D. Gill ed., 2013); Geoffrey S. Corn, *Geography of Armed Conflict: Why It Is a Mistake to Fish for the Red Herring*, 89 INT'L L. STUD. 77 (2013).

35. See, e.g., *The Concept of Legalization*, 52 INT'L ORG. 613 (Kenneth Abbott et al. eds., 2015); JOSE E. ALVAREZ, INTERNATIONAL ORGANIZATIONS AS LAW-MAKERS (2005).

36. See, e.g., 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. 241; Ken Anderson, *The Ottawa Convention Banning Landmines, the Role of International Non-Governmental Organizations and the Idea of International Civil Society*, 11 EUR. J. INT'L L. 91 (2000); Stuart Maslen & Peter Herby, *An International Ban on Anti-Personnel Mines: History and Negotiation of the "Ottawa Treaty"*, 325 INT'L REV. RED CROSS 1 (1998), <https://www.icrc.org/eng/resources/document/article/other/57jpn.htm>.

37. The international legal order continues to lack universal, centralized, legislative and adjudicatory bodies that could definitively delineate the sources of law and judge their content. As Leo Gross noted a half century ago, we are left in a situation where, in the absence of such authorities, "each state has a right to interpret the law, the right of autointerpretation, as it might be called." Leo Gross, *States as Organs of International Law and the Problem of Autointerpretation*, in *Essays on international law and organization* volume 1, 386 (1984).

reach different conclusions on how IHL approaches autonomous weapon systems (*cf.*, the U.S. and U.K. positions³⁸) without necessarily favoring any particular outcome.³⁹

In between these poles lie a host of transnational and international regimes for legal discussions on IHL and autonomy. Domestic regulatory officials could cooperate transnationally to regulate autonomous technologies in addition to IHL itself (e.g., harmonizing export controls or government contracting rules to limit the development or deployment of indiscriminate autonomous weapon systems).⁴⁰ Or, States could pursue clarifications or alterations to IHL in bilateral, plurilateral, or multilateral contexts. Alternatively, actors may “graft” the conversation into an existing institutional setting as they have already done with the CWC. But the CWC has no exclusive claim to this topic. Other fora (e.g., the U.N. Human Rights Council) could easily take up the issue with different goals and participants (not to mention outcomes) than those on display at the CWC so far.⁴¹

In addition to formal settings, non-legal processes may take center stage, whether among non-State actors (e.g., Non-Governmental Organizations) alone or in public-private partnerships. Even without the formal imprimatur of “law,” these settings can create “epistemic communities” which devise common understandings of existing law or new norms.⁴² New norms can, in turn, serve as a precursor to the adoption of international law by States via treaties or practice, or they may have an extra-legal character (e.g., moral or professional norms) that still constrain or empower what various actors consider appropriate boundaries of behavior.⁴³ Of

38. Compare DoD DIRECTIVE 3000.09, *supra* note 244, at 4(a) (“Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force.”), with U.K. Ministry of Defence, *Written Evidence from the Ministry of Defence Submitted to the House of Commons Defence Committee Inquiry ‘Remote Control: Remotely Piloted Air Systems – Current and Future UK Use’* (2013) (noting current U.K. policy against “autonomous release of weapons” and adequacy of existing IHL in regulating new autonomous technologies).

39. Provided, of course, that States stay within the acceptable boundaries of international legal discourse. See, e.g., Andrea Bianchi, *The Game of Interpretation in International Law: The Players, the Cards, and Why the Game Is Worth the Candle*, in INTERPRETATION IN INTERNATIONAL LAW 34, 55 (Andrea Bianchi, Daniel Peat & Matthew Windsor eds., 2015) (“Social consensus exists nowadays amongst the players on some fundamental aspects of the game of interpretation in international law.”).

40. See, e.g., ANNE-MARIE SLAUGHTER, *A NEW WORLD ORDER* (2004); Kenneth W. Abbott & Duncan Snidal, *Strengthening International Regulation through Transnational New Governance: Overcoming the Orchestration Deficit*, 42 VANDERBILT J. TRANS’L L. 501 (2009).

41. See, e.g., Amitav Acharya, *How Ideas Spread: Whose Norms Matter? Norm Localization and Institutional Change in Asian Regionalism*, 58 INT’L ORG. 239 (2004); Richard Price, *Reversing the Gun Sights: Transnational Civil Society Targets Land Mines*, 52 INT’L ORG. 613 (1998).

