2012

Law and Ethics for Robot Soldiers

Kenneth Anderson
ANDERS@WCL.AMERICAN.EDU

Matthew C. Waxman
Columbia Law School, mwaxma@law.columbia.edu

Law and Ethics for Robot Soldiers

Kenneth Anderson
American University – Washington College of Law

Matthew Waxman
Columbia Law School

Policy Review, 2012

American University, WCL
Research Paper No. 2012-32

Columbia Law School
Public Law & Legal Theory Working Paper Group
Paper Number 12-313

http://ssrn.com/abstract=2046375

(Draft for SSRN with footnotes)

Law and Ethics for Robot Soldiers

Kenneth Anderson[*] & Matthew Waxman[**]

April 5, 2012

Policy Review
(Forthcoming, 2012)[1]

A lethal sentry robot designed for perimeter protection, able to detect shapes and motions, and combined with computational technologies to analyze and differentiate enemy threats from friendly or innocuous objects – and shoot at the hostiles.[2] A drone aircraft, not only unmanned but programmed to independently rove and hunt prey, perhaps tracking enemy fighters who have been previously "painted and marked" by military forces on the ground. Robots individually too small and mobile to be easily stopped – but capable of swarming and assembling themselves at the final moment of attack into a much larger weapon.[3] These are among the ripening fruits of automation in weapons design. Some are close at hand, such as the lethal sentry robot designed in South Korea, while others lie ahead in a future that is, however, less and less distant.[4]

---

[*] Professor, Washington College of Law, American University; Non-Resident Senior Fellow, The Brookings Institution; Member of the Hoover Institution Task Force on National Security and Law.
[**] Associate Professor, Columbia Law School; Adjunct Senior Fellow, Council on Foreign Relations; Member of the Hoover Institution Task Force on National Security and Law.

[1] We are grateful to friends and colleagues who generously provided comments on the draft and allowed us to consult unpublished manuscripts and research. In particular, we thank Michael Adams, Ronald Arkin, Ryan Calo, Mark Hagerott, Shane Harris, Tod Lindberg, Sonia McNeil, William Marra, Noel Sharkey, Micah Zenko, and several others who prefer to remain unnamed, as well as members of the Hoover Institution Task Force on National Security and Law, where this essay was first presented, and the Harvard New Weapons Task Force. We are of course solely responsible for any views or errors in the essay.

[2] South Korea has designed and deployed such a robot, although limited at this point to alerting human operators who make the decision to fire a weapon. AFP, "S. Korea Deploys Sentry Robot Along N. Korea Border," July 13, 2010, available at http://www.defensenews.com/article/20100713/DEFSECT02/7130302/S-Korea-Deploys-Sentry-Robot-Along-N-Korea-Border.

[3] The current non-weapons drone swarm technology is demonstrated in a series of videos embedded in this article: Farhad Manjoo, "I Love You, Killer Robots: Quadrotor Robots Are Amazingly Cute and Will Probably Kill Us All," Slate, March 5, 2012, available at http://www.slate.com/articles/technology/technology/2012/03/quadrotor_drones_are_amazing_and_cute_and_they_will_probably_destroy_us_all_.html.

[4] The starting point for these and many other automated and autonomous battlefield weapons scenarios is Peter Singer, Wired for War: The Robotics Revolution and Conflict in the 21st Century (Penguin Press 2009). As Singer himself has noted, however, the field has advanced faster and further than even he anticipated and a book based on interviews conducted a mere four or five years ago is now already looking a little dated. For a dissenting view, that removing

Lethal autonomous machines will inevitably enter the future battlefield – but they will do so incrementally, one small step at a time. The combination of inevitable and incremental development raises not only complex strategic and operational questions but also profound legal and ethical ones. The inevitability of comes from both supply-side and demand-side factors. Advances in sensor and computational technologies will supply "smarter" machines that can be programmed to kill or destroy, while the increasing tempo of military operations and political pressures to protect one's own personnel and civilian persons and property will demand continuing research, development, and deployment. The process will be incremental because non-lethal robotic systems (already proliferating on the battlefield, after all) can be fitted in their successive generations with both self-defensive and offensive technologies. As lethal systems are initially deployed, they may include humans in the decision-making loop, at least as a fail-safe – but as both the decision-making power of machines and the tempo of operations potentially increase, that human role will likely but slowly diminish.[5]

Recognizing the inevitable but incremental evolution of these technologies is key to addressing the legal and ethical dilemmas associated with them; U.S. policy toward resolving those dilemmas should be built upon these assumptions. The certain yet gradual development and deployment of these systems, as well as the humanitarian advantages created by the precision of some systems, make some proposed responses — such as prohibitory treaties — unworkable as well as ethically questionable.[6] Those features also make it imperative, though, that the United States resist its own impulses toward secrecy and reticence with respect to military technologies, recognizing that the interests those tendencies serve are counterbalanced here by interests in shaping the normative terrain — the contours of international law as well as international expectations about appropriate conduct — on which it and others will operate militarily as technology evolves. Just as development of autonomous weapon systems will be incremental, so too will development of norms about acceptable systems and uses be incremental. The United States must act, however, before international expectations about these technologies harden around the views of those who would impose unrealistic, ineffective or dangerous prohibitions or those who would prefer few or no constraints at all.[7]

Incremental Automation of Drones

---

humans from military targeting is unlikely for the foreseeable future, see Wener J.A. Dahm, "Killer Drones Are Science Fiction," Wall Street Journal, Feb. 15, 2012.

[5] This is approximately the view of a leading US Air Force strategy document on the future evolution and use of drones – "remotely-piloted vehicles" or "unmanned aerial vehicles" (UAVs) in more technical language. See United States Air Force, Unmanned Aircraft Systems Flight Plan, 2009–2047, at 16 (2009), at http://www.fas.org/irp/program/collect/uas_2009.pdf ("Air Force Flight Plan").

[6] For a survey of possible modes of international governance of lethal autonomous robotics, see Gary E. Marchent, et al, "International Governance of Autonomous Military Robotics," 12 Columbia Science & Technology Law Review (2011).

[7] Markus Wagner, "Taking Humans Out of the Loop: Implications for International Humanitarian Law," 21 Journal of Law, Information and Science (2011), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1874039.

The incremental march toward automated lethal technologies of the future, and the legal and ethical challenges that accompany it, can be illustrated by looking at today's drone aircraft.[8] Unmanned drones piloted from afar are already a significant component of the United States' arsenal. At this writing, close to one in three U.S. Air Force aircraft is remotely piloted (though this number also includes many tiny tactical surveillance drones), and the drone proportion will only grow.[9] Current drone military aircraft are not autonomous in the firing of weapons – the weapon must be fired in real-time by a human controller – and so far there are no known plans or, apparently in the view of military, reasons today to take the human out of the weapon firing loop.

Nor are today's drones truly autonomous as aircraft – they require human controllers in real-time, even when they are located far away.[10] They *are*, however, increasingly automated in their flight functions – self-landing capabilities, for example, and particularly automation to the point that a single controller can run many drone aircraft at once, increasing efficiency considerably.[11] The automation of flight is gradually increasing as sensors and aircraft control through computer programming improves.

