Law and Ethics for Robot Soldiers

By KENNETH ANDERSON *&*
MATTHEW WAXMAN

LETHAL SENTRY ROBOT designed for perimeter protec-
tion, able to detect shapes and motions, and combined
with computational technologies to analyze and differen-
tiate enemy threats from friendly or innocuous objects —
and shoot at the hostiles. A drone aircraft, not only
unmanned but programmed to independently rove and hunt prey, perhaps
even 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. These
(and many more) are among the ripening fruits of automation in weapons
design. Some are here or close at hand, such as the lethal sentry robot
designed in South Korea. Others lie ahead in a future less and less distant.

Lethal autonomous machines will inevitably enter the future battlefield —
but they will do so incrementally, one small step at a time. The combination

*Kenneth Anderson is a law professor at American University. Matthew Waxman
is a professor at Columbia Law School and adjunct senior fellow at the Council
on Foreign Relations. Both are members of the Hoover Institution Task Force on
National Security and Law.*

of "inevitable" and "incremental" development raises not only complex strategic and operational questions but also profound legal and ethical ones. Inevitability 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 nonlethal 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 slowly diminish.

Recognizing the inevitable but incremental evolution of these technologies is key to addressing the legal and ethical dilemmas associated with them; U.S. policy for resolving such 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.

Those same features also make it imperative, though, that the United States resist its own impulses toward secrecy and reticence with respect to military technologies and recognize that the interests those tendencies serve are counterbalanced here by interests in shaping the normative terrain — i.e., the contours of international law as well as international expectations about appropriate conduct on which the United States government 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.

Incremental automation of drones

*T*HE 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. 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). The drone proportion will only grow. Yet current drone

military aircraft are not autonomous in the firing of weapons — the weapon must be fired in real time by a human controller. So far there are no known plans or, apparently in the view of military, reasons to take the human out of the weapon firing loop.

Nor are today's drones truly autonomous as aircraft. They require human pilots and flight support personnel in real time, even when they are located far away. 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. The automation of flight is gradually increasing as sensors and aircraft control through computer programming improves.

Looking to 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. Drone aircraft might gradually become capable of higher speeds, torques, g-forces, and other stresses than those 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.

Some believe that the next generations of jet fighters won't be manned. Or that manned fighters will join with unmanned.

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, or anti-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). In similar ways, 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. Some military operators regard many emerging automated weapons systems as merely 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, 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 some overlapping 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 from its own personnel onto remote-area civilian popula-

tions) 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 otherwise quite different.

Automated-arms racing?

*T*HESE 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 nonstate 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.

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, 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, thus making the robot dependent upon executing its own programming, or even genuinely autonomous.

Aside from conventional war on conventional battlefields, covert or special operations will involve their own evolution toward incrementally autonomous systems. Consider intelligence gathering in the months preceding the raid on Osama bin Laden's compound. Tiny surveillance robots equipped with facial recognition technology might have helped affirmatively identify bin Laden much 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 fail-safe, but with very little time to override programmed commands.

These examples have all been stylized to sound precise and carefully controlled. At some point in the near future, however, someone — China, Russia, or someone else — will likely design, build, and deploy (or sell) an autonomous weapon system for battlefield use that is programmed to target something — say a person or position — that is firing a weapon and is positively identified as hostile rather than friendly. A weapon system programmed, that is, to do one thing: identify the locus of enemy fire and fire back. It thus would lack the ability altogether to take account of civilian presence and any likely collateral damage.

Quite apart from the security and war-fighting 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. Yet 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 such as development of smaller, more mobile, armed robots able to "hide" as well as "hunt" on their own.

The implication is that the arms race in battlefield robots will be more than simply a race for ever more autonomous weapons systems. More likely, it will mostly be a race for ways to counter and defend against them — partly through technical means, but also partly through the tools of international norms and diplomacy, provided, however, that those norms are not over-invested with hopes that cannot realistically be met.

Legal and ethical requirements

*T*HE 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. 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. But arguments over the legitimacy of particular weapons (or their legitimate use) go back to the beginnings of debate over the laws and ethics of war: the legitimacy, for example, of poison, the crossbow, 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.

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 ethi-

cally and legally abhorrent and should be prohibited by law. 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 perceiving they can benefit from it, too. In some cases, legal prohibitions on the weapon system as such erode, as happened with submarines and airplanes; what survives is typically legal rules for the *use* of the new weapon, with greater or lesser specificity. In a few cases (including some very important ones), legal prohibitions on the weapon as such gain hold. The ban on poison gas, for example, has survived in one form or another with very considerable effectiveness throughout the 20th century.

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?

Where in the long history of new weapons and their ethical and legal regulation will autonomous robotic weapons fit?