42. See, e.g., Peter M. Haas, *Introduction: Epistemic Communities and International Policy Coordination*, 46 INT’L ORG. 1 (1992).

43. See Marchant et al., *supra* note 233, at 291–93 (discussing internal professional codes and norms); see generally Martha Finnemore & Kathryn Sikkink, *International Norm Dynamics and Political Change*, 52 INT’L ORG. 887 (1998).

course, IHL also has a *sui generis* setting where the International Committee of the Red Cross can play host to meetings of experts like the one on which this symposium was based.⁴⁴

For now, the autonomous weapons conversation has prioritized substantive discourse over its location. The lessons from other global governance contexts suggest, however, that the latter issue deserves closer scrutiny. States and scholars alike need to appreciate the trade-offs involved in employing different international fora for discussions of law and autonomous weapons technology.

WHEN? THE TEMPORAL QUESTIONS ON IHL AND AUTONOMOUS SYSTEMS

The temporal question is among the most prominent and long-standing issue in the existing discourse. Putting it simply, *when* should IHL deal with autonomous weapons: now or later? Proponents of a ban on autonomous lethal weapons want regulation immediately, rather than waiting to see how existing IHL deals with the particular manifestations of the technology in practice.⁴⁵ Others (e.g., Professors Kenneth Anderson and Matthew Waxman) argue for a more incremental approach with a “gradual evolution and adaptation of long-standing law of armed conflict principles—to regulate what seems to many like a revolutionary technological and ethical predicament.”⁴⁶

But the temporal issues for autonomy and IHL need not be limited to the binary choice of now versus later. The question can be reformulated to ask more generally about *when* existing IHL applies. So far, I have discussed autonomous weapon systems in light of the conventional temporal conditions for IHL’s application, namely when (i) an armed conflict exists; and (ii) a weapon is used, designed, or intended to be used in an “attack.” Traditionally, attacks occur when actions produce violent consequences (e.g., injury, death, damage, or destruction) and operate as a prerequisite for applying core IHL principles such as distinction, proportionality, and precaution.⁴⁷ Thus, when an autonomous system qualifies as a weapon capable of being used in an attack, IHL’s rules clearly apply.

Does it follow, however, that when an autonomous system does not operate as

44. See, e.g., ICRC Experts Meeting; News Release, ICRC, *Cluster Munitions: ICRC Hosts Meeting of Experts* (Apr. 17, 2007) (on file with author), <https://www.icrc.org/eng/resources/documents/news-release/2009-and-earlier/cluster-munition-news-170407.htm>.

45. See, e.g., *Autonomous Weapons: An Open Letter from AI & Robotics Researchers*, FUTURE OF LIFE INST., <http://futureoflife.org/open-letter-autonomous-weapons/> (last visited Mar. 7, 2016).

46. Anderson & Waxman, *supra* note 6, at 27; see also Kenneth Anderson, Daniel Reisner & Matthew Waxman, *Adapting the Law of Armed Conflict to Autonomous Weapon Systems*, 90 INT’L L. STUD. 386, 410 (2014) [hereinafter “Anderson et al.”].

47. See, e.g., Michael N. Schmitt, *Cyber Operations and the Jus in Bello: Key Issues*, 87 INT’L L. STUD. 89, 91 (2011); Robin Geiss & Henning Lahmann, *Cyber Warfare: Applying the Principle of Distinction in an Interconnected Space*, 45 ISR. L. REV. 381 (2012). For a view that IHL is not so limited, see HEATHER HARRISON DINNISS, *CYBER WARFARE AND THE LAWS OF WAR 196–202* (2012).

a weapon (i.e., it does not injure, kill, damage, or destroy), it operates free from the constraints of IHL? This might occur in the cyber context, for example, depending on how damage is defined. A majority of the *Tallinn Manual*'s experts suggested that damage "requires replacement of physical components."⁴⁸ Thus, if an autonomous system temporarily shut down a factory (or a stock exchange) via cyber means without any direct, physical harm, it would not generate any damages and thus avoid the "attack" label. And if a cyber-operation is not an attack, it would lie outside IHL's constraints; no legal review would be needed to deploy it, nor would the principle of discrimination preclude it from targeting civilians. Indeed, if the system engaged in isolation, it could avoid triggering an armed conflict, and thus escape IHL's regulation entirely.