Looking into the future, some observers believe that one of the next generations of jet fighter aircraft will no longer be manned, or at least that manned fighter aircraft will be joined by unmanned aircraft.[12] Drone aircraft might gradually become capable of higher speeds, torques, g-forces, and other stresses than a human pilot can endure (and perhaps at a cheaper cost as well). Given that speed in every sense – including turning and twisting in flight, reaction and decision times – is an advantage, design will emphasize automating as many of these functions as possible, in competition with the enemy's systems.[13]

Just as the aircraft might have to be maneuvered far too quickly for detailed human control of its movements, so too the weapons – against other aircraft, drones, anti-

[8] See generally, Shane Harris, Out of the Loop: The Human-free Future of Unmanned Aerial Vehicles (Hoover Institution 2012), at http://media.hoover.org/sites/default/files/documents/EmergingThreats_Harris.pdf.

[9] Spencer Ackerman and Noah Shachtman, "Almost 1 in 3 US Warplanes Is a Robot," Wired Danger Room, January 9, 2012, at http://www.wired.com/dangerroom/2012/01/drone-report/.

[10] A still useful popular-audience reference guide to the best-known military drone, the Predator, is Robert Valdes, "How the Predator UAV Works," How Stuff Works, April 1, 2004, at www.science.howstuffworks.com/predator.htm.

[11] The jumbo-size Global Hawk drone is already able to perform automated take-off and landing. Singer, Wired for War, at 36. Given that nearly all crashes of the Predator's weaponized cousin, the Reaper, have occurred in human-controlled landings, increasing the automation of take-off and landing is an important priority. Noah Shachtman, "US Drone Goes Down Over Pakistan. Again," Wired Danger Room, January 25, 2010, available at http://www.wired.com/dangerroom/2010/01/us-drone-goes-down-over-pakistan-again/.

[12] Economist, "Unmanned Aerial Warfare: Flight of the Drones: Why the future of air power belongs to unmanned systems," October 8, 2011 (Flight of the Drones)

[13] This amounts to the famous "OODA Loop" by any other name. OODA refers to the techno-strategic concept first developed by fighter pilot and strategist John Boyd – "observe, orient, decide, and act." See generally Frans P.B. Osinga, Science, Strategy and War: The Strategic Theory of John Boyd (Routledge 2006). The Air Force Flight Plan explicitly sees increased reliance on drones because of their ability to "rapidly compress the observe, orient, decide, and act (OODA) loop." Increased automation is a crucial part of that compression. Air Force Flight Plan, at 16. It adds that increasingly "humans will no longer be 'in the loop' but rather 'on the loop' – monitoring the execution of certain decisions." Id. at 41.

aircraft systems – might have to be utilized at the same speeds in order to match the beyond-human speed of the aircraft's own systems (as well as the enemy aircraft's similarly automated counter-systems).[14] In similar respects, defense systems on modern U.S. naval vessels have long been able to target incoming missiles automatically, with humans monitoring the system's operation, because human decision-making processes are too slow to deal with multiple, inbound, high-speed missiles.[15] Some military operators regard many emerging automated weapons as a more sophisticated form of "fire and forget" self-guided missiles. And because contemporary fighter aircraft are designed not only for air-to-air combat, but for ground attack missions as well, the design changes that reduce the role of the human controller of the aircraft platform may shade into automation of the weapons directed at ground targets, too.

Although current remotely-piloted drones, on the one hand, and future autonomous weapons, on the other, are based on different technologies and operational imperatives, they generate similar concerns about their ethical legitimacy and lawfulness. Today's arguments over the legality of remotely-piloted, unmanned aircraft in their various missions (especially targeted killing operations, and concerns that the United States is using technology to shift risk to its own personnel onto remote-area civilian populations) presage the arguments that already loom over weapons systems that exhibit emerging features of autonomy. Those arguments also offer lessons to guide short and long-term U.S. policy toward autonomous weapons generally, including systems that are quite different.

Automated-Arms Racing?

These issues are easiest to imagine in the airpower context, but in other battlefield contexts, too, the United States and other sophisticated military powers – and eventually unsophisticated powers and non-state actors, as such technologies become commodified and offered for licit or illicit sale – will find increasingly automated lethal systems more and more attractive. Moreover, as artificial intelligence improves, weapons systems will evolve from robotic "automation" – the execution of precisely pre-programmed actions or sequences in a well defined and controlled environment – toward genuine "autonomy" – meaning the robot is capable of generating actions to adapt to changing and unpredictable environments.[16]

[14] This is, then, another expression of the OODA Loop, because, in air-to-air combat, "Boyd's insight was that … the advantage lay with the fighter pilot whose OODA Loop was faster and more accurate than his opponent's, and who was able to throw his opponent's loop out of sync." William C. Marra & Sonia K. McNeil, "Automation and Autonomy in Advanced Machines: Understanding and Regulating Complex Systems," Lawfare Research Paper Series 1-2012 (April 2012), at 9. The fastest OODA Loop of the future combat plane is likely to be an automated one – automated in both flight and weapons functions, and unmanned as well.

[15] The best known of these systems, the Aegis sea defense system, has four levels of human control; in fully automated "casualty" mode, the system acts on its own without human intervention. Singer, Wired for War, at 124.

[16] The distinction between "automation" and "autonomy" in robotics and artificial intelligence is an enormous topic; we mark it here without seeking fully to explain it, because it is part of the continuum along which lethal weapons gradually "slide" toward genuine autonomy. We commend to readers Marra & McNeil, "Automation and Autonomy in Advanced Machines," at 15-28, including the notes and references, for a background discussion of the distinction as currently used in the engineering and military discussions.

Take efforts to protect peacekeepers facing the threat of snipers or ambush in an urban environment: small mobile robots with weapons could act as roving scouts for the human soldiers, with "intermediate" automation – the robot might be pre-programmed to look for certain enemy weapon signatures and to alert a human operator of the threat, who then decides whether or not to pull the trigger. In the next iteration, the system might be set with the human being not required to give an affirmative command, but instead merely deciding whether to override and veto a machine-initiated attack. That human decision-maker also might not be a soldier on site, but an off-battlefield, remote robot-controller.

It will soon become clear that the communications link between human and weapon system could be jammed or hacked (and in addition, speed and the complications of the pursuit algorithms may seem better left to the machine itself, especially once the technology moves to many small swarming, lightly armed robots). One technological response will be to reduce the vulnerability of the communications link by severing it - making the robot dependent upon executing its own programming, or even genuinely autonomous.[17]

Aside from conventional war on conventional battlefields, covert or special operations will involve their own evolution toward incrementally autonomous systems. Consider the Osama bin Laden-compound raid — tiny surveillance robots equipped with facial recognition technology might have helped affirmatively identify Bin Laden earlier. It is not a large step to weaponize such systems, and then perhaps go the next step to allow them to act autonomously, perhaps initially with a human remote-observer as a failsafe, but with very little time to override programmed commands.