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.

But that wait-and-see view is shortsighted and mistaken. Not all the important innovations in autonomous weapons are so far off. 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.

Moreover, looking to the long term, technology and weapons innovation does not take place in a vacuum. The time to take into account law and ethics to inform and govern autonomous weapons systems is now, before technologies and weapons development have become "hardened" in a particular path and their design architecture becomes difficult or even impossible to change. Otherwise, the risk is that technology and innovation alone, unleavened by ethics and law at the front end of the innovation process, let slip the robots of war.

This is also the time — before ethical and legal understandings of autonomous weapon systems likewise 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. What might such a framework look like? Consider the traditional legal and ethical paradigm to which autonomous weapons systems must conform, and then the major objections and responses being advanced today by critics of autonomous weapons.

A legal and ethical framework

*T*HE BASELINE LEGAL and ethical principles governing the introduc-
tion of any new weapon are distinction (or discrimination) and pro-
portionality. 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. Any autonomous robot
weapon system will have to possess the ability to be aimed, or aim itself, at
an acceptable legal level of discrimination.

Proportionality adds that even if a weapon meets the test of distinction,
any actual use of a weapon must also involve an 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. While easy to state in the
abstract, this evaluation for taking into account civilian collateral damage is
difficult for many reasons. 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 excess can be
determined in truly gross cases, there is no accepted formula that gives deter-
minate outcomes in specific cases; it is at bottom a judgment rather than a
calculus. Nonetheless, it is a fundamental requirement of the law and ethics
of war that any military operation undertake this judgment, and that must
be true of any autonomous weapon system's programming as well.

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," however, is to ask what we would require of an ethical
human soldier performing the same function.

Some leading roboticists have been studying ways in which machine pro-
gramming might eventually capture the two fundamental principles of dis-
tinction and proportionality. As for programming distinction, one could the-
oretically 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 pur-
poses, is a relative judgment: Measure anticipated civilian harm and measure
military advantage; subtract and measure the balance against some deter-
mined standard of "excessive"; if excessive, do not attack an otherwise law-
ful target. Difficult as these calculations seem to any experienced law-of-war
lawyer, they are nevertheless 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. The ethical and legal engineering matter every bit as much as the mechanical or software engineering.

Four objections

*I*F THIS IS the optimistic vision of the robot soldier of, say, decades from now, it is subject already to four main grounds 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. Artificial intelligence has over-promised before. Once into the weeds of the judgments that these broad principles imply, the requisite intuition, cognition, and judgment look ever more marvelous — if not downright chimerical when attributed to a future machine.

This skepticism is essentially factual, a question of how technology evolves over decades. Noted, it is quite possible that fully autonomous weapons will never achieve the 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, by turning decision chains over to machines, might indeed reduce risks to civilians by making targeting more precise and firing decisions more controlled, especially compared to human soldiers whose failings might be exacerbated by fear, vengeance, or other emotions.

It is true that relying on the promise of computer analytics and artificial intelligence risks pushing us down a slippery slope, propelled by the promise of future technology to overcome human failings rather than addressing them directly. If forever unmet, it becomes magical thinking, not technological promise. Even so, 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 categorical moral one which 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). In that regard, the title of this essay is deliberately provocative in pairing "robot" and "soldier," because, on this objection, such a pairing is precisely what should never be attempted.

This is a difficult argument to engage, since it stops with a moral principle that one either accepts or 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. If the decision to fire is made by a machine, who should be held responsible for mistakes? The soldier who allowed the weapon system to be used and make a bad decision? The commander who chose to employ it on the battlefield? The engineer or designer who programmed it in the first place?

This is an objection particularly salient to those who put significant faith in laws-of-war accountability by mechanisms of individual criminal liability, whether 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 its global effectiveness is far from clear. Devotion to individual criminal liability as the presumptive mechanism of accountability risks blocking development of machine systems that would, if successful, reduce actual harm to civilians on or near the battlefield.

One objection holds that autonomous weapons systems undermine the possibility of holding anyone accountable.

Finally, the long-run development of autonomous weapon systems faces the 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. The result might be a greater propensity to use military force and wage war.

As a moral matter, this objection is subject to a moral counter-objection. Why not just forgo all easily obtained protections for civilians or soldiers in war for fear that without holding these humans "hostage," so to speak, political leaders would be tempted to resort to war more than they ought? Moreover, as an empirical matter, this objection is not so special to autonomous weapons. Precisely the same objection can be raised with respect to remotely-piloted drones — and, generally, with respect to any technological development that either reduces risk to one's own forces or, especially perversely, reduces risk to civilians, because it invites more frequent recourse to force.