Defining autonomous systems entirely by analogy to the effects of existing weapons risks creating behavioral incentives that are at odds with IHL's core principles. Prudence suggests further consideration is needed as to the extent of these risks. In particular, we need to evaluate whether the novel aspects of autonomous systems might require the adjustment of existing interpretations rather than reflexively perpetuating existing legal boundaries by analogy to when IHL has applied in the past.⁴⁹

How? By what means can IHL regulate autonomous systems?

Much of the legal reasoning surrounding autonomous systems in armed conflicts has analyzed the *how* question, asking how IHL restricts what autonomous systems might do. In doing so, conventional wisdom has treated IHL as an undifferentiated set of international legal obligations. But IHL's regulatory means can (and should) be disaggregated. For example, IHL is not an entirely proscriptive legal regime—it regulates as much by permission and empowerment as it does by prohibition and limitation.⁵⁰ As such, assessing autonomous systems' potential only in terms of their relationship to IHL's proscriptive aspects may miss the potential for these systems to operate in as yet unexplored ways (e.g., as medical care units on a battlefield).

Perhaps even more importantly, IHL regulations arise with different levels of precision, which may be grouped into rules, standards, and principles.⁵¹ Rules seek

48. *Tallinn Manual*, *supra* note 32, at 108–109, ¶ 10.

49. I have written extensively elsewhere about the costs and benefits of analogic reasoning in the face of new technologies, most notably cyber capabilities. *See, e.g.*, Duncan B. Hollis, *Re-Thinking the Boundaries of Law in Cyberspace: A Duty to Hack?*, in *CYBERWAR: LAW & ETHICS FOR VIRTUAL CONFLICTS* 129 (Jens David Ohlin et al. eds., 2015).

50. For example, IHL may prohibit indiscriminate weapons and limit the use of lawful weapons to proportionate attacks, but it also allows (without requiring) the killing of enemy combatants, just as it empowers those who would provide medical care to victims. *See, e.g.*, AP I, *supra* note 10, at arts. 15–16 (protections for civilian medical personnel and persons engaged in medical activities); art. 42(3) (combatant right to participate in hostilities).

51. Here, I am drawing on Dan Bodansky's work as well as the earlier scholarship of Kathleen Sullivan. *See, e.g.*, Daniel Bodansky, *Rules vs. Standards in International Environmental Law*, 98 AM. SOC'Y INT'L L. PROC. 275 (2004); Kathleen M. Sullivan, *The*

to bind their subjects to respond in specific, determinate ways when certain facts exist. Rules thus operate *ex-ante*; once the facts are clear, so too is the expected behavior. Standards, in contrast, work best after the fact, as an *ex-post* evaluation encompassing a wider range of circumstances or background values and policies. Principles, meanwhile, set forth broad considerations for evaluating future behavior without delineating a precise norm for the behavior itself. All three types of regulations exist in IHL. Compare, for example, the rule of perfidy that prohibits killing, injuring or capturing an adversary after pretending to seek negotiations under a flag of truce.⁵² Compliance with this rule in an operation is much simpler to evaluate in advance than for those who must apply the standard of proportionality since the latter involves evaluating and balancing an array of circumstances that may not be fully clear until after the fact.⁵³ The principles of military necessity and humanity, meanwhile, operate as fundamental guidelines in conducting *all* military operations.⁵⁴

The differing mechanisms by which rules, standards, and principles operate may have important implications for the relationship between autonomous systems and IHL. For example, requiring such systems to obey rules may pose less risk of unpredictable behavior by an autonomous system than directives that apply a principle. On the other hand, principles by their breadth retain a flexibility to accommodate novel technologies that rules, with their rigidity, often lack. Standards, meanwhile, involve judgments that appear best made not by the autonomous system alone, but by third parties authorized to evaluate whether the system adequately accounted for factors such as “foreseeability” or “reasonableness.” Such differences suggest a need for more robust analysis of *how* IHL regulates autonomous systems than has occurred to date. It will be important to differentiate, for example, those circumstances in which rules, standards, and principles can effectively regulate autonomous weapons versus those cases where one or more of these mechanisms may prove ill-suited to that task.