These examples have all been stylized to sound precise and carefully controlled. Consider also, though, that at some point in the not-distant future, someone – China, Russia, or someone else – will likely design, build, deploy and sell an autonomous weapon system for battlefield use that is programmed to target only something – say a person or position – that is firing a weapon and is positively identified as hostile rather than friendly. It might lack the ability altogether to take account of civilian presence and likely collateral damage. Quite apart from the security and warfighting implications, the U.S. government would have grave legal and humanitarian concerns about such a foreign system offered for sale on the international arms markets, let alone deployed and used. The United States would then find itself in a peculiar situation – potentially facing a weapon system on the battlefield that conveys significant advantages to its user, but which the United States would not deploy itself because (for reasons described below) it does not believe it is a legal weapon. The United States will have to come up with technological counters and defenses – which might entail development of smaller, more mobile, armed robots that are able to hide as well as hunt on their own.

[17] "Stealth" versions of surveillance drones, such as the RQ-170 Sentinel reportedly used over Iran (including the one that went down in 2011), already operate in non-continuous communications contact with the ground base, receiving spurts of command signals that trigger automatic and pre-configured programming, such as surveillance routines. The short bursts of signals reduce the likelihood of detection or interference with the communications links. See Michael Hastings, "The Rise of the Killer Drones: How America Goes to War in Secret," Rolling Stone, April 16, 2012, at http://www.rollingstone.com/politics/news/the-rise-of-the-killer-drones-how-america-goes-to-war-in-secret-20120416. Such "burst" communications that trigger pre-set automated routines are used in many contexts, including such robotic devices as the Mars rovers.

The implication is that the arms race in battlefield robots will be more than simply a race for ever more autonomous weapons systems. It will mostly be a race for ways to counter and defend against them – partly through technical means and partly through international norms and diplomacy.

Legal and Ethical Requirements of Weapons

The legal and ethical evaluation of a new weapons system is nothing new; it is a long-standing requirement of the laws of war, one taken seriously by U.S. military lawyers.[18] In recent years, U.S. military judge advocates have rejected proposed new weapons as incompatible with the laws of war, including blinding laser weapons and, reportedly, various cutting edge cyber-technologies that might constitute weapons for purposes of the laws of war.[19] But arguments over the legitimacy of particular weapons (or their legitimate use) go back to the beginnings of laws and ethics of war – the legitimacy, for example, of poison, the cross-bow, submarines, aerial bombardment, antipersonnel landmines, chemical and biological weapons, and nuclear weapons. In that historical context, debate over autonomous robotic weapons – the conditions of their lawfulness as weapons and the conditions of their lawful use – is nothing novel.[20]

Likewise, there is nothing novel in the sorts of responses autonomous weapons systems will generate: on the one hand, emergence of a new weapon often sparks an insistence in some quarters that the weapon is ethically and legally abhorrent and should be prohibited by law while, on the other hand, the historical reality is that if a new weapon system greatly advantages a side, the tendency is for it gradually to be adopted by others that perceive they can benefit from it as well. In some cases, legal prohibitions on the weapon system as such erode, as happened with submarines and airplanes; what

[18] Article 36 of 1977 Additional Protocol I provides that in the "study, development, acquisition or adoption on a new weapon, means or method of warfare," a party to Protocol is "under an obligation to determine whether its employment would, in some or all circumstances, be prohibited," either by Protocol I or by "any other rule of international law applicable" to such party. Although the United States is not a party to Protocol I and has not indicated whether it formally accepts Article 36 as a rule of customary international law, but it almost certainly the case that the United States does see it as customary law, binding on it and others. The United States in its actual practice undertakes extensive legal review of new weapons systems, with elaborate protocols. For a recent example of the US Air Force directive on procedures to be followed for review of the legality of weapons systems and cyber capabilities, see Order of the Secretary of the Air Force, "Legal Reviews of Weapons and Cyber Capabilities," Air Force Instructions 51-402, 27 July 2011, at http://www.fas.org/irp/doddir/usaf/afi51-402.pdf.

[19] Personal interviews with the authors.

[20] Historian of the laws of war Stephen C. Neff notes that virtually the only specific rule on the conduct of warfare newly laid down in the Middle Ages, in the Second Lateran Council of 1139, was a "flat prohibition against the use of 'that murderous art of crossbowmen and archers'," at least within Christendom, for they were permitted against infidels (and the ban on poison predated this period). Stephen C. Neff, War and the Law of Nations (Cambridge UP 2005), at 65. Neff goes on to observe that the First Hague Peace Conference (1899) adopted three specific weapons declarations – prohibitions on expanding bullets, asphyxiating gases, and the launching of projectiles from balloons (which can be taken as an early pre-airplane form of aerial bombardment). Id. at 187-88. As with the Middle Ages ban on the crossbow, these prohibitions were of mixed effectiveness at best in the face of changing notions of necessity and technological innovation.

survives is a set of legal rules for the *use* of the new weapon, with greater or lesser specificity.[21] In other cases, legal prohibitions gain hold.[22] The ban on of poison gas, for example, has survived in one form or another with very considerable effectiveness over the history of the 20th century.[23]

Where in this long history of new weapons and their ethical and legal regulation will autonomous robotic weapons fit? What are the features of autonomous robotic weapons that raise ethical and legal concerns? How should they be addressed, as a matter of law and process – by treaty, for example, or by some other means?

One answer to these questions is – wait and see. It is too early to know where the technology will go, so the debate over ethical and legal principles for robotic autonomous weapons should be deferred until a system is at hand. Otherwise it is just an exercise in science fiction and fantasy.

That wait-and-see view is short-sighted and mistaken. Not all of the important innovations in autonomous weapons are far-off on the horizon; some are possible now or will be in the near-term, and some of them raise serious questions of law and ethics even at their current research and development stage.[24]

Moreover, looking to the long-term, technology and weapons innovation does not take place in a vacuum. This is the time – before technologies and weapons development have become "hardened" in a particular path and whose design architecture is difficult to change – to take account of the law and ethics that ought to inform and govern autonomous weapons systems, as technology and innovation let slip the robots of war.

---

[21] For example, before and after the First World War, diplomatic attempts were made to ban submarines as weapons systems altogether. When that proved not possible, the interwar London Naval Treaty (1930) attempted to require the Prize Rules (including removal of the ship's crew to a "place of safety" before sinking the vessel) as the basis for naval warfare; in effect substituting a set of conduct rules for a weapons ban. Those, too, proved impracticable in the Second World War, and the rules evolved still further as conduct rules. See generally, Natalino Ronzitti, The Law of Naval Warfare: A Collection of Agreements and Documents With Commentaries (Martinus Nijhoff 1988). As is often the case with weapons, evolving modern technology has made identification of lawful targets by submarines far easier and more accurate. A broadly similar account could be given of aerial warfare, starting with the Hague ban on balloon-launched projectiles of 1899, followed by the introduction of air war in the First World War, the unsuccessful Hague Rules of Air Warfare (1923), the explosion of air warfare and technology in the Second World War, and – accelerating through many decades – the current emphasis on precision weaponry and air platforms, including drones, as a sort of technological fix to the limitations of the normative rules in constraining actors.