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: the incremental way autonomous weapon systems will develop. After all, these objections are often voiced as though there was likely to be some determinate, ascertainable point when the human-controlled system becomes the machine-controlled one. It seems far more likely, however, that the evolution of weapons technology will be gradual, slowly and indistinctly eroding the role of the human in the firing loop. And crucially, the role of real-time human decision-making will be phased out in some military con-

texts 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 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. 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 and in any case under greater restraint.

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. They are already doing so faster than many observers of technology would have guessed. And the decision architectures that would govern firing a weapon are not so completely removed from those of, say, an elder-care robot engaged in home-assisted living programmed to decide when to take emergency action.

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 U.S. efforts to find technological "fixes" to successive forms of violation of the basic laws of war committed by its adversaries. A challenge for the United States and its allies is that it is typically easier and faster for nonstate adversaries to come up with new behaviors that 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, 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 for responding to other states' or actors' use for military ends. That strategy necessarily includes a role for normative constraints.

These observations — and alarm at the apparent development of an arms

race around these emerging and future weapons — lead many today to believe that an important part of the solution lies in some form of multilateral treaty. A proposed treaty might be "regulatory," restricting acceptable weapons systems or regulating their acceptable use (in the manner, for example, that certain sections of the Chemical Weapons Convention or Biological Weapons Convention regulate the monitoring and reporting of dual use chemical or biological precursors). Alternatively, a treaty might be flatly "prohibitory"; some advocacy groups have already moved to the point of calling for international conventions that would essentially ban autonomous weapons systems altogether, along the lines of the Ottawa Convention banning antipersonnel landmines.

Ambitions for multilateral treaty regulation (of either kind) 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 nonfighting states and some nongovernmental groups and actors, will have little traction with states whose practice matters most, whether they admit to this or not. Israel might well be the first state to deploy a genuinely autonomous weapon system, but for strategic reasons not reveal it until actually used in battle. Some states, particularly Asian allies worried about a rising and militarily assertive China, may want the United States to be more aggressive, not less, in adopting the latest technologies, given that their future adversary is likely to have fewer scruples about the legality or ethics of its own autonomous weapon systems. America's key Asian allies might well favor nearly 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 workable 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 likely be unattainable. Moreover, no one should forget that 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 endemic 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 international norms we do not mean new binding legal rules only — whether treaty rules or customary international law — but instead widely-held expectations about legally or ethically appropriate conduct, whether formally binding or not. Among the reasons the United States should care is that such norms are important for guiding and constraining its internal practices, such as R&D and eventual deployment of autonomous lethal systems 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. They assist in establishing common standards among the United States and its partners and allies to promote cooperation and permit 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 than treaties is the gradual development of internal state norms and best practices.

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

To be successful, the United States government would have to resist two extreme instincts. It would have to resist its own instincts to hunker down behind secrecy and avoid discussing and defending even guiding principles. It would also have to refuse to cede the moral high ground to critics of autonomous lethal systems, opponents demanding some grand international treaty or multilateral regime to regulate or even prohibit them.

The United States government should instead carefully and continuously 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 — distinction — but how good must that capability be in any particular circumstance? The legal threshold has historically depended in part upon the general state of aiming technology, as well as the intended use. Proportionality, for its part, 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 once again, what is the standard of care for an autonomous lethal system in any particular "proportionality" circumstance? This is partly a technical issue of designing systems capable of discerning and estimating civilian harm, but also partly an ethical issue of attaching weights to the variables at stake.

These questions move from overarching ethical and legal principles to processes that make sure these 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, to 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 but also 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.

The U.S. must develop a set of principles to regulate and govern advanced autonomous weapons.

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. 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.

To be sure, this proposal risks papering over enormous practical and policy difficulties. The natural tendency of the U.S. national security community — likewise that of other major state powers — will be to discuss little or nothing, for fear of revealing capabilities or programming to adversaries, as well as inviting industrial espionage and reverse engineering of systems.

Policy statements will necessarily be more general and less factually specific than critics would like. Furthermore, one might reasonably question not only whether broad principles such as distinction and proportionality can be machine-coded at all but also whether they can be meaningfully discussed publicly if the relevant facts might well be distinguishable only in terms of digital ones and zeroes buried deep in computer code.

These concerns are real, but there are at least two mitigating solutions. First, as noted, 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 global standards than let other states or groups set them.

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 too public about the results of such tests, it should be prepared to share them with its close 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. 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 autonomous weapons available on international markets.

A traditional approach to a new challenge

*I*N 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 through a set of normative commitments, given vast uncertainties about the technology and future security environment. Better cautiously to wait, the argument might go, and avoid binding itself to one or another legal or ethical interpretation until it needs to. This fails to appreciate, however, 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 long-standing law-of-war principles. The challenges are scarcely novel.

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