WHY? EXPLORING THE REASONS FOR IHL’S REGULATION OF AUTONOMOUS SYSTEMS

Of all the circumstances that Aquinas defined for evaluating acts “properly called human,” he listed “why” as the “most important of all.”⁵⁵ The same may be true in why IHL needs to evaluate autonomous systems. There are at least three different justifications offered: (i) a *military* rationale; (ii) a *humanitarian* rationale; and (iii) an account focused on *practical necessity*. For starters,

Justices of Rules and Standards, 106 HARV. L. REV. 22, 57–59 (1992).

52. AP I, *supra* note 10, art. 37.

53. *See id.* art. 51(5)(b) (stating that the standard of proportionality prohibits attacks *expected* to cause loss of civilian life, injury to civilians, damage to civilian objects, or a combination of these, which would be *excessive* in relation to concrete and direct military advantage *anticipated*).

54. *See, e.g.,* Michael N. Schmitt, *Military Necessity and Humanity in International Humanitarian Law: Preserving the Delicate Balance*, 50 VA. J. INT’L L. 795, 796 (2010).

55. Aquinas, *supra* note 7, Q7(4).

autonomous systems may provide a distinct military advantage, allowing militaries to pursue their necessary (and lawful) ends more efficiently. Hence, proponents emphasize the advantages of autonomous systems over human actors in terms of their enhanced sensory and processing capabilities; their ability to operate free of human emotions that cloud judgement; and their potential to operate free from the need for self-preservation that dominates human behavior.⁵⁶

On the flip side, autonomous systems may be justified for their humanitarian potential: their capacity to improve the effectiveness of IHL regulations aimed at minimizing human suffering. Several scholars have emphasized how autonomous systems may operate “to foster compliance with international humanitarian law.”⁵⁷ Autonomous systems might, for example, prove more effective at discriminating among military and civilian objects than the *status quo* given existing human error rates.⁵⁸ Or, they might provide capabilities with less harmful effects than extant means and methods of warfare, requiring a recalibration of the requisite precaution and proportionality analyses.

These two reasons—military necessity and humanity—constitute the core justifications for IHL and balancing these principles has long served as its chief objective.⁵⁹ The advent of autonomous systems requires renewed attention, therefore, not simply to explaining why these systems advance one goal or the other, but to understanding, in a more holistic sense, how they affect the balance between them.

Beyond these two traditional rationales, a third one—*practical necessity*—looms large. Simply put, the technology is coming (albeit incrementally) whether States want it or not.⁶⁰ Indeed, I have already noted the linkage between cyber operations and autonomy that suggests one cannot exist without the other. To the extent that technologies operate at speeds or scales in excess of human capacities, autonomous systems may be the only practical response.

To be clear, I am not suggesting that these technologies come without risks, nor am I trying to minimize the significant legal and ethical questions they

56. Marchant et al., *supra* note 233, at 279–80 (“[F]or a variety of reasons . . . in the future autonomous robots may be able to perform better than humans.”). For a discussion of autonomous weapons and military necessity more generally, see Walker, *supra* note 31.

57. Schmitt, *supra* note 10, at 25; *see also* Anderson et al., *supra* note 46, at 410.

58. *See* Newton, *supra* note 6, at 20–21 (suggesting that autonomous weapons may be able to focus lethal violence only on appropriately identified targets); Gregory P. Noone & Diana C. Noone, *The Debate over Autonomous Weapons Systems*, 47 CASE WESTERN RES. J. INT’L L. 25, 31–32 (2015) (discussing human errors in armed conflicts).

59. Schmitt, *Military Necessity and Humanity in International Humanitarian Law*, *supra* note 544, at 796. For a review of the rise of humanitarian concerns in IHL, *see* Theodor Meron, *The Humanization of Humanitarian Law*, 94 AM. J. INT’L L. 239 (2000).

60. This is a point one of the symposium’s organizers, Gary Brown, makes in this volume. *See* Gary Brown, *Out of the Loop*, 30 TEMP. INT’L & COMP. L. J. (2016) (noting that arguments against developing autonomous weapon systems (AWS) come too late, as AWS already exist); *see also* Anderson & Waxman, *supra* note 6, at 27 (“[I]ncremental development and deployment of autonomous weapon systems is inevitable . . .”).

generate. Rather, I argue that an understanding of *why* IHL needs to account for autonomy may simply lie in the realities of future conflicts. Even if one believes that a prohibition on certain fully autonomous technology is the best regulatory outcome, past experience cautions against ending the legal analysis there; bans did not work for the submarine or the crossbow, while the campaign to ban landmines remains incomplete.⁶¹ Sadly, even those weapons IHL does ban (e.g., chemical weapons, biological weapons) still remain in use.⁶² Accordingly, IHL and the lawyers who practice it need to give further analysis not only to the menu of regulatory options IHL has on offer for autonomous weapon systems today, but also keep in mind *why* it selected these options in the first place. The reasons IHL exists will be as, if not more important, than its precise contents in discerning whether these autonomous systems must conform to IHL and whether, perhaps instead, IHL must be the one to conform and adjust to the existence of such systems.