[22] In general, the cautious restrictions upon weapons found in the Convention on Conventional Weapons (CCW) (1980) and its operative protocols (at least I-V) addressing separate weapons have been reasonably effective. Similarly the Chemical Weapons Convention (CWC) (1997) and the Biological Weapons Convention (BWC) (1972).

[23] The most well-known breaches included the lapses of the First World War, the Italian interwar campaign in Ethiopia, and Saddam Hussein's widespread use of gas in the Anfal campaign in 1988 against the Kurds and in the Iran-Iraq war. See generally Joost Hiltermann, A Poisonous Affair: America, Iraq, and the Gassing of Halabja (Cambridge UP 2007), including an introduction to the history of poison gas use and attempts to outlaw it from the late 19th century up through its use by the regime of Saddam Hussein in Iraq.

[24] One useful overview of the combined legal, ethical, and technological debate is Armin Krishnan, Killer Robots: Legality and Ethicality of Autonomous Weapons (Ashgate 2009).

This is also the time – before ethical and legal understandings of autonomous weapon systems become hardened in the eyes of key constituents of the international system – to propose and defend a framework for evaluating them that advances simultaneously strategic and moral interests.[25] After surveying the traditional legal and ethical paradigm into autonomous weapons systems could fit, and the major objections and responses being advanced by critics of such systems, we offer an approach to do so.

A Legal and Ethical Framework for Lethal Autonomous Weapons

The baseline legal and ethical principles governing the introduction of any new weapon are distinction (or discrimination) and proportionality. Distinction says that for a weapon to be lawful, it must be capable of being aimed at lawful targets, in a way that discriminates between military targets and civilians and their objects. Although most law-of-war concerns about discrimination run to the *use* of a weapon – is it being used with no serious care in aiming it? – in extreme cases, a weapon itself might be regarded as inherently indiscriminate.[26] Any autonomous robot weapon system will have to possess the ability to be aimed, or aim itself, at an acceptable legal level of discrimination.[27]

---

[25] A recent and widely-circulated report from the UK Ministry of Defense on the future of unmanned systems made this point forcefully, noting that as "technology matures and new capabilities appear, policy-makers will need to be aware of the potential legal issues and take advice at a very early stage of any new system's procurement cycle." UK Ministry of Defense, Joint Doctrine Note 2/11: The UK Approach to Unmanned Aircraft Systems, Developments, Concepts and Doctrine, 30 March 2011 (MOD 2011 Report), at para. 508.

[26] See, e.g., 1977 Additional Protocol I, Art. 51 (4) and (5), which describe prohibited "indiscriminate attacks." These provisions combine indiscriminate uses of an otherwise lawful weapon together with prohibitions on means or methods of attack – which might include technological means – that "cannot be directed at a specific military objective" or the "effects of which cannot be limited as required by this Protocol." Viewed historically, a prohibition on a weapon as such has been much less frequent than prohibitions on indiscriminate uses or effects, although the contemporary period has shown greater attention to weapons systems as such. Part of this is simply the effect of technological innovation in weaponry – some of those innovations have emphasized precision, while others have emphasized the extension of firepower.

[27] The extent of technical capability of weapons systems is inevitably a part of the equation; forms of aerial bombardment that were arguably acceptable under the law of the Second World War would be indisputably illegal today, for example. But that is mostly due to advances in technology widely available to virtually any party able to undertake aerial attack. Two deeply divisive issues – which robotic weapons of many sorts put unavoidably on the table, because most of them involve increases in the precision of weaponry – are: First, is a party to a conflict that has available to it as a matter of superior resources precision weapons obligated as a matter of law to use them or face charges of indiscriminate attack? Second, is a party to a conflict, such as a guerrilla non-state actor insurgency, prohibited from using the only means of attack available to it, but which are widely considered indiscriminate, such as sowing terrain with "dumb" landmines without self-destruct mechanisms or firing rockets normally used in military operations as antipersonnel fragmenting weapons and which cannot be properly aimed solely at military targets, even as professional militaries would do as a matter of course? Do differences in availability of precision technology dismantle the historical standard of parity of rules – the same absolute rules binding one side reciprocally bind the other – in favor some relative standard of "parity of capabilities" instead? The US military, so far as we are aware, holds that using the most discriminating weapon available, over less discriminating but nonetheless lawful weapons is a matter of humanitarian policy, and not law, and always subject to resource constraints, availability, and other matters of military necessity. See Matthew C. Waxman, International Law and the Politics of Urban Air Operations (RAND 2000) at 13-14, 57, available at

Proportionality says that even if a weapon meets the test of distinction, any use of a weapon must also involve evaluation that sets the anticipated military advantage to be gained against the anticipated civilian harm (to civilian persons or objects). The harm to civilians must not be excessive relative to the expected military gain.[28] This calculus for taking into account civilian collateral damage is difficult for many reasons.[29] While everyone agrees that civilian harm should not be excessive in relation to military advantages gained, the comparison is apples and oranges. Although there is a general sense that such excess can be determined in truly gross cases, there is no accepted formula that gives determinate outcomes in specific cases. Nonetheless, it is a fundamental requirement of the law and ethics of war that any military operation undertake this calculus, and that must be true of any autonomous weapon system's programming.

---

http://www.rand.org/pubs/monograph_reports/MR1175.html. Moreover, our understanding is that the US holds to a "parity of rules," rather than a relative "parity of capabilities," as a matter of binding legal obligation. These legal issues around sometimes-claimed obligations to use precision weaponry will matter insofar as robotic systems come to be understood as more precise.

[28] The customary formulation as found, for example, in 1977 Additional Protocol I, Art. 51(5)(b).

[29] The literature on proportionality jus in bello is vast; however, two examples perhaps illustrate the difficulty of comparisons. One is International Criminal Tribunal for Yugoslavia (ICTY), Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia (2000), at paragraphs 48 et seq. (proportionality), which notes the difficulties of establishing objective criteria assessing disproportionality, concluding as to law that it "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." Id., at para. 48.) The second is Laurie R. Blank, Operational Law Experts Roundtable on the Gotovina Judgment: Military Operations, Battlefield Reality and the Judgment's Impact on Effective Implementation and Enforcement of International Humanitarian Law (Emory Public Law Research Paper 12-186, January 28, 2012), at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1994414, which criticizes a recent judgment of the ICTY for having seriously misapplied the principle of proportionality its Gotovina judgment. Prior to the Gotovina judgment, the authors are not aware of any court holding a commander liable post hoc for a good faith application of the proportionality calculus, turning on the court's finding that the commander's judgment of proportionality was so clearly wrong as to be liable – indeed, we are otherwise unaware of any *prosecution*, let alone liability, turning strictly on the commander's good faith judgment of proportionality as such, so much a matter of judgment of incommensurables it has been thought to be. The introduction of machine judgment to this calculus, however, is likely to make legal challenge more common, not a rare exception.

Indeed, as machines incrementally enter the process of making decisions to fire weapons, and are used to make automated or algorithmic data-driven estimations of likely civilian harm or military advantage, then the possibility that more and more proportionality determinations will be legally challenged seems likely. Consider what might be demanded for discovery in either a possible war crimes prosecution in an international tribunal someday, or a civil trial for damages for civilian harm: the code for how the machine made estimations of likely harms, to start with, but also the background methods for setting the algorithm. Incorporated into artificial intelligence machines for helping commanders do a better job of estimating and mitigating collateral damage as part of a proportionality calculus, the effect might well be to open what had previously been the untouched black box of a commander's good faith discretion into a hotly and routinely contested legal area.

These are daunting legal and ethical hurdles if the aim is to create a true "robot soldier." One way to think about the requirements of the ethical robot soldier is to ask what we would require of a human soldier performing the same function.

Some leading roboticists have been working on creating algorithms to capture the two fundamental principles of distinction and proportionality. As for programming distinction, one could theoretically start with fixed lists of lawful targets – for example, programmed targets could include persons or weapons that are firing at the robot – and gradually build upwards toward inductive reasoning about characteristics of lawful targets not already on the list. Proportionality, for programming purposes, is a relative judgment: measure anticipated civilian harm and measure military advantage; subtract and measure the balance against some determined standard of "excessive"; if excessive, do not attack an otherwise lawful target. Difficult as these seem to any experienced law-of-war lawyer, these are the fundamental conditions that the ethically designed and programmed robot soldier would have to satisfy and therefore what a programming development effort must take into account.[30] The ethical and legal engineering matter every bit as much as the mechanical or software engineering.[31]

Four Objections

If this is the optimistic vision of the robot soldier of, say, decades from now, however, it is subject already to four main categories of objection. The first is a general empirical skepticism that machine programming could ever reach the point of satisfying the fundamental ethical and legal principles of distinction and proportionality.[32] Artificial intelligence has overpromised before, and once into the weeds of the judgments that these broad principles imply, the requisite intuition, cognition, and judgment look ever more marvelous.[33]

This skepticism is essentially factual, a question of how technology evolves over decades. It is quite possible that fully autonomous weapons will never achieve the

[30] One of the most ambitious of these efforts is by roboticist Ronald C. Arkin, who describes his work on both distinction and proportionality in his "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture," Technical Report GIT-GVU-07-11, Georgia Tech Robotics Lab (2007), at 43-53, available at http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf. For a general discussion, see also Ronald C. Arkin, Governing Lethal Behavior in Autonomous Robots (Chapman and Hall 2009).

[31] "In order to ensure that new unmanned aircraft systems adhere to present and future legal requirements, it is likely that a systems engineering approach will be the best model for developing the requirement and specification … the legal framework for operating the platform would form a list of capability requirements that would sit alongside the usual technical and operational requirements." MOD 2011 Report, at para. 505.

[32] Economist, Flight of the Drones (Oct. 8, 2011), noting that some "question whether artificial intelligence, which always seems just a few years away will ever work well enough … [a]nd decisions about what is proportionate often require fine distinctions and sophisticated judgment … there will be an almost infinite combination of contingencies facing drones."

[33] See, e.g., Noel Sharkey, "March of the Killer Robots," London Telegraph, June 15, 2009; Noel Sharkey, "The Automation and Proliferation of Military Drones and the Protection of Civilians," 3 Law, Innovation and Technology (2011), at 236-37; Robert Sparrow, "Building a Better WarBot : Ethical Issues in the Design of Unmanned Systems for Military Applications," 15 Science and Engineering Ethics (June 2009), at 178.

ability to meet these standards, even far into the future. Yet we do not want to rule out such possibilities – including the development of technologies of war that might reduce risks to civilians by making targeting more precise and firing decisions more controlled (especially compared to human-soldier failings that may be exacerbated by fear, vengeance, or other emotions). True, relying on the promise of computer analytics and artificial intelligence risks pushing us down a slippery slope, propelled by the future promise of technology to overcome human failings rather than addressing them directly. But articulation of the tests of lawfulness that autonomous systems must ultimately meet helps channel technological development toward the law of war's protective ends.

A second objection is a moral one, and says that it is simply wrong per se to take the human moral agent entirely out of the firing loop. A machine, no matter how good, cannot completely replace the presence of a true moral agent in the form of a human being possessed of a conscience and the faculty of moral judgment (even if flawed in human ways).[34] In that regard, the title of this Essay is deliberately provocative in pairing "robot" and "soldier," because, on this objection, that is precisely what should never be attempted.[35]

This is a difficult argument to address, since it stops with a moral principle that one either accepts or does not. Moreover, it raises a further question as to what constitutes the tipping point into impermissible autonomy given that the automation of weapons functions is likely to occur in incremental steps.

The third objection holds that autonomous weapons systems that remove the human being from the firing loop are unacceptable because they undermine the possibility of holding anyone accountable for what, if done by a human soldier, might be a war crime.[36] If the decision to fire is taken by a machine, who should be held responsible for

[34] The direct moral case against autonomous weapons as such is made in Dr. Matthew Bolton, Thomas Nash, and Richard Moyes, "Ban autonomous armed robots," Article 36, March 5, 2012, available at http://www.article36.org/statements/ban-autonomous-armed-robots/ ("Whether static or mobile, simple or complex, it is the automated violent response to a signal that makes landmines and fully autonomous weapons fundamentally problematic – it is killing by machine … certain fundamentals seem strong. Decisions to kill and injure should not be made by machines and, even if at times it will be imperfect, the distinction between military and civilian is a determination for human beings to make.")

[35] At its most capacious, this is an argument against machines ever becoming genuinely autonomous agents in a legal or moral sense, whether for weapons purposes or anything else; for a discussion of the general philosophical arguments, see Samir Chopra and Laurence F. White, A Legal Theory for Autonomous Artificial Agents (University of Michigan Press 2011), and Wendell Wallach and Colin Allen, Moral Machines: Teaching Robots Right from Wrong (Oxford UP 2009). One need not carry the argument against artificial intelligence "agency" in order to express caution, and indeed draw a line, at autonomous "robot soldiers" firing weapons, but it sometimes appears that the two arguments are run together.

[36] See W.J. Hennigan, "New Drone Has No Pilot Anywhere, So Who's Accountable?*"* L.A. Times, Jan. 26, 2012, at http://articles.latimes.com/2012/jan/26/business/la-fi-auto-drone-20120126. The MOD 2011 Report, at para. 510, states that legal responsibility "for any military activity remains with the last person to issue the command authorizing a specific activity." It goes on to add that this assumes that a "system's basic principles of operation have, as part of its release to service, already been shown to be lawful, but that the individual giving orders for use will ensure its continued lawful employment throughout any task." Para. 510. We are doubtful that this will work quite as anticipated; the speed and opacity of such systems will surely make most serious attempts to hold the "last person" responsible will either result in a standard of due care and

mistakes: The soldier who allowed the weapon system to be used and make a bad decision?[37] The commander who chose to employ it on the battlefield? The engineer or designer who programmed it in the first place?[38]

This is an objection particularly salient to those who put significant faith in laws of war accountability through mechanisms of individual criminal liability, through international tribunals or other judicial mechanisms. But post-hoc judicial accountability in war is just one of many mechanisms for promoting and enforcing compliance with the laws of war, and devotion to individual criminal liability as the presumptive mechanism of accountability risks blocking development of machine systems that would, if successful, reduce actual harms to civilians on or near the battlefield.

Finally, the long-run development of autonomous weapon systems faces an objection that by removing one's human soldiers from risk and reducing harm to civilians through greater precision, the disincentive to resort to armed force is diminished.[39] The result might be a greater propensity to use military force and wage war.[40]

---

negligence that in effect gives a legal safe harbor for reliance upon design absent something truly extraordinary, or else a strict liability standard for the "last person" that raises questions as to whether it will actually be used. Moreover, there is always the cautionary tale of the semi-automated Aegis system that, in the early 1980s, mistakenly identified a civilian Iranian jet aircraft, which was then shot down by human operators; although the system was not acting in an automated firing mode, human operators relied upon the machine's judgment, which turned out to be tragically wrong. See Singer, Wired for War, at 124-25. There are reasons to believe that human psychology will often lead to trusting the machine judgment as being better than one's own in pressured and jumbled circumstances. Finally, the MOD 2011 Report says that the "authorized entity that holds legal responsibility will be required to exercise some level of supervision ... any fielded system employing weapons will have to maintain a 2-way data link between the aircraft and its controlling authority." Para. 510. Since the 2-way data link will quite possibly be one of the weakest and most attackable links in the control chain, future systems might instead operate as though they were highly sophisticated "fire and forget" systems, responding with automated sub-systems triggered by bursts of less-hackable communications signals.

[37] See Mary L. Cummings, "Automation and Accountability in Decision Support System Interface Design," 32 Journal of Technology Studies (2006), at 23-31; see also M.L. Cummings, "Creating Moral Buffers in Weapon Control Interface Design," IEEE Technology and Society (Fall 2004), at 28-33.

[38] See Sparrow, "Building a Better WarBot," at 178-79. We have profited from presentations by Oren Gross and Ashley Deeks at a panel, Gregory McNeal (Moderator), "Future Weapons, Past Laws," University of Santa Clara Law School International Humanitarian Law Symposium, Feb 3-4, 2012, available at http://digitalcommons.law.scu.edu/humanitarian/symposium/track/6/.

[39] Economist, Flight of the Drones: "Looking farther ahead, there are fears that UAS and other robotised killing machines will so lower the political threshold for fighting that an essential element of restraint will be removed. Robert E. Lee said 'it is well that war is so terrible, otherwise we would grow too fond of it.' Drones might make leaders fonder of war." See also, MOD 2011 Report, at para. 517.

[40] Singer, Wired for War, at 431-433; Paul W. Kahn, "The Paradox of Riskless Warfare," Philosophy & Public Policy Quarterly, Vol. 22, No. 2 (2208), at 2-8; Robert Sparrow, "Predators or Plowshares? Arms Control of Robotic Weapons," IEEE Technology & Society Magazine (Spring 2009), at 26. Some of these issues are considered in relation to Walzer's account of just war theory by Peter M. Asaro, "How Just Could a Robot War Be?" in Philip Brey, Adam Briggle and Katinka Waelbers (eds), Current Issues in Computing and Philosophy (Amsterdam: IOS 2010), available at www.cybersophe.org/writing/Asaro%20Just%20Robot%20War.pdf.

As a moral matter, this objection is subject to a moral counter-objection, in that this would entail foregoing easily obtained protections for civilians or soldiers in war for fear that without, in effect, holding these humans hostage, political leaders would resort to war more than they ought.[41] Moreover, as an empirical matter, it is not so clear that this concern is special to autonomous weapons, since precisely the same objection is already raised with respect to remotely-piloted drones – and, generally, with respect to any technological developments that either reduce risk to one's own forces or, especially perversely, reduce risk to civilians.[42]

These four objections run to the whole enterprise of building the autonomous robot soldier, and important debates could be held around each of them. Whatever their merits in theory, however, they all face a practical difficulty in the incremental way autonomous weapon systems will develop. These objections are often voiced as though there was likely to be some determinable break-point between the human-controlled system and the machine-controlled one. 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. And the role of real-time human decision-making will be phased out in some military contexts in order to address some technological or strategic issue unrelated to autonomy, such as the speed of the system's response. "Incrementality" does not by itself render any of these universal objections wrong per se – but it does suggest that there is another kind of discussion to be had about regulation of weapons systems undergoing gradual, step-by-step change.

International Treaties and the Challenge of Incremental Evolution

Critics sometimes portray the United States as engaged in relentless, heedless pursuit of technological advantage – whether in drones or other robotic weapons systems – that will inevitably be fleeting as other countries mimic, steal, or reverse engineer its technologies.[43] According to this view, if the United States would quit pursuing these technologies, the genie might remain in the bottle or at least emerge much more slowly.

This is almost certainly wrong, in part because the technologies at issue – drone aircraft or driverless cars, for example – are going to spread with respect to general use far outside of military applications. The decision architectures that would govern firing a weapon are not so very different from those of, say, an elder-care robot engaged in home-assisted living, and which is programmed to decide when to take emergency action.[44]

---

[41] We reiterate that this argument, while in vogue at this moment, seems to us morally problematic, relying as it does upon a premise of moral hostage-taking – holding innocent civilians and even combatants' lives hostage by deliberately foregoing the use of the most precise technology available for targeting, for the express purpose of putting pressure on political leadership not to use force in the first place.

[42] Kenneth Anderson, "Efficiency Jus ad bellum and in bello: Making the Use of Force Too Easy?" in Claire Finkelstein, Jens David Ohlin, Andrew Altman, eds., Targeted Killings: Law and Morality in an Asymmetrical World (Oxford UP 2012), at 374-399.

[43] See, e.g., Scott Shane, "Coming Soon: The New Drones Arms Race," N.Y. Times, October 8, 2011, at http://www.nytimes.com/2011/10/09/sunday-review/coming-soon-the-drone-arms-race.html?pagewanted=all.

[44] Sparrow discusses this "dual use" challenge for international regulation of military robotics. "Predators or Plowshares?" at 28. For a sense of the extraordinary range of activities in ordinary

Moreover, even with respect to militarily-specific applications of autonomous robotics advances, critics worrying that the United States is spurring a new arms race overlook just how many military-technological advances result from efforts to find a technological solution to yet another form of violation committed by adversaries of the basic laws of war. A challenge for the United States and its allies is that it is typically easier and faster for non-state adversaries to come up with new behaviors to violate the laws of war to gain advantage than it is to come up with new technological counters.

In part because it is also easier and faster for states that are competitively engaged with the United States to deploy systems that are, in the U.S. view, ethically and legally deficient, the United States *does* have a strong interest in seeing that development and deployment of autonomous battlefield robots be regulated, legally and ethically. Moreover, some critics are right to argue that even if U.S. abstention from this new arms-race alone would not prevent the proliferation of new destructive technologies, it would nonetheless be reckless for the United States to pursue them without a strategy – including a role for normative constraints – for responding to other states' or actors' use of them for military ends.

These observations – and alarm at the apparent development of an arms race around these emerging and future weapons – lead many to believe that an important part of the solution lies in some form of multilateral treaty.[45] A proposed treaty might be a regulatory one, restricting acceptable weapons systems or restricting acceptable use, perhaps something in the manner of the Chemical Weapons Convention or Biological Weapons Convention.[46] Alternatively, a treaty might be prohibitory; some advocacy groups have already moved to the point of calling for international conventions that

---

life into which robotics appears gradually to be entering – from driverless car technologies to eldercare – see the conference program of Michael Froomkin, Program Chair, We Robot Conference, University of Miami School of Law, April 19-20, 2012, at http://robots.law.miami.edu/. Quite apart from those forms of robotics, no one doubts today that remotely-piloted, and increasingly automated, aircraft will transform civilian aviation worldwide, and that in the end, military remotely-piloted vehicles will be a relatively modest part of a sea-change in aviation.

[45] See, for example, the work of the International Committee for Robot Arms Control: http//www.icrac.co.uk. Their campaign is described in Nik Fleming, "Campaign Asks for International Treaty to Limit War Robots," New Scientist, Sept. 30, 2009, available at http://www.newscientist.com/article/dn17887-campaign-asks-for-international-treaty-to-limit-war-robots.html.

[46] Each of these arms control treaties has been effective in considerable part, however, because in each case the final object of the treaty – a prohibition on the creation or use of certain weapons – is well understood, defined and accepted. Thus the work of the treaty is not to define the final prohibited weapon, but instead to regulate the use of the precursors and intermediate materials that have perfectly lawful, indeed industrially indispensable, uses. The case of autonomous lethal weapons is different in that the function of a treaty would be to seek to determine that which is prohibited – and in an environment of incremental advances, it is likely to subject to many, many debates and failures of agreement or interpretation that are not at issue in the CWC or BWC. For this reason among others we believe that the CWC and BWC are inapposite models for regulation of autonomous weapons systems.

Marchant, et al point out that there are many other arms control models that could be drawn upon, including formal international agreements as well as "soft-law" approaches. "International Governance of Autonomous Military Robotics," 289-314.

would essentially ban autonomous weapons systems altogether, along the lines of the Ottawa landmines convention.[47]

Ambitions for multilateral treaty regulation in this context are misguided for several reasons. To start with, limitations on autonomous military technologies, although quite likely to find wide superficial acceptance among non-fighting states and some non-governmental groups and actors, will have little traction. Some states, particularly Asian allies worried about a rising and militarily assertive China, may want the United States to be more aggressive in adopting the latest technologies, given that its adversary is likely to have far fewer compunctions about its own autonomous weapon systems, and others are likely to favor any technological development that extends the reach and impact of U.S. forces or enhances their own ability to counter adversary capabilities.

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. Moreover, there are serious humanitarian risks to prohibition, given the possibility that autonomous weapons systems could in the long run be more discriminating and ethically preferable to alternatives; blanket prohibition precludes the possibility of such benefits. And, of course, there are the general challenges of compliance - the collective action problems of failure and defection that afflict all such treaty regimes.

Principles, Policies, and Processes

Nevertheless, the dangers associated with evolving autonomous robotic weapons are very real, and the United States has a serious interest in guiding development in this context of international norms, by which we mean not binding legal rules, whether treaty rules or customary international law – but instead widely-held expectations about legally or ethically appropriate conduct. Among other reasons, such norms are important to the United States for guiding and constraining its internal practices, such as R&D and eventual deployment of autonomous lethal systems that it regards as legal; they help earn and sustain necessary buy-in from the officers and lawyers who would actually use or authorize such systems in the field.[48] They help establish common standards among

[47] For example, the UK-based NGO "Article 36" (referring to the provision of 1977 Additional Protocol I requiring legal review of new weapons systems) (at http://www.article36.org/) has called for a ban on autonomous lethal weapons systems. See, e.g., Matthew Bolton, Thomas Nash, and Richard Moyes, "Ban Autonomous Armed Robots," Article 36, March 5, 2012, available at http://www.article36.org/statements/ban-autonomous-armed-robots/. This article explicitly argues that autonomous lethal weapons are morally the equivalent of an indiscriminate antipersonnel landmine and should be banned on the same logic. See also Jürgen Altmann, "Preventive Arms Control for Uninhabited Military Vehicles," in R. Capurro & M. Nagenborg (eds.), Ethics and Robots (AKA Verlag Heidelberg 2009), at 69-82.

[48] The success of the NGO campaign, led by Human Rights Watch and the International Committee of the Red Cross, to ban blinding laser weapons, for example - which finally resulted in a new Protocol IV to the CCW (entry into force 1998; US ratification 2009), depended in considerable part on the resistance of military officers to blinding lasers as a weapon of war. See Ann Peters, "Blinding Laser Weapons: New Limits on the Technology of Warfare," 18 Loy. L.A. Int'l & Comp. L. Rev. 733 (1996), at http://digitalcommons.lmu.edu/ilr/vol18/iss4/3.

the United States and its partners and allies to promote cooperation and joint operations. And they raise the political and diplomatic costs to adversaries of developing, selling, or using autonomous lethal systems that run afoul of these standards.

A better approach to treaties for addressing these systems is the gradual development of internal state norms and best practices that, worked out, debated, and applied to the United States' own weapons development process, can be carried outwards to discussions with others around the world. This requires a long-term, sustained effort combining internal ethical and legal scrutiny – including specific principles, policies, and processes – and external diplomacy.

To do so, the United States government should resist two extreme instincts – its own instincts among officials to hunker down behind secrecy and avoid discussing and defending even guiding principles, on the one hand, and the instincts of critics or skeptics of autonomous lethal systems favoring the idea of some grand international treaty to regulate or even prohibit them, on the other. Instead the United States government should carefully and continually develop internal norms, principles, and practices that it believes are correct for the design and implementation of such systems. It should also prepare to articulate clearly to the world the fundamental legal and moral principles by which all parties ought to judge autonomous weapons, whether those of the United States or those of others.

The core, baseline principles can and should be drawn and adapted from the customary law-of-war framework: distinction and proportionality. A system must be capable of being aimed at lawful targets; but how good must that capability be in any particular circumstance? The legal threshold has always depended in part upon technology as well as intended use, and the refinement of these standards should consider advancing technological capability against non-autonomous capabilities. Proportionality requires that any use of a weapon must take into account collateral harm to civilians; this rules out systems that simply identify and aim at other weapons without taking civilians into account, but what is the standard of care for an autonomous lethal system in any particular circumstance? This is partly a technical issue of designing systems capable of discerning and calculating civilian harm, but also partly an ethical issue of attaching weights to the variables at stake.[49]

These questions move from overarching ethical and legal principles to processes that make sure principles are concretely taken into account – not just down the road at the deployment stage but much earlier, during the R&D stage. It will not work to go forward with design and only afterwards, seeing the technology, decide what changes need to be made in order to make the system's decision-making conform to legal requirements. By then it may be too late. Engineering designs will have been set for both hardware and software; significant national investment into R&D already undertaken that will be hard to write off on ethical or legal grounds; and national prestige might be in play. This would be true of the United States or other states developing such systems; legal review by that stage would tend to be one of justification at the back end, rather than seeking best practices at the front end.

[49] See generally, Christopher Coker, Ethics and War in the 21st Century (Routledge 2008), at 143-152.

The United States must develop a set of principles to regulate and govern advanced autonomous weapons not just to guide its own systems, but also to effectively assess the systems of other states. This requires that the United States work to bring along its partners and allies – including NATO members and technologically advanced Asian allies – by developing common understandings of norms and best practices as the technology evolves in often small steps. Just as development of autonomous weapon systems will be incremental, so too will development of norms about acceptable systems and uses.

Internal processes should therefore be combined with public articulation of overarching policies.[50] Various vehicles for declaring policy might be utilized over time – perhaps directives by the Secretary of Defense – followed by periodic statements explaining the legal rationale behind decisions about R&D and deployment of weapon technologies. The United States has taken a similar approach in the recent past to other controversial technologies, most notably cluster munitions and landmines, by declaring commitment to specific standards that balance operational necessities with humanitarian imperatives.[51]

To be sure, this proposal risks papering over enormous practical and policy difficulties. The natural instinct of the US national security community – likewise that of other major state powers – will be to discuss little or nothing, for fear of revealing capabilities or programming details to adversaries, as well as to invite industrial espionage and reverse engineering of systems. Policy statements will necessarily be more general and less factually specific than critics would like.[52] Furthermore, one might reasonably question

[50] In March 2011, as an example of a similar sort of public articulation, the White House announced that adherence to Protocol I's detainee treatment "principles is also an important safeguard against the mistreatment of captured U.S. military personnel. The U.S. Government will therefore choose out of a sense of legal obligation to treat the principles set forth in Article 75 as applicable to any individual it detains in an international armed conflict, and expects all other nations to adhere to these principles as well." Fact Sheet: New Actions on Guantanamo and Detainee Policy, March 7, 2001, available at http://www.whitehouse.gov/the-press-office/2011/03/07/fact-sheet-new-actions-guant-namo-and-detainee-policy. We acknowledge that this and the cluster munitions policy, below, present different situations than that of our preferred policy mechanism here. In the landmines and cluster munitions cases, the United States was responding to a soft-law initiative undertaken by NGOs and others in the international community, rather than initiating one; moreover, the United States makes no claims that its internal practices that it urges on others are anything more than desirable or best practices, and that matters are a very long way from anything that could remotely be called "customary law." Nonetheless, the situations share the feature of the United States seeking through internal practices that are not themselves characterized as "law" to influence norms on a broader basis, and to push others in the international community to accept certain pragmatic normative constraints, whether or not they ever result in "law" as such.

[51] See, for example, DoD Press Release, Cluster Munitions Policy Released, July 9, 2008, available at http://www.defense.gov/releases/release.aspx?releaseid=12049; U.S. Department of State, Landmine Policy White Paper, Feb. 27, 2004, available at http://www.fas.org/asmp/campaigns/landmines/FactSheet_LandminePolicyWhitePaper_2-27-04.htm.

[52] This is similar to a debate currently underway over targeted killing through drone warfare and whether the US government can rest with general statements of legal principles governing the practice, or whether it ought to release actual internal legal opinions or other secret government documents. For a summary of that on-going argument, see Kenneth Anderson, "Readings: The Canonical National Security Law Speeches of the Obama Administration Senior Official and General Counsels," Lawfare Blog, April 19, 2012, at

whether broad principles such as distinction and proportionality can meaningfully be applied and discussed publicly with respect to technological systems distinguishable only in terms of digital ones and zeroes buried deep in programmed computer code.

These concerns are real, but there are at least two mitigating solutions. First, the United States will need to resist its own impulses toward secrecy and reticence with respect to military technologies, recognizing that the interests those tendencies serve are counterbalanced here by interests in shaping the normative terrain on which it and others will operate militarily as technology quickly evolves. The legitimacy of such inevitably controversial systems in the public and international view matters too. It is better that the United States work to set the global standard by actively explaining its compliance with it than to let other states or groups set it – whether those who would impose unrealistic, ineffective or dangerous prohibitions or those who would prefer few or no constraints at all.

Of course, there are limits to transparency here, on account of both secrecy concerns and the practical limits of persuading skeptical audiences about the internal and undisclosed decision-making capacities of rapidly evolving robotic systems. A second part of the solution is therefore to emphasize the internal processes by which the United States considers, develops, and tests its weapon systems. Legal review of any new weapon system is required as a matter of international law; the U.S. military would conduct it in any event. Even when the United States cannot disclose publicly the details of its automated systems and their internal programming, however, it should be quite open about its vetting procedures, both at the R&D stage and at the deployment stage, including the standards and metrics it uses.

Although the United States cannot be very open publicly with the results of its tests, for fear of disclosing details of its capabilities to adversaries, it should be prepared to share them with its military allies as part of an effort to establish common standards. Looking more speculatively ahead, the standards the United States applies internally in developing its systems might eventually form the basis of export control standards, in sharing technologies with these allies and, as other countries develop their own autonomous lethal systems, the United States can lead in forging a common export control regime and standards of acceptable weapons available on international markets.[53]

A Traditional Approach to a New Challenge

In the end, one might still raise an entirely different objection altogether to these proposals: that the United States should not unnecessarily constrain itself in advance to a set of normative commitments, given vast uncertainties about the technology and future security environment. Better that the United States cautiously wait, the argument might go, and avoid binding itself to one or another legal or ethical interpretation until it

---

http://www.lawfareblog.com/2012/04/readings-the-national-security-law-speeches-of-the-obama-administration-general-counsels/.

[53] For background on the evolution of conventional weapons export control regimes, which might serve as a model, see Richard F. Grimmett, CRS Report for Congress, Military Technology and Weapons Export Controls: The Wassenaar Arrangement (2006). For comparison to UK concerns regarding export controls, see MOD 2011 Report, at para. 514.

needs to. This objection, however, fails to appreciate that, while significant deployment of highly-autonomous systems may be far off, R&D decisions are already upon us. Moreover, shaping international norms is a long-term process, and unless the United States and its allies accept some risk in starting it now, they may lose the opportunity to do so later.

In the end, all of this is a rather traditional approach – relying on the gradual evolution and adaptation of longstanding law-of-war principles – to regulate what seems to many like a revolutionary technological and ethical predicament. That is in part because the challenge of regulating radical innovations in weaponry within a long-standing legal and ethical framework is hardly novel.

Some view these automated technology developments as a crisis for the laws of war. To the contrary, provided we start now to incorporate ethical and legal norms into weapons design, the incremental movement from automation to genuine machine autonomy can be made to serve the ends of law on the battlefield.