CONCLUSION

Thomas Aquinas interrogated the application of divine law to human acts by asking about their circumstances in terms of *who, what, where, when, how, and why*. A similar formulation may be employed to evaluate the application of IHL to the advent of autonomous systems in armed conflicts. Doing so provides a critical lens for gauging the current scope (and state) of international legal discourse on this topic. Each question allows us to chart islands of agreement and contestation that deepen our understanding of how IHL regulates this technology. They also serve as inflexion points, suggesting new lines of inquiry, such as closer attention to the relationship between cyber operations and autonomy, the settings in which IHL discourse occurs, the temporal conditions for its application, the trade-offs involved in deploying rules, standards and principles, and, above all, why IHL needs to regulate this technology in the first place. Such questions do not come with ready answers; whether (let alone how) to extend the social conventions of humans to an autonomous technology is a profoundly difficult inquiry. Yet, given the stakes, it is a necessary one.

In the end, this paper seeks to do no more than to set the stage; to illustrate the breadth and depth of questions that autonomous systems pose for the future of

61. See, e.g., Anderson & Waxman, *supra* note 6, at 8–9. For a more detailed discussion of whether autonomous weaponry is suited for regulation, see (in this volume) Sean Watts, *Autonomous Weapons: Regulation Tolerant or Regulation Resistant*, 30 TEMP. INT'L & COMP. L. J. (2016).

62. See, e.g., Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction, Sept. 3, 1992, 1974 U.N.T.S. 45; see Int'l Committee of the Red Cross, *Chemical weapons use: An unacceptable repeat of history that demands attention*, STATEMENT OF THE ICRC: NINETEENTH SESSION OF THE CONFERENCE OF THE STATES PARTIES TO THE CHEMICAL WEAPONS CONVENTION, 1 – 5 DECEMBER 2014, THE HAGUE, NETHERLANDS (Dec. 2, 2014) <https://www.icrc.org/en/document/chemical-weapons-conference-states-parties-convention> (discussing the ICRC's condemnation of Syria's use of chemical weapons).

conflict. Answering such questions will likely require new plays and new players. So much of the autonomous conversation to date has been segregated. Analyses of cyber operations, drones, and semi-autonomous lethal weapons have largely occupied separate tracks.⁶³ Researchers from law, political science, ethics, engineering and military doctrine have prioritized *intra*-disciplinary analyses of what their particular areas of expertise have to say about the new technology.⁶⁴ IHL and its lawyers are no exception, focusing largely on what the existing law has to say about autonomous weapon systems.⁶⁵

Nothing about the technology, however, suggests that such silos are static, let alone impermeable. Autonomous technology's capacity to operate (and maybe even evolve) with limited-to-no human intervention clearly warrants more cross-cutting and inter-disciplinary research than witnessed to date. International lawyers may hold center stage on what IHL has to say about autonomous weapon systems. But lawyers hold no monopoly on what the technology may have to say about IHL, let alone law more generally. Questions about whether law can (or must) remain a quintessentially "human" enterprise in lieu of it applying to (and being applied by) autonomous systems reach across disciplines and subject-matter. Only the future will tell how the answers play out, but it will undoubtedly be a play worth watching.

63. See, e.g., Kenneth Anderson, *Comparing the Strategic and Legal Features of Cyberwar, Drone Warfare, and Autonomous Weapon Systems*, HOOVER INSTITUTION (Feb. 27, 2015), <http://www.hoover.org/research/comparing-strategic-and-legal-features-cyberwar-drone-warfare-and-autonomous-weapon-systems>.

64. See, e.g., Lin et al., *supra* note 17; Scharre & Horowitz, *supra* note 15.

65. See, e.g., Anderson & Waxman, *supra* note 6; Schmitt, *Military Necessity and Humanity in International Humanitarian Law*, *supra* note 